newgen

NewgenONE OmniXtract
for Intelligent Data Extraction

NewgenONE OmniXtract

NewgenONE OmniXtract enables enterprises to extract critical information from documents of multiple layouts. It allows data capture from structured and semi-structured documents generated through various channels and formats, including PDFs, TIFF, MS Word, MS Excel, and RTF, and makes it available for further processing.

Using Newgen OmniXtract, organizations can automate their document-centric processes, reducing human intervention and enhancing operational efficiency. By leveraging its AI/ML capabilities and pre-built industry-specific models, enterprises can enable intelligent decision-making.

Business Benefits

- **Smarter Decision-making**

  Enable smarter decision-making, driven with contextual data. Deliver personalized, enhanced customer experience

- **Higher Operational Efficiency**

  Ensure first-time-right with human-in-the-loop verification. Enable accurate, faster data entry through AI/ML-enabled data extraction

- **Minimized Business Risks**

  Mitigate business risks with access to error-free data in real-time. Witness quick, accurate data extraction and verification via document layout learning

- **Higher ROI**

  Avoid cost-intensive multiple data extraction utilities. Yield higher returns and bring down investment costs with a single extraction product

- **Faster Results**

  Extract data from invoices instantly using the AI/ML model. Eliminate the hassle to configure templates

OmniXtract Features

Ingest

▪ **Document Ingestion:**

Ingest documents and images in multiple formats, from different content sources
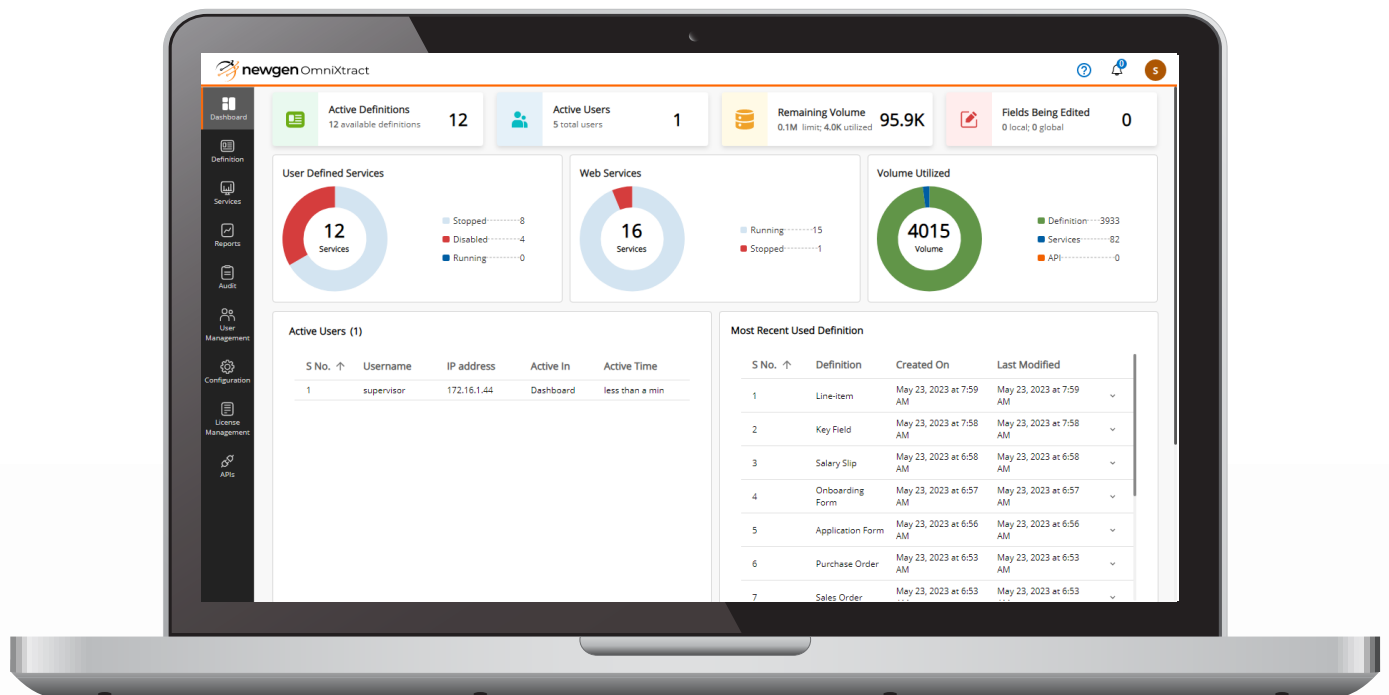
- Support for multiple formats, such as PDF, TIFF, MS Word, MS Excel, and RTF
- Support for various content sources,including multi-functional printers, scanners, enterprise applications, and repositories

Process

▪ **Document Pre-processing:**

Improve quality of a single page or multiple pages scanned documents in real-time with automatic detection and correction of image distortions

- Enable color and grayscale imaging
- Allow auto-correction of document orientation
- Enable skew correction
- Remove background noise automatically

Snapshot of key performance indicators

Extract

- **AI/ML-based Extraction:**

  Ensure automated data extraction from structured and semi-structured documents using AI/ML

  - Extract data using pre-built industry-specific models
  - Enable continuous verification and training leveraging AI/ML models with human-in-the-loop review system

- **Multiple Data Formats Extraction:**

  Extract complex data formats from documents of different layouts

  - Recognize data from multiple table structures
  - Extract hand-written text
  - Enable multi-line data extraction
  - Support for multiple languages
  - Leverage predefined data formats, including URL, address, and date

- **Web-based Verification:**

  Assist users with automated interfaces for speedy extraction and verification of unrecognized and recognized data

  - Enable one-click extraction even for undefined labels
  - Ensure high-speed manualverification with intuitive navigation

- **Data Definition:**

  Leverage multi-user definition studio with a no-code interface for defining multiple document types with different layouts

  - Predefine data types and plain text labels
  - Fine-tune data definition and labels
  - Replicate data definitions and labels to other environments
  - Provide pre-configured documenttypes from various verticals to enable quicker implementation
  - Collaborate simultaneously on a single definition while avoiding conflicts and duplication
  - Visualize applied parameters, labels, and section headers to create a simplified definition

- **Auto-identification and Recognition Technologies:**

  Ensure seamless data extraction by leveraging advanced technologies, including Intelligent Character Recognition (ICR), Optical Mark Recognition (OMR), and Optical Character Recognition (OCR)

  - Recognize barcodes (USS 39, CODE 128, QR code, PDF 417, and data matrix)
  - Support for Magnetic Ink Character Recognition (MICR)
  - Support for Machine Readable Zone (MRZ) detection

- **Extraction Services:**

  Extract data from large volume of documents across business processes

  - Allow scalable extraction at production stage
  - Run multiple extraction services
  - simultaneously
  - Push extracted data into iBPS in the required format
  - Leverage diagnostic capabilities to instantly troubleshoot in case of extraction failure

- **Security and Audit:**

  Ensure secure content access while complying with regulatory requirements

  - Track users' activities across all extraction modules for better audit compliance
  - Support for audio and visual CAPTCHA for enhanced security
  - Restrict data access with roles and rights management

Enhance

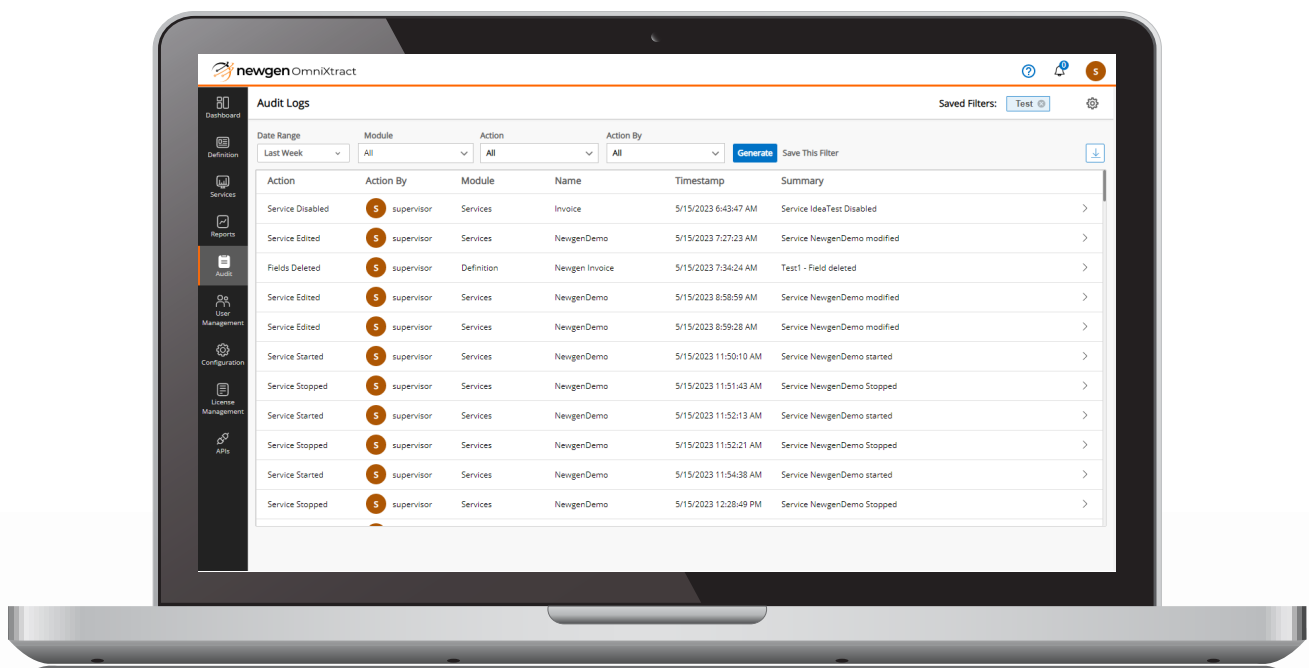- **Document Layout Learning:**

  Enhance accuracy of data extraction continuously by learning through processed documents

  - Optimize definitions to improve accuracy of results

- **Data Extraction Analytics:**

  Monitor extraction process and accuracy rates. Get visual indications to find data inaccuracy and make corrections

  - Enhance extraction performance through historical data analysis
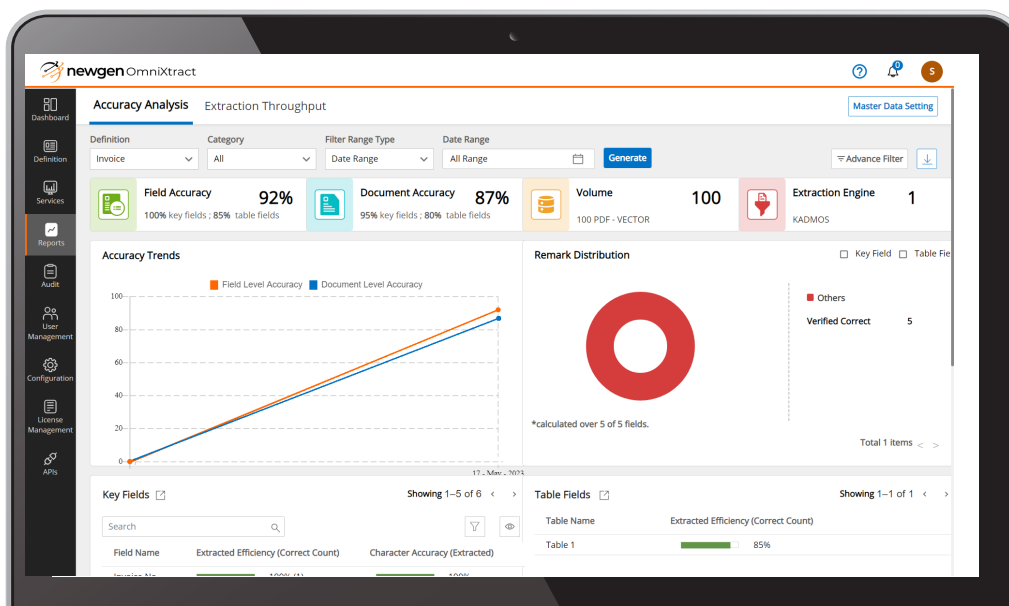 - Check accuracy of extracted data via system generated confidence rate

Audit logs to track user activities

Reports and Visualization:

Gain critical extraction insights across various stages of the extraction journey

- Gain access to contextual reports and dashboard
- Get image-assisted output analysis with visualization of actionable items for guided definition optimization
- Monitor throughput and trend reports
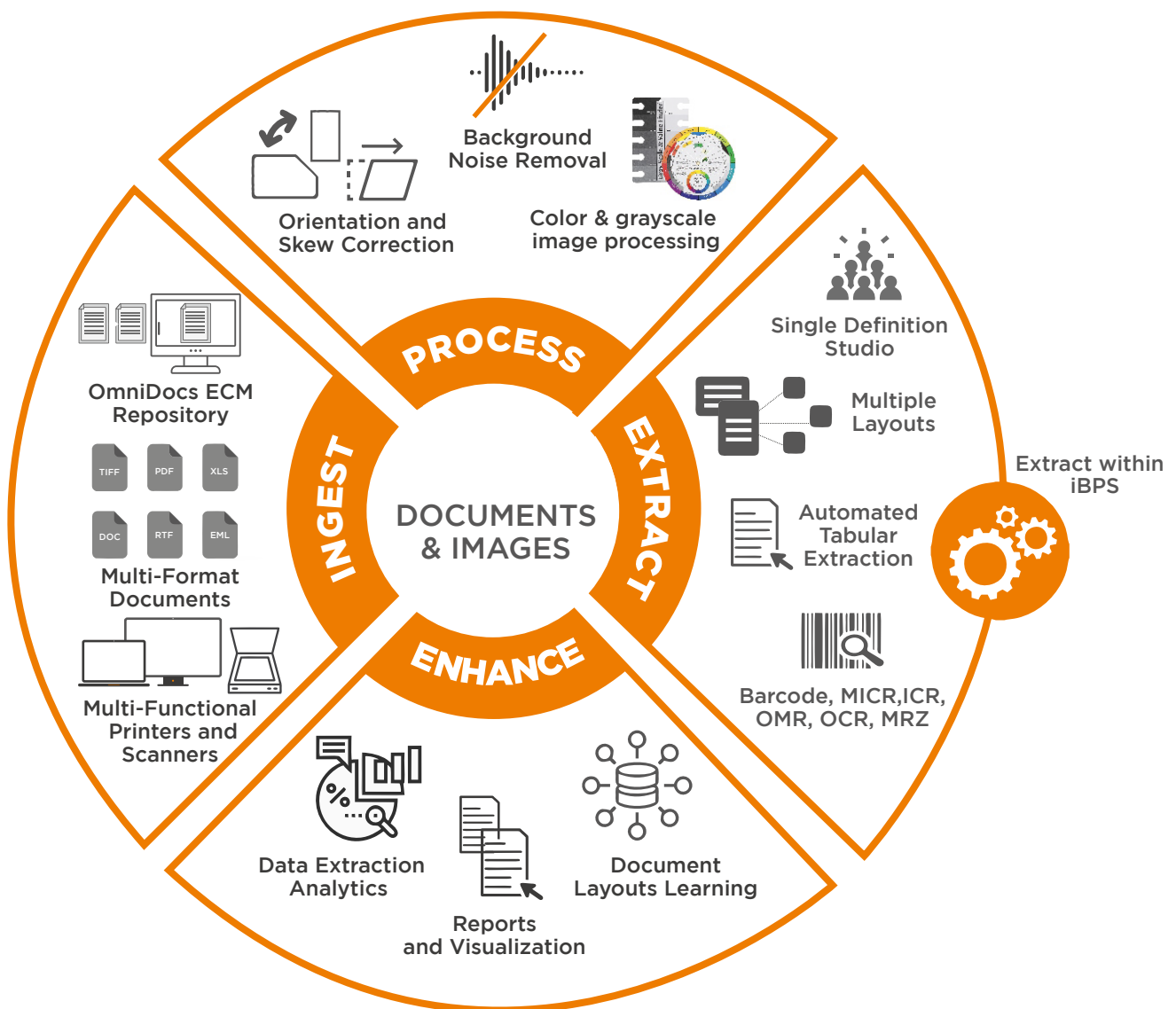- Monitor real-time accuracy and system performance through dashboards

Real-time accuracy reports and trend analysis

Why NewgenONE OmniXtract?

- Digitize documents of multiple layouts through a single interface

- Accelerate business operations with instant data extraction and verification

- Ensure accurate data extraction through continuous learning

- Scale rapidly to handle the burgeoning volume of documents

- Extract data from invoices using pre-trained AI/ML models

- Improve performance through a hybrid data extraction approach

Background Noise Removal

Orientation and Skew Correction

Color & grayscale image processing

Single Definition Studio

Multiple Layouts

Extract within iBPS

OmniDocs ECM Repository

TIFF PDF XLS

DOC RTF EML

Multi-Format Documents

Automated Tabular Extraction

PROCESS

EXTRACT

INGEST

ENHANCE

DOCUMENTS & IMAGES

Multi-Functional Printers and Scanners

Barcode, MICR,ICR, OMR, OCR, MRZ

Data Extraction Analytics

Reports and Visualization

Document Layouts Learning

NewgenONE OmniXtract Framework

"Our clients process documents to the scale of 10-50 million documents per year. OmniXtract is scalable and can handle the growing volumes of documents.

OmniXtract Product Head

NewgenONE OmniXtract for Industry Specific Document Types

Cross Industry

- Invoices
- Purchasing orders
- Appraisal forms
- Sales orders
- Expense forms
- Surveys
- Questionnaires
- Salary slips
- Answer sheets
- Onboarding forms
- Claims

Banking and Insurance

- Loan processing forms
- Cheques
- Bank statements
- KYC forms
- Account opening forms
- Credit card applications
- Insurance forms
- Proposal forms

Government

- Application forms
- Passports
- Payment receipts

About Newgen

Newgen is the leading provider of a unified digital transformation platform with native process automation, content services, communication management, and AI/ML capabilities. Globally, successful enterprises rely on Newgen's industry-recognized low code application platform to develop and deploy complex, content-driven, and customer-engaging business applications on the cloud. From onboarding to service requests, lending to underwriting, and for many more use cases across industries, Newgen unlocks simple with speed and agility.

newgen

May 2023