

# DATAMATICS

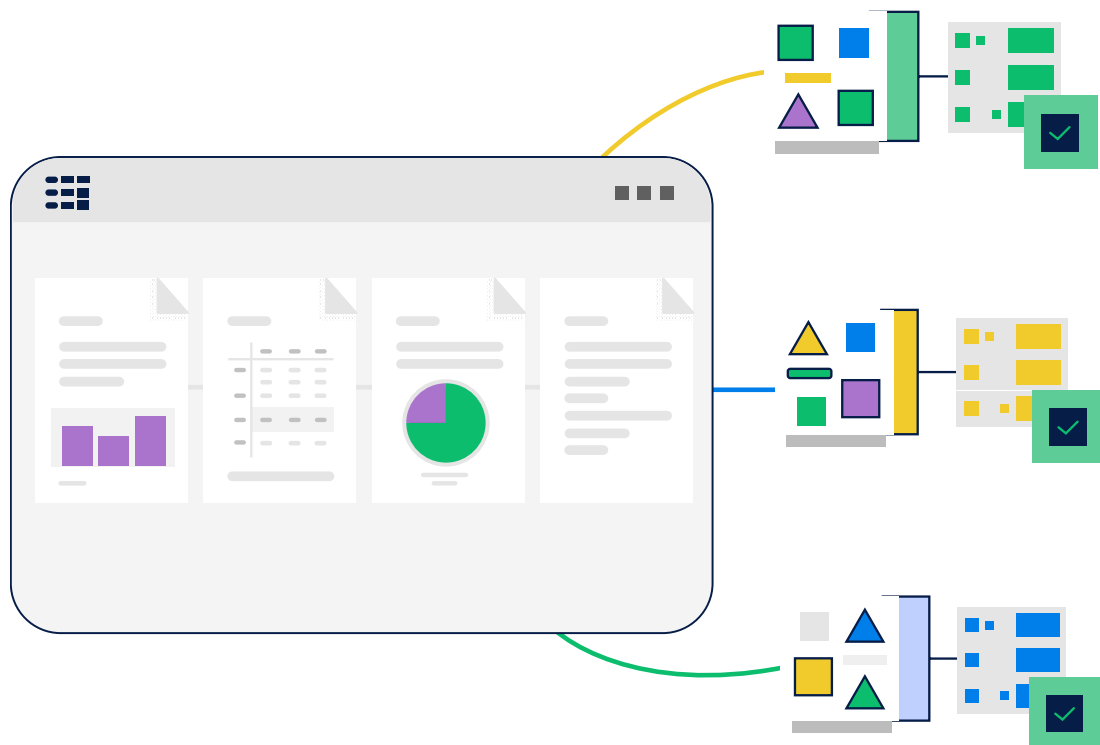


**WHITEPAPER**

---

## **INTELLIGENT DOCUMENT PROCESSING WITH TEMPLATE-FREE APPROACH**

As businesses look forward to drive innovation-led growth and yet be ready to tide over the next disruption, process automation is the new frontier. Intelligent Document Processing and Robotics driven Intelligent Automation is set to automate all paper-intensive and repetitive manual tasks. Most businesses that focus on employee engagement, talent retention, and scaling up operations with existing resources in disruptive markets have already adopted process automation in order to positively influence both the internal and external customers. This paper details how Intelligent Document Processing with its template-free approach brings free-text documents in the ambit of automation.



Intelligent Document Processing, which is an inherent component of Intelligent Automation or Cognitive Automation, improves enterprise agility and productivity. It accelerates and scales up operations involving documents with free-text or unstructured data. It supports business agility through an innovative approach and provides an improved business user experience and customer satisfaction.

## Why is Intelligent Document Processing – led Intelligent Automation important today?

Intelligent Document Processing-led Intelligent Automation is essentially a low code/no code automation platform. It brings unstructured data and free-text in the purview of end-to-end automation. It extracts unstructured data from documents and makes it usable for integration with business systems for further processing. It improves the sense of employee engagement as business users are able to automate their own processes. The solution is scalable and deployable on-premise as well as on cloud. The technology enables teams to work in collaboration across geographies that in effect exponentially improves business productivity and efficiency. Intelligent Document Processing is an important component of the Intelligent Automation ensemble that makes end-to-end automation a business reality.

## What does Intelligent Document Processing encompass?

Intelligent Document Processing leverages technologies such as Artificial Intelligence/Machine Learning (AI/ML) and Natural Language Processing (NLP) to read and categorize different types of unstructured or free-text documents, extract data from them, validate the extracted data, convert it into structured or tabular data that is easily usable for downstream processing and integration with core business applications. Free-text documents include handwritten or printed paper documents and electronic documents, such as email, PDF, web pages in public domain, and social media that contains audio, video, text, and images.

## What are the key tasks undertaken by Intelligent Document Processing?

- 1 Propagate a template-free approach to process multiple document types
  - 2 Pre-process images of the documents to improve image quality
  - 3 Post-process the extracted data to improve accuracy of data capture
  - 4 Deliver multi-OCR compatibility to select the best probable image output and improve data accuracy
-

- 5 Convert images to text and then to information
- 6 Use human-in-the-loop capability to validate data
- 7 Monitor the overall data extraction process with Analytics-powered dashboard
- 8 Automate paper-based workflows
- 9 Recognize handwritten text and bring it in the automation ambit

## Why template-free approach is important in Intelligent Document Processing?

Task of creating and maintaining different templates for different documents is tedious and time-consuming. Further each time a document layout changes, the system requires changing the template itself, thus making the entire effort of automation redundant. This scenario necessitates the requirement of an AI/ML-driven, template-free approach for document processing, thus making it intelligent.

**Intelligent Document Processing** uses AI/ML algorithms and models to recognize keywords, and their synonyms, usually found at a particular location in a document and extract the data associated with that keyword/field. The models become more intelligent with each batch-load processing thus improving accuracy and speed. The template-free approach eliminates the need to update the rules associated with the templates frequently. It enables end-to-end automation in the **Intelligent Automation** paradigm by bringing paper and unstructured data in the automation purview.

AI/ML-enabled low code/no code technology platforms, such as Intelligent Document Processing, allows businesses to empower their employees to automate their repetitive tasks. These intelligent platforms become more intelligent over the time through repeated batch processing and exception handling. They auto-correct errors and improve the accuracy, straight through processing, and workload throughput. The AI/ML-powered template-free approach eliminates the requirement for creation of new templates and updation of business rules every time the document format changes.

## How does Human-in-the-Loop capability enable Intelligent Document Processing?

Human-in-the-Loop (HITL) capability empowers the business users to train the Intelligent Document Processing platform through exception handling. In case the confidence level of the extracted data is less than the pre-configured threshold, then the HITL interface allows the business users to key in the right data input. The Intelligent Document Processing platform continuously learns through the HITL interface and generates a highly accurate and a faster output in the next batch processing cycles. HITL is an apt interface to continuously train the AI/ML models so that it tends to achieve 99.99% accuracy for the extracted data and improves the end-to-end straight through processing of the entire Intelligent Automation ensemble.

## What are the drivers for Intelligent Document Processing adoption?

Nearly 90% of the data that businesses generate or use is in unstructured format. The businesses need to repeatedly key in this data as and when required for other enterprise solutions. Similarly, manual classification of the documents in unstructured format takes much time.

Similarly, data is searchable, routable across stakeholders, and capable of insights generation only when it is in structured format or a tabular format. Data in unstructured format is unusable and it loses correctness due to errors while rekeying it in different systems. These and many such reasons necessitate the institutionalization of Intelligent Document Processing.

Intelligent Document Processing enables businesses to improve employee, stakeholders, and customer satisfaction by manifolds due to faster turnaround time and closure of processes. It improves auto-routing of data between different stakeholders for decision-making and approval management through integration with workflows. It increases data visibility and accountability. It helps to fulfill compliances and reduce business risk. It improves productivity and efficiency and optimizes time-to-value.

## What are the primary benefits of Intelligent Document Processing over OCR?

AI/ML-enabled, template-free Intelligent Document Processing eliminates the requirement to maintain multiple templates and their upgrades due to document format change. It frees up employees from repetitive data-entry, maintenance of templates and business rules. It improves employee engagement and customer satisfaction due to elimination of long winding cycle times and improved straight through processing in important business processes.

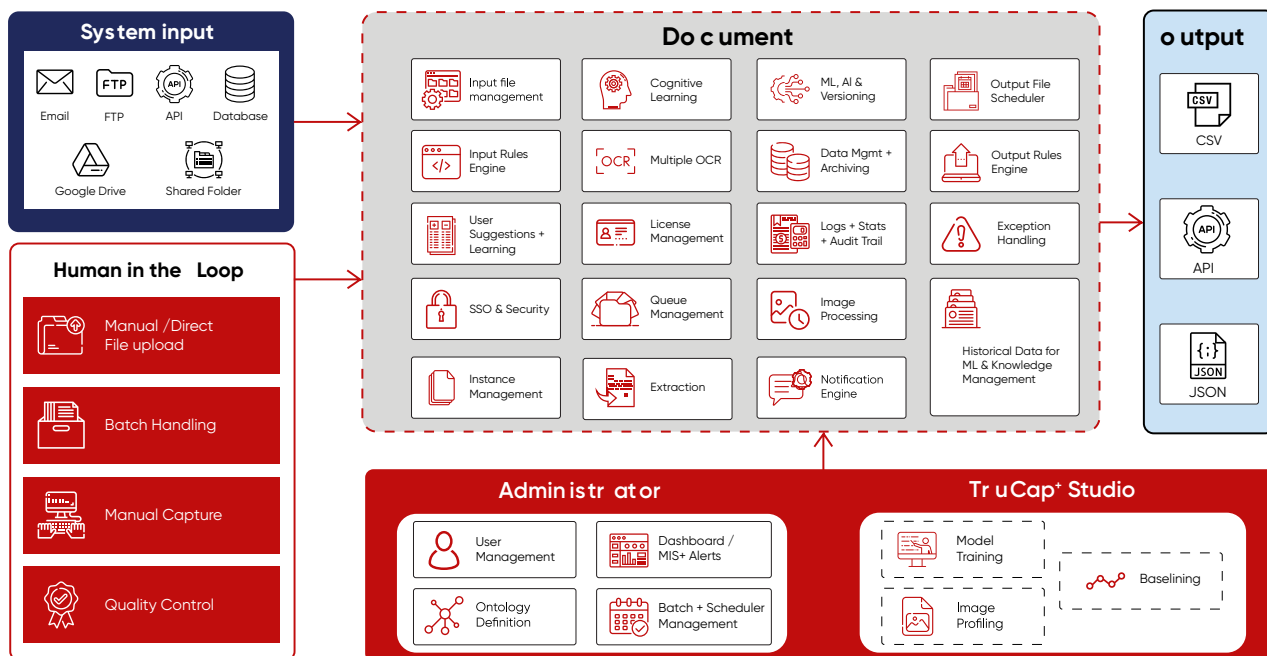
With faster onboarding and improved business agility, Intelligent Document Processing reduces the total cost of ownership of the enterprise infrastructure and process management. It enables data sanitization and validation within a stack of documents through both supervised and unsupervised learning models. It's preprocessing, post-processing, and multi-OCR features improve the data extraction accuracy by manifolds with both electronic and non-electronic/handwritten documents.

## What is TruCap+ Intelligent Document Processing?

TruCap+ is an enterprise-grade, AI/ML-powered, and template-free Intelligent Document Processing proprietary solution. With a facility to deploy on-premise and on-cloud, it offers features for different personas – admin, subject matter expert, operations manager, and operator, with different levels of access. It allows business users to incorporate popular business keywords and their synonyms to configure the solution to read different types of documents.

The convergence of the AI/ML models, keywords, and business rules into one Intelligent Document Processing solution enables the TruCap+ to extract unstructured data from documents in a template-free mode. TruCap+ enables faster time to value and lowers the total cost of ownership of the entire business infrastructure.

TruCap+ has a user friendly interface that allows business users to build functions atop the pre-configured functions. It has high data extraction accuracy at field-level and a HITL functionality to manually intervene when the pre-defined confidence-level is low for the extracted data fields. Each HITL intervention trains the AI/ML model to improve accuracy and hence the throughput in the next cycle.



## What are the benefits of TruCap+ Intelligent Document Processing?

TruCap+ Intelligent Document Processing brings automation solution to the finger-tips of business users thus improving employee engagement by manifolds. The intuitiveness of the TruCap+ interface supports faster time-to-value and hence a better adoption across industry. The important benefits of TruCap+ Intelligent Document Processing are:

### 01 | Provides faster time-to-automation

With a user friendly interface and drag-and-drop features, TruCap+ supports faster deployment and adoption and an equally faster RoI.

### 02 | Improves straight through processing (STP)

TruCap+ has in-built pre-processing and post-processing features for image enhancement and extracted data accuracy improvement, respectively. Optimization of the image and extracted data eliminates data entry, manual supervision/intervention, and rework thus enabling faster downstream processing.

## **03 | Improves engagement and satisfaction**

With improved speed of workload execution and hence the reduced turnaround time, both the internal and external customers/stakeholders promote the business in their circles thus exhibiting ownership, engagement, and satisfaction.

## **04 | Improves specialized processing**

TruCap+ enables automated logo detection to support data/document classification, table detection and data extraction, and much more.

## **05 | Offers a universal solution**

The template-free of TruCap+ enables data extraction and processing from structured, semi-structured, unstructured, and other different document types, including free-text in images, video, audio, social media, streaming data, hand-written forms, etc.

## **06 | Convergence of multi-OCR output towards higher accuracy**

TruCap+ has an in-built OCR and also supports other multiple OCR solutions, such that it picks the most probable OCR output and offers higher data-extraction accuracy.

## **07 | Validates inter and intra-document sanctity**

AI/ML models have the unique ability to check the sanctity of personal identities and other parameters within a stack of documents related to a particular person or a case. For example: Congruence of the Social Security number across the driver's license, passport, and wage stub.

## **08 | Reduces time-to-value**

The intuitive interface supported by a low code/no code interface empowers business users to automate complex and paper-intensive workflows within a matter of hours thus reducing time-to-value.

## **09 | Lowers total cost of ownership of the entire IT infra**

TruCap+ reduces the TCO of the entire IT infra and processes through elimination of data entry and paper-forms with propagation of end-to-end automation at enterprise-level.



## **10 | Continuous training of AI/ML models**

Each HITL instance improves the algorithmic intelligence further towards a more intelligent output in the next cycle, with more throughput and higher accuracy across the automation ensemble.

## **11 | Eliminates manual intervention over the time**

TruCap+ supports auto-splitting of lengthy consolidated documents at logical nodes and auto-classification of the document fragments/documents for easy storage and retrieval in data management solutions.

## **12 | Offers in-built analytical overview of data extraction program**

TruCap+ has in-built analytics that provides program-level insights through scheduled reports. It offers integration with high-end analytics solutions through APIs for continuous evaluation and better decision making towards process improvement.

## **13 | Flexible deployment**

TruCap+ is deployable on-cloud and on-premise. It can be leveraged across a hybrid environment to automate processes involving distant geographies.

## **14 | Flexible tiered pricing**

Offers pricing based on number of pages with the option to process additional volumes, as required.

## **15 | Offers end-to-end automation**

TruCap+ brings free-text in the purview of automation as well as integration with enterprise systems and automation solutions, such as RPA, workflow, etc.

# What are the important use cases that TruCap+ Intelligent Document Processing delivers?

TruCap+ enables businesses to dynamically handle high volume and document-intensive processes, categorization of documents, operator training, and institutionalize industry best practices.

With its template-free approach, TruCap+ enables a number of enterprise automation use cases thus driving end-to-end automation across horizontals and verticals –

## 01 | **Accounts Payable/Procure-to-Pay Automation**

It enables data extraction from unstructured and semi-structured documents, such as invoices of different vendors and purchase orders, and integration with Enterprise Accounts Payable systems.

## 02 | **Insurance Life-cycle Automation**

It enables extraction of data from lengthy documents, such as contracts, agreements, policies, claims, etc., thus reducing the cycle time.

## 03 | **KYC Automation**

It enables extraction of data from lengthy documents, such as contracts, agreements, policies, claims, etc. thus reducing the cycle time.

## 04 | **Order management/Order-to-Cash Automation**

It enables data extraction from paper-form orders and integrates the data with the core order management system for processing.

## 05 | **EXIM Order Processing**

It facilitates data capture from EXIM documents to integrate it with the core Trade Finance automation solutions of banks and non-banking financial institutions.

## 06 | **Loan and Mortgage Processing Automation**

It enables data extraction from loan document forms and mortgage documents in semi-structured and handwritten formats, converts it to a structured format, and integrates it with the core systems for further processing.

## 07 | **Vendor Platform Automation**

It extracts data from the vendor quotes and applications and automates the vendor reconciliation for small to large enterprises.

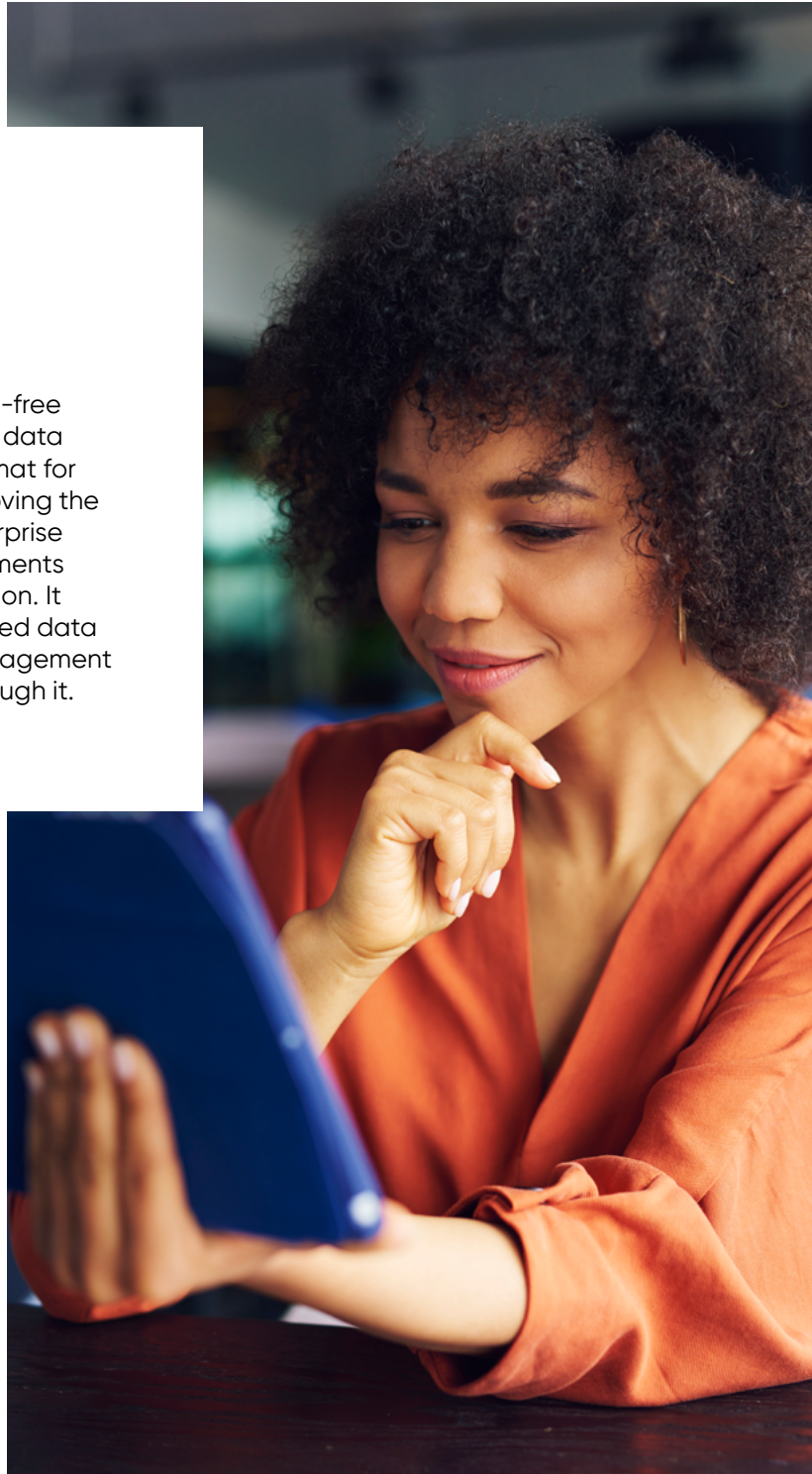
## 08 | **Bill of Lading Processing**

It automates the data extraction process in document-intensive environments, such as bill of lading processing, manufacturing, supply chain and logistics, etc.

# SIMPLY PUT

---

Intelligent Document Processing with its template-free approach offers agility in extracting unstructured data from documents, converting it into structured format for integration with core business systems, and improving the cycle time. It is important to the end-to-end enterprise automation programs as it brings free-text documents and unstructured data in the purview of automation. It eliminates manual data entry and makes structured data available for automation, enterprise content management for insights generation, and decision-making through it.



# ABOUT THE AUTHOR

---

## **SHASHI BHARGAVA**

Executive Vice President  
& Head - Solutions Group

Shashi heads the product team at Datamatics. He is responsible for proactively identifying and defining IT solutions encompassing the latest technologies in new areas for clients across verticals. He is also leading the product team for IP development in the areas of RPA, IDP and BI & Analytics.

He has over 33 years of extensive experience in managing offshore centers, delivering software services to clients across the world on multiple technologies and various domains. He has successfully led organizational growth through P&L management, client engagement, pre-sales support, product design & development, and building high-performance teams with demonstrable productivity. In his previous role, Shashi has been instrumental in new areas of growth and setup multiple "Centre of Excellence" to drive revenue and business success.

Shashi holds a Bachelor's Degree in Electronics and Communication Engineering from IIT Varanasi.

# ABOUT DATAMATICS

---

Datamatics provides intelligent solutions for data-driven businesses to increase productivity and enhance the customer experience. With a complete digital approach, Datamatics portfolio spans across Information Technology Services, Business Process Management, Engineering Services and Big Data & Analytics all powered by Artificial Intelligence.

It has established products in Robotic Process Automation, Intelligent Document Processing, Business Intelligence and Automated Fare Collection.

Datamatics services global customers across Banking, Financial Services, Insurance, Healthcare, Manufacturing, International Organizations, and Media & Publishing.

The Company has presence across 4 continents with major delivery centers in the USA, India, and Philippines. To know more about Datamatics, visit [www.datamatics.com](http://www.datamatics.com)

## FOLLOW US ON

© Copyright 2023 Datamatics Global Services Limited and its subsidiaries (hereinafter jointly referred as Datamatics). All rights reserved.  
Datamatics is a registered trademark of Datamatics Global Services Limited in several countries all over the world. Contents in this document are proprietary to Datamatics.  
No part of this document should be reproduced, published, transmitted or distributed in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, nor should be disclosed to third parties without prior written approval from the marketing team at Datamatics.

website: [datamatics.com](http://datamatics.com) | email: [business@datamatics.com](mailto:business@datamatics.com)

USA

UK

UAE

India

Philippines