OmniXtract

User and Administration Guide

Version: 4.0 SP1

Disclaimer

This document contains information proprietary to Newgen Software Technologies Ltd. User may not disclose or use any proprietary information or use any part of this document without written permission from Newgen Software Technologies Ltd.

Newgen Software Technologies Ltd. makes no representations or warranties regarding any software or to the contents or use of this guide. It also specifically disclaims any express or implied warranties of merchantability, title, or fitness for any particular purpose. Even though Newgen Software Technologies Ltd. has tested the hardware and software and reviewed the documentation, it does not guarantee or imply that this document is error free or accurate regarding any particular specification. As a result, this product is sold as it is and user, the purchaser, is assuming the entire risk as to its quality and performance. Further, Newgen Software Technologies Ltd. reserves the right to revise this publication and make changes in its content without any obligation to notify any person, of such revisions or changes. Newgen Software Technologies Ltd. authorizes no Newgen agent, dealer or employee to make any modification, extension, or addition to the above statements.

Newgen Software Technologies Ltd. has attempted to supply trademark information about company names, products, and services mentioned in this document. Trademarks indicated below were derived from various sources.

Newgen Software, Registered Office, New Delhi
E-44/13
Okhla Phase - II
New Delhi 110020
India
Phone: +91 1146 533 200
info@newgensoft.com

Contents

Preface

The *OmniXtract User and Administration Guide* explains how to use the different modules of OmniXtract such as Definition, Services, Reports, Audit Log, User Management, Configuration, License Management, and application programming interface (API).

To ensure you are referring to the latest and most recent revision of this guide, download it from one of the following locations:

- Newgen Internal Doc Portal, if you are a Newgen employee.
- Newgen Partner Portal, if you are a Newgen partner.

Revision history

Revision Date	Description
August 2024	Initial publication

Intended audience

This guide is intended for resources who use Intelligent Document Extraction solutions to create and optimize extraction definition, monitor extraction performance, perform output analysis, and configure and scale the extraction services. The user must have basic knowledge of working with enterprise-level extraction solutions, an understanding of extractable quality guidelines, and an understanding of types of documents and design variations within the scope of OmniXtract. To perform different operations, the readers must have read and write permissions on the respective OmniXtract modules. The users are responsible for maintaining the application services that require root or administrative access to start or stop the daemon or Windows service.

Documentation feedback

To provide feedback or any improvement suggestions on technical documentation, write an email to docs.feedback@newgensoft.com.

To help capture your feedback effectively, share the following information in your email:

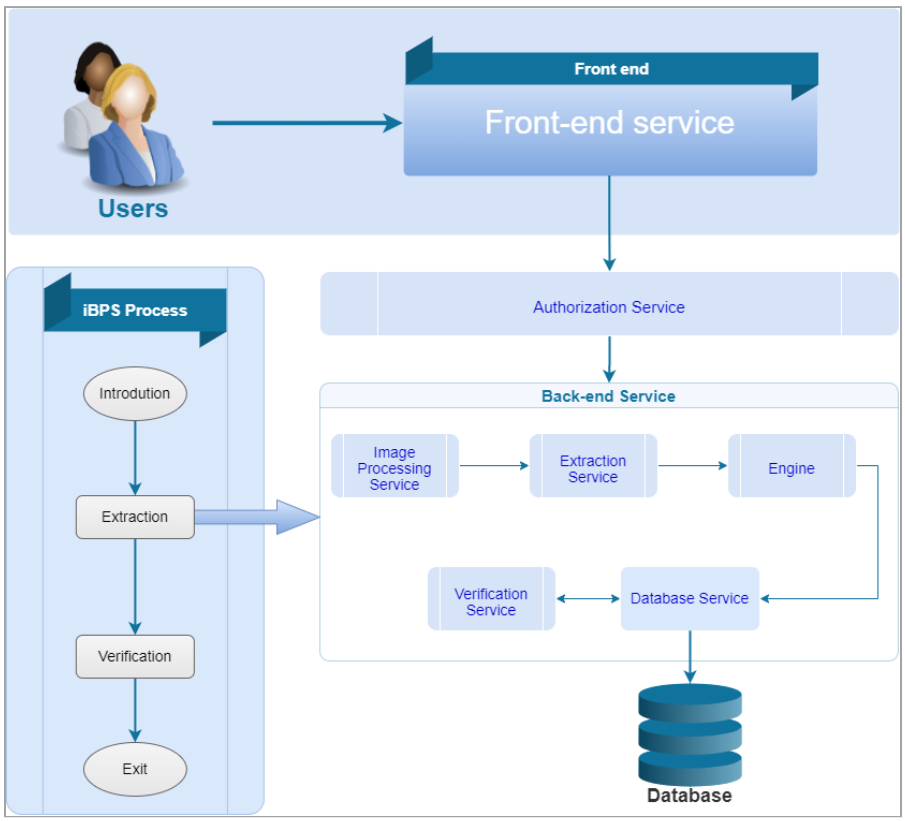
- Document name
- Version
- Chapter, topic, or section
- Feedback or suggestions

About OmniXtract

Newgen OmniXtract is an efficient and scalable data extraction system. It is designed to support multiple file formats and design variations. OmniXtract enables the extraction of tabular and non-tabular data from both structured and semi-structured documents.

Managing manual extraction of business-critical data from documents is a challenging and lengthy task. The complexity of this task multiplies with the increase in the number of documents. Both structured and semi-structured documents offer their respective set of data extraction challenges. OmniXtract handles all these challenges using dedicated data extraction tools and technologies.

The various modules or components in the OmniXtract system are shown in the following figure:



The below table describes the different modules of OmniXtract:

Module	Description
Definition	It allows you to create a Definition that comprises fields for extraction and their attributes such as field name and field type. OmniXtract supports extraction on structured as well as semi-structured documents.
Services	Automatically extract documents at run time as per the created Definition. Services can run 24 x 7, without any manual intervention.
Reports	Reports give an insight into the performance of the overall extraction solution for the period. Reports are used to monitor the extraction performance in real-time.
Audit Log	The audit log monitors all the actions performed by users across different modules of OmniXtract in the signed-in cabinet.
User Management	It is used for providing and managing access rights assigned to users on different modules of OmniXtract.
Configuration	Configuration management is a process for maintaining functional attributes of the system as per the requirement.
License Management	The License Management allows you to manage the license of the OmniXtract system that uses secure and industry standards for data extraction.
APIs	API is a set of definitions and protocols used for building and integrating an application software.

The primary features of OmniXtract are as follows:

- Data extraction based on deep learning
- Collaborative web-based definition studio
- Control center and performance monitoring
- High productivity and scalability
- Enhanced user experience
- Data extraction and processing from different document formats
- Line-item data extraction using an improvised extraction model
- Structured and semi-structured document-type processing

- Multi-page document processing
- Preprocessing of image
- Color and grayscale image processing
- OCR, OMR, Barcode, QR code, MRZ, MICR, and ICR recognition technologies
- Import and export labels in plain text
- Pre-defined data format
- Multi-line data extraction
- Automatic recognition of tabular data
- Image clip extraction
- Database lookup creation and synchronization
- Output data formatting
- Create local definition
- Import and export of definition
- Automated and scalable extraction services
- Generation of the following reports:
    - Extraction throughput report
    - Accuracy analysis report
- Automatic language identification and multi-language support
- Support of multiple extraction engines like Kadmos, Tesseract, Google Vision, Amazon Textract, and Microsoft Azure Computer Vision.
- Pre-defined definition for extracting key-fields and line-items from invoices.
- Support for Azure and Amazon form parser within the invoice domain.

Signing into OmniXtract

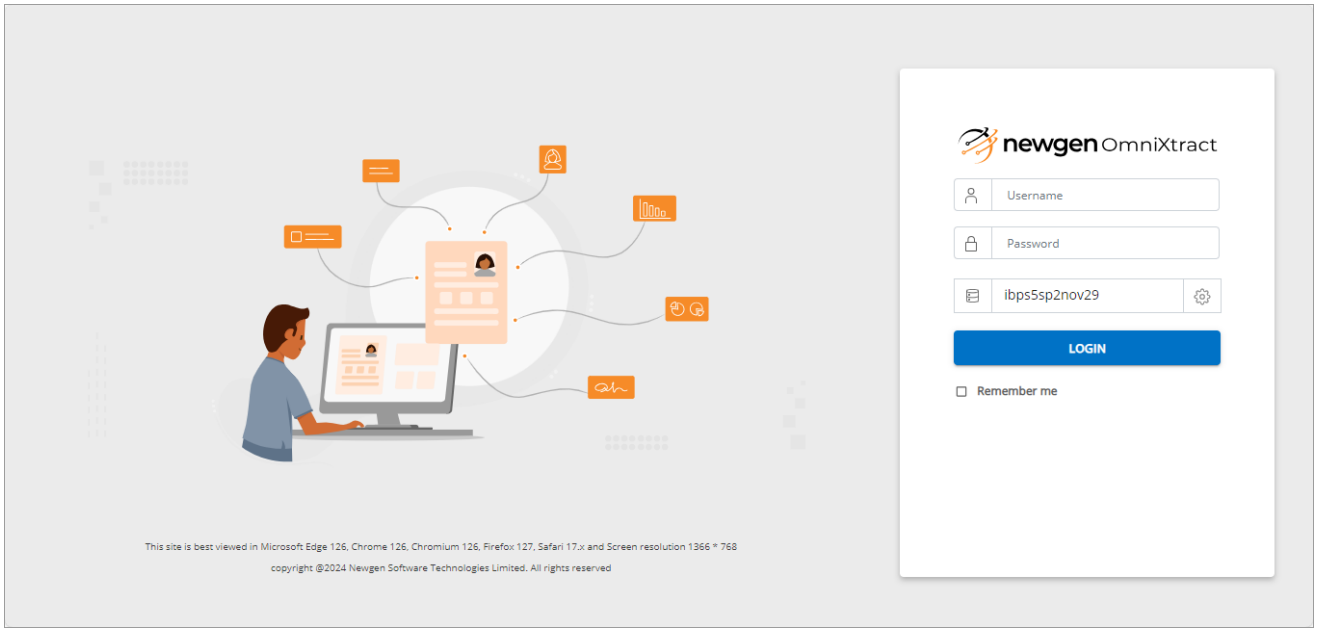
To sign into OmniXtract, perform the following steps:

1. Open a web browser of your choice.

   OmniXtract is best viewed in Microsoft Edge 126.0.2592.56, Chrome 126.0.6478.115, Chromium 82, Firefox 126, and Safari 12.x with a screen resolution of 1366 * 768.

2. In the address bar, enter the URL in the below format:
   http://<IP or Machine name of the OmniXtract Server>:<Port number>
   The default port number is 9102 (UI Service Port).
3. Enter your **Username** and **password**, and select the **cabinet**.
4. Enter the **captcha**.
5. Click **LOGIN**.

   Select the **Remember me** checkbox to allow the system to remember your username and cabinet. When you sign in the next time your username and cabinet are pre-filled.

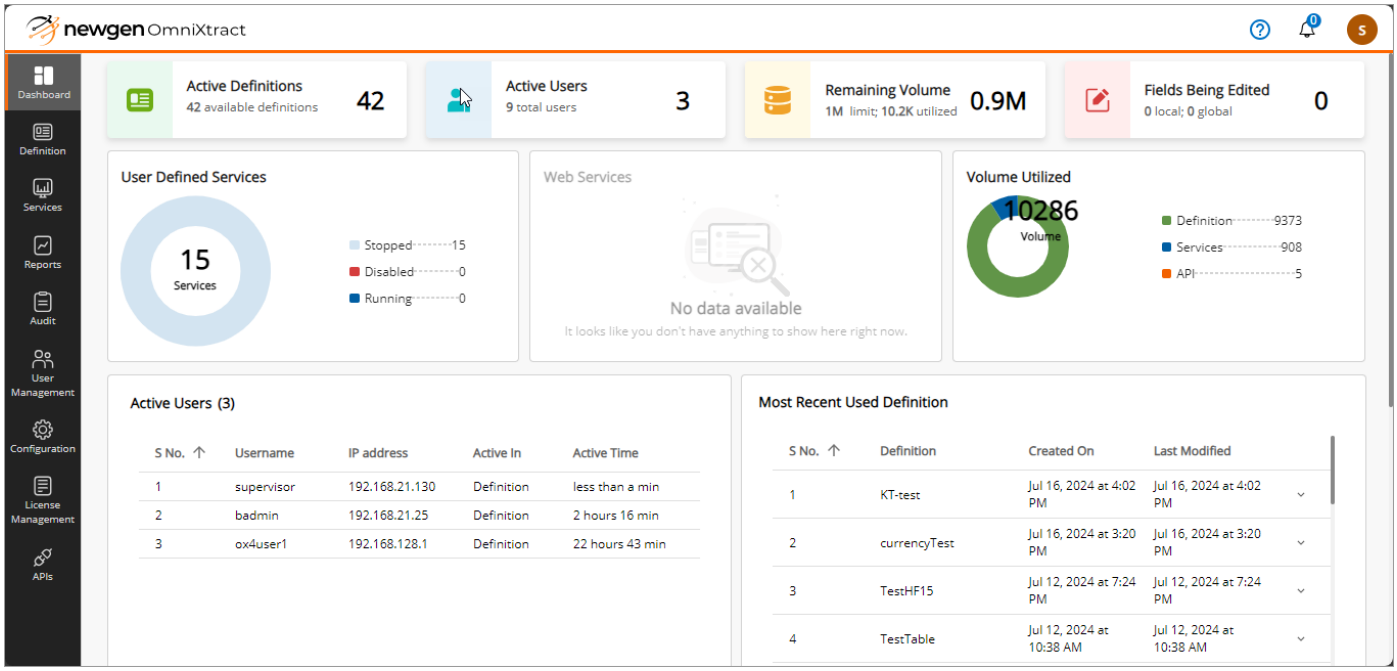


On successful sign-in, the OmniXtract home page appears. By default, the Dashboard screen appears.

The OmniXtract menu bar appears in the left side bar of the screen. The menu bar options include:

- Dashboard
- Definition
- Services
- Reports
- Audit
- User Management
- Configuration
- License Management
- APIs

Your access to different modules depends on the rights assigned to you.



Dashboard

The Dashboard displays the live tracking of active Definitions, active users, and remaining volume in terms of pages, and fields currently being edited. It also displays graphical representations of user-defined services, web services, and total volume utilized in terms of pages by Definition, Services, and APIs.

Using Dashboard, you can get detailed information on active users, most recently used Definitions, User-defined Services, and Web-services in tabular views.

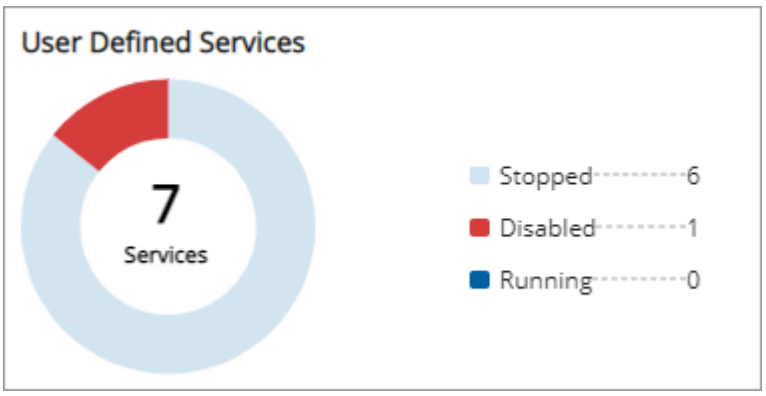
Live tracking

Live tracking is available at the top of the dashboard screen. It displays the overview of the Definitions, the number of logged-in users, the remaining volume, and the number of edited fields.

User-defined services

It gives a graphical representation of the status of user-defined services. A user-defined service can have one of the following statuses:

- Running
- Stopped
- Disabled



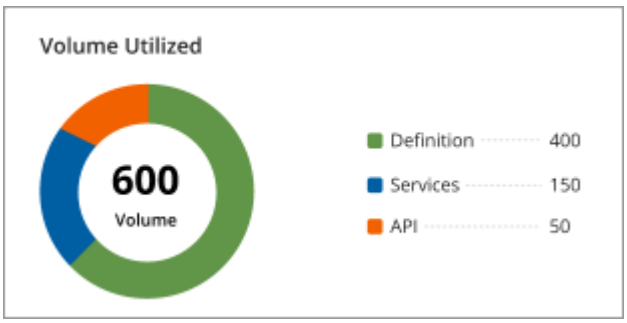
Web services

It gives a graphical representation of the status of web services. A web service can have one of the following statuses:

- Running
- Stopped

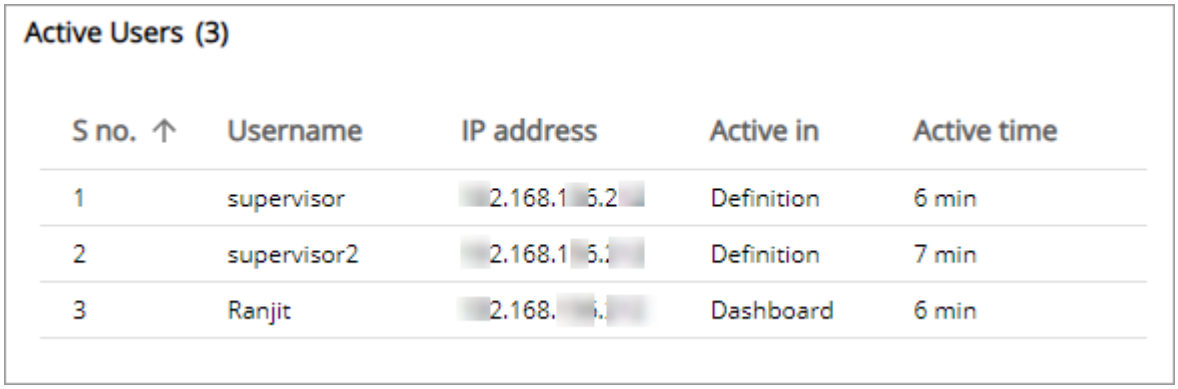
Volume utilized

The volume utilization gives a glimpse of extraction volume consumed in terms of pages by Definition, Extraction Services, and extraction APIs.



Active users

Active users give a glimpse of the logged in users. This area of the dashboard lists the details such as username, IP address, module the user is currently active in, and the active time.



Most recent used definition

It gives a glimpse of the recently used Definitions. It gives details of the Definition along with its date of creation and when it was last modified.

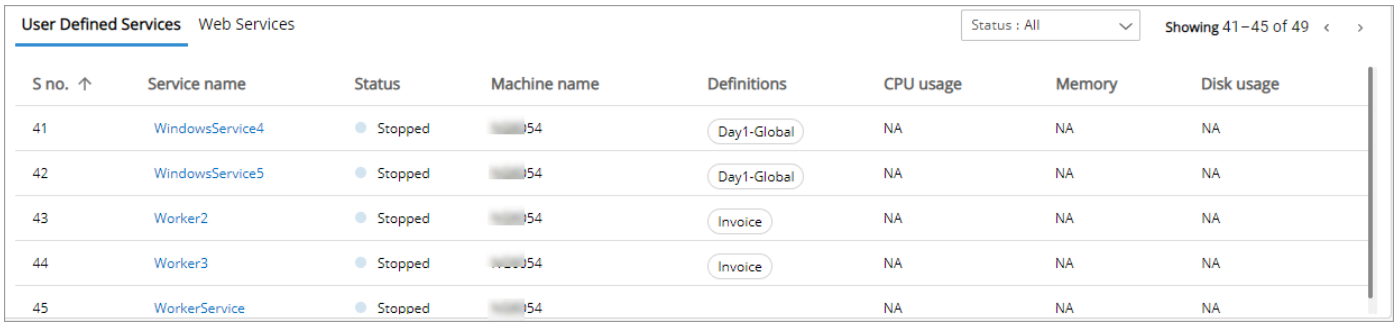
Click the accordion button ˅ to view details such as the lookup table the Definition is associated with and the date and time on which the associated lookup table was last synchronized.

User-defined services

User-defined services give the following information:

- Service name
- Service status
- Machine name where it is running
- List of Definitions it is utilizing

The CPU usage, Memory, and Disk usage does not apply to User Defined Services.



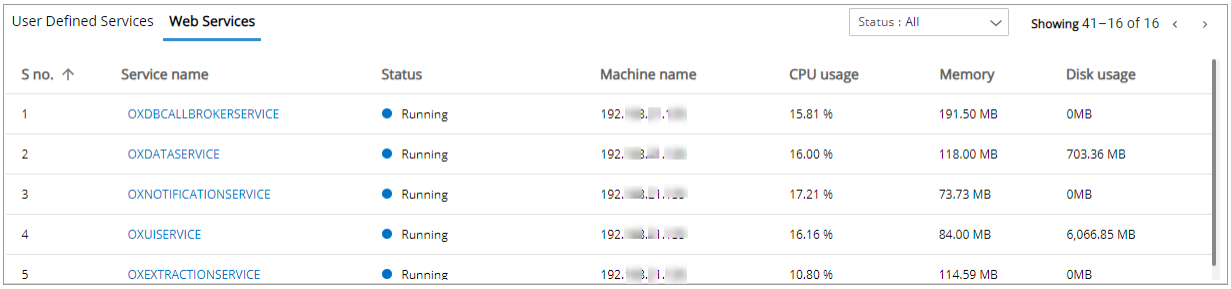
Click the service name to view its detailed information, as described in the following table:

Details Tab	CPU Usage Tab	Memory Usage Tab
• Average CPU usage in percentage • Average memory consumption • Average disc utilization in KB/S • Average network utilization in MB/S • Service name • Service type • Machine name where the service is running • Status whether the service is running, stopped, or disabled • Service running time • Service sleep time or the duration of time for which the service was paused • Definitions where this service is being utilized • Service requests in terms of the number of completed and failed requests	The CPU Usage information is not available for User Defined Services.	The Memory Usage information is not available for User Defined Services.

Web services

The Web Services tab gives the following details:

- Service name
- Service status
- Machine name where it is running
- List of mapped Definitions
- CPU usage
- The total consumed memory



Click the service name to view its general details, CPU usage, and memory usage.

Working with definition

The Definition module of OmniXtract allows you to create a definition for the automatic extraction of data from both structured and semi-structured documents. The Definition module of OmniXtract leverages a single definition studio with a no-code interface for defining multiple document types with different layouts.

The Definition module is accessible by the Supervisor or users with a Definition role. Only users with sufficient rights can access the definitions.

Features

The various features of the OmniXtract Definition module are:

- Predefined data types and plain text labels.
- Fine-tune data definition and labels.
- Support to replicate data definitions and labels to other environments.
- Provide pre-configured document types from various verticals to enable quick implementation.
- Diagnostic capability in case of extraction failure.
- Image pre-processing with a granular level of control.
- Adding non-tabular fields.
- Creating Global and Local Definitions.
- Adding a lookup source for automatic vendor detection by defining primary key identification and management.
- Defining properties for extraction of table fields.
- Associating lookup source to a field.
- Navigating to the next images and documents present in the selected folder.
- Defining the page from where extraction must happen.
- Defining the length of the data that requires extraction.
- Defining the search area on the sample image.
- Showing field results on the screen.
- Defining the format of the extracted data.
- Processing multi-page documents.

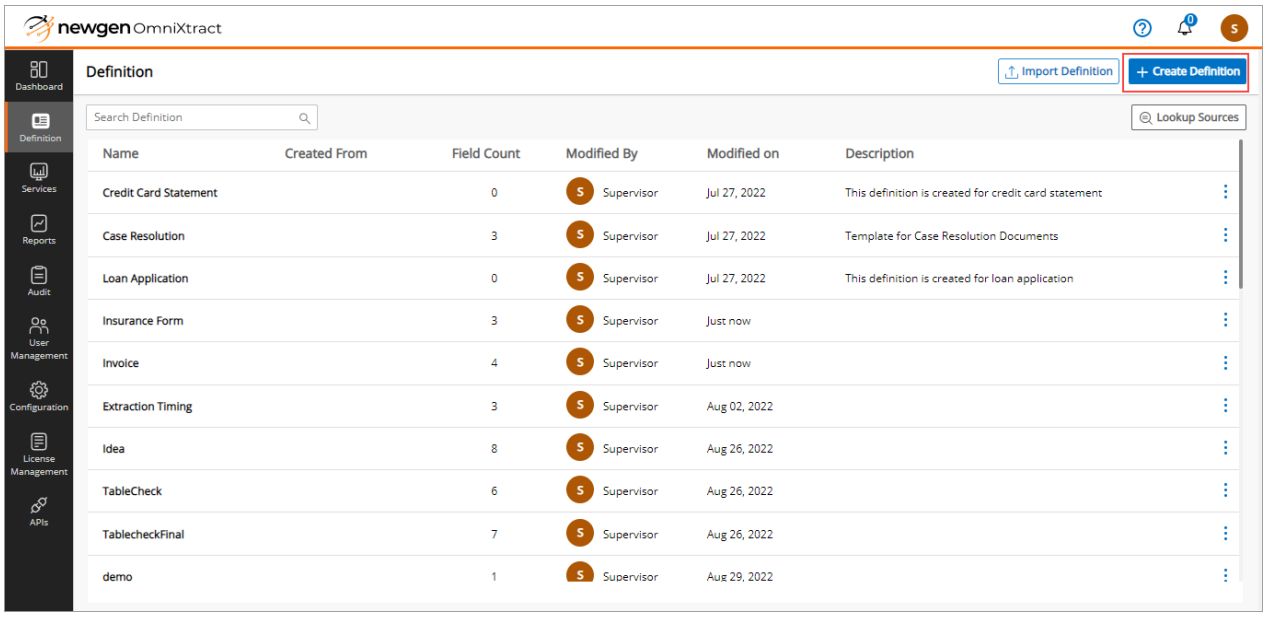
- Import and export label library from text files.
- Extracting labels by zoning data.
- Extract data by creating a zone around the data.
- Setting the priority of the labels.
- Refinement of Definition from time to time to enhance recognition capability.
- Defining Section Header and Section Footer.
- Pattern-based local Definition for alphabetic, alphanumeric, and numeric fields.
- Contextual reports for visualization and analysis of extraction results.

Creating a new definition

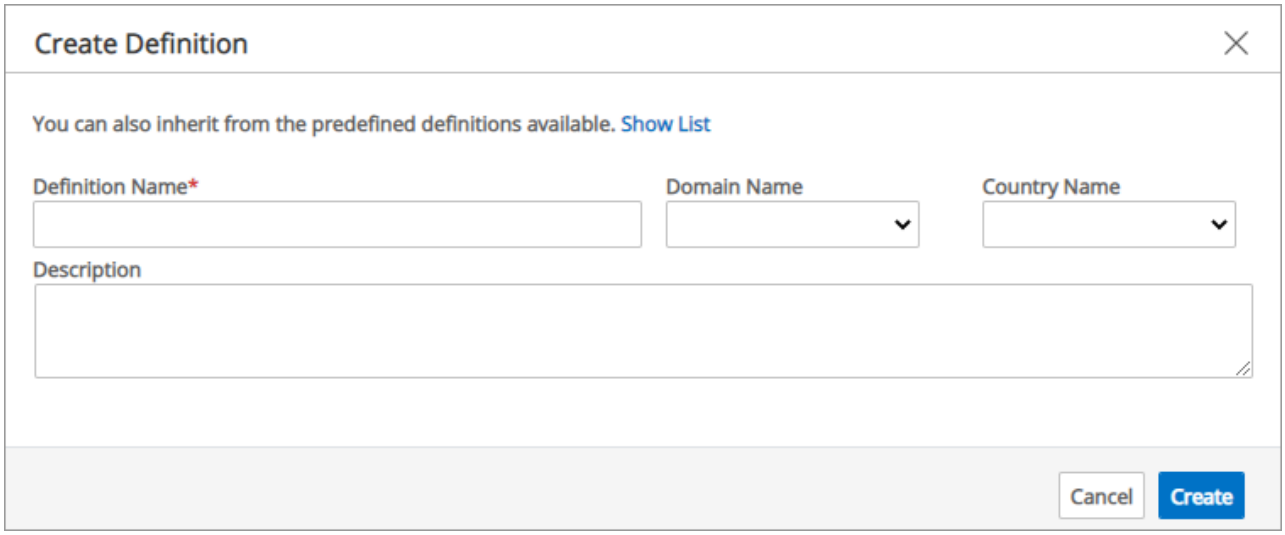
Data extraction can be performed on fully structured and semi-structured documents. A vendor invoice is an example of a semi-structured document, while a bank account opening form is an example of a fully structured document.

To create a new Definition, perform the following steps:

1. Select **Definition** from the menu bar. The Definition screen appears.
2. Click **Create Definition** in the upper-right of the Definition screen.
   The Create Definition link also appears in the middle of the Definition screen when no definition is created.



The Create Definition dialog appears.



3. Enter the following details:
   - **Definition Name** – Specify the name of the Definition. It is mandatory to provide a Definition name.
   - **Domain Name** – Select the name of the domain that your business belongs to such as telecom, agriculture, healthcare, and insurance.
   - **Country Name** – Select the country name.
   - **Description** – Specify a short description of the definition.
4. Click **Create**. The specified definition gets added to the Definition list.

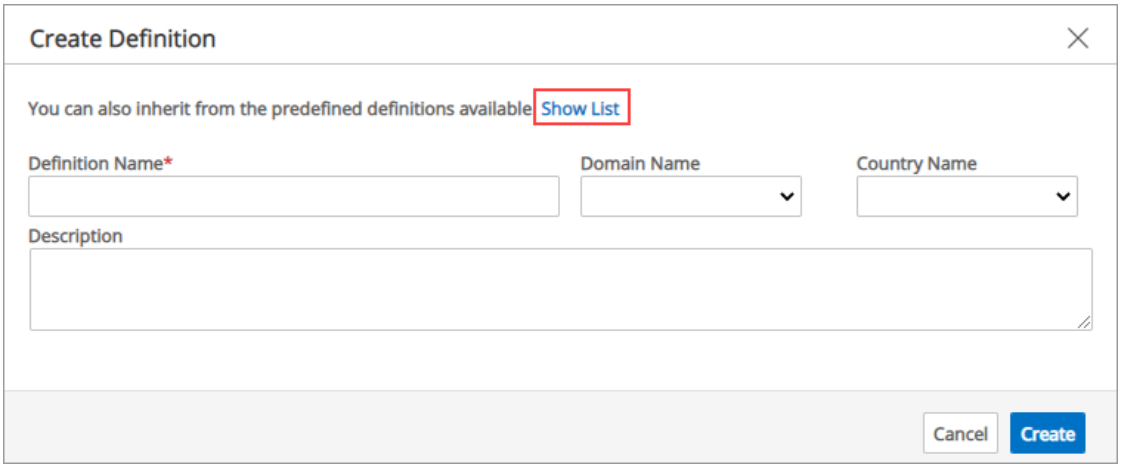
> After you have created the definition, you must add data fields to it for data extraction. To create data fields, refer to the section Working with fields.

Creating a definition using the predefined definitions

You can also create a new definition with fields using the predefined definitions available in OmniXtract.

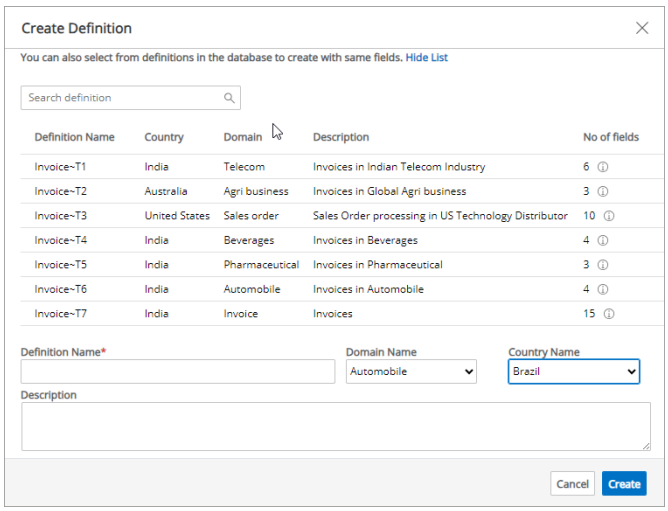
To use the predefined definitions, perform the following steps:

1. Open the Definition screen.
2. Click **Create Definition**. The Create Definition dialog appears.



3. Click the **Show List** link. A list of predefined definitions showing the following details appear:
   - Definition Name
   - Country
   - Domain
   - Description
   - Number of fields – Hover the mouse pointer over the tooltip to view the names of the fields.

> To learn about fields, refer to the section Working with fields.



4. Select a definition from the list of definitions. The Domain Name, Country Name, and Descriptions fields are filled automatically as you select a definition.

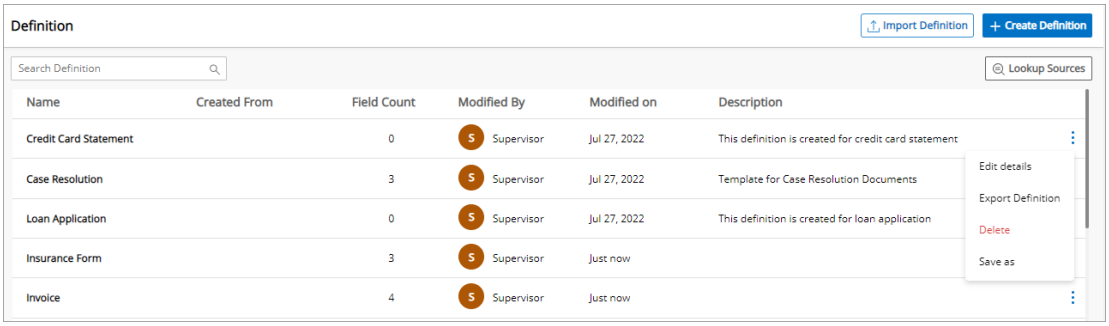
5. Enter the **Definition Name**.
6. Click **Create** to create the definition. The created definition appears in the list of definitions.

Editing a definition

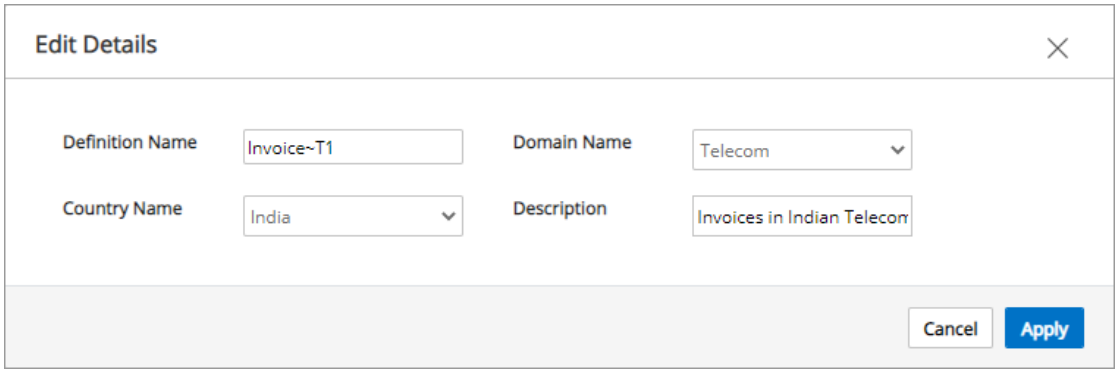
Edit details allow you to modify the description of an existing definition.

To edit an existing definition, perform the following steps:

1. Select **Definition** from the menu bar. The list of existing definitions appears.



2. Click the **More options** ⋮ icon next to the required Definition name and select **Edit details** from the list. The Edit Details dialog appears.



3. Modify the required details and click **Apply** to save the changes.

Exporting and importing a definition

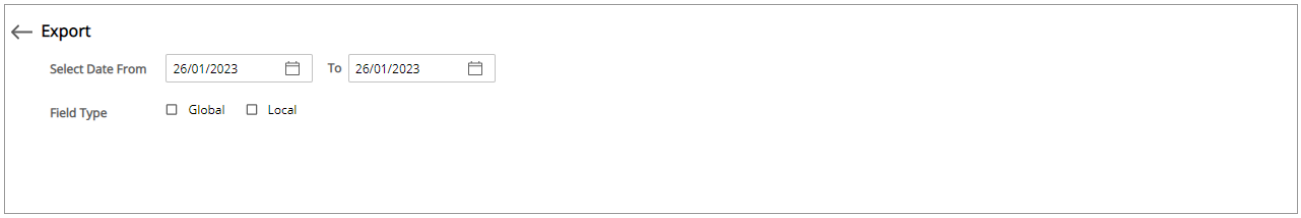
The export and import definition features allow you to copy a Definition from one environment to another, for example, from development to UAT. When you export a Definition, it gets downloaded in the form of a script file. This script file can then be imported into the target environment.

Exporting a definition

Export Definition is enabled when no image is in the open state.

To export a definition, perform the following steps:

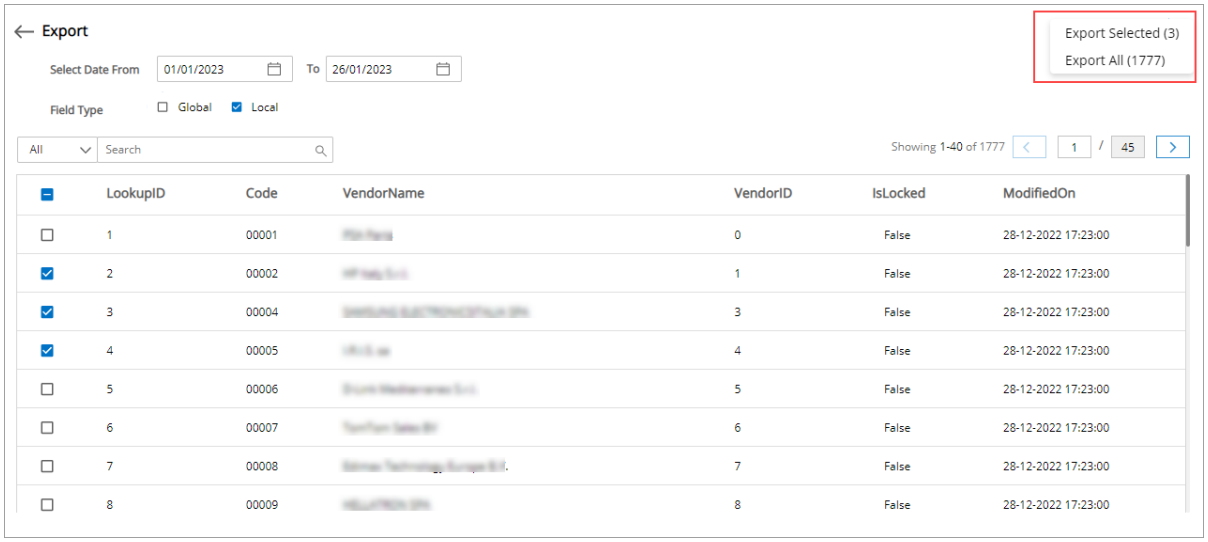
1. Go to the definition list and open the definition.
2. Click ⋮ (**More options**) icon in the upper-right corner of the screen.
3. Select the **Export Definition** option. The Export screen appears.



4. Set the start (**From**) and end (**To**) dates to shortlist the created and updated Definitions during this period.

   If the modification date filter is not selected, then all records are available for export.

5. Select **Global**, **Local**, or both checkboxes, as required.
   - On selecting the **Global** checkbox, a Definition with global attributes is exported.
   - On selecting the **Local** checkbox, a Definition with local attributes is exported.
   - On selecting both, a Definition having both global and local attributes is exported.
   - On selecting the **Local** checkbox, the **Export Selected** and **Export All** options appear in the Export dropdown list in the upper-right corner.



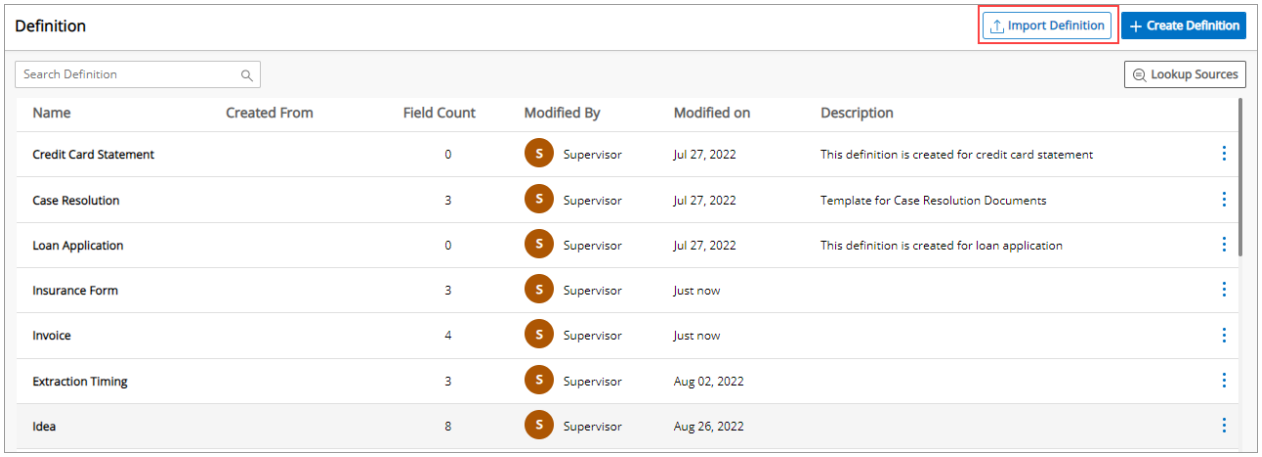
6. Select **Export All** to export the local Definition for all the records or select the **Export Selected** option to export the local Definition of only the selected records.
    a. In the case of large records, you can search them based on different search criteria such as Code, Vendor Name, Vendor ID, IsLocked, and Modified on. Select **All** to perform a search on any columns.
    b. Type the required keyword in the **Selected** option and select a column from the dropdown list where the search is required.
7. in the upper-right corner of the screen, click the **Export** dropdown and select **Export Selected** or **Export All** as required.

Importing a definition

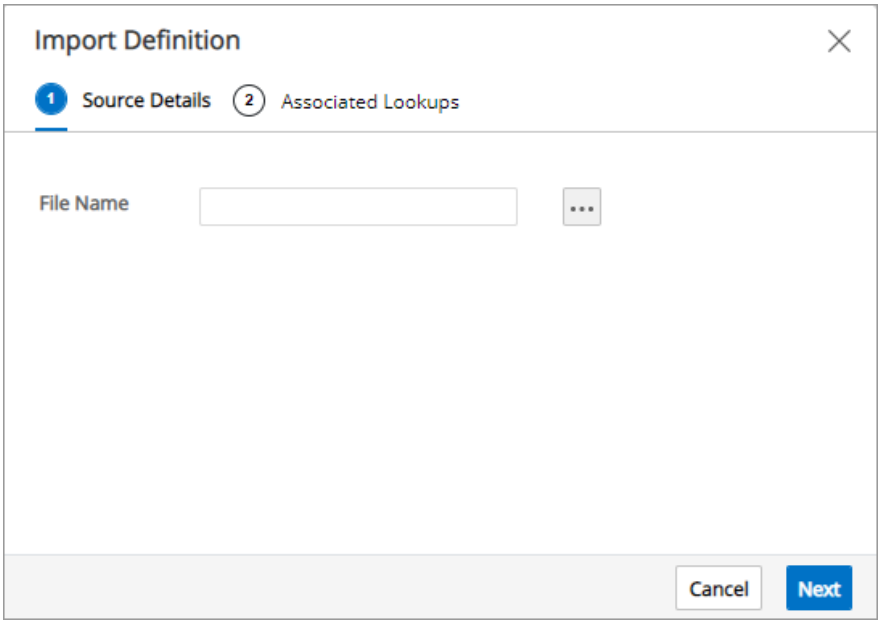
The Import Definition imports an exported Definition created on other OmniXtract environments. The source and target environments must have the same OmniXtract version configured. Make sure the target environment does not have the same Definition name as that of the source environment.

To import a definition, perform the following steps:

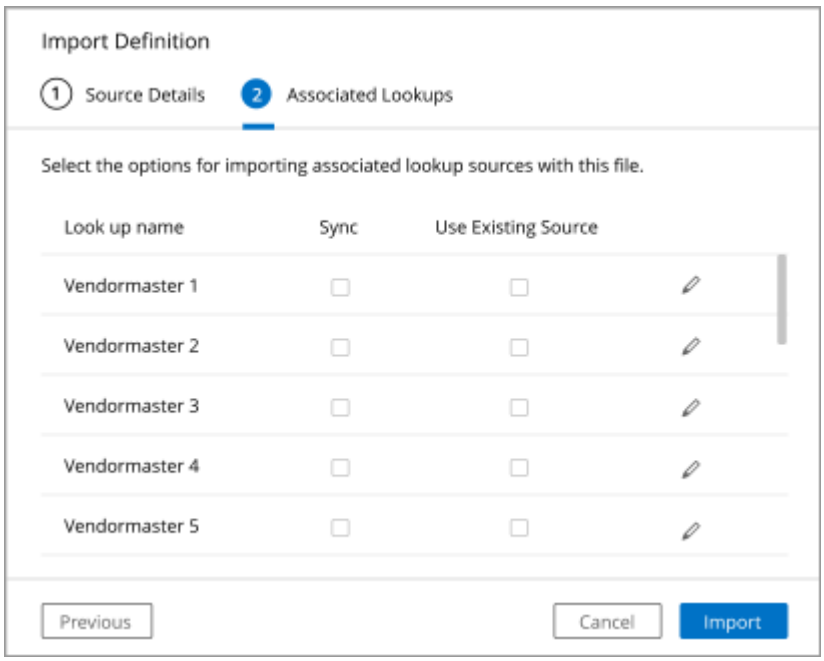
1. Select **Definition** from the menu bar. The list of definitions appears.



2. Click **Import Definition** in the upper-right of the Definition list screen. An Import Definition link appears in the middle of the screen if there is no definition. The Import Definition dialog appears.



3. Click the ... browse icon and select the definition script file.
4. Click **Next** to go to the Associated Lookups tab.
   It allows you to associate Lookup sources with the definition. You must select the options for importing associated lookup sources with the file.
5. Specify the following details for the required lookup:
   - **Sync** – Select the checkbox if you want to synchronize the associated lookup.
   - **Use Existing Sources** – Select the checkbox to use existing sources.
   - **Edit** – Click to modify the lookup.



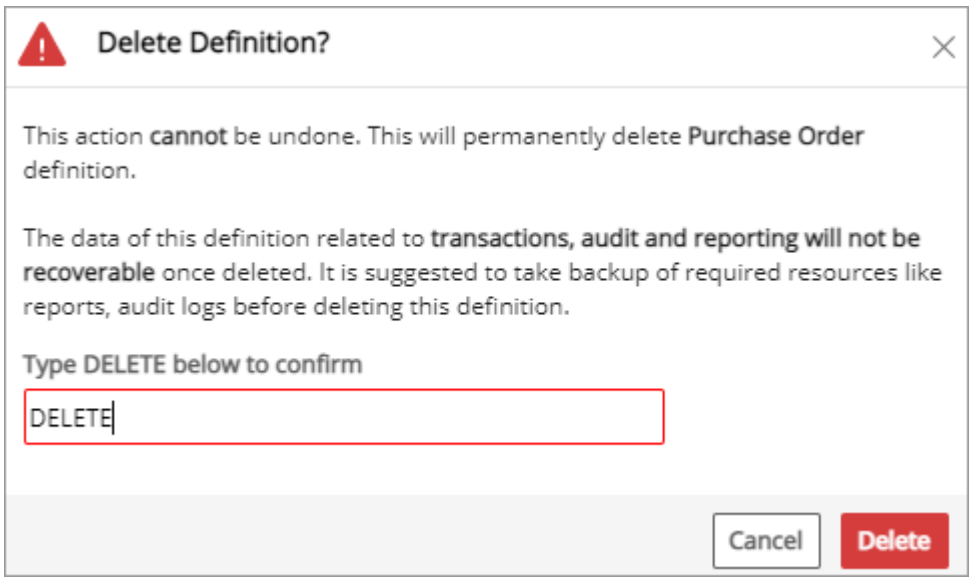
6. Click **Import**. The imported Definition appears.

Deleting a definition

When the definition gets deleted, you cannot recall or undo it. Also. you cannot recover transactions, audits, and reports of the deleted definition. If these are required, you can take their backups before deleting the definition.

To delete a definition, perform the following steps:

1. Go to the definitions list.
2. Click ⋮ (**More options**) against the required definition.
3. Select the **Delete** option. The Delete Definition dialog appears.



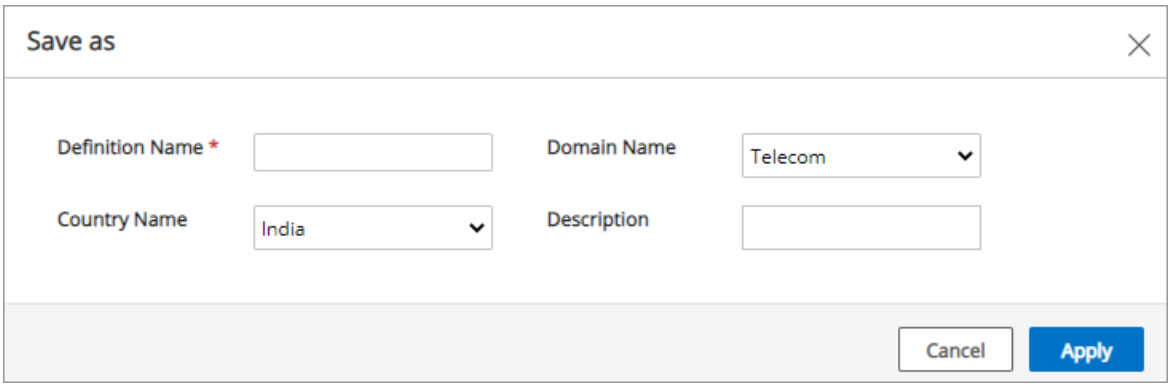
4. Enter **DELETE** in the box and click **Delete** to confirm. The selected definition gets deleted and removed from the list of definitions.

Save as

The Save as option allows you to create a copy of the existing definition with a different name.

To save the created definition with a different name, perform the following steps:

1. Go to the definitions list.
2. Click ⋮ (**More options**) against the required definition.
3. Select the **Save as** option. The Save as dialog appears.

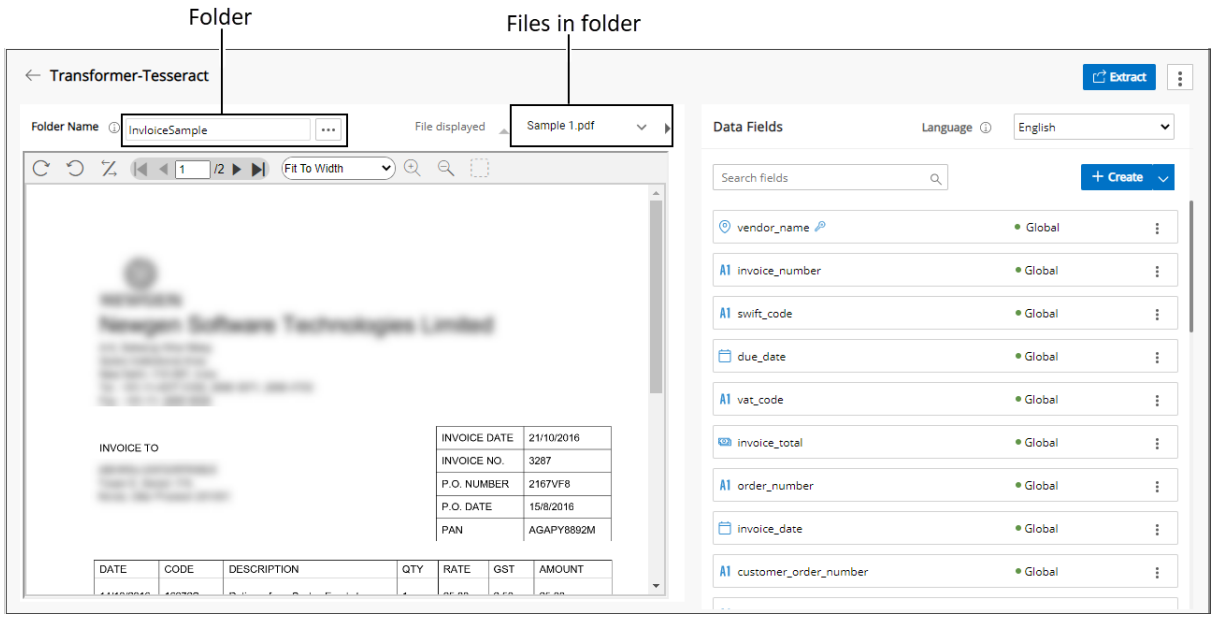


4. Modify the details as required and click **Apply**.

Previewing file

When there is a file or data to show up, its preview appears in the left pane. Though when there is nothing to show up, you can browse the file or folder. To add a file or folder, click the browse icon ••• against the Folder Name.

In the preview pane, you can see the folder name and the files that belong to that folder.



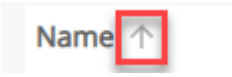
Sorting a definition list

Sorting allows you to sort the list of definitions based on Name, Created From, Field Count, Modified By, Modified On, and Description.

To sort the list of definitions, perform the following steps:

1. Go to **Definition** from the menu bar.

2. Click the up arrow against the column heading to sort the definition list in ascending order and the down arrow to sort the definition list in descending order.
   For example:

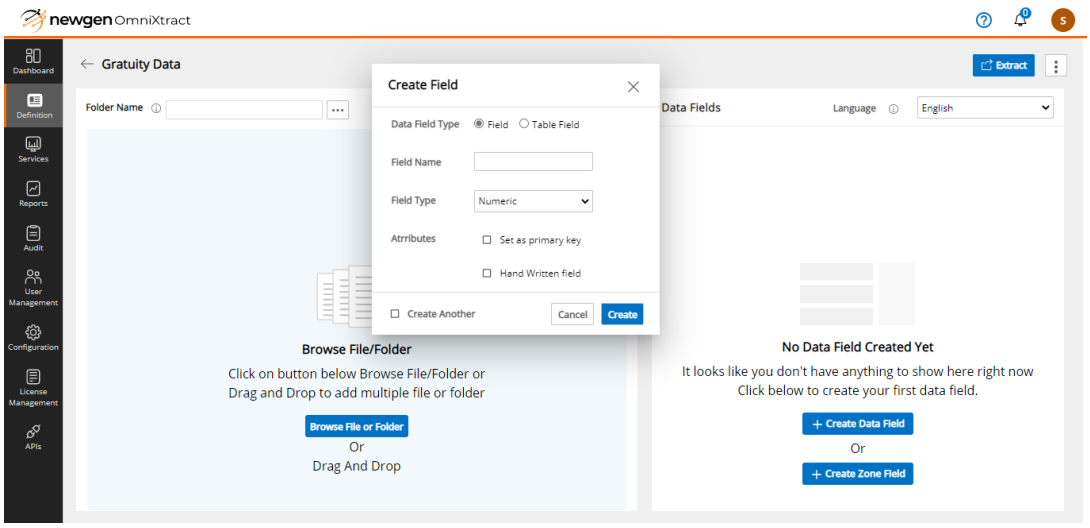


Working with fields

The Create Data Field option allows you to create a field to extract data from a particular image or document. Data fields associated with the definition are shown in the right pane. If there is no data field associated with the definition, you can create it by clicking Create Data Field. If data fields are already associated with the definition, you can add more data fields by clicking **Create** in the upper-right corner of the Data Fields pane.

The field option appears by following the below steps:

1. Click **Create Data Field** or **Create**. The Create Field dialog appears. By default, **Field** is selected.
   Select **Table Field** to create a table field, otherwise.
   The Create Field dialog appears. It shows two options for Data Field Types:
   - **Field** – To learn how to create a field, refer to the section Creating a field.
   - **Table Field** – To learn how to create a field, refer to the section Creating a table field.



Creating a field

Fields are created to specify data that requires extraction for further processing. An invoice or a purchase order contains many fields that together capture the details of the order. A field must be created as per the uploaded document for proper identification and extraction.

To create a new Field, perform the following steps:

1. In the Create Field dialog, enter the following details:
   - **Field Name** – Enter the name of the field.
   - **Field Type** – Select the field type. The supported field types are:
     - Barcode
     - Address
     - Email
     - Phone
     - Date
     - URL
     - Amount
     - Currency code
     - Alphanumeric
     - Numeric
     - Pattern
     - Strict alphanumeric
     - MICR
     - MRZ
     - OMR
     - Image clip
     - Non-recognition zone
     - Alphabetic
   - **Attributes** – Select the following attributes, as required:
     - Select the Setting a primary key attribute to create a unique identification field.
     - Select the Hand written field attribute if the field is of handwritten type.
2. Click **Create**.
3. Select **Create Another** to iterate the creation of the field. It minimizes the number of iterations performed while creating multiple fields.

Setting a primary key

The primary field within a document is a field that uniquely identifies the document. For example, the vendor name is the primary field for invoices, the company name is the primary field for purchase order documents. The primary field identification is always based on its master table.

The following factors affect the identification rate of the primary key:

- Customers must provide correct master details. The details must match with corresponding details printed on the image.
- The vendor master must contain details such as name, address, city, state, PIN, PAN/GSTIN, email, phone 1, phone 2, fax 1, fax 2, website, and other details better identification.
- The table column, marked as the primary key, must uniquely identify the vendor. For example, if the state is the same for all vendors, then you must not use it as Lookup Column for vendor identification.

To add a primary key, perform the following steps:

- The Global option appears selected by default.
- A field can have a global definition and a local definition specific to the primary key.
- The Local option is disabled until the Primary key is created.

Select **Set as Primary Key** to make the field primary. If the Primary key is selected, the Lookup tab appears in the Properties pane allowing you to associate it with a primary key. Consider the following points while defining a Primary key:

- Only an address field is allowed for the primary key.
- Only one field is allowed for a primary key.
- The Lookup source association is mandatory for the primary non-tabular field.
- Local definitions get saved against this primary field.

Once a primary key is added, the return value is picked from the attached database for primary fields.

Handwritten field

On selecting the following field types, the **Handwritten** checkbox gets enabled:

- Alphabetic
- Address
- Alphanumeric
- Amount
- Date
- Email
- Numeric
- Pattern
- Phone
- Strict Alphanumeric
- URL

Local and global definitions

A field can have a Global Definition and a Local Definition. The Global option is selected by default. Select the Local or Global option as required:

- Select the **Global** option to create a Global Definition.
- Select the **Local** option to create a Local Definition. A local definition can be created only after the vendor is identified after extraction.

> A field can have a global definition and a local definition specific to the primary key field.

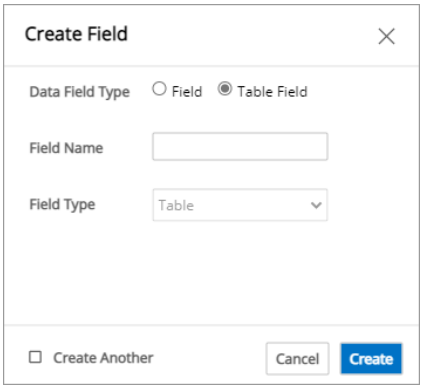
Creating a table field

The table field is used for the extraction of tabular data. Table detection and extraction happen automatically based on trained AI/ML models specific to the document type. If it fails or a model is not available for a specific document type, then a definition must be created manually.

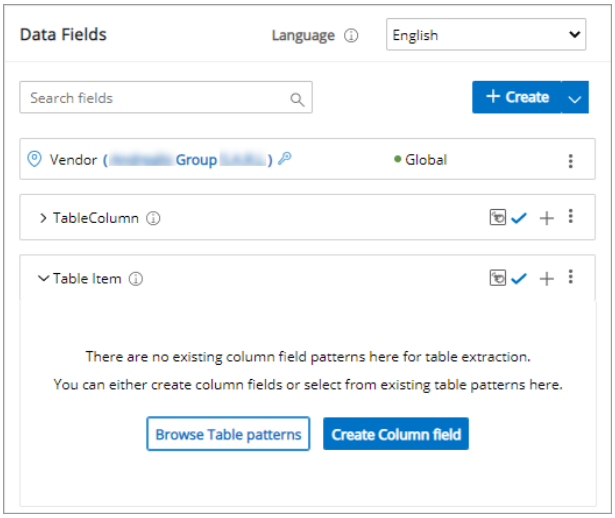
> The tabular model is supported only for invoices and not for other document types.

To create a new table field, perform the following steps:

1. Select Data Field Type as **Table Field** and provide the following details.
   • **Field Name** – Enter the field name.
   • **Field Type** – The field type for a table field is always a Table. It cannot be edited.
2. Select the **Create Another** checkbox to allow creating another field. It minimizes the number of iterations performed while creating multiple fields.
   Clear the checkbox if you do not want to create another field.

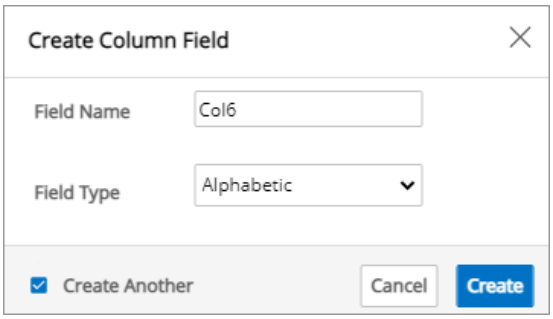


3. Click **Create**. A success message appears on the creation of the field.
4. Click the added table field to add table columns.



5. Click **Create Column Field** or + (**Add column**). The Create Column Field dialog appears.

Click **Browse Table Patterns** to view the table preview. The table preview gets generated only after running the extraction.



6. Enter the **Field Name** and select its **Field Type**.
7. Click **Create**.

Select the **Create Another** checkbox to iterate the creation of the field. It minimizes the number of iterations performed while creating multiple columns. Clear the checkbox if you don't want to create another column.

8. After adding the required column fields, click **Cancel** to close the dialog. Clicking Cancel to close the dialog is required only when you have selected the Create Another checkbox.

Editing a field

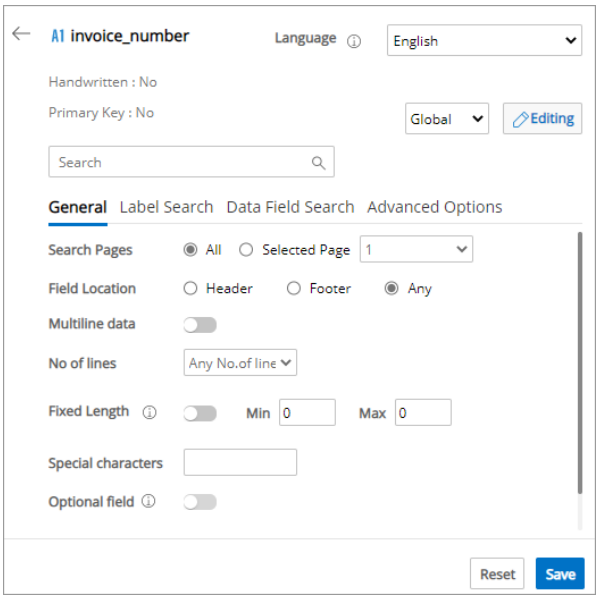
Editing allows you to modify the details of an existing field.

To edit a field, perform the following steps:

1. Open the required definition.
2. Open the field that you want to edit.
3. Click **Edit** to enable editing of the field. Each field has four tabs:
   - General tab
   - Label Search tab
   - Data Field Search tab
   - Advanced Options tab

Adding general details

In General Tab, necessary information is provided so that general data properties such as location, length, and presence of special characters if any can be extracted properly. Here, you can also specify the page number of the image from which the data must be extracted. The general tab allows you to apply basic-level settings to the field.



The following table describes more about the Fields and their description present in this tab:

Fields	Description
Search pages	Select **All Page** to define search on all pages of the document. Select **Selected Page** to specify the page number from which data extraction is required. The above-mentioned options are disabled, and the selected page is by default selected with the current page as the value of page number if any of the following types are selected: • Image Clip • Non-recognition Zone • OMR • Hand-written checkbox

Fields	Description
Field location	The following options decide the field location: • If **Header** is selected, then the search is performed in the top-half of the first page of the document. • If **Footer** is selected, then the search is performed in the bottom half of the last page of the document. • If **Any** is selected, then the search is performed across the image. The above options are disabled: • If the Table and Handwritten checkbox is selected. • On selecting Set Page. • If Currency Code, Image Clip, Non-Recognition Zone, and OMR type is selected.
Multiline data	Select this option when the field that requires extraction has data in multiple lines. On selecting, it enables the number of lines option to select between any of the below options: • Lines from 1-10, by default • Any number of lines. This option is enabled if any of the following types are selected: ◦ Address (by default selected) ◦ Alphabetic ◦ Alphanumeric ◦ Strict Alphanumeric • On selecting the Multi-line option, the Fixed and Special Characters fields are disabled.
Number of lines	Specify the number of lines required in the extraction.
Fixed length	Specify the fixed length of data of a particular length from the image.
Email address	It is available only when the Type is selected as **Email**. Enter the required **Account Name** and/or **Domain**, in this box.
URL address	It is available only when the Type is selected as the **URL**. Enter any specific Domain/ Extension.

Fields	Description
Special characters	Specify the special characters expected in that field. This option gets enabled if any of the following types are selected: • Alphanumeric • Numeric • Strict Alphanumeric • Alphabetic
Optional field	It is enabled for fields with table properties. Optional fields show that the field can or cannot have data.
Overlap in	It is enabled for fields with table properties. Select the **Left** option if the table field data that requires extraction overlaps in the left column. Select the **Right** option if the table field data that requires extraction overlaps in the right column. Select **None** if there is no overlapping of data.

Getting label variations

The Label Search allows you to get the label variations from the user in the following ways:

- Manually typing
- Importing labels from a text file
- Creating an area from where the label must be extracted.

It ensures that the various representations of the field name are captured. For example, the invoice number might appear as Invoice Number, Invoice No., Inv. No., Invoice, INVOICE No., and similar other representations. So, all the representations must be added as labels.

Label options:

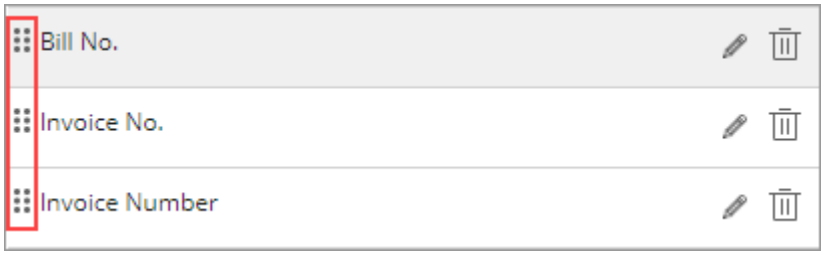
Lable options	Description
Adding a label	To add a new label, enter the label in the label box. Click **Add**.
Importing a label	To import a new label, click **Import** and browse the file.

Lable options	Description
Drawing a label	To draw a new label, click **draw**.
Section Header	
Name	Name of the section header.
Margin	Set margins of the header section. It is used to determine the position of the data that requires extraction by locating different labels. It is enabled for a local definition only. For a global definition, the margin is always set to default.
Add	Click **Add** to add the entered section header.
Import	It is used to import the section header by importing a text (.txt) file containing different labels.
Draw	It is used to add a section header by drawing a zone on the document.
Section Footer	
Name	Name of the section footer.
Margin	Set the margins of the footer section. It is used to determine the position of the data that requires extraction by locating different labels. It is enabled for a local definition only. For a global definition, the margin is always set to default.
Add	Click **Add** to add the entered section footer.
Import	It is used to import the section footer by importing a text (.txt) file containing different labels.
Draw	It is used to add a section footer by drawing a zone on the document.
Icons	
Edit icon	The edit icon against the label name allows you to edit the label details. In the case of a local definition, you can edit both the label name and margin. In the case of a global definition, you can edit only the label name.
Delete icon	The delete icon against the label name allows you to delete the added label.

Lable options	Description
Export	The export icon , appearing in the Label List, Section Header, and Section Footer, allows you to download the defined labels. You can use these downloaded or exported labels to create labels in other data fields.

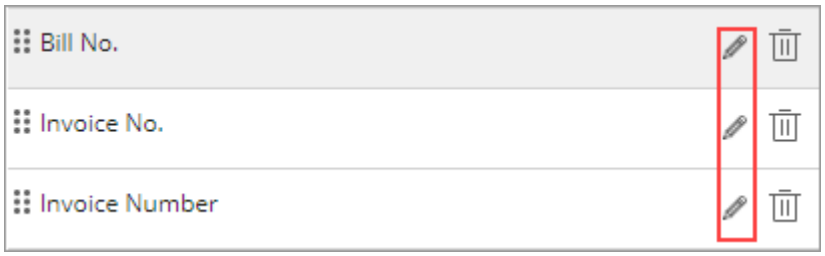
Ordering labels: Ordering labels allow you to set the sequence of labels that must be searched for data extraction. To order labels, perform the following steps:

1. Click the icon associated with the required label.
2. Drag and drop it at the required position.



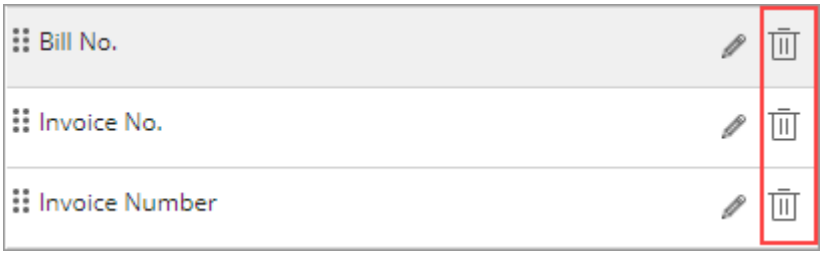
Editing labels: You can edit the labels at the Label Search tab. To do this, perform the following steps:

1. Select the required label to edit it.
2. Click the **Edit** icon. Edit mode gets enabled where you can specify the label.



Deleting labels: You can delete labels at the label search tab. To do this, perform the following steps:

1. Select the label to delete it.
2. Click the **Delete** icon.



Defining Data Position: Data Position is used to define the location of the data that requires extraction against the selected label. For example, on one label, data might appear on the right side, while on another label; data might appear on the left side. Declaration of data position aids in faster data extraction.

The Data Position option gets enabled only for the Local Definition.

The following table describes the meanings of the different data positions:

Data Position	Description	Example
Default	The position is defined in the *IPS.config* file.	
Right	It means that data is located on the right side of the selected label.	**Invoice** 443914
Bottom	It means that data is located just below the selected label.	**Tax Invoice** ABN 15 364 527 724
Left	It means that data is located on the left side of the selected label.	443914 **Invoice**
Top	It means that data is located just above the selected label.	443914 **Invoice**

Adding data field search details

The Data Field Search tab allows you to add validations, lookups, and define the output format of the field.

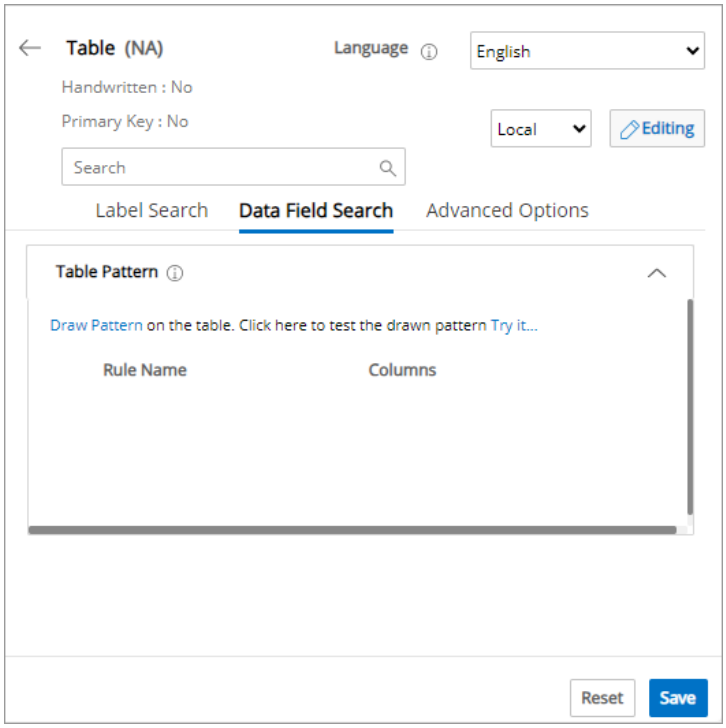
The Output Format is used to add words or symbols that must be removed from the output data (the data saved in the database after extraction). You can also define a format in which you want to save the extracted data in the database.

For example, the date of the invoice might follow a pattern of dd/mm/yy, mm/dd/yyyy, mm/dd/yy, and other similar patterns, but you want to save the extracted dates in a pattern that is common for all. So, if the output format for the date is defined as mm/dd/yyyy, then all the dates get saved in this format only irrespective of the different formats of the extracted dates. For more details on patterns, refer to the section Patterns.

For a tabular field:

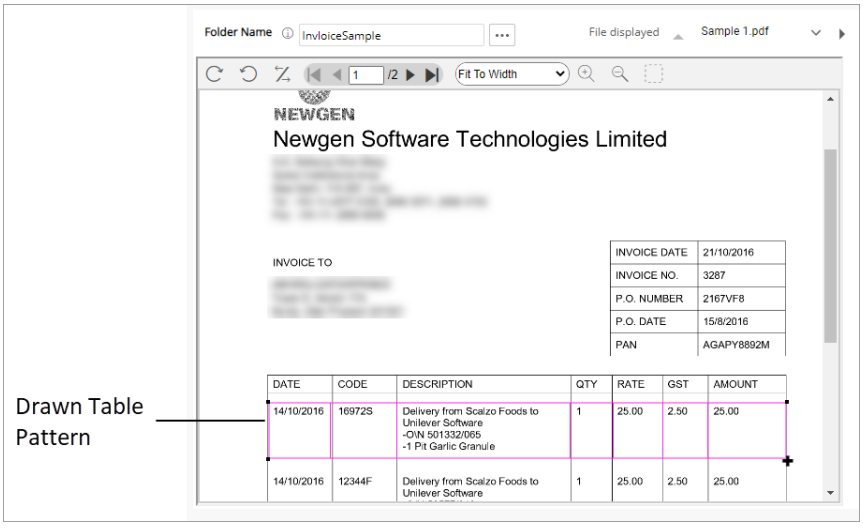
To edit the Data Field Search tab and define manual extraction settings for a Table Field, perform the following steps:

1. Open the required definition.
2. Click **Extract** in the upper-right of the definition screen to perform extraction and identify the vendor details.
3. Open the required table field.
4. Click **More Options** and select **Edit details**. The details of the selected table appear in read-only mode.
5. Click **Edit** to edit the table details.
6. Make the field **Local** if it is a Global field. A field can be made local only if the vendor is identified.
7. Click the **Data Field Search** tab and expand the table pattern section.



8. Click the **Draw Pattern** link. The link gets enabled only if the vendor is identified.
9. Draw a rectangular box over the required table columns.
10. Move your mouse pointer to the end of the first column.
11. Right-click and select **Draw Zone Partition** to define a partitioning area.
    You can adjust the position of the drawn partition line by selecting and moving it left or right using your mouse, as required.
12. Similarly, draw zone partitioning for other columns.

> To delete the drawn pattern and partitioning, right-click on it and select **Delete**.



13. Click the **Try it** link to extract and view table columns as per the drawn partitions.
14. Click **Save** to save the modified table field definition.

For a non-tabular field: To define a data field search on an image, perform the following steps:

1. Open the required definition.
2. Click **More Options** against the required data field and select **Edit details**. The details of the selected data field appear in read-only mode.
3. Click **Edit** to modify the details.
4. Click the **Data Field Search** tab. Depending on the Field Type of the selected data field, the following options appear:
   - Pattern – Refer to the section Patterns for details.
   - Validations – Refer to the section Validations for details.
   - Lookup – Refer to the section Lookup for details.
   - Output Format – Refer to the section Output Format for details.
5. After defining the above details, click **Save** to save the changes made to the field.

Adding patterns

The Pattern option appears in the Data Field Search tab for the Pattern type data field.

A pattern is used to specify the pattern in which the field data appears in the image. Pattern matching aids in accurate, fast, and reliable data extraction. In OmniXtract, a pattern can be generated using a combination of A N X * C keys. Here:

- **A** refers to Alphabetic.
- **N** refers to Numeric.
- **X** refers to Alphanumeric.
- ***** refers to Any. It refers to something that might be an Alphabetic, Numeric, or Special Character.
- **C** refers to Constant. On clicking the C button, a dropdown list appears. Click on any character to add it to the pattern box.

For example – if an invoice number is ABC12345/001M for a vendor and the vendor follows an arrangement where the invoice number is 13 digits, having the first three positions as alphabets followed by the next five as numbers followed by a '/' followed by next three positions as numbers followed by the last position as Alphabet, then its

pattern can be AAANNNNN*NNNA, where A refers Alphabetic, N refers Numeric, / refers the selected special character.

A unique number of five characters starting with MZ followed by three numeric characters, for example, MZ985 the pattern can be [M][Z]NNN. Wherein, [M] represents that the first position will always have M. Likewise, [Z] represents that the second position will always have Z in the pattern. N in the pattern stands for a numeric.

The Pattern option is available for:

- **Global Definition** – When the type is selected as Pattern.
- **Local Definition** – When type is selected as any of the following:
  - Alphabetic
  - Alphanumeric
  - Numeric
  - Pattern
  - Strict Alphanumeric

Date pattern

The Date pattern allows you to define vendor-specific date patterns for dynamic detection and extraction. The input date pattern allows you to define the output date pattern which formats the extracted date pattern into the required output format.

A date pattern can be generated using a combination of DMY and C keys. The different representations of D, M, and Y are:

- **D** denotes a day value between 1 and 31.
- **DD** denotes a day value between 01 and 31.
- **DDD** denotes a short representation of a day. Example – Sun, Mon.
- **DDDD** denotes a verbose representation of a day. Example – Sunday, Monday.
- **M** denotes a month value between 1 and 12.
- **MM** denotes a month value between 01 and 12.
- **MMM** denotes a short representation of a month. Example – Jan, Feb.
- **MMMM** denotes the verbose representation of a month. Example – January, February.
- **YY** denotes a short representation of the year of the date. For example, 01/01/19 represents DD/MM/YY.
- **YYYY** denotes the full representation of a year of the date. For example, 01/01/2019 represents DD/MM/YYYY.

- **C** refers to constant. On clicking the C button, a dropdown list appears. Double-click the required character to select it.

Time pattern

Time pattern allows you to define vendor-specific time patterns for dynamic detection and extraction.

To define a time pattern, perform the following steps:

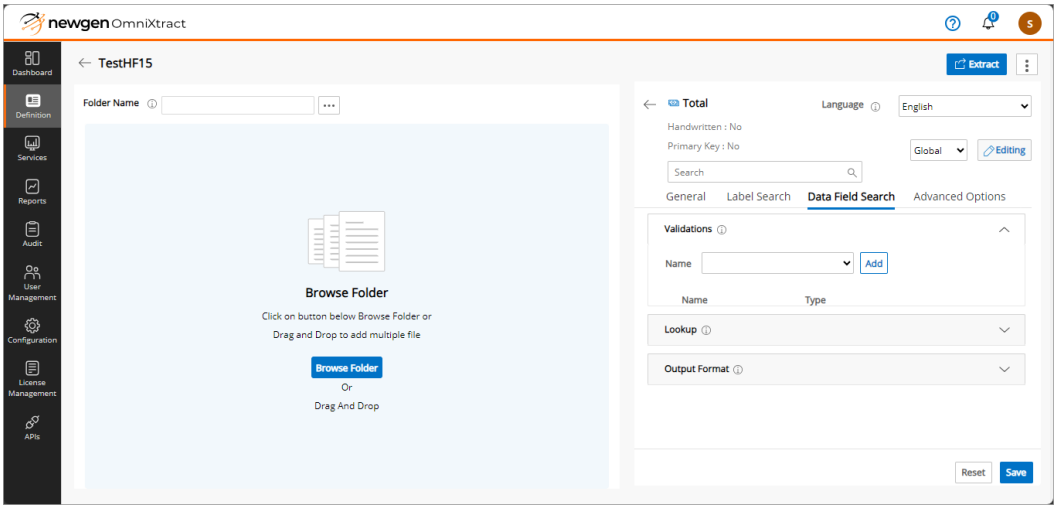
1. Click **Time** to select a time format. A dropdown list containing different time formats appears.
2. Click the desired format to include it in the date pattern.
   - **h** means a 12-hour clock time. Example – 4.
   - **hh** means a 12-hour clock time, with a leading 0 when the hour is in a single digit. Example – 06.
   - **H** means a 24-hour clock time. Example – 15.
   - **HH** means a 24-hour clock time, with leading when the hour is in a single digit. Example – 02.
   - **M** means minutes.
   - **mm** means minutes, with a leading zero when the minute is in a single digit.
   - **s** means seconds.
   - **ss** means seconds, with a leading zero when the second is in a single digit.
   - **t** is used for abbreviated AM or PM. Example – A or P.
   - **tt** is used for AM or PM. Example – AM or PM.

Associating validations

Associate validations are helpful for any field to ensure data correctness after extraction. You can use it to ensure that extracted data remains error-free and used efficiently.

To associate validations to a field, perform the following steps:

1. Select the required validation from the **Name** dropdown list. It contains all the existing Validations defined previously. To learn how to define a Validation, refer to the section Working with validations.
2. Click **Add**.



Associating lookup sources

It allows you to associate the lookup source to a field and specify the column against which matching must happen.

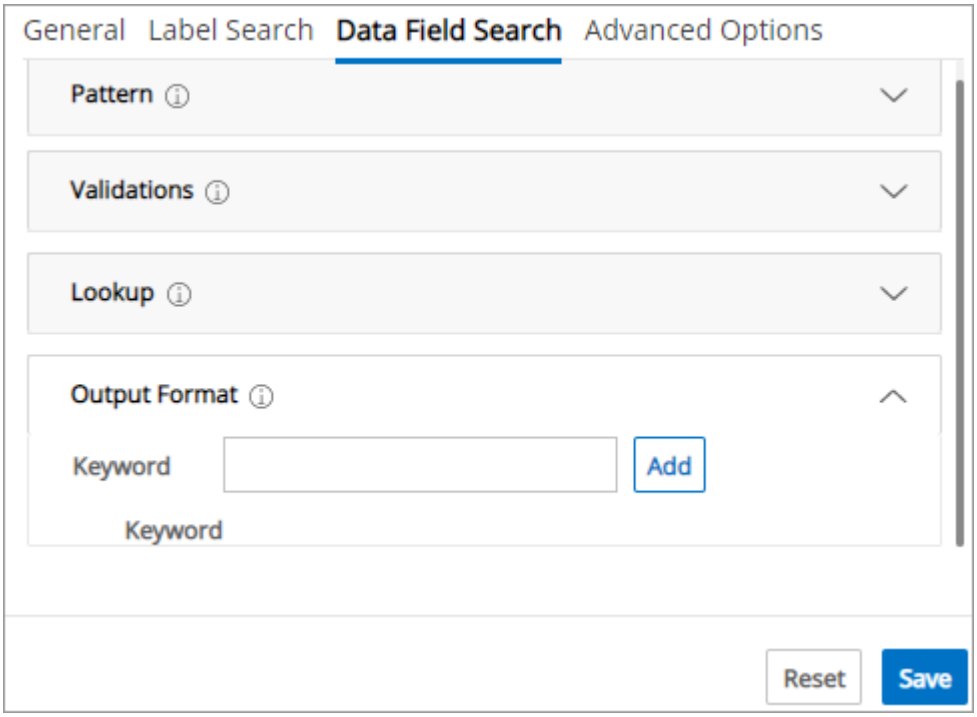
To associate Lookup Source to a field, perform the following steps:

1. Select **Search Type** from the following options:
   - **Exact** – It exactly matches the defined parameters.
   - **Fulltext** – It matches the whole text parameters that are defined.
   - **Fuzzy** – It matches the parameters partially or as per the defined phrases. It is faster than others, but the only drawback is that it doesn't filter the exact text.
2. Select **Search Priority** from the following options:
   - **FTS** – FTS or Full-Text Search searches for the complete defined text.
   - **Regex** – It searches for the text defined using regular expression.
   - **Hybrid** – It is the combination of FTS and Regex.
3. Select the **Source Name**. This dropdown list contains the existing lookup sources. To learn how to define a lookup source, refer to the section Lookup sources.
4. Select the columns against which matching must happen. On successful matching, the value of the selected Return column is given.
5. Select the Filter column to further refine the database search using the selected search options in case multiple records are matched against the selected lookup column.

Defining output format

The Output Format allows you to add words or symbols that must be removed from the output data (the data saved in the database after extraction). It also allows you to define a format to save the extracted data in the database.

For example, the date of the invoice might follow a pattern of dd/mm/yy, mm/dd/yyyy, mm/dd/yy, and other similar patterns, but you want to save the extracted dates in a pattern that is common for all. So, if the output format for the date is defined as mm/dd/yyyy, then all the dates get saved in this format irrespective of the different formats of the extracted dates.



To add a keyword, enter the **Keyword** and click **Add**. The added keyword gets added to the list of keywords.

Defining advanced options

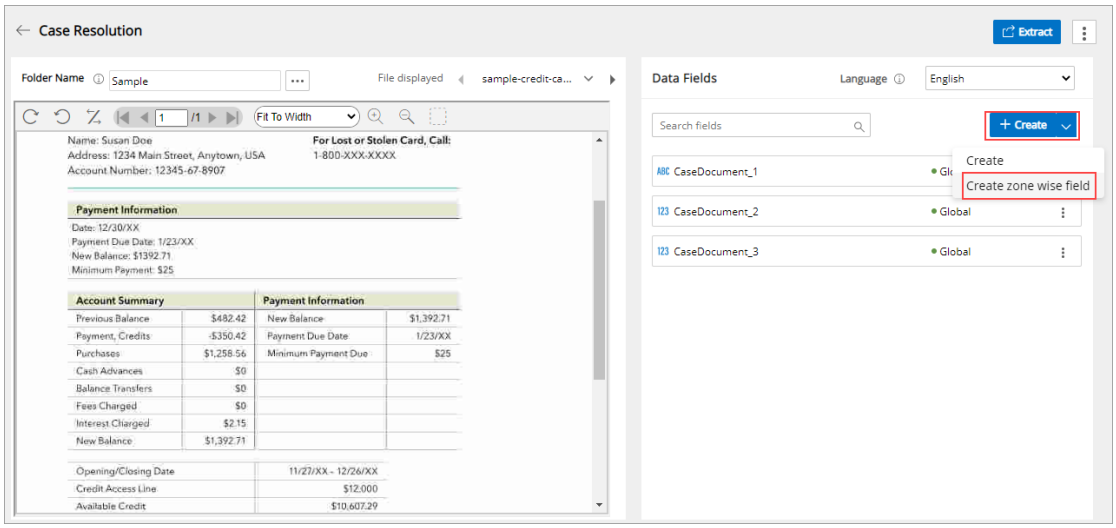
The Advanced Options tab allows you to specify the accuracy analysis and pre-processing settings of the field.

Creating zone-wise field

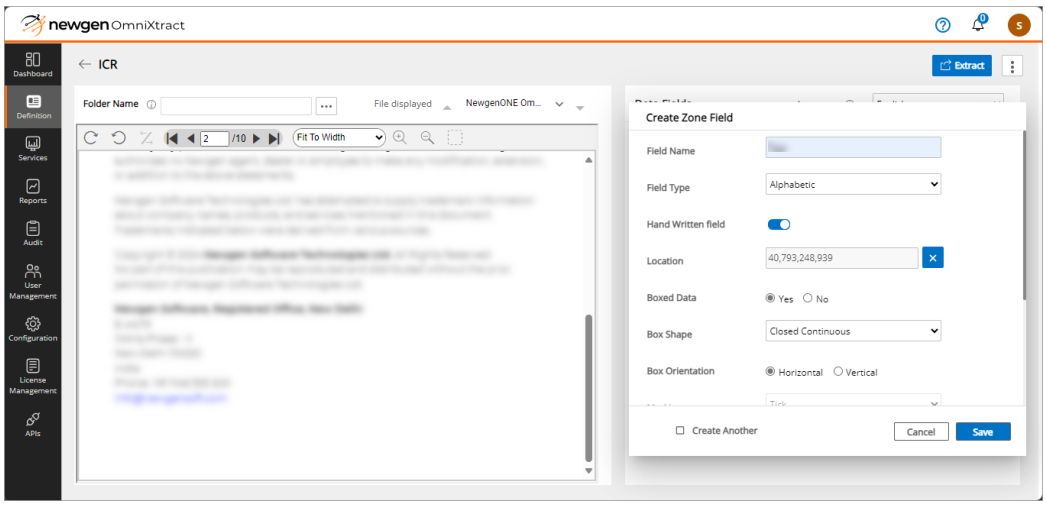
Zone fields are created on structured documents. It allows you to mark a field for extraction by drawing a zone.

To create a new zone field, perform the following steps:

1. Open the required definition.



2. Click the **Create** dropdown in the Data Fields section.
3. Select **Create zone wise field**. The Create Zone Field dialog appears.
4. Enter the following details:
   - **Field Name** – Enter the name of the field. The character limit for entering a field name is 50.
   - **Field Type** – Select the data type of field.
5. Click **Set Zone Properties** to specify additional properties of the zone. Some additional fields appear.
6. Enter the following details:
   - **Location** – It's the coordinate of the data on the uploaded file. To add a location, click the + and draw a zone around the data that requires extraction. Once the zone is created, the coordinates are automatically filled in the Location box.



- **Boxed Data** – It's to specify whether the data that requires extraction is inside a box or not.
    - Select **Yes**, if the defined data is inside a box.
    - Select **No**, if the defined data is not inside a box.
- **Box Shape** – Click the dropdown option and select one of the following options:
    - Closed Continuous
    - Comb Continuous
    - Closed Discrete
    - Comb Discrete
    - Rectangle
- **Box Orientation** – This is primarily used for OMR options. Click the drop-down option and select one of the following options:
    - **Horizontal** – If the box orientation is horizontal.
    - **Vertical** – If the box orientation is vertical.
- **Marking** – It is enabled in the case of OMR.
    - For the box shape, the OMR selection can be **Ticked** (✓) or **Crossed** (x).
    - For ellipses and circles, the OMR selection must be **Filled**.

Refer to the following table to know more about the Fields and their description:

Field	Description
Add zone	It allows you to create a zone around the data that requires extraction.
Delete the existing	Click on this button to delete the defined zone coordinates.

Field	Description
Location	This non-editable box consists of the Defined Zone location.
Boxed	Click on the dropdown option and select one of the following options: • **Yes** – If the defined data is inside a box. • **No** – If the defined data is not inside a box.
Box Shape	Click on the dropdown option and select one of the following options: • Closed Continuous • Comb Continuous • Closed Discrete • Comb Discrete • Rectangle
Box Orientation	Click on the dropdown option and select one of the following options: • **Horizontal** – If the Box Orientation is horizontal. • **Vertical** – If the Box Orientation is vertical.
Box Selection	Click on the dropdown option and select one of the following options: • **Single** – If one cell of the box contains a single value, then select single. • **Multiple** – If one cell of the box contains multiple values, then select multiple.
Marking	It is enabled in the case of OMR. For a rectangle, OMR selection can be taken as **Ticked** and **Crossed**. For ellipses and circles, the selection can be **Filled**.

Field	Description
Value	It is enabled in the case of OMR. For each OMR marking variable, select and give the value that the respective variable represents. You can assign a value from the dropdown list or enter your custom value in the box. The dropdown list contains the following options: • 0-9 – It assigns sequential numerical values, starting from zero till 9, to the marking variable. • A-Z – It assigns sequential capitalized Alphabetical values, starting from A till Z to the marking variable. • A-Z, 0-9 – It assigns both sequential capitalized Alphabetical and Numerical values, starting from A till Z and 0 till 9, to the marking variable. • a-z – It assigns sequential lowercase Alphabetical values, starting from a to z, to the marking variable. • a-z, 0-9 – It assigns sequential lower case Alphabetical and Numerical Values, starting from A till Z, and 0-9 to the marking variable.
Delimiter	It is enabled for the OMR field type. Give the separator for the extracted OMR values in the delimiter box.

Refer to the following table to understand the availability of properties for different field types:

Field	Location	Boxed	Box Shape	Box Orientation	Box Selection	Marketing	Value	Delimeter
Address	Yes	Yes	Yes	Yes	Yes	No	No	No
Alphabetic	Yes	Yes	Yes	Yes	Yes	No	No	No
Alphanumeric	Yes	Yes	Yes	Yes	Yes	No	No	No
Amount	Yes	Yes	Yes	Yes	Yes	No	No	No
Barcode	Yes	No	No	No	No	No	No	No

Field	Location	Boxed	Box Shape	Box Orientation	Box Selection	Marketing	Value	Delimeter
Currency Code	No	No	No	No	No	No	No	No
Date	Yes	Yes	Yes	Yes	Yes	No	No	No
Email	Yes	Yes	Yes	Yes	Yes	No	No	No
Image Clip	Yes	No	No	No	No	No	No	No
MICR	Yes	No	No	No	No	No	No	No
MRZ	Yes	No	No	No	No	No	No	No
Non-Recognition Zone	Yes	No	No	No	No	No	No	No
Numeric	Yes	Yes	Yes	Yes	Yes	No	No	No
OMR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pattern	Yes	Yes	Yes	Yes	Yes	No	No	No
Phone	Yes	Yes	Yes	Yes	Yes	No	No	No
Strict Alphanumeric	Yes	Yes	Yes	Yes	Yes	No	No	No
URL	Yes	Yes	Yes	Yes	Yes	No	No	No

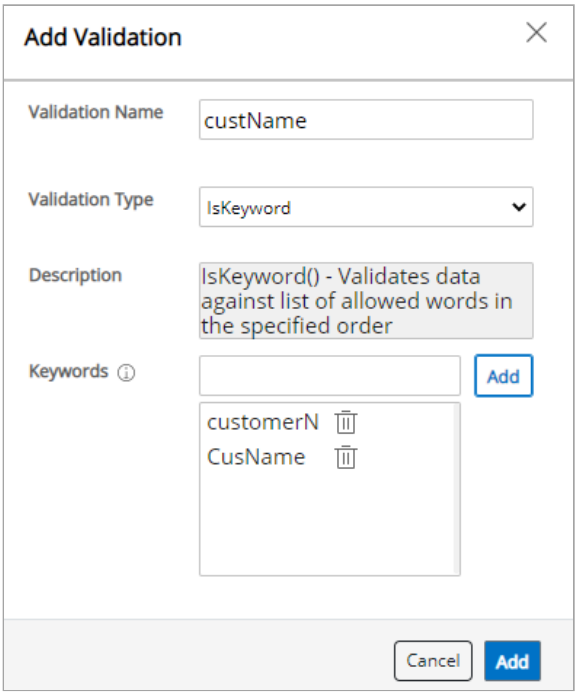
7. Select the **Create Another** checkbox to allow creating another field. It minimizes the number of iterations performed while creating multiple fields.
   Clear the checkbox if you don't want to create another field.
8. Click **Create**.
   Once the box properties are saved, you can update them by clicking the View and Update link.

Working with validations

Validations are associated with the fields to ensure the correctness of the extracted data. For example, the *IsNumeric* function checks whether the data is a valid numeral with the relaxation of specified special characters.

To associate validations to a Field, perform the following steps:

1. Open the definition where you want to add validation.
2. Select **Validations** from the ellipsis in the upper-right. The Validations screen appears.
   If there are no validations, The Add New Validation button appears in the middle of the screen.
   If some validations are already created, The Add New Validation button appears in the upper-right of the screen.
3. Click **Add New Validation**. The Add Validation dialog appears.



4. Specify the following details:
   - **Validation Name** – Name of the validation
   - **Validation Type** – Data type of validation
   - **Description** – Description of the validation
   - **Keywords** – Keywords to identify validation. Enter the keywords first and then click **Add** to add that keyword to the list of keywords.
5. Click **Add**. Added validation appears.
   You can **View or Edit Details** and **Delete** the added validation.
6. Go to the validation and click the ellipsis to perform the desired action.

Image processing

Image processing in OmniXtract refers to a technique that involves applying certain procedures to an image with the aim of obtaining an improved image or extracting valuable information from it.

Image processing can be done at:

- Definition level
- Field level

At the global level, image processing allows two libraries to process images – OpenCV and NewgenIpLib. The following options are available for image processing:

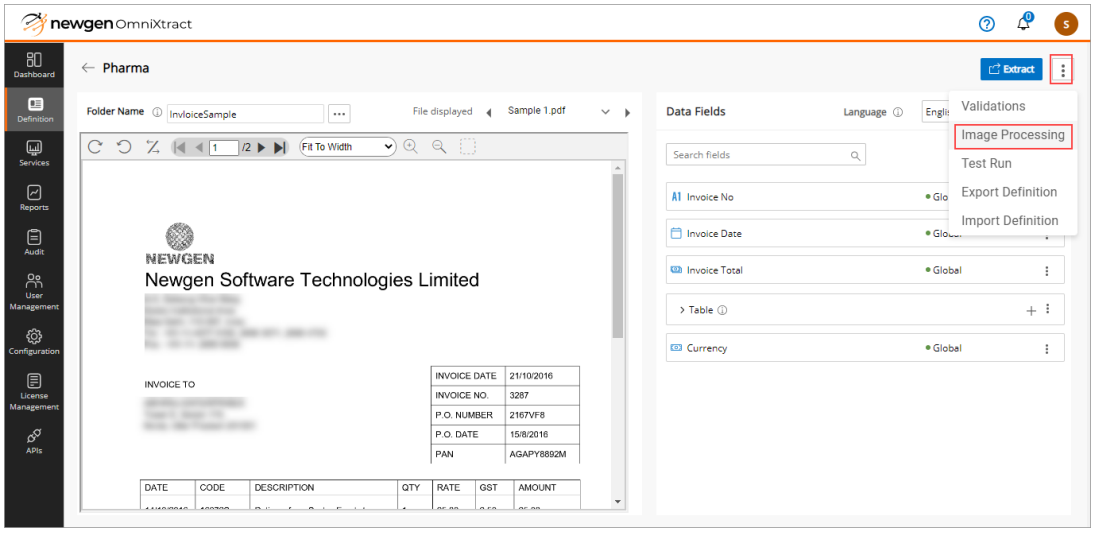
- NewgenIplib
  - ConvertAnyToBW
  - Adaptive Binarization
  - Auto Deskew
  - Remove NoiseEx
  - AutoDetect Oriertation
  - AutoDeskew ByLines
  - Remove Tick
  - Detect RemoveLines
- OpenCV
  - Auto Deskew
  - ColourToBW
  - Noise Removal
  - Line Removal
  - Deskew
  - Gaussiar Blur
  - Median Blur
  - Threshold
  - Orientation
  - Perspective Correction
  - Gamma Correction
  - Binarize NativeAdaptive
  - Denoise
  - SaltPepper

- RemoveLines

Image Processing at the Definition Level

To configure settings for image processing, perform the following steps:

1. Open the definition that requires image processing.



2. Click **More options** and select **Image Processing**. The Image Processing screen appears.
3. Select the **Page Number** option as **All** or **Selected Page**.
   - **All** – When All is selected, the image processing settings are applied to all the pages of the image.
   - **Selected Page** – When the Selected Page is selected, the image processing settings are applied to only the selected pages.
4. Click **Add Rule**. A row containing Rule Name and Library Name gets added.
5. Select the **Rule Name** and the corresponding **Library Name**.
6. Click ✔ to save the selection.
   Add more rules as required.
7. Click **Save** to save the defined settings. In addition to the Save button, the following buttons are also given:
   - **Reset Changes** – It resets the field values to the last saved state.
   - **Preview** – It generates the image preview based on the defined settings.
   - **Cancel** – It cancels the changes made to the image processing settings and closes the Image Processing screen.

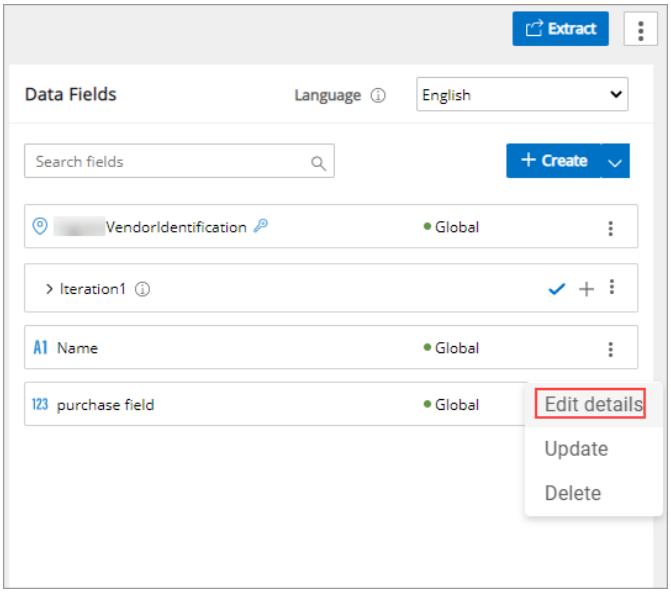
Actions on the added rules – The following actions are available on the added rules. These actions appear on clicking the rule name.

Action	Description
Reset	It resets the field values to the last saved state.
Preview	It generates the image preview based on the selected rule and library name.
Delete	It deletes the rule.
Expand	It expands the rule rows to show additional settings. It appears for only those rule libraries where additional settings are provided.
Collapse	It collapses the expanded rule rows.

Image Processing at the Field Level

To configure settings for image processing, perform the following steps:

1. Open the required definition.
2. Click ⋮ (**More Options**) against the required field.



3. Select **Edit details**. The field definition appears.
4. Click Edit. The field definition appears in the edit mode.

5. Open the **Advanced Options** tab. The remaining steps are the same as those described for the Definition level.
6. You can also rearrange the added rules by dragging and dropping the rule through .
7. Click **Save** to save the image processing settings.

Test run

Test Run is used for extraction on multiple files in one go. It performs data extraction from the images sequentially from the uploaded folder. The extraction results are displayed one by one for all the images.

To perform Test Run, perform the following steps:

1. Open the required definition.
2. Click (**More options**) in the upper-right of the opened definition screen.
3. Select **Test Run**. The test run gets started and the extraction details are displayed under Field Details comprising Key Fields and Table fields tabs.

Extraction

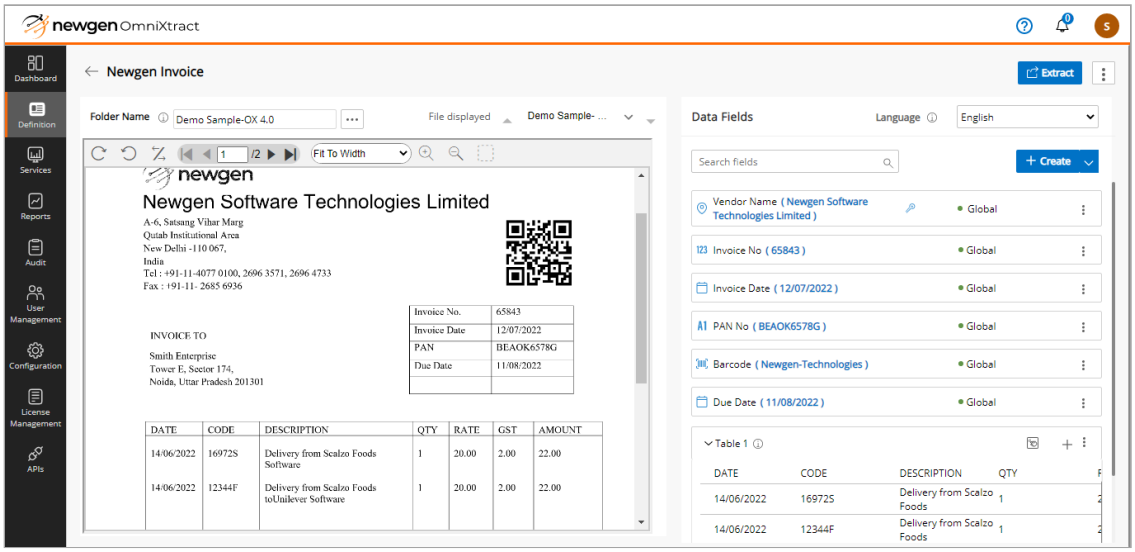
Extraction is used to extract data from non-tabular as well as tabular fields. It is performed after adding the fields and their properties are defined. The extracted data is displayed against the respective fields appearing in the fields' pane.

Data extraction is done once the required fields are added and their properties are defined. Based on the defined fields, data can be extracted in one go. Data extraction of all the fields is done simultaneously.

To extract data from all the fields simultaneously, perform the following steps:

1. Select **Definition** and then click **Extract**.
   The data of all the fields are extracted and appear below the respective fields.
2. Hover over any extracted data to view its location in the image. The selected data gets highlighted in green color, while the label that was used for extraction is highlighted in yellow color.

On processing table fields, the tabular area gets highlighted with dotted lines, and the first row and first column are highlighted by default. You can traverse to the next rows on the popup screen.



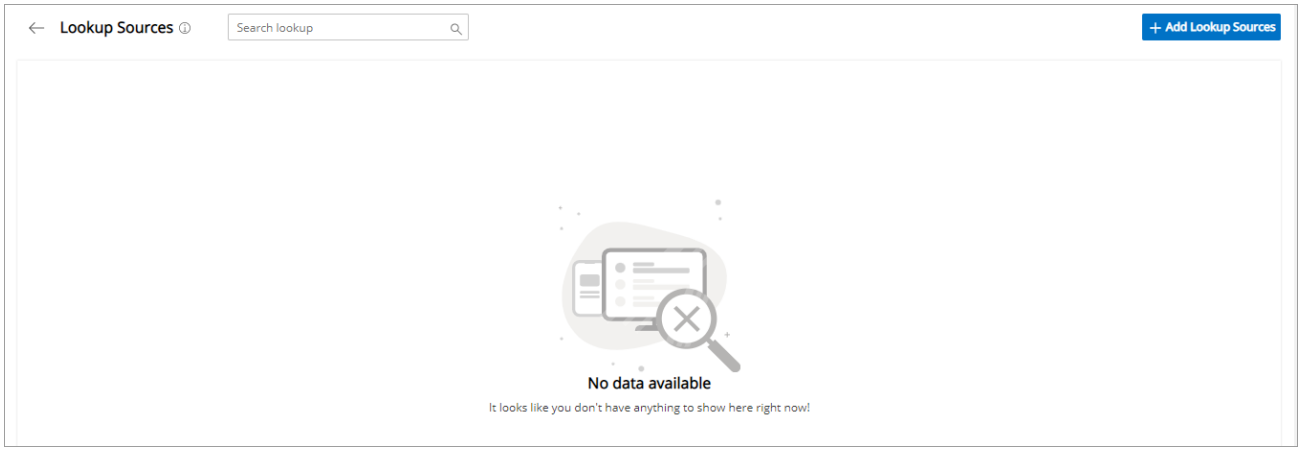
Lookup sources

It allows you to create Lookup sources from your master tables present in SQL, Oracle, or PostgreSQL databases. These are required for the Primary Field and other field types where an extraction must happen using a lookup match with master tables.

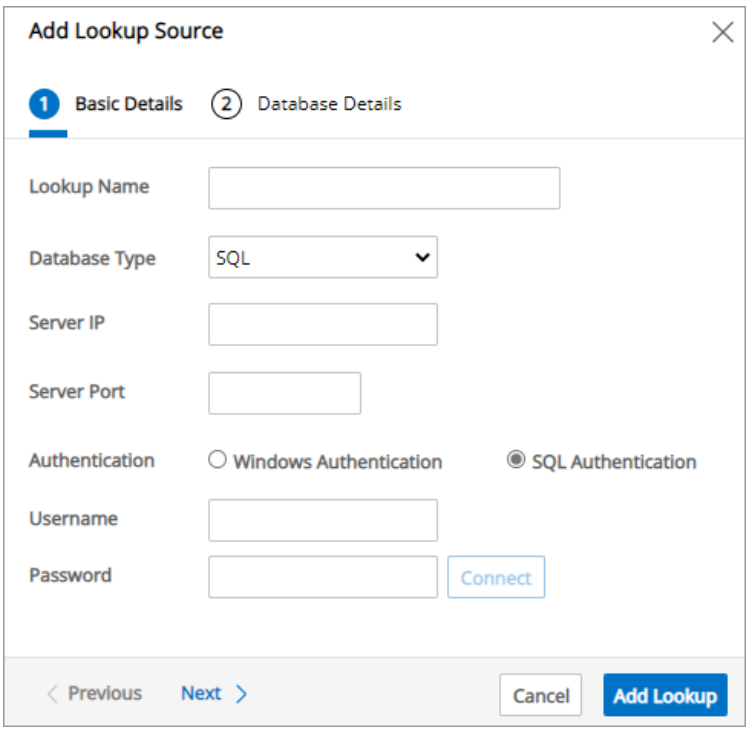
Adding a lookup source

To add a new lookup source, perform the following steps:

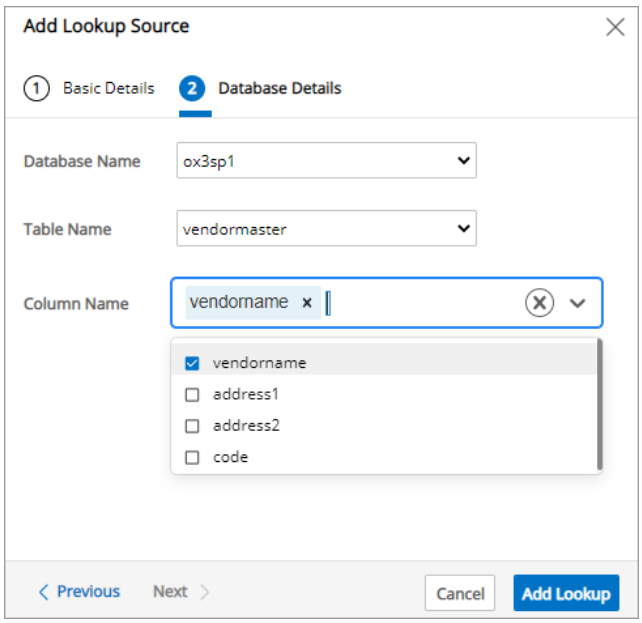
1. Go to the **Definition** list and click **Lookup Sources**. The Lookup Sources screen appears.



2. Click **Add Lookup Sources**. The Add Lookup Source dialog appears.



3. Enter the following **Basic Details** details:
   - **Lookup Name** – Specify the lookup name.
   - **Database Type** – It shows the type of database such as SQL/NPGSQL/ORACLE. If the database is selected, then it asks to add the server details of the database.
   - Specify the server details including **Server IP**, **Server Port**, **Service Name** (for Oracle only), **Username**, and **Password**.
     Ensure the database is connected to the server. If the database is not connected, then click **Connect** to connect to the database server. The connection gets established if the specified database details are correct.
4. Click **Next**. The Database tab appears. It allows you to associate the Lookup Source. You are required to select the options you will use whilst importing associated lookup sources with the file.



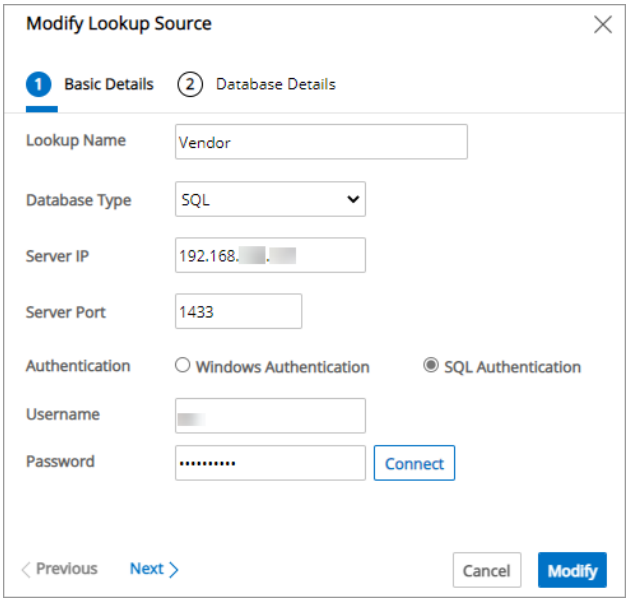
5. Enter the following details.
   - **Database Name** – Select the required database name from the dropdown list.
   - **Table Name** – Select the required table name from the dropdown list. It shows the name of the table from where the records are retrieved. It must have a unique key (based on single or multiple columns) defined and must not have any auto-increment column.
   - **Column Name** – It allows you to add the columns from where data is extracted. The Select Columns section lists all the columns defined in the selected table. Select the columns as per the requirement.
     The **Database** and **Table** fields get enabled once the connection is established.

6. Click **Add Lookup** to add the new Lookup Source. The defined Lookup Source gets added to the Lookup Sources List. The following details are shown:
   - **Database Name** shows the name of the database.
   - **Database Type** shows SQL/NPGSQL/ORACLE.
   - **Database IP** shows the IP address on which the selected MSSQL/Oracle database is running.
   - **Table Name** shows the name of the table selected as the lookup source.
   - **Fields** show the column names selected from the lookup table.

Editing a lookup source

To edit an existing lookup source, perform the following steps:

1. Go to the **Lookup Sources** list.
2. Click ⋮ (**More options**) against the required lookup source.
3. Select **Edit Details**. The Modify Lookup Source dialog appears.



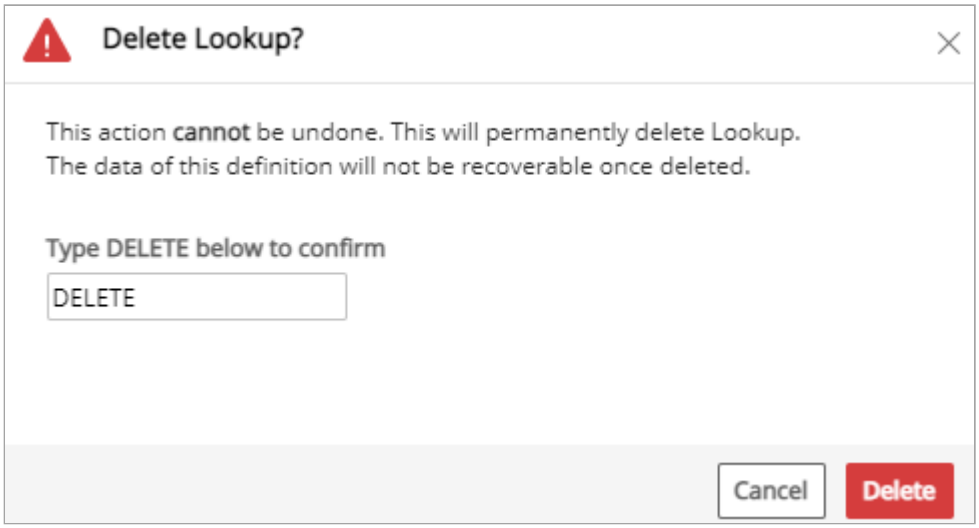
4. Modify the details as required and click **Modify** to save the details.

Deleting a lookup source

You cannot delete a lookup if it is associated with any definition.

To delete an existing lookup source, perform the following steps:

1. Go to the Lookup Sources list.
2. Click ⋮ (**More options**) against the required lookup source.
3. Select **Delete**. The Delete Lookup dialog appears.



4. Enter **DELETE** in the box.
5. Click **Delete**. On confirmation, the lookup source gets deleted and removed from the Lookup Sources list.

Synchronizing a lookup source

Synchronization allows you to maintain consistency and update the existing lookup sources. Synchronization of a lookup source must be performed at a periodical interval, preferably during off-business hours.

Lookup source tables are modified frequently due to the addition of new vendors, changes in address, mobile number, or any other attribute. Therefore, the lookup created in OmniXtract must be synchronized with the lookup source table to get the latest data.

To synchronize an existing lookup source, perform the following steps:

1. Go to the **Lookup Sources** list.
2. Click ⋮ (**More options**) against the required lookup source.
3. Select **Sync**. The lookup source starts synchronizing with the database table and shows the progress of the synchronization.

Services

A service fetches data from an iBPS or NewgenONE server and transfers it to OmniXtract for further processing. An OmniXtract service serves the following purposes:

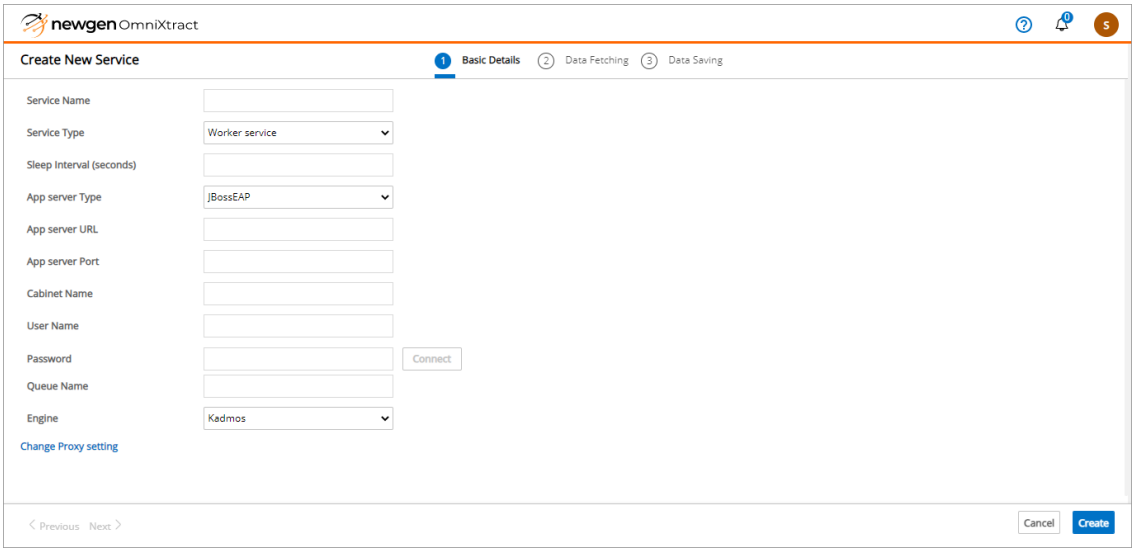
- Fetching the document location information from the workflow server and registering the document for extraction in OmniXtract.
- Automatically extracting the required information from documents as per the OmniXtract definition.
- Saving data to Newgen's BPM solution iBPS or NewgenONE after extraction. OmniXtract polls iBPS or NewgenONE at a fixed interval for getting new workitems.

Creating a service

The Service Manager must be running before creating a new service.

To create a service, perform the following steps:

1. Select **Services** from the menu bar.
2. Click **Create Service** in the upper-right of the Services list screen.
   If no service is created, the service list screen appears blank and shows a Create Service button in the middle.
   The Basic Details tab appears.

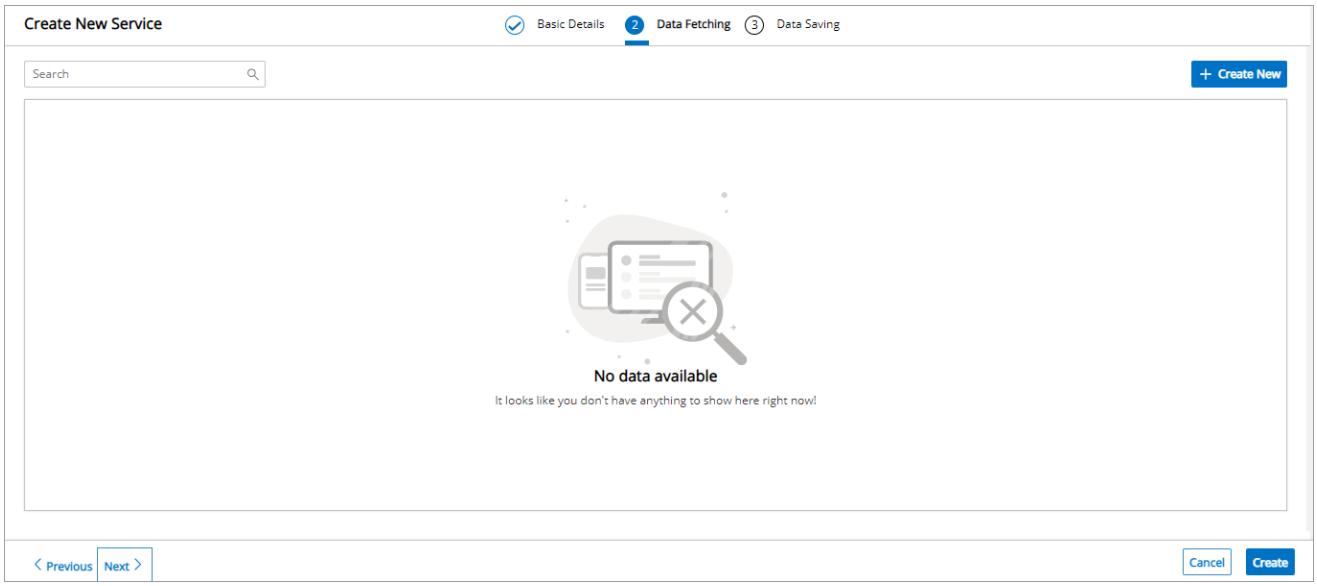


3. Specify the basic details as explained in the following table:

Field	Description
Service Name	Name of the service you want to register.
Service Type	Type of service. The available options are Windows and Worker services.
Sleep Interval (sec)	Time interval (in seconds).
App Server Type	The type of application server – JBoss, WebLogic, or WebSphere.
App Server URL	The URL of the selected app server. The format of this URL is: *http://IP:Port/oxejbintegrator/rest/api/execute*
App Server Port	The port on which the App Server is running (JNDI port).
Cabinet Name	Cabinet name to which you want to process.
Username	Name of the authorized user.
Password	The password of the user.
Connect	Click Connect to connect to the cabinet.
Queue Name	Name of the queue from which an item is fetched.

Field	Description
Engine	The available options are Kadmos, Google Vision, and Tesseract.
Change Proxy Settings	This is used to configure proxy settings. It has the following options: • **Proxy option** – Select No Proxy, System Defined Proxy, and Manual Proxy. • **Server IP** – IP of the Proxy Server. • **Server Port** – Port number of the Proxy Server. • **Username** – User ID to access the Proxy Server. • **Password** – The password of the user

4. Click **Next**. The Data Fetching tab appears. This tab provides the settings that you want to fetch from the environment.



5. Click **Create New**. The Edit Details dialog appears.
6. Specify the following details:

Field	Description
Basic details	
Workflow Definition	Name of the Workflow document that must be fetched.

Field	Description
Mapping Type	Select the mapping type between the definition and queue variable.
Language	
Identification Type	Specify identification type between the definition and queue variable. • In case of definition, specify OmniXtract definition name. • In the case of the queue variable, specify the queue variable.
Fetch Page Range	Select page range to fetch from the definition.
Fetch Country	Select the Fetch Country checkbox if the value from the workflow must be fetched in a country-specific format. Example – date input format, and amount input format. On selecting Fetch Country, the Fetch Value Type dropdown list gets enabled. Select the following: • **Fixed Value Based** – On selecting Fixed Value Based, a dropdown list containing a list of countries appears. Select the country as per the requirement. • **Field Value Based** – On selecting Field Value Based, the Queue Variable box appears. Enter the name of the queue variable.
Fetch Primary	Select the Fetch Primary checkbox if the primary field value must be fetched from the workflow. On selecting Fetch Primary, the grid below gets enabled to input the name of Queue Variable in workflow against the Primary Field Entity in OmniXtract. The primary field can be extracted in OmniXtract by creating a definition or else data can be fetched from iBPS or NewgenONE.
Fetch Filter	On selecting the Fetch filter, the grid below gets enabled to input the name of the Queue Variable in workflow against the Primary Field filter column in OmniXtract. The primary field can be further filtered by using the specific column mentioned in the primary field filter column.

7. Click **Save** to save the settings. The Fetching Details are added to a list.
   You can add as many definitions to the services as you want. To do this, click **Create New** in the upper-right of the Data Fetching tab.
8. To edit or delete the Data Fetching details, click ⋮ (**More options**) against the definition and select Edit Details to modify the details and Delete to delete the details.
9. Click **Next**. The Data Saving tab appears.
10. Enter the following details:

Fields	Description
OmniXtract Definition	Select a definition.
Upload Type	Select the upload type as: • **Upload** – Upload a new document. • **New Version** – Upload a new version of the document. • **Overwrite** – Overwrite the original document. • **None** – Don't upload a preprocessed image.
Workflow Document Type	Enter the iBPS or NewgenONE workflow document type.
Code	Enter the iBPS or NewgenONE field that contains the code of the extraction.
Description	Enter the iBPS or NewgenONE field that contains a description of the extraction.

11. Click ✓ at the end of the row to save the details.
12. Click **Create New** to add more rows to the Data Saving details.
13. Click ⌄ at the end of the row to expand the row and define field mapping details.
14. Click **Add Text Field**. The Text Field Details section appears.
15. Click **Add Row** and specify the following details:
    • **Field Name** – Select the field name of the selected definition for mapping.
    • **Queue Variable** – Select the iBPS or NewgenONE queue variable.
16. Click ✓ at the end of the row to save data mapping details.
17. To modify or delete the added Text Field Details, click ✎ to modify and 🗑 to delete the row.
18. Click **Create** to create the OmniXtract service.

Viewing and editing service details

To view and edit details of any service, perform the following steps:

1. Go to the menu bar and open **Services**.
2. Click ⋮ (**More options**) and select **View/Edit Details**. The details of the selected service appear in edit mode.
3. Modify the details as required and click **Save** to save the changes made.

Exporting a service

The export feature allows you to export any configured service to a different environment. The exported service can later be imported to register a new service on some other OmniXtract environment.

To export a service, perform the following steps:

1. Go to the menu bar and open **Services**.
2. Click ⋮ (**More options**) and select **Export**. The selected service gets downloaded in the form of a text file.

Disabling a service

The disable option allows you to disable an enabled or registered service. Once a service is disabled, you cannot start it.

To disable a service, perform the following steps:

1. Go to the menu bar and open **Services**.
2. Click ⋮ (**More options**) and select **Disable**. The selected service gets disabled.

Enabling a service

The enable option allows you to enable a disabled service.

To enable a service, perform the following steps:

1. Go to the menu bar and open **Services**.
2. Click ⋮ (**More options**) and select **Enable**. The selected service gets enabled.

> You can enable only disabled services.

Creating a copy of the service

This feature is used to make a copy of an already-defined service.

To create a copy of the service, perform the following steps:

1. Go to the menu bar and open **Services**.
2. Click ⋮ (**More options**) and select **Create a Copy**. The properties of the selected service appear in an edit mode.
3. Enter the **Service Name**. Make sure this name is different from the original service.
4. Modify the other details if required.
5. Click **Create**. The service gets created and appears in the list of services.

Importing a service

Import is used to import an exported service already defined on some other OmniXtract environment. Make sure the target environment does not have the service with the same name.

To import a service, perform the following steps:

1. Go to the menu bar and open **Services**.
2. Click **Import Service** in the upper-right of the screen. The Open dialog appears.
3. Select the exported service text file. Create New Service screen appears.

4. Modify the details if required. Refer to the section Creating a service for details.
5. Click **Create** to create the service.

Starting a service

You must start a service after creating it to process the workitems.

To start a service, perform the following steps:

1. Go to the menu bar and open **Services**. The Services screen appears.
2. Click the (**Start service**) icon against the required service.

> You can start only those services that are in a stopped but enabled state.

Stopping a service

To stop a service, perform the following steps:

1. Go to the menu bar and open **Services**. The Services screen appears.
2. Click the (**Stop service**) icon against the required service.

> You can stop only those services that are already started.

Reports

Reports give an insight into the performance of the overall extraction solution for the period. Reports are used to monitor the extraction performance in real time.

The Reports module is accessible by the Supervisor or users having the Maintenance role assigned to them.

Reports are of the following two types:

- Accuracy Analysis
- Extraction Throughput

To open the Reports module, select **Reports** from the menu bar.

Accuracy analysis report

It provides you with actionable insights through the visualization of key metrics required to observe and analyze extraction performance.

This section comprises the following subsections:

- Filters
- Data cards
- Accuracy trends
- Key fields
- Table fields
- Remark distribution
- Top volume vendors
- File format distribution

Filters

The insights for Accuracy Analysis can be populated by applying various filter options as listed below:

- **Definition** – To select the definition for Extraction Insights.
- **Category** – Combo box to select various pre-defined categories of key volume contributors.
- **Filter Range Type** – To filter the data by either providing Date Range or Workitem Range.
- **Start Date or Start Workitem** – Based on the Filter type selection, you must provide the Start Data or Start Workitem number.
- **End Date or End Workitem** – Based on the Filter Range Type selection, you must provide the End Data or End Workitem number.

After specifying the filters, click **Generate** to generate the Accuracy Analysis report.

Data cards

On applying filters and generating the report, the Data Cards are populated with insights on:

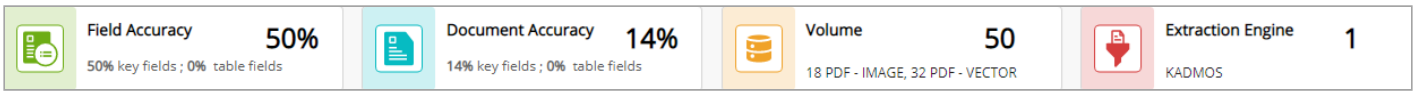
- **Field Accuracy** – Refers to the average accuracy percentage of all the non-tabular fields for the selected filter parameters.

> A Field is considered correctly extracted only if all the characters in the field are recognized correctly. If there is even a single character incorrectly recognized, then that field-level accuracy for the field is considered 0%. If the field is empty and the system also recognizes that as empty, then that field is considered as correct.

- **Document Accuracy** – Refers to the average accuracy percentage of documents extracted without requiring human intervention to correct the extracted data, for the selected filter parameters.

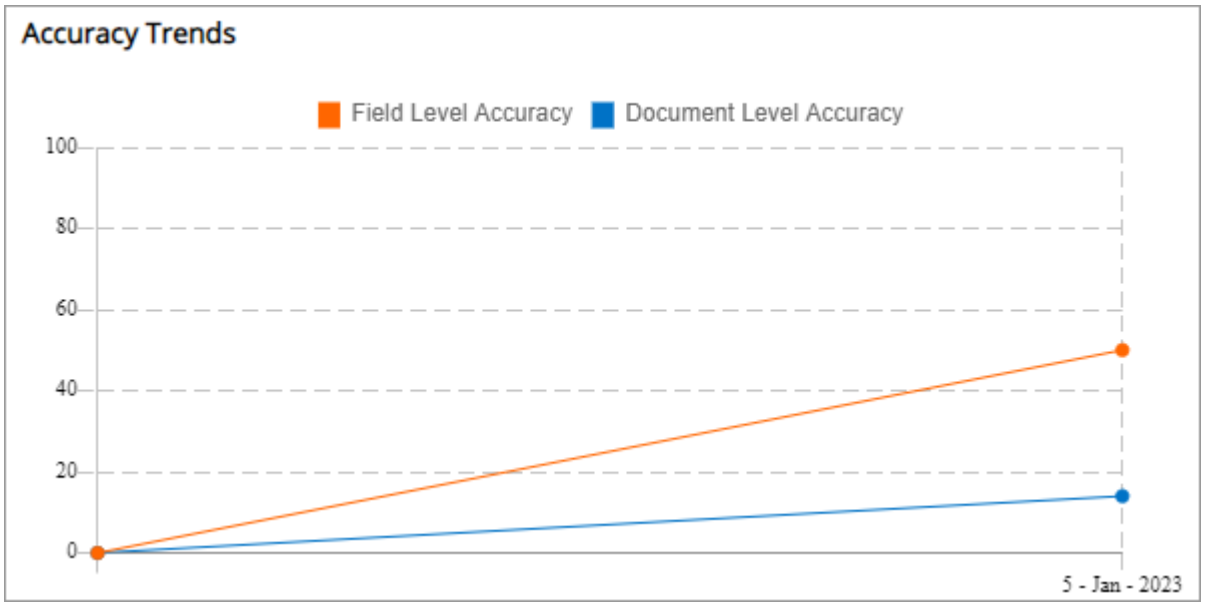
> The Document-level accuracy is considered as 100% of all the fields in the document are extracted correctly. If there is an error in the recognition of a single field/character, then the Document-level accuracy is 0% for the document analyzed.

- **Volume** – Refers to the total number of documents extracted for the selected filter parameters.
- **Extraction Engine** – Provides insight into the extraction engines used for extracting data.



Accuracy trends

Accuracy Trends give a graphical presentation of the extraction performance over a period. It shows field-level and document-level accuracies.



Key fields

It gives accuracy for each of the non-tabular fields configured in Definition for extraction, at a field-level and character-level, with an in-built search to find the required field. On the Character level, the character-level accuracy of the non-tabular fields can be viewed.

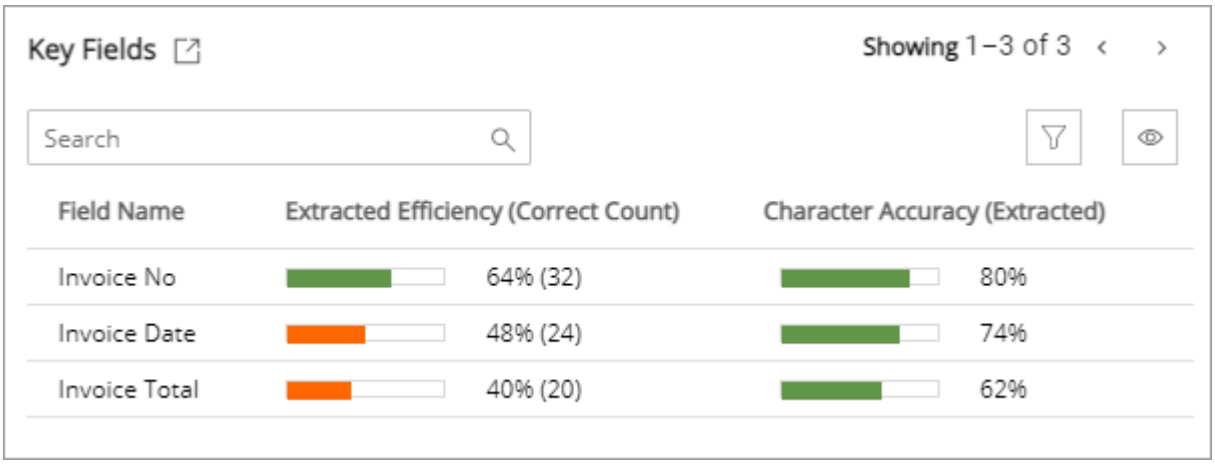
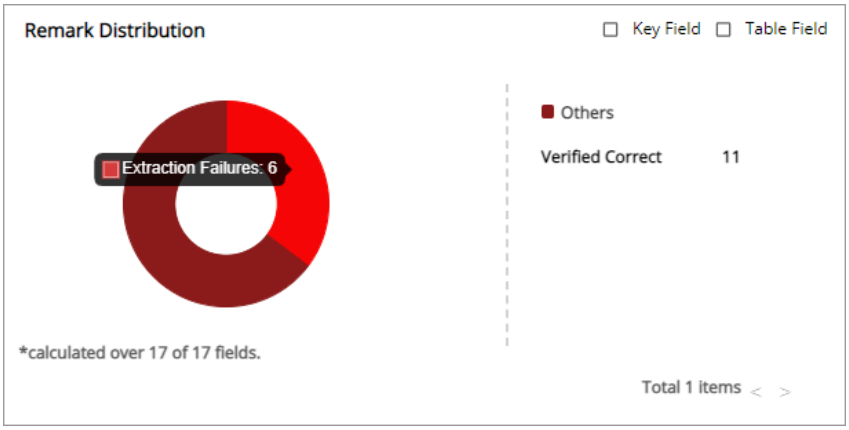


Table fields

It gives the accuracy of the tabular data for each of the table columns configured in Definition for extraction, at a field-level and character-level, with an in-built search to find the required column.

Remark distribution

It shows the actionable insights derived from the Image-assisted output analysis.



Top volume vendors

It provides insights into the extraction performance of the top contributors of volume in the system. It gives details such as Vendor ID, Vendor Name, and Document Count. The Field Level and Document Level percentage show accuracies for the invoices processed.

File format distribution

It provides insight into various file formats of documents ingested by the system with their volume count, distribution, and accuracy level for non-tabular fields and tabular data.

File Format

Key Field Table Field

File Type	Document Count	Distribution%	Field Level%	Document Level%
PDF - IMAGE	18	36	29	5
PDF - VECTOR	32	64	62	18

Extraction throughput report

The Extraction Throughput report gives insight into the performance of the overall extraction solution for a period. This interface can be used to monitor the extraction performance in real-time and take appropriate actions on observing outliers.

This section comprises the following subsections:

- Filter
- Weekly or monthly statistics
- Overall statistics
- Volume statistics
- Turnaround time statistics

Filter

The Definition dropdown allows you to select a definition and generate its extraction throughput.

Weekly or monthly statistics

On selecting the Document Type, the statistics on Last Week/Last Month extraction performance can be toggled and viewed. The following insights are provided:

- **Document processed** – Number of documents processed during the timeframe.
- **Document-level success count** – The number of documents extracted without requiring human intervention to correct the extracted data.

  The Document-level success count is incremented by 1 if all the fields in the document are extracted correctly. If there is an error in the recognition of a single field or character, then the Document-level success count is considered 0.

- **The manual effort required** – The number of documents requiring human intervention for verifying and correcting the extracted data.

Overall statistics

With the selection of year from the dropdown list, yearly statistics on performance can be viewed. The insights provided are:

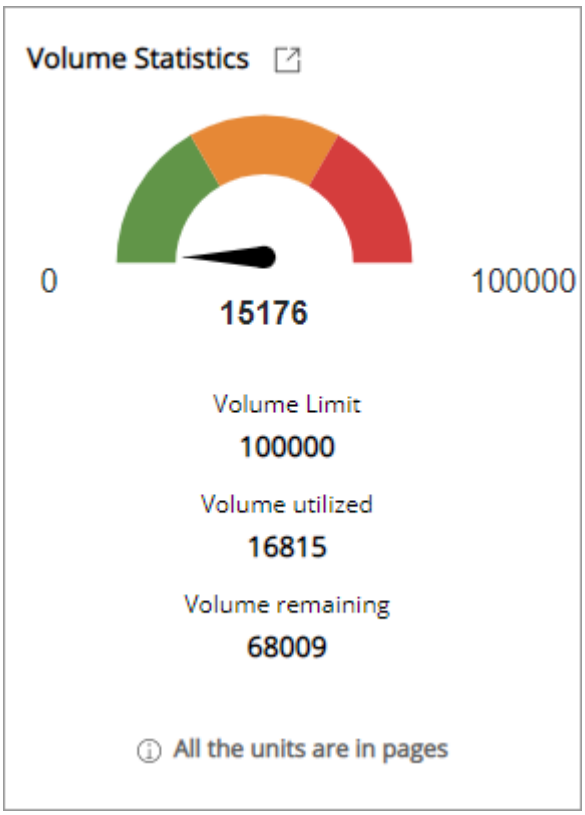
- **Workitem count** – Number of documents processed for the selected year.
- **Exception count** – Number of documents with extraction failure.
- **Field accuracy** – Refers to the percentage accuracy in the extraction of non-tabular fields from a document.

> A Field is considered correctly extracted only if all the characters in the field are recognized correctly. If there is even a single character incorrectly recognized, then that field-level accuracy for the field is considered 0%. If a field is empty and the system also recognizes that as empty, then that field is considered as correct.

- **Document-level** – Refers to document-level accuracy percentage.
- **OmniXtract TAT** – The average time taken by extraction services
- **Verification TAT** – The average time taken in verifying the extracted data.
- **Total fields count** – The total number of non-tabular fields extracted for the selected year.
- **Corrected fields** – The total number of fields requiring human intervention in correcting the extracted data.

Volume statistics

Volume Statistics provides insight to have an estimate of the usage of extraction services.



The key insights provided are:

- **Volume limit** – Refers to the total number of pages that can be extracted before requiring renewal of the license.
- **Volume utilized** – Refers to the number of pages extracted to date.
- **Volume remaining** – Refers to the number of pages that can be further extracted.

You can view volume consumption for every extraction engine configured in the system.

Turnaround time statistics

It provides insights into the average time taken by the Extraction services in processing a document and the average time taken by the human in the loop in verifying the extracted data.

The key insights provided are:

- **Definition TAT**
  - **Document Type** – Refers to the selected Document Type.
  - **Service Name** – Refers to the configured extraction service.
  - **Time Taken** – Refers to the average time taken by the individual services.
- **OmniXtract TAT**
  - **Average Idle Time** – Refers to the average time of inactivity on the Verification interface of the human in the loop verifying the extracted data.
  - **Average WI Time** – Refers to the average time taken to manually verify a document.

Master data settings

The Master Data Settings allow uploading of actual or master data using the standard template present in the system for all existing definitions.

To access the Master Data Settings, click **Master Data Settings** in the upper-right corner of the Reports screen. The Master Data Setting dialog appears.

Downloading Actual Data Template

The Download Template allows you to download an excel file containing the format of the document type. Once downloaded, you must fill it with workitem-related data. The updated excel file is then uploaded as a Master Data file that you must use for comparison with extracted data.

To download the actual data template, perform the following steps:

1. Select **Document Type** from the dropdown. It contains a list of all the definitions available in OmniXtract.
2. Click **Download Template**. The actual data template gets downloaded in the form of an excel file.

Uploading Master Data File

You must upload the actual or master data that you want to compare with extracted data to get insights into the current level of definition optimization.

To upload the Master Data File, perform the following steps:

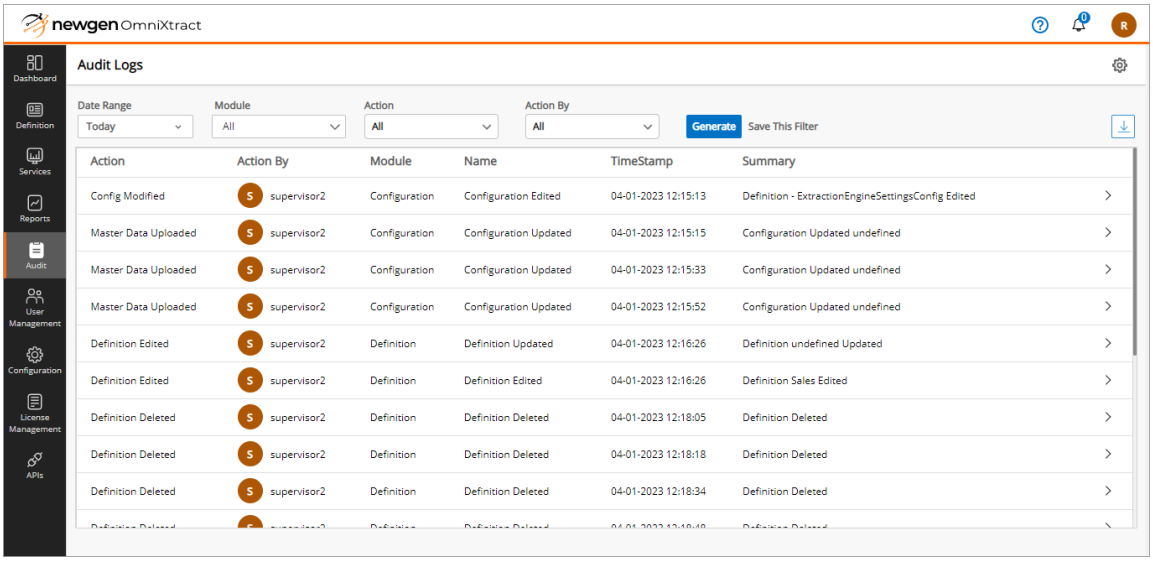
1. Click the ••• (browse) icon against the Master Data File to browse for and select the actual data file that you had downloaded and updated with relevant workitem-related data.
2. Click **Upload**. The Upload Status shows the status of the file being uploaded.
   The Upload Status comprises Initiate, File, File Validation, DB Updation, and Complete steps.
   On successful upload, the Upload Status appears as Upload Successfully. In the case of a failure, the Upload Status appears as Upload Failed.
3. Click **Continue** upload and close to continue upload in the background and close the dialog.

Audit

The audit log is an account of the operations performed on the specified object by any of the users of the cabinet used for signing into OmniXtract. Logs are generated for all actions that users perform on different modules of OmniXtract.

The Audit module is accessible by the Supervisor or users having the Maintenance role assigned to them.

Select **Audit** from the menu bar. The Audit Logs screen appears.



This section comprises the following subsections:

- Generating audit logs
- Viewing audit details
- Downloading audit logs
- Audit settings

Generating audit logs

To generate audit logs, perform the following steps:

1. Specify the filters as described below:

Filter	Description
Date Range	You can filter audit logs as per the date range. The filter by date range options include: • Today • Yesterday • Last Week • Last Month • Last 3 Months • Last Year • Year • Custom Range – It allows you to define custom date ranges. For example, from 1st December 2022 to 7th January 2023.
Module	You can filter audit logs as per the modules. Select the required modules. You can select All to view the audit logs of all the modules. The options include: • All • API • Configuration • Definition • License Management • Lookup • Reports • Rights Management • Services
Action	You can filter audit logs as per the actions. Select the required actions. Click Select All to view audit logs of all the actions. You can select single or multiple actions as well.
Action By	You can filter audit logs based on different users. Select the required users to view the audit logs of these users. Click Select All to select all the users.

2. Click **Generate** to generate the audit logs. The audit logs appear based on the selected filters.

Saving a filter

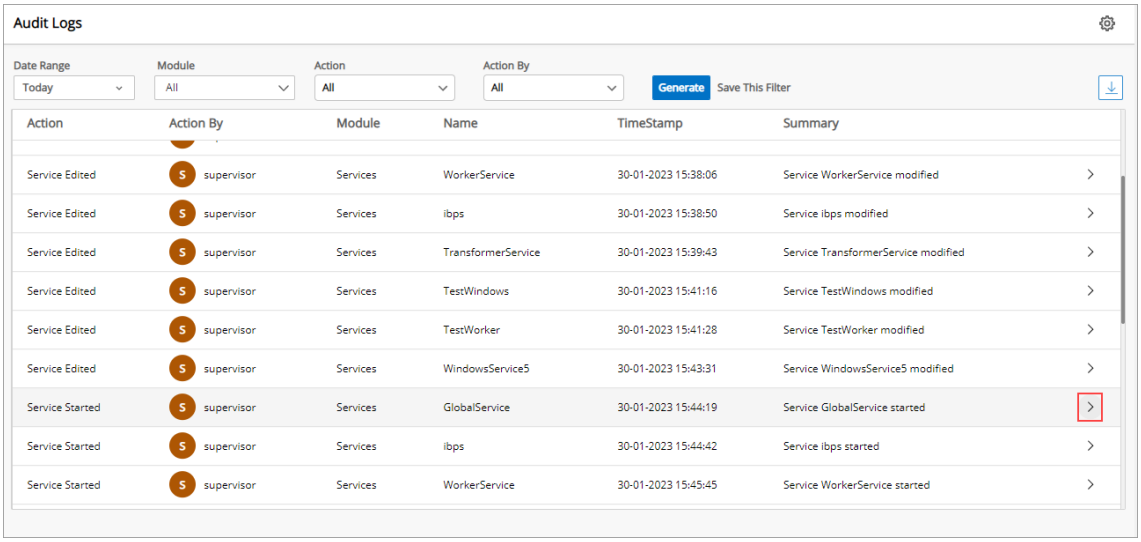
The **Save This Filter** option allows you to save a filter so that you can use it again. You can save only one filter. When you save another filter with different criteria, the existing saved filter gets replaced by the new one.

To save a filter, perform the following steps:

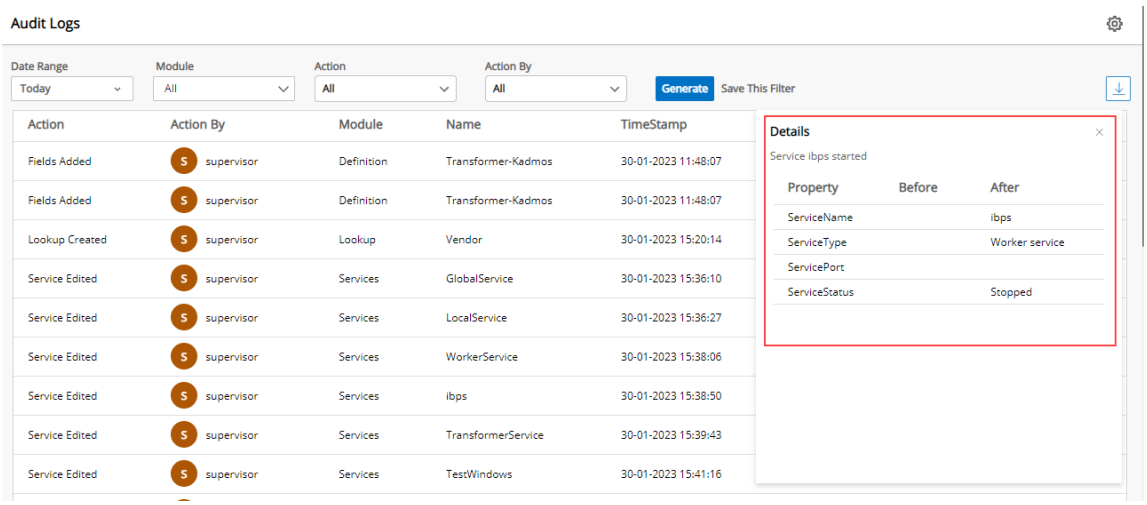
1. Specify the filter criteria as required.
2. Click **Generate**.
3. Click **Save This Filter**. The Save Filter dialog appears.
4. Enter the **Filter Name** and click **Save**. The saved filter appears on the right side, on the Audit Logs title bar.
5. Click the saved filter name to generate audit logs based on the saved filter criteria.
6. To delete the saved filter, click the cross icon associated with the saved filter name.

Viewing audit details

You can view the details of a particular audit log by clicking **>** at the end of the audit log row.



It opens the Details dialog showing details of the action.



Downloading audit logs

To download the audit logs, click the icon appearing to the right of the filter options.

Audit logs are downloaded in an Excel sheet format.

Audit settings

Audit settings allow you to define settings for Purge audit logs.

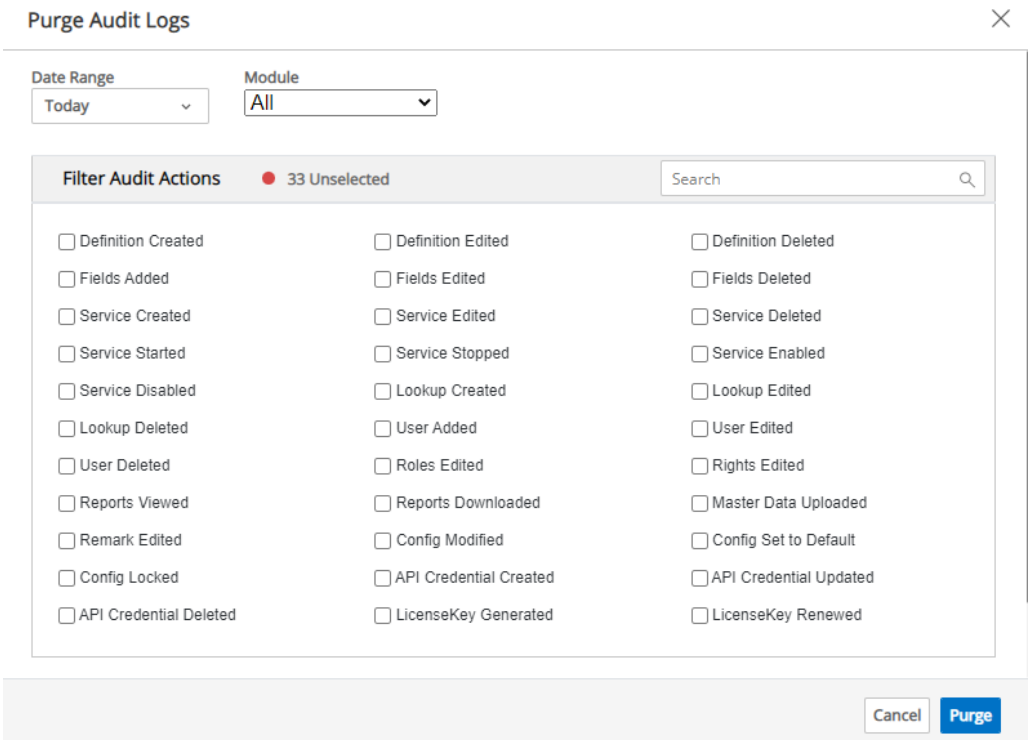
To access Audit Settings, click the gear icon in the upper-right corner of the Audit Logs screen. Click Purge to define purge settings.

Purge audit logs

The Purge Audit Logs feature allows you to remove or delete audit actions that are no longer required. Purging removes the audit logs permanently and you cannot get them back.

To purge audit logs, perform the following steps:

1. Click **Audit Settings** and select **Purge**. The Purge Audit Logs dialog appears.



2. Select the **Date Range**. The date range options are:
   - Today
   - Yesterday
   - Last Week
   - Last Month
   - Last 3 Months
   - Last Year
   - Year
   - Custom Range – The Custom Range allows you to select two dates. The audit logs generated between these two dates are purged.
3. Select the required **Module**. To purge audit actions of all the modules, select **All**.
4. Select the required **Audit Actions** that you want to purge. The list of Audit Actions depends on the selected Module.
   If the list of audit actions is long, then you can search and select them. For more details, refer to the section Searching for audit actions.
5. Write **PURGELOGS** in the box given just below the info icon.
6. Click **Purge** to purge the selected audit actions.

Searching for audit actions

To search for an audit action, perform the following steps:

1. Enter the search text in the **Search** box. The search results appear once you stop typing the text.
2. Select the required actions from the search results.
3. To clear the search results, click the ⊗ icon.

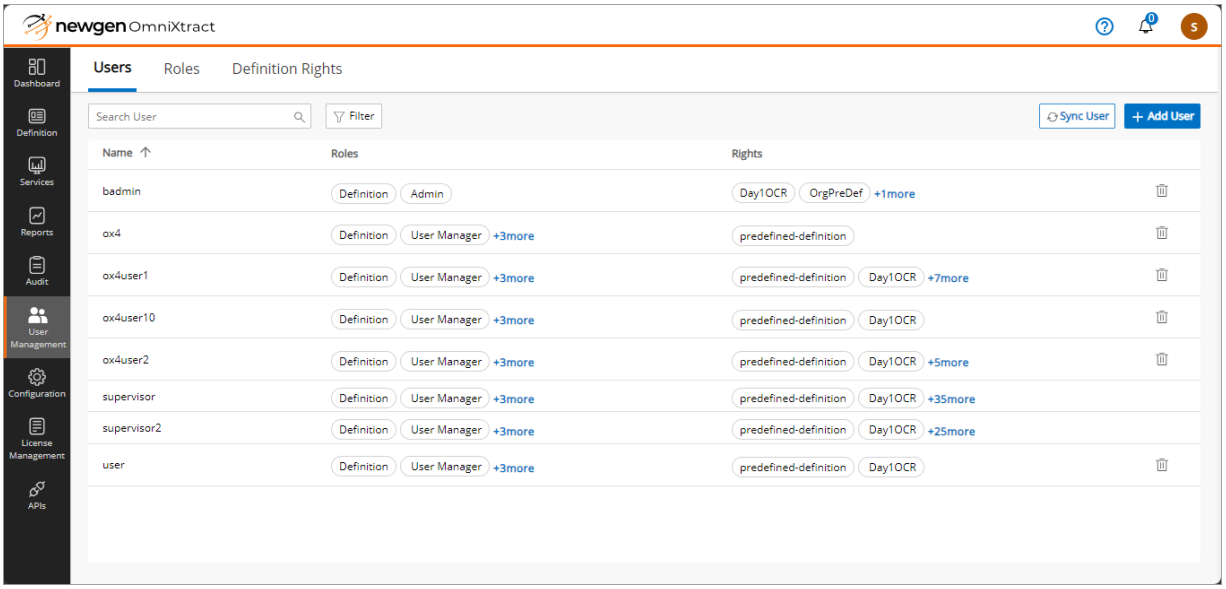
User management

User Management allows you to provide appropriate rights to users for accessing and working on different modules of OmniXtract. OmniXtract imports users from OmniDocs and you can assign roles and rights to those users.

OmniDocs must be preinstalled, and its cabinet must be associated with OmniXtract.
Refer to the OmniDocs Administration Guide to learn how to create and manage users in OmniDocs.

To access the User Management module, select **User Management** from the menu bar. User Management comprises the following tabs:

- Users
- Roles
- Definition rights



Users

Depending on the business needs, you must add different users with different roles and privileges to work on OmniXtract. You can add only those users to OmniXtract who are existing users of OmniDocs.

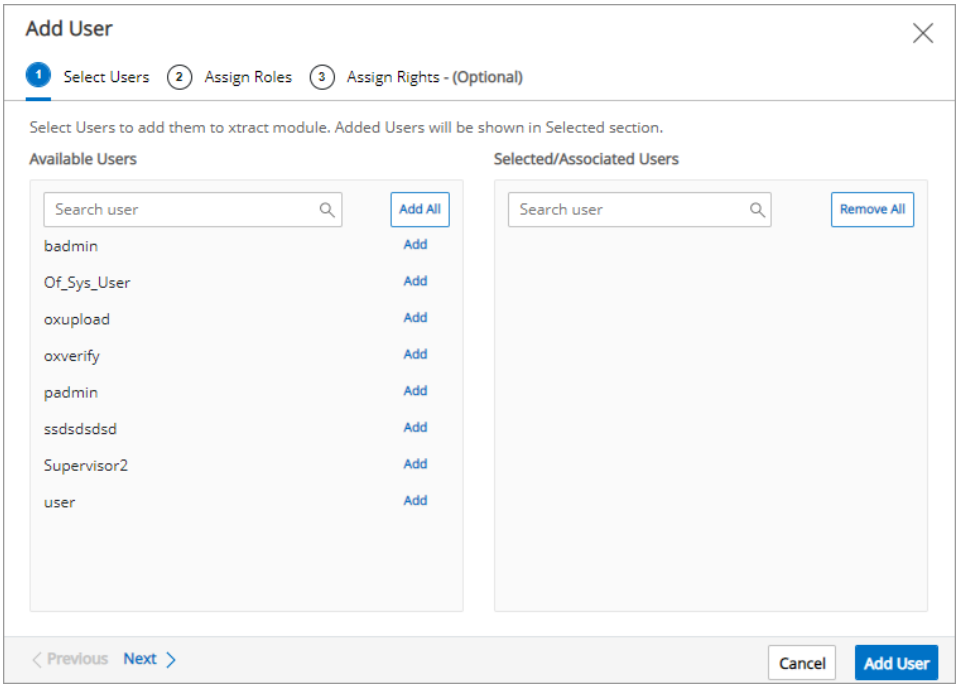
User management operations such as addition, deletion, and modification of users are performed through OmniDocs.

The Users tab shows a list of existing users with their email IDs, roles, and rights assigned to them.

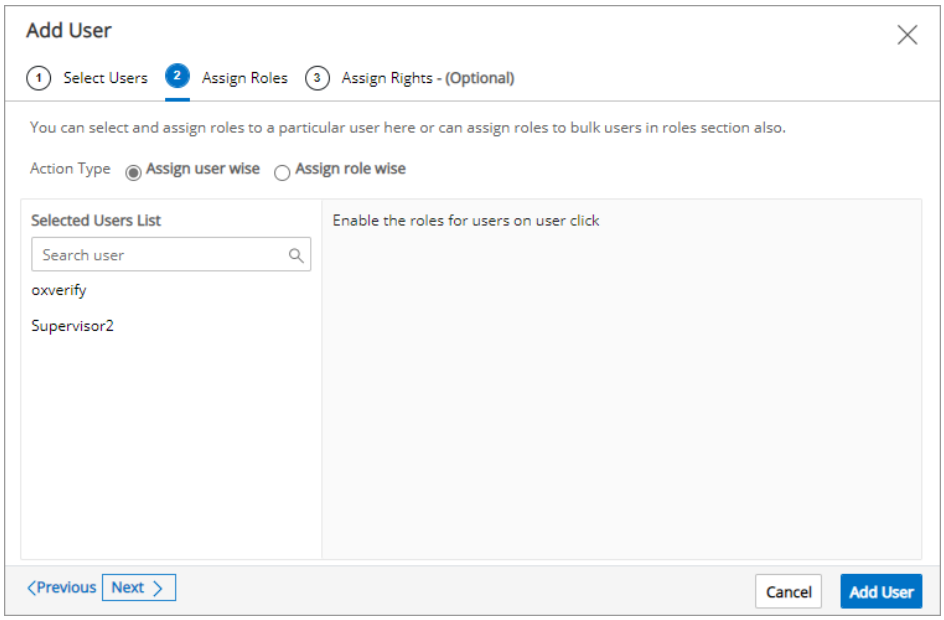
Adding users

To add users to OmniXtract, perform the following steps:

1. Open **User Management** from the menu bar.
2. Click **Add User** in the upper-right corner of the screen. The Add User dialog appears, showing a list of OmniDocs users.
   The first step involves adding OmniDocs users to OmniXtract.



3. Click **Add** against the required user to move it from the Available Users to Selected/Associated Users section.
   To add all the available users, click **Add All**.
   To search for a particular user from the list of Available Users, enter the username in the Search user box and click the **Search** icon.
   The added users appear in the Selected/Associated Users section.
4. To remove any added user, click **Remove** against the required user.
   Click **Remove All** to remove all the added users.
   Skip this step, if you do not want to remove any users.
5. Click **Next**. The Assign Roles tab appears. It allows you to select and assign roles to the selected OmniDocs users.



6. Select Action Type as **Assign user wise** or **Assign role wise** option.

   For the **Assign user wise** option, perform the following steps:

   a. Select the **Assign user wise** option.
   b. Click the required username from the Selected User List. A list of predefined roles appears in the right pane. You can assign the following roles:

Roles	Description
Definition	If the Definition role is assigned to any user, then that user can access Definition, Reports, and Configuration modules after signing in to OmniXtract.

Roles	Description
User Manager	If the User Manager role is assigned to any user, then that user can access the User Management module after signing in to OmniXtract.
Maintenance	If the Maintenance role is assigned to any user, then that user can access Services and API modules after signing in to OmniXtract.
Configurator	If the Configuration role is assigned to any user, then that user can access the Configuration module after signing in to OmniXtract.
Admin	If the Admin role is assigned to any user, then that user can access all the modules after signing in to OmniXtract.

c. Enable the desired roles for assigning them to the user.

For the **Assign role wise option**, perform the following steps:

a. Select the **Assign role wise** option. A list of predefined roles appears in the left pane.
b. Select a role from the **Role List**. As you click a role, a list of users that were selected in the Select Users step appears.
c. Select the desired users to assign them the selected role.

7. Click **Next**. The Assign Rights – (Optional) tab appears. It allows you to assign rights to the existing definitions to the selected users.
8. Select Action Type as **Assign user wise** or **Assign role wise option**.

For the **Assign user wise** option, perform the following steps:

a. Select the **Assign user wise** option.
b. Click the required username from the Selected User List.
c. Select the required definition from the **Rights** dropdown. This dropdown contains all the OmniXtract definitions. The selected user can work on only the definitions selected here.

For the **Assign right wise** option, perform the following steps:

a. Select the **Assign right wise** option. A list of all the defined definitions appears in the left pane.

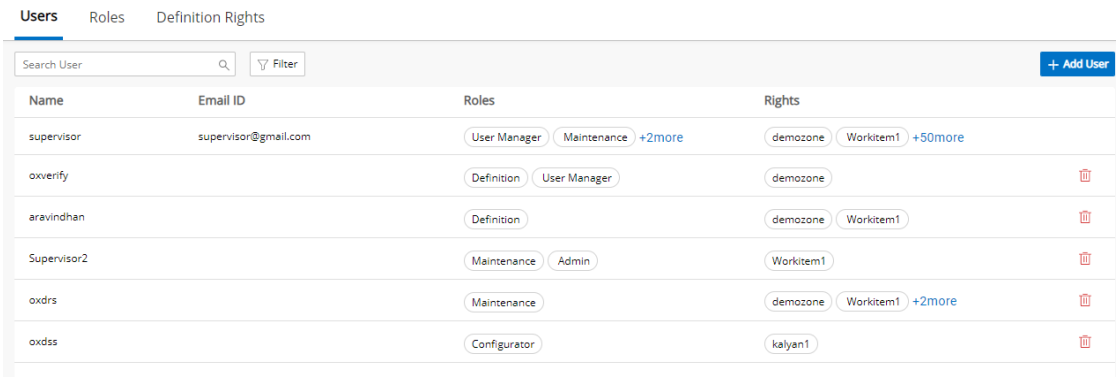
b. Select a right or definition from the **Rights List**. As you click the right, a list of users that you had selected in the Select Users step appears.
c. Select the desired users to assign them the selected right.

9. Click **Add User**. A success message for the addition of users with assigned rights appears.
   The added users appear in the list of Users.

Deleting users

To delete any user, perform the following steps:

1. Click the **Users** tab of User Management.



2. Click the (**Delete**) icon against the desired user. The Delete User dialog appears.
3. Click **Delete** to confirm the user deletion. On confirmation, a success message for the user deletion appears, and that user is removed from the list of users.

Roles

The Roles tab allows you to view the predefined roles, their descriptions, and the number of users assigned to those roles. It also allows you to add or remove users from roles.

Adding users to roles

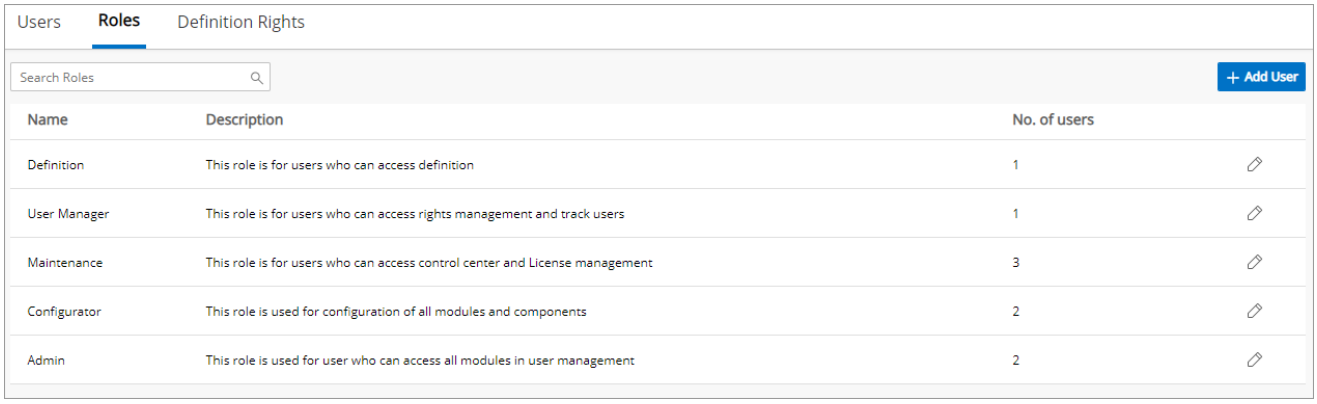
The steps for adding users to a role are the same as those of adding a user for OmniXtract. Refer to the section Adding users.

Editing user role association

Editing a role allows you to add or remove users from a role.

To update a role, perform the following steps:

1. Click the **Roles** tab of User Management. The Roles tab appears.



2. Click the (**Edit**) icon against the desired role. The Update Role dialog appears.
3. From the list of Available Users, click **Add** against the required user to add it to the selected role. You can click **Add All** to add all the users simultaneously.
4. To remove a user from the list of Selected/Associated users, click **Remove** against the desired user. You can click **Remove All** to remove all the users simultaneously.
5. Click **Update** to update the role. The role gets updated, and a success message appears for the same.

Definition rights

The Definition Rights tab allows you to view a list of all the definitions, their field counts, the last modified date, the description of the definitions, and the number of users

assigned to the definitions. It also allows you to add new users to the definitions and update the existing definitions.

Adding users to definition rights

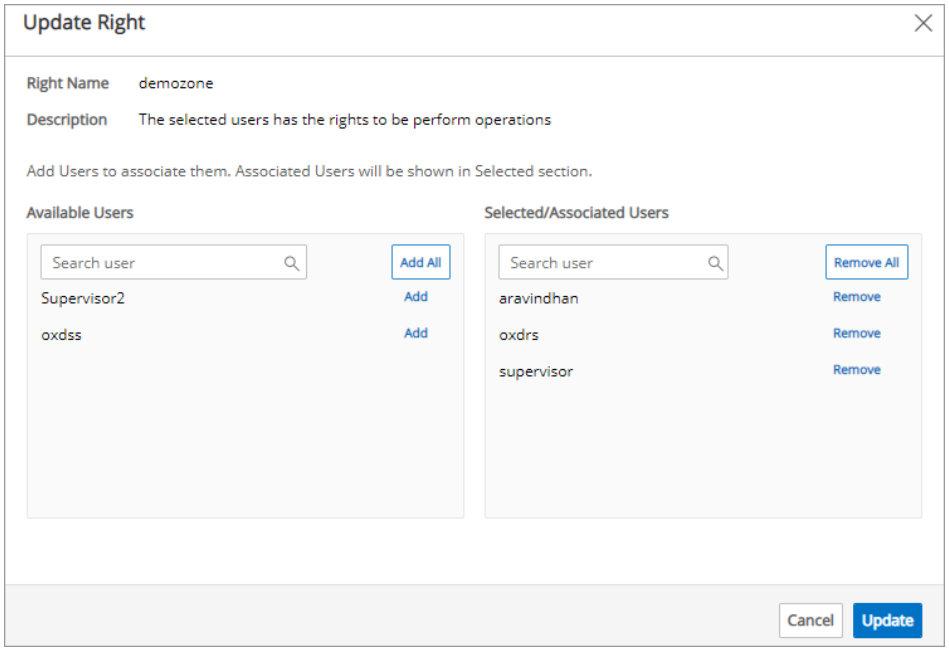
The steps of adding users to a definition are the same as those of adding a user for OmniXtract. Refer to the section Adding users.

Editing user definition rights association

Editing of definition rights allows you to add or remove users from a definition.

To edit the definition rights, perform the following steps:

1. Click the **Definition Rights** tab of User Management. The Definition Rights tab appears.
2. Click the (**Edit**) icon against the desired definition. The Update Right dialog appears.



3. From the list of Available Users, click **Add** against the required user to add it to the selected definition. You can click **Add All** to add all the users simultaneously.

4. To remove a user from the list of Selected/Associated users, click **Remove** against the desired user. You can click **Remove All** to remove all the users simultaneously.
5. Click **Update** to update the definition rights. The definition rights get updated, and a success message appears for the same.

Configuration

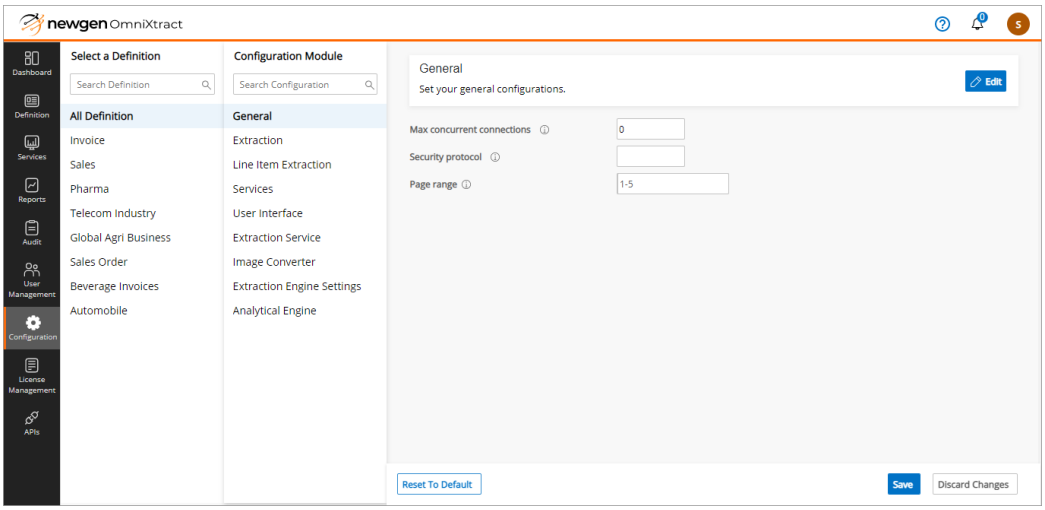
Configuration management is a process of maintaining functional attributes of the system as per the requirement.

Select **Configuration** from the menu bar. The Configuration Module screen appears.

The Configuration Module comprises the following configurations:

- General
- Extraction
- Line-item extraction
- Services
- User interface
- Extraction service
- Image converter
- Extraction engine settings
- Analytical engine
- AI model

The Configuration module is accessible by the Supervisor, users having the Configurator role, or users having the Definition role assigned to them.



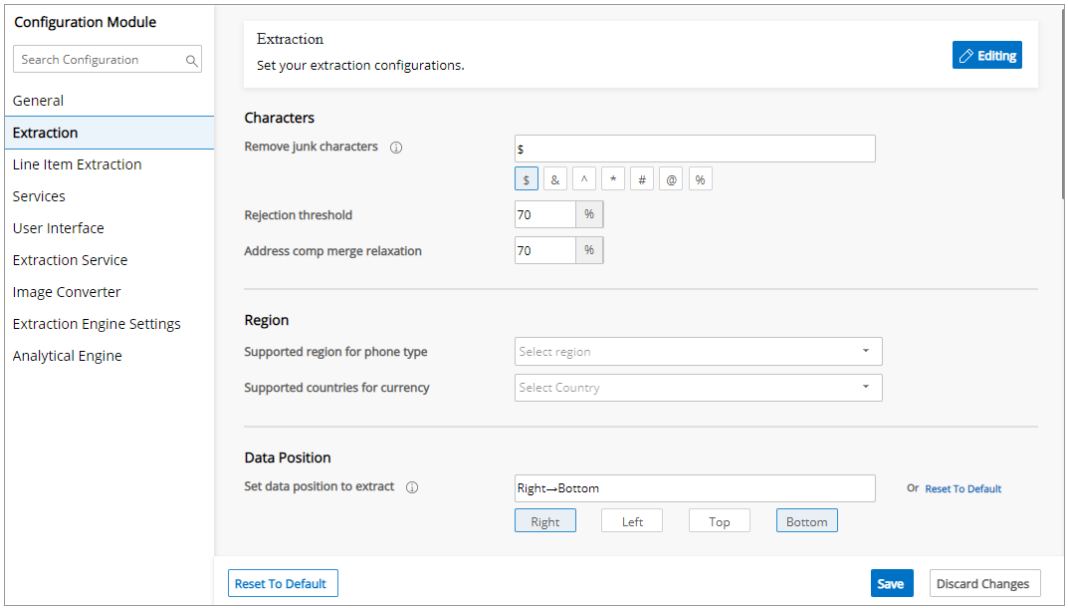
General

The General configuration allows you to define product-level configurations.

Define the General configuration as described in the table below:

Fields	Description
Max concurrent connections	It is used to specify the maximum number of connections that can be made to the OmniXtract webservice in online integration.
Security protocol	It is used to specify the security protocol. The supported security protocols include SSL, TLS 1.0, TLS 2.0, and TLS 3.0.
Page range	Mention the page range for extraction. Example of a page range – 1-8, 11, 13. If this is the page range, then extraction is performed on pages 1 to 8, 11, and 13 only.

Once you have defined the properties of the General configuration, click **Save** to save the properties. To clear the entered details and restore the last saved values, click **Reset to Default**.



Extraction

Extraction helps you to set your extraction configurations. After you have added the fields and defined their properties, you need to extract non-tabular fields. The extracted data is displayed against the respective fields appearing in the Fields pane.

Extraction includes the following configurations:

- **Characters** – Character configuration can be done using the following fields:
  - **Remove junk characters** – It allows you to add what junk characters can be removed while performing an extraction. You can choose from the pre-defined junk characters – $, &, ^, #, @, %.
  - **Rejection threshold** – A page is considered a poor-quality image and gets rejected if junk characters on that page cross the defined percentage of the rejection threshold.
  - **Address component merge relaxation** – Specify the range of blank spacing that can be used. Suppose you have an address block Newgen Software Pvt. Ltd. In the address block if there is more than a single space then it gets rejected. But it can be accepted if there is a single space given that is acceptable. The acceptable merge relaxation is provided in this field to ensure that the defined range is a part of the single word/address.
- **Region** – Specify the regions where you are targeting extraction. For example, select +91 as an Indian phone code and select INR from the currency dropdown to predefine the region.
- **Data Position** – Set the data position to extract. It specifies from where the system must start looking for data to be extracted and where it must go next.
- **Detection** – Detection includes the following configuration:
  - **Label detection** – Choose any one label detection method among the given properties:
    - **Regex** – for label detection using regular expression.
    - **LD** – for using a label detection mechanism.
    - **Hybrid** – for label detection using both Regex and LD.
  - **Label detection match percent** – The defined percentage is related to the label detection mechanism percentage with the exact match. It shows what percentage level you can match data.

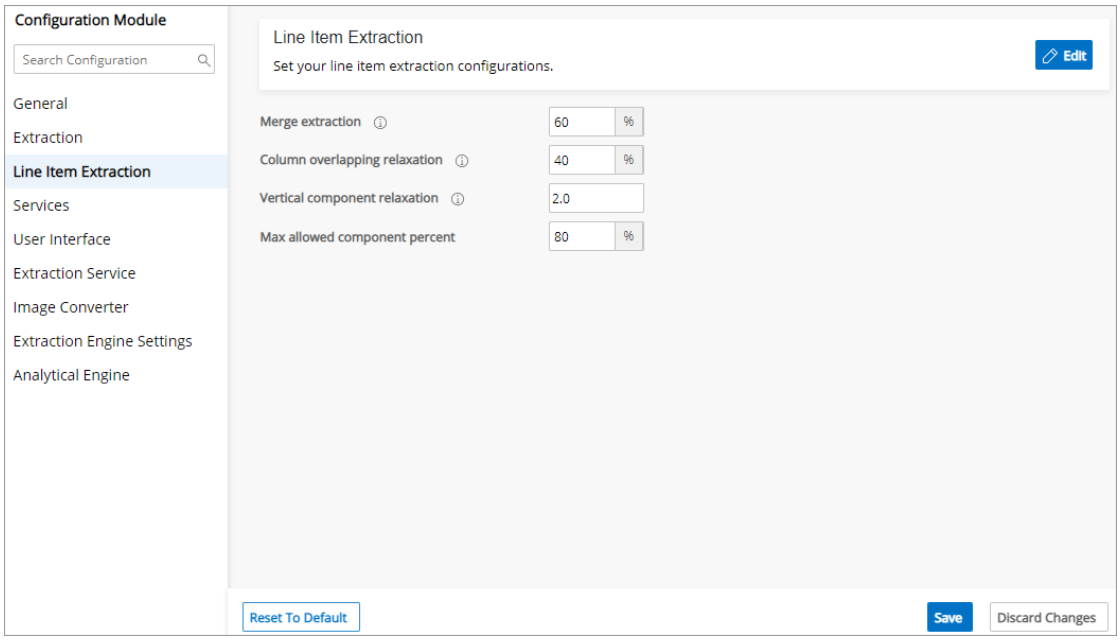
- **Right-aligned block detection** – The defined percentage is related to the label detection mechanism percentage with the block detection. It shows what percentage level you can match data.
- **Address intersect percent** – The defined percentage is related to the label detection mechanism percentage with the address intersect. It shows what percentage level you can match data.
- **OMR fill percent** – Specifies how much percentage of a filled cell must be considered as filled.

- **Batch size** – It defines the number of records that must be considered during that operation's iterations. It includes the following properties:
  - **Extraction batch size** – Specify the batch size of data for extraction.
  - **Lookup batch size** – Specify the batch size of data that must be looked up from the vendor database.
- **Engine** – The engine section allows you to configure the extraction engine details. By default, the extraction engine is Kadmos. It includes the following properties:
  - **Extraction engine** – Select the extraction engine that must be used for extracting data from the document. The extraction engine options include Kadmos, Google Vision, Tesseract, Computer Vision, and Amazon Textract.
  - **PDF engine** – Select the PDF engine that must be used for extracting data from the document. The PDF engine options include Very PDF, Poppler, and Apache PDFBox.
  - **Very PDF processing wait sec.** – Specify the time in seconds for processing the PDF.
- **Language** – It allows you to set language details of data that requires extraction. It includes the following properties:
  - **Language** – Select the language for data extraction.
  - **Character information** – Enable it to show character information.
  - **Multi-language support** – Enable it for the multi-language support of the data.
- **Others** – This section defines the generic settings for extraction. The other generic settings include:
  - **Use find data cell** – Enable this option to extract an OMR field.
  - **Use NER** – Enable this option to extract an address field.
  - **Use global on local failure** – Disable this option to restrict the usage of the global definition if the local definition fails.

- **Image components from pdf required** – It defines whether the extraction of image objects from PDF is required or not. If enabled, image objects from PDF get extracted.

Line-item extraction

Line-Item Extraction allows you to configure properties for the table data extraction. After adding fields and defining their properties, you need to extract line-item data. The extracted data gets displayed against the respective fields appearing in the Fields pane.



Line-Item Extraction includes the following configurations:

Fields	Description
Merge extraction	It shows until which percentage level you can merge data extraction.
Column overlapping relaxation	It shows until which percentage level you can overlap column relaxation.
Vertical component relaxation	It shows the level of vertical component relaxation for the line-item data extraction.
Maximum allowed component percent	It shows the maximum percentage of the component that you can acquire.

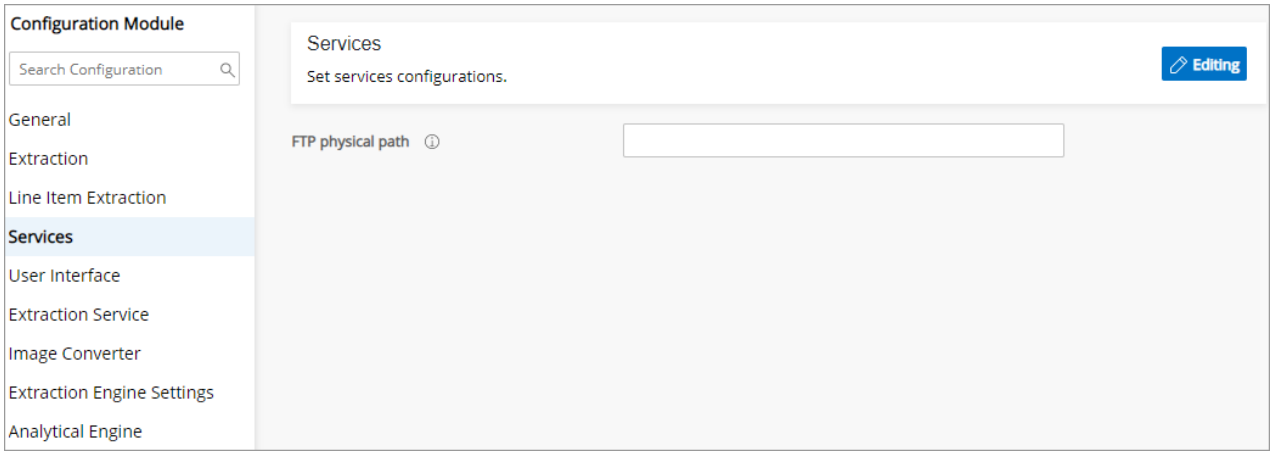
Services

It allows you to define a physical or cloud storage path where you can store preprocessed images for extraction.

You can define a path for storing preprocessed images in the following ways:

- **When the file upload mode is FTP Path**:

1. Create a folder *OMNIXTRACTFTP* in *<OmniXtract 4.0 installation folder>* for storing preprocessed images.
2. Specify the full path of the above-created folder in the **FTP Physical Path** box.
3. Click **Save**.



- **When the file upload mode is Amazon S3**:

1. Enter the **Amazon S3 Access Key**, **Secret ID**, and select **Region Endpoint**.
2. Enter **Amazon S3 Bucket Name**.
3. Click **Save**.

User interface

The User Interface configuration allows you to configure authentication settings for users.

The User Interface configuration includes the following settings:

- Enable or disable HTTPS for user authentication. HTTPS is used to secure communication over the network.
- Enable or disable the captcha in a user interface.

Extraction service

The Extraction Service allows you to define settings for the engine that is used to skew or unskew the image. Specify the service name and detection type using the extraction service configuration.

The Extraction Service configuration includes the following settings:

- **Detect name** – Select the text detection engine. The options are:
  - IPLIB
  - OpenCV
  - None
- **Detect type** – The Detect type option appears when the detection engine is OpenCV. It allows you to select the detection model for the OpenCV engine. The options are:
  - Detect
  - Detect East

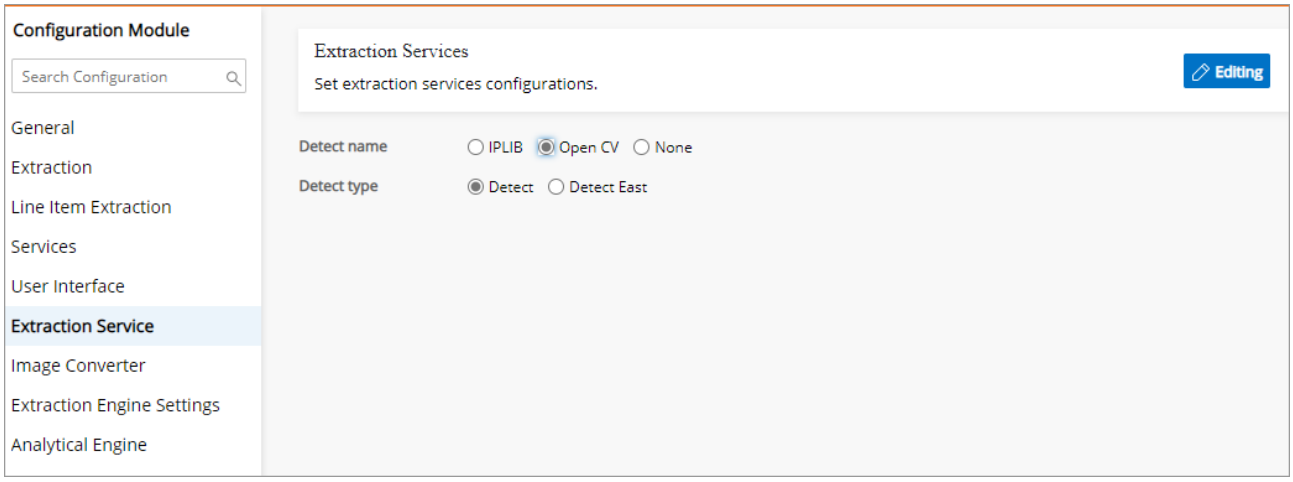
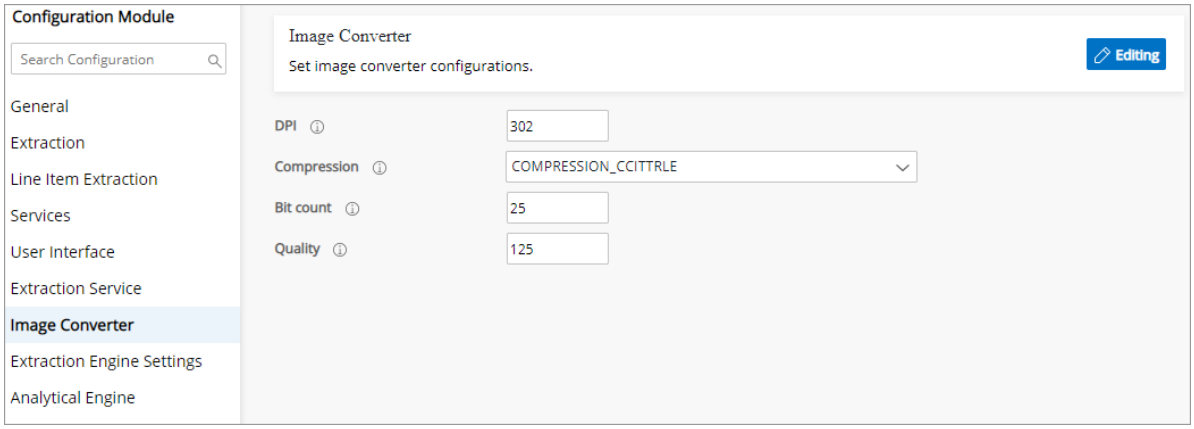


Image converter

While performing extraction of the image, the following image conversion configurations must be specified:

Fields	Description
DPI	Specify the dot per inch property of the image. An image must have at least a DPI of 25.
Compression	Specify the image compression service.
Bit count	Specify the bit count of the image. It must be at least 25.
Quality	Specify the image quality in pixels.



Extraction engine settings

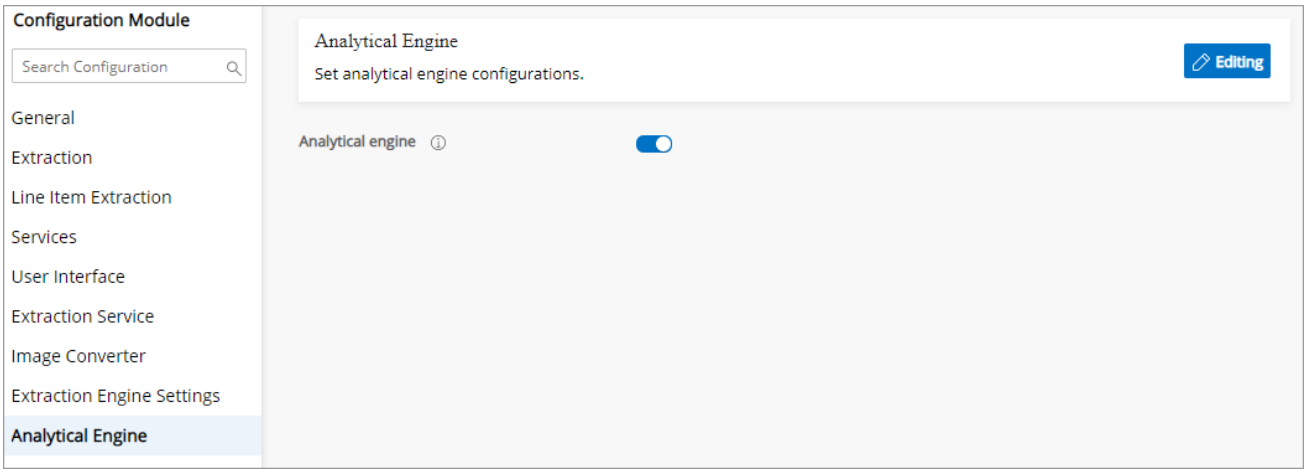
The Extraction Engine Settings allow you to define configurations for the following extraction engines:

Extraction Engine	Configuration
Google Vision	Configuration related to the Google Vision extraction engine can be performed here. Mention the details related to the Google Vision engine such as API, API key, Grant type, and other details as required.

Extraction Engine	Configuration
Azure Vision	Configuration related to the Azure Vision extraction engine can be performed here. Mention the details related to the Azure Vision engine such as API, Subscription key, End key, and other details as required.
Amazon Extract	Configuration related to the Amazon Extract engine can be performed here. Mention the details related to the Amazon Extract engine such as API, Access key, Secret key, and other details as required.
Tesseract	Configuration related to the Tesseract extraction engine can be performed here. Mention the details related to the Tesseract engine such as Engine mode, page segment mode, and zone page segment mode.
Kadmos	Configuration related to the Kadmos extraction engine can be performed here. For Kadmos, you can enable Use alternate engine for single digit data.

Analytical engine

You must enable analytical engine configuration to utilize an analytical engine for data extraction.



AI model

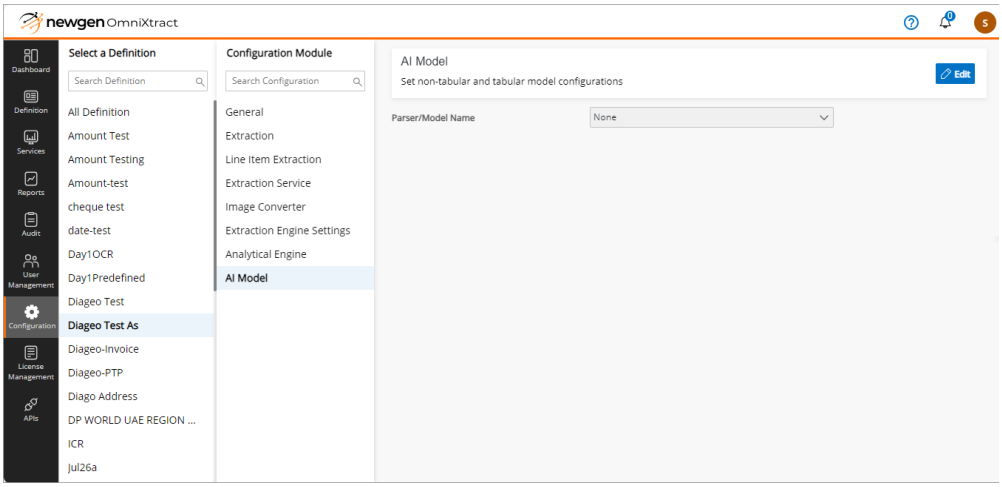
The AI Model allows you to define configurations of the trained AI model that is running in the background to extract correct data from a document. It leverages the deep learning model to minimize definition creation effort for tabular and non-tabular data extraction. To use AI Model in OmniXtract, the Transformer Service must be already installed.

The AI model for tabular and non-tabular data supports only invoices.

The AI Model must be configured at the definition level, not the global one that applies to all definitions. The AI Model option appears on selecting a definition.

To configure the AI Model for a definition, perform the following steps:

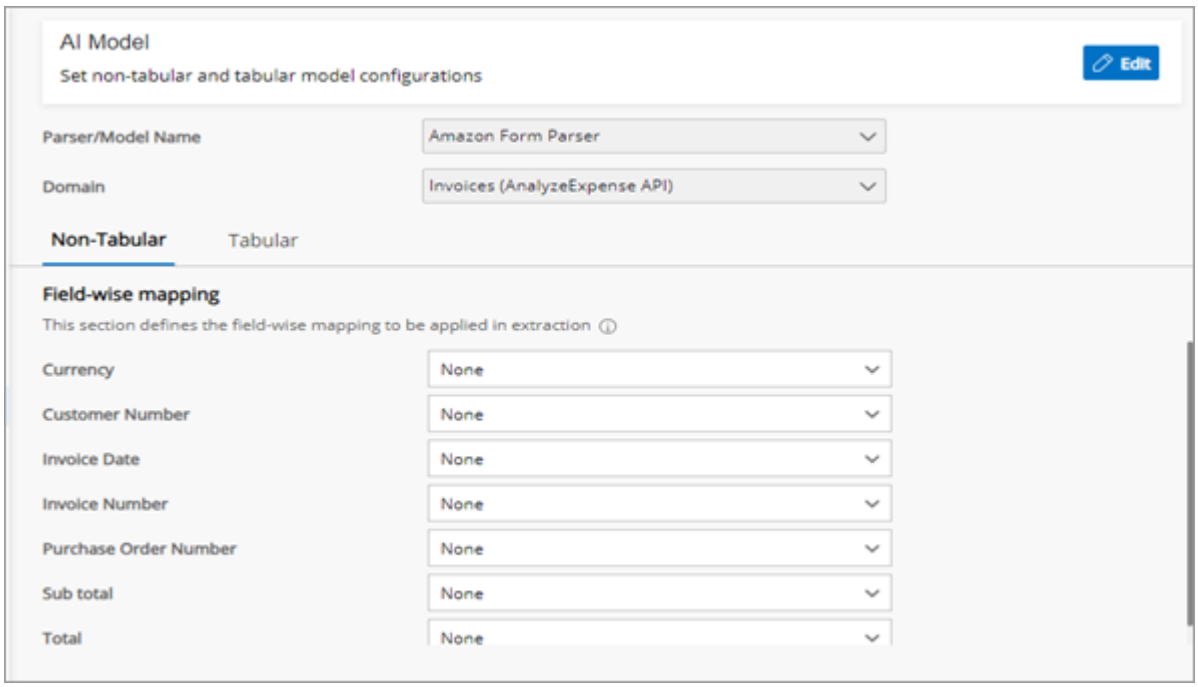
1. Select the required definition from the list of definitions and click **AI Model**. The screen to configure the AI Model for the non-tabular and tabular fields appears.



2. Click **Edit** to enable editing and set configurations for non-tabular and tabular fields on the document.
3. Select the **Parser/Model Name**, from the following options:
   - None
   - Transformer
   - Amazon Form Parser
   - Azure Form Parser
4. If the **Transformer** is selected as parser or model name then on the Non-tabular tab, specify the following details and then proceed to step 5:

Fields	Description
Connect end point	Enter the connection endpoint of the Transformer Service, that is, the IP and Port of the machine where the Transformer API is running. For example, *http://127.0.0.1:8181*. Click **Connect** to connect to the machine.
AI/ML model for non-tabular fields	Select the **AI/ML model for non-tabular fields**. This dropdown gets enabled on the successful connection to the Transformer API machine.
Field-wise mapping	This section appears on selecting the required AI/ML model. The left side shows the non-tabular fields created in the selected definition. The right side shows the field names defined in the selected AI/ML model. Use this section to map fields of the selected definition corresponding to the fields of the selected AI/ML model. Only supported field types such as Alphabetic, Alphanumeric, Amount, Date, Numeric, and StrictAlphanumeric based on the Transformer model are listed for mapping.
Utilize OCR	This function is used for sending data from OmniXtract to the transformer application. This feature is applicable only for Google Vision or Tesseract OCR engines.

In case **Amazon Form Parser** or **Azure Form Parser** is selected. Specify the domain parser invoices and in the Field-wise mapping section for the **Non-tabular tab**, select the AI/ML field against the equivalent definition field. Select **None** if you don't want to map the definition field with the model field.
The OmniXtract rule gets applied on all such fields that are selected as **None**.



In the case of Amazon or Azure form parser, the field-wise mapping are different for each definition.

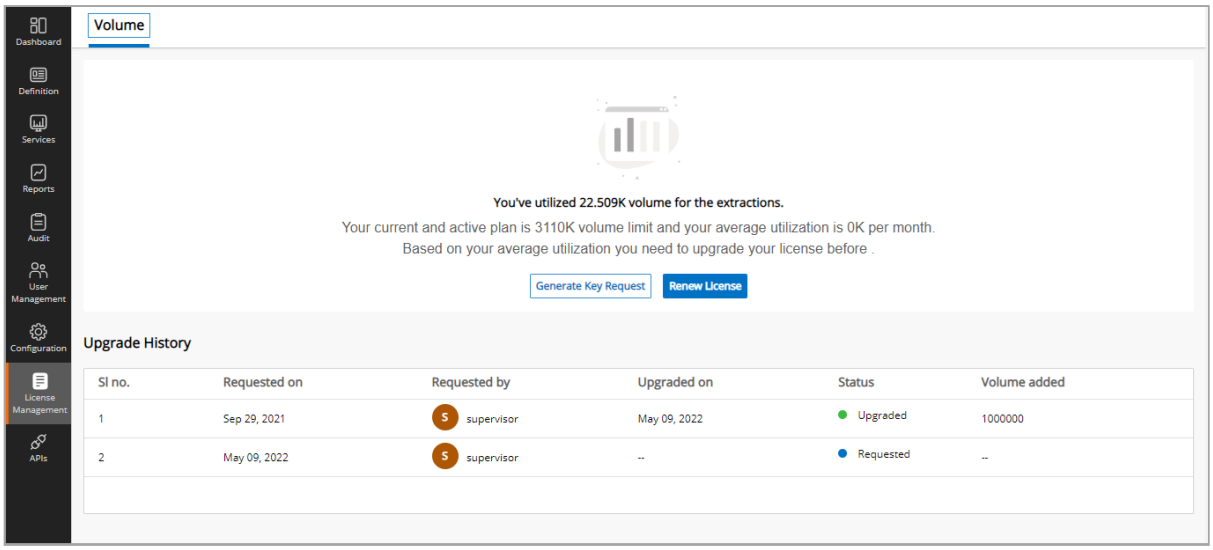
5. Click the **Tabular** tab to set AI/ML configuration for extracting tabular data. Once the configuration is defined, the table detection and extraction happen automatically based on trained AI/ML models specific to the definition.
   The left side of the Tabular tab shows the list of tables defined in the selected definition. The right side shows a dropdown containing table model names defined in the AI/ML model.
6. Select a **Table Model** from the dropdown against the table created in the selected definition. As you select a model type, a list of associated versions for that model appears.
7. Select the required version for the selected Model Type. By default, the latest version is selected.
8. Once you have defined the configurations, click the following options as required:
   - **Save** to save the defined configuration.
   - **Reset to Default** to clear the entered details and restore the last saved values, if any.
   - **Discard Changes** to reject the changes.

License management

The License Management module is accessible by the Supervisor or users having the Admin role assigned to them.

The License Management allows you to manage the license of the OmniXtract system that uses secure and industry standards for data extraction. It allows you to generate a key request for a new license and renew your existing license.

Select **License Management** from the menu bar. The license management screen appears.



The License Management allows you to perform the following actions:

- **View Volume Consumption** – It provides information on the consumption of total volume. It also gives information about your current and active plan of volume.
- **Generate Key Request** – Click here to generate a key request for the license. Enter the purchased license key and click Submit.
- **Renew License** – Click here to renew your purchased license. Enter the Secret Key and click Upgrade to renew the license.

API

The API module is accessible by the Supervisor or users having the Admin role assigned to them.

An application programming interface or API is a set of definitions and protocols used for building and integrating an application software. The configuration of the API provides a central place for modules to store configured data. This data includes the configuration of simple details such as a site name and complex details such as entities, views, and content types.

Select **API** from the menu bar. The API screen appears, showing the following tabs:

- Summary
- Credentials

Summary

The Summary tab shows the summary of APIs with the following details:

- **APIs Traffic and Errors** – Displays the graphical presentation of the total traffic generated by the API and the errors that were hit by the API.

- **API Details** – Displays the API details in a tabular format. It shows the number of requests made and the percentage of errors of a particular API. You can filter the range of data by:
  - Today
  - Yesterday
  - Last week
  - Last month
  - Last 3 months
  - Last year
  - Year
  - Custom range

Credentials

Credentials are used to utilize a particular API. The credentials tab displays the list of credentials of the API.

To create credentials, perform the following steps:

1. Open the Credentials tab and click 

. The Restrict and rename API key screen appears.
2. Enter a **Name** for the API key.
3. Select **Scopes**. You can select all the scopes or only the required scopes.
   As you select the scope, the endpoints assigned for the selected scopes appear in the Endpoints assigned for the scope box.
4. Click 

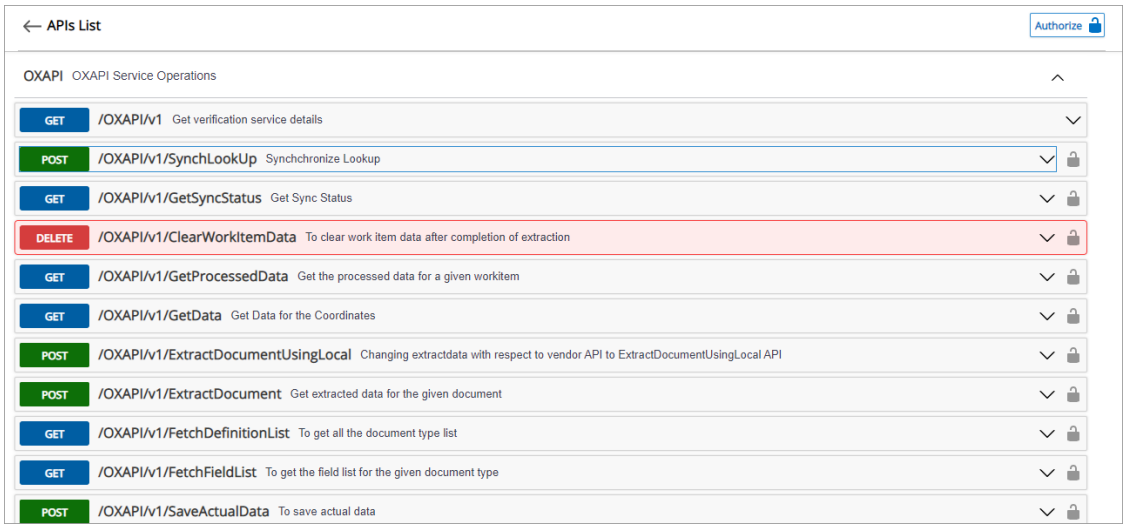
to generate a **Client ID** and **Secret Key**.
5. Click **Save Credentials** to confirm the entered API credentials. A confirmation message for the successful creation of the credentials appears.
   You can perform the following operations on the added credentials:

Operation	Description
Edit	Click **Edit** to modify the API credentials. When you open the API credentials in edit mode, you get an option to regenerate the Secret Key. On clicking **Regenerate**, the existing Secret Key gets replaced by a new key.
Delete	Click **Delete** to delete the API credentials.
Download	Click **Download** to download the API credentials.
Copy	Click **Copy** to copy the credentials on a clipboard.
	This icon means that the credential is not restricted.
	This icon means that the credential is restricted.

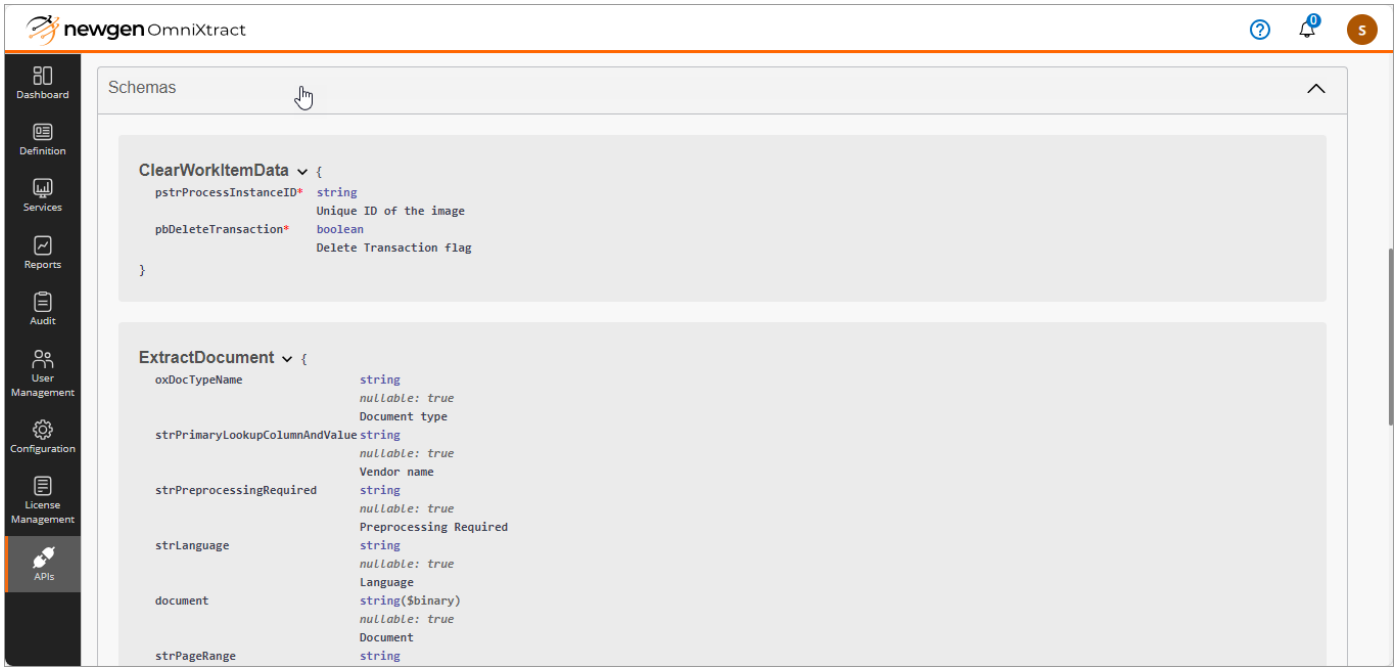
API list

Click the **API List** link at the upper right to get a list of APIs associated with OmniXtract. The APIs List screen appears.

To view details of an API, click the desired API.



You also get an option to try out the API. In addition to the API details, you can view the API schema.



Authorize

The Authorize feature allows you to provide credentials to ensure the secure communication of an API.

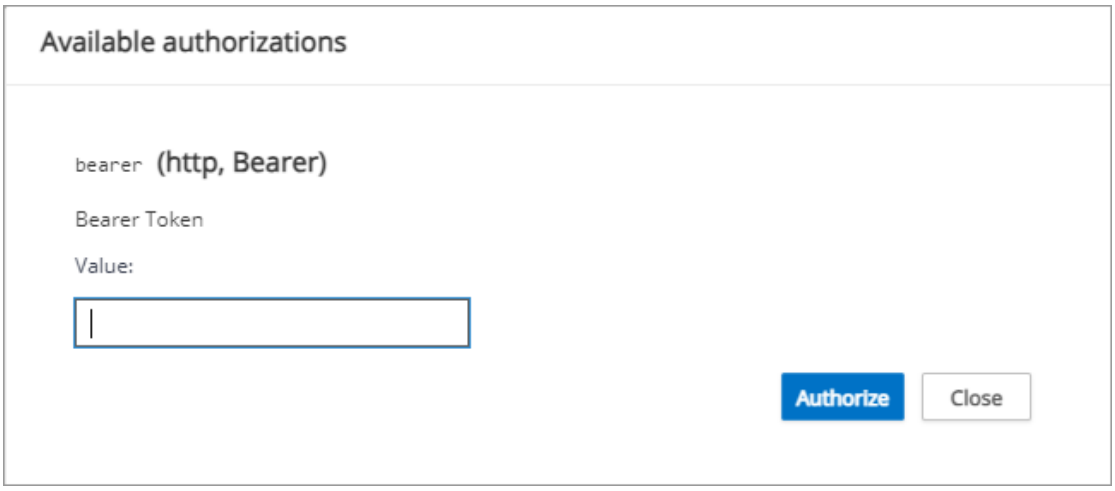
To authorize an API, perform the following steps:

1. Go to the **API List** screen.
2. Click on any of the listed API.
3. Enter the **Client ID, Client Secret Key, Scope,** and **Grant type**.
4. Click **Execute** to run the API.
   On successful running of API, you get a token response.
   For example:

Token response:

```
{
        "access_token":
        "eyJhbGciOiJIUzI1NiIsInR5cCI6ImF0K2p3dCJ9.eyJzdWIiOiI4MjhmNTc4Ny02YzZlLTRlZWMtYjNjM",
        "token_type": "Bearer",
        "expires_in": 3599,
        "refresh_token":
        "eyJhbGciOiJSU0EtT0FFUCIsImVuYyI6IkEyNTZDQkMtSFM1MTIiLCJraWQiOiJVS1BZNVU0TUFISVJV
        NF
        CqEUjFfFRYB54CxECWB5U2X2s.NKbQaHZVLeX_76VzBysiSRJKcEp40ey-06G_NLUINIM"
}
```


5. Copy the access_token value and click **Authorize**. A summary of the authorization details appears.
6. Enter the access token value and click **Authorize**.



7. Click **Close** to close the dialog or click **Logout** to sign out of the authorized access.

Notification

The bell icon  , appearing on the right side of the OmniXtract title bar, acts as a link to the notification. It contains a list of notifications generated on the occurrence of the following events:

- Lookup synchronization
- Accuracy report download
- Master data upload

When there is any notification related to the above-mentioned events, the bell icon appears as  , where the number denotes the total number of uncleared notifications.

To view and clear notifications, perform the following steps:

1. Click the  (**Notification**) icon. The Notification drawer appears.
2. Once you have viewed the notification, click × against that notification to clear it.
3. To clear all notifications at once, click **Clear All**. The cleared notifications are removed from the list and the notification counter gets decreased.
4. To close the Notification drawer, click the  icon again.

Appendix

The information in this appendix contains the frequently asked questions and their answers related to OmniXtract.

Frequently asked questions

1. **What is OmniXtract 4.0 SP1?**
   OmniXtract 4.0 SP1 is a web-based application that swiftly extracts critical data from documents of multiple layouts. It ingests structured and semi-structured documents from various channels and different formats, including PDFs, Tiff, Word, Excel, RTF. It also provides additional support for documents scanned in PNG and JPEG format. It extracts the static data present in documents and makes it readily available for further processing. Additionally, it enables organizations to automate their business processes across various domains while minimizing human intervention, mitigating business risks, achieving higher ROI, and significantly enhancing operational efficiency.
   For the complete details of new features, refer to **OmniXtract 4.0 SP1 Release Notes**.

2. **What type of images that are supported in OmniXtract 4.0 SP1?**
   The image formats that are supported in OmniXtract 4.0 SP1 are TIFF, JPEG, and PNG.

3. **How is OmniXtract 4.0 SP1 different from OmniXtract 3.0?**
   OmniXtract 4.0 SP1 is a web-based application that can be hosted on the cloud or on-premises. The new version brings in features such as creating a collaborative web-based definition, deep learning-based tabular and non-tabular extractions, user activity tracking, roles and rights management, productivity, and scalability enhancement, improved user experience and, integration with Amazon and Azure parser among others.

4. **How is OmniXtract 4.0 SP1 different from OmniXtract 4.0 SP1?**

   OmniXtract 4.0 SP1 brings new features such as integration with Amazon and Azure parser, pre-defined templates for invoice key fields, and line-item extraction. It also provides additional support for documents scanned in PNG and JPEG format along with the latest version of PDFBox and Tesseract OCR.

5. **What type of document can OmniXtract 4.0 SP1 process?**
   OmniXtract can process structured and semi-structured documents. For example, account opening forms, examination answer sheets, invoices, purchase orders, sales orders, and so on.

6. **How is OmniXtract different from OmniScan as both support extraction?**
   OmniScan is a specialized solution for the bulk scanning of documents that has a limited capability of extraction based on a fixed location given for every field. On the other hand, OmniXtract is a specialized solution for the extraction of data from structured and semi-structured documents. It does vendor identification and extracts normal fields as well as tables. Besides this, it has in-built data types to support all your extraction requirements.
7. **How skewed or tilted images are handled?**
   OmniXtract provides an integrated environment for skew correction, auto-orientation, removal of noise, and removal of lines.

8. **Does OmniXtract support enhancement to images? Can OmniXtract convert a scanned image to an acceptable DPI?**
   No, OmniXtract does not support image enhancements. However, some corrections can be made to the uploaded documents. OmniXtract expects scanned images in 300 DPI.
9. **Can OmniXtract extract handwritten text?**
   Yes, OmniXtract supports the recognition of boxed or guided boxed hand-printed data from a structured document. Cursive or running hand-printed data is not recommended for ICR recognition. For more information on handwritten field extraction, refer to the handwritten field.

10. **Can we process vector PDFs with a shaded background?**
    Yes, OmniXtract can extract data from a vector PDF with a shaded background. An image PDF or TIFF file with shaded background is not supported.

11. **Does OmniXtract support password-protected PDFs?**
    Direct processing of any password-protected PDF is not supported. OmniXtract can extract data only from an unprotected PDF.

12. **Can OmniXtract extract data from mobile-captured images?**
    No, OmniXtract cannot extract data from mobile-captured images.
13. **Can OmniXtract extract the signature from the image?**
    Yes, the signature and photo can be captured using zone-based Definition through OmniXtract and can be uploaded as a document in iBPS or NewgenONE. Auto signature or photodetection and recognition are not supported.

14. **Can an automatic extraction model be used on new use cases?**
    Yes, an automatic extraction module can be trained for new use cases. The AI/ML model can be used or trained only on invoice documents.

15. **Is an expert required to create a Definition?**
    Technical expertise is not required to create a Definition in OmniXtract.
16. **Can more than one user create a Definition?**
    Yes, OmniXtract 4.0 SP1 offers a collaborative web-based Definition studio that allows different users to collaborate simultaneously on a single Definition while avoiding conflicts and duplication. Multiple users can work simultaneously on the same Definition but not in the same field. Visit creating a new definition section for more information.

17. **How can the automatic extraction model be trained for new use cases?**
    OmniXtract is shipped with a base model pre-trained on invoices for tabular data extraction. You can train the new use cases with many representative samples

related to the customer to get the best result out of the deep learning model. The AI/ML model can be used or trained only on invoice documents.

18. **What are the factors that affect automatic tabular extraction?**
    A table with a header is required for automatic data extraction. Multiple tables are not supported in automatic tabular extraction. The extraction model comes pre-equipped with an extraction model for invoices. The model can be easily trained to adapt to different use cases.

19. **Can I monitor the changes in OmniXtract?**
    Yes, all the changes made in any module of OmniXtract are logged and can be viewed in the Audit module.

20. **What if the automatic extraction from the table fails?**
    In case the automatic tabular extraction fails, you can always fall back to the traditional pattern-based algorithm to extract data.
21. **Do I have to create a Definition for each vendor?**
    You need to create a global Definition and if needed, a local vendor-specific Definition can be created for a field.

22. **Is there any limit on the number of pages to be extracted from a document?**
    No, there is no limit to number of pages to be extracted. However, there is a provision to set the page limit for extraction, to optimize the processing of documents more efficiently.

23. **Is data extraction possible from line items?**
    Yes, you can extract data from line items with line-item data stretching to multiple pages.

24. **Is there any limit on the number of line-item records for extraction?**
    No, there is no such limit on the number of records of line items. You can extract line items from all pages.

25. **Is there any default limit on the number of characters within a column of a table?**
    No, there is no such limit, but you need to make sure that your buffer (Queue variable or external table column) size in iBPS or OmniFlow is enough to accommodate it as it is finally pushed to iBPS or OmniFlow.

26. **What recognition types are supported in OmniXtract?**
    The following recognition types are supported in OmniXtract:
    - Machine printed (OCR)
    - Hand-printed (ICR) (zone-based)
    - Tick marks or OMR (zone-based)
    - MICR
    - MRZ
    - Barcode (1D and 2D barcode)
    - Image clips

27. **Can data be extracted from multiple tables present in the document?**
    You can use pattern-based extraction to extract data from multiple table structures in the document.

28. **Can I format data using OmniXtract?**
    Yes, you can format extracted data based on its data type. For example, the Date and Amount can be displayed in the required format.

29. **Can the data be corrected using the OmniXtract interface?**

    No, OmniXtract extracts data from the image. Data correction is done via iBPS or NewgenONE.

30. **Can one create a global or local Definition using iBPS or NewgenONE?**
    You cannot create a global or local Definition using iBPS or NewgenONE. You can create a local rule automatically and save it for the field in OmniXtract based on

the data extracted. This rule is used for processing the field when another document for the same vendor is processed next time by OmniXtract.

31. **What are the broad-level challenges in vector PDFs?**
    The quality of recognizing the non-OCR documents is a broad-level challenge. It is due to the extraction of data from non-OCR documents mainly depends on document complexity and the availability of labels to identify the key-value pair.

32. **How to process multiple documents in a single workitem?**
    In OmniXtract, different Definitions coexist for each document type. Multiple document types can be uploaded in a single work item that is mapped with OmniXtract document types in OmniXtract services.