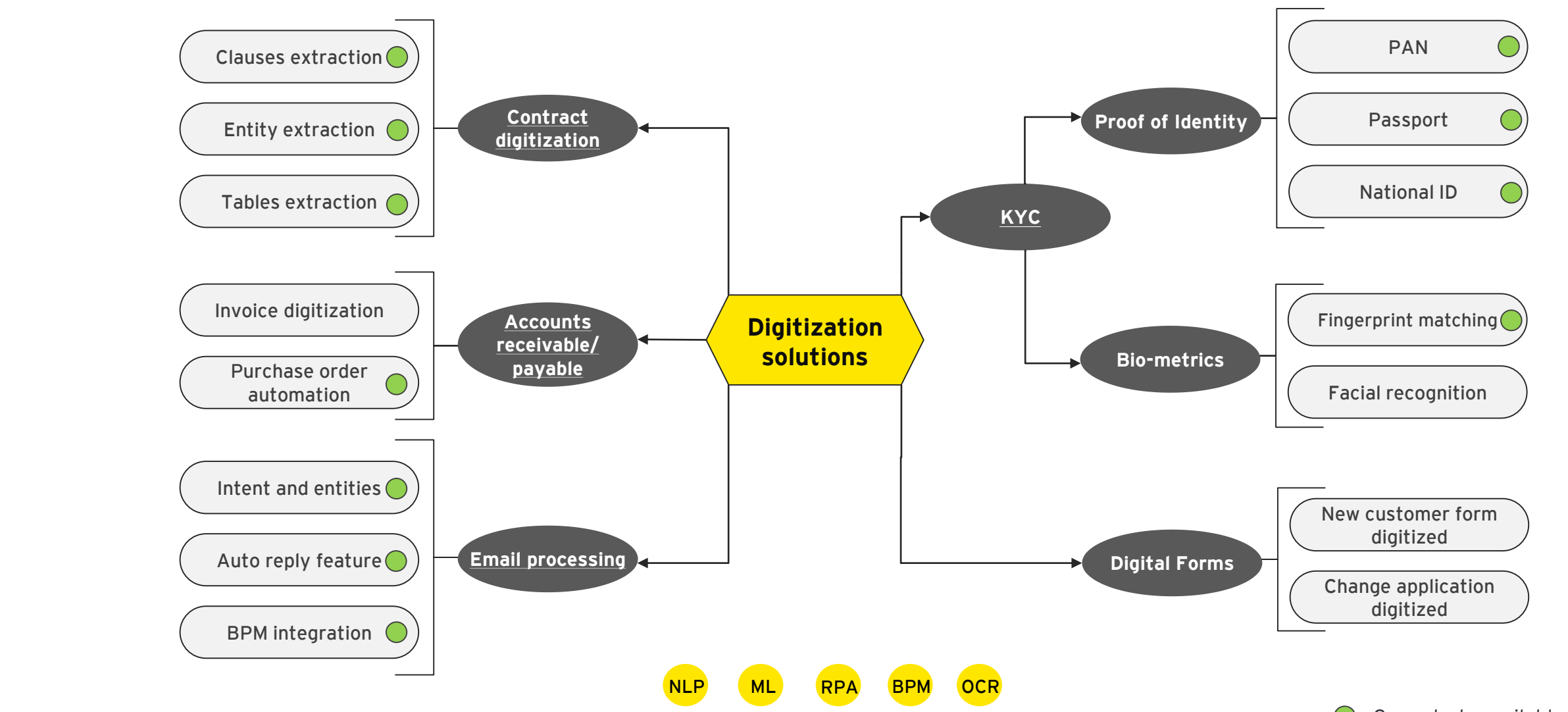
Invoice / Purchase order processing

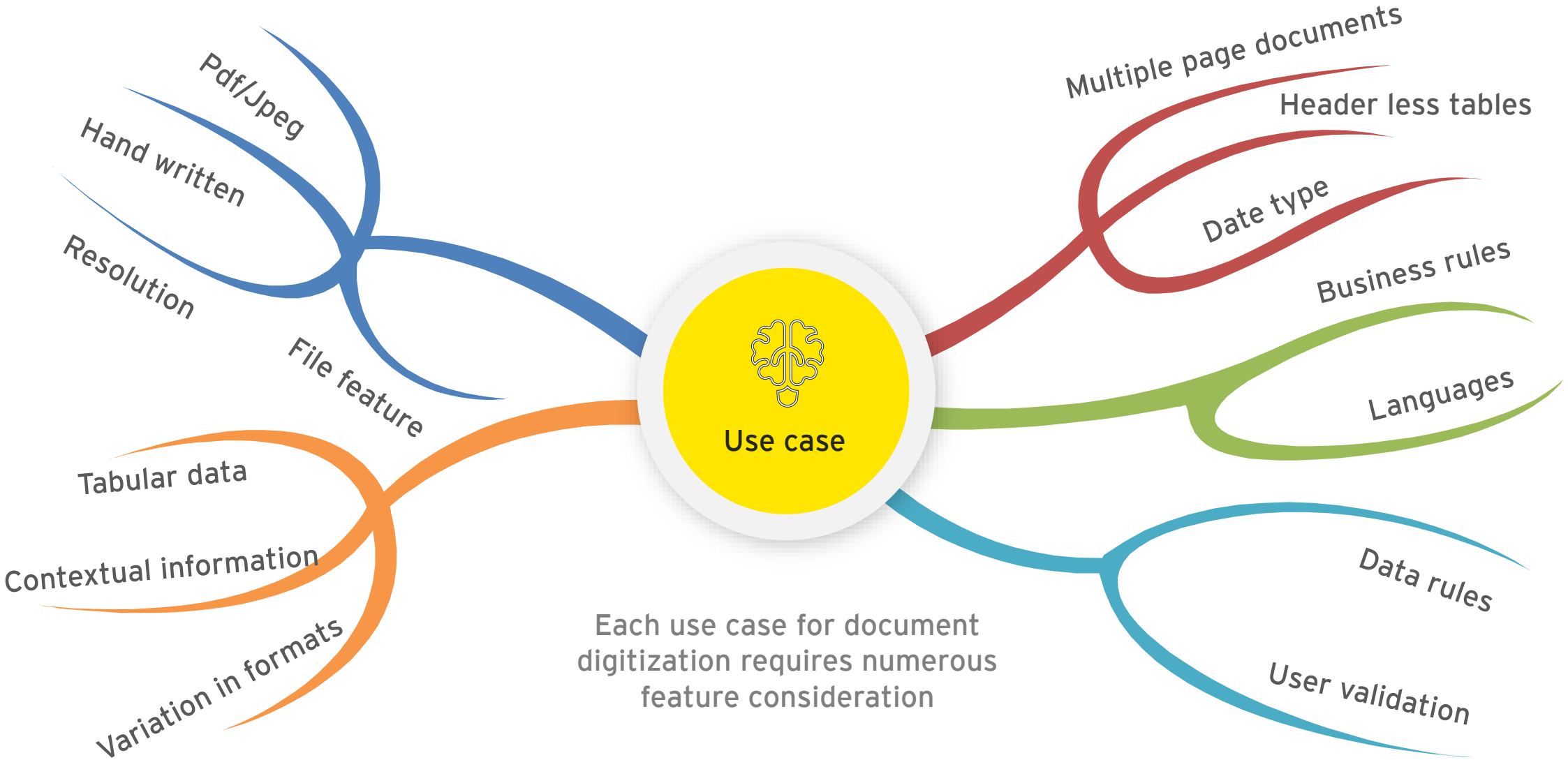
The better the question. The better the answer.
The better the world works.

Building a better working world

Data digitization solutions

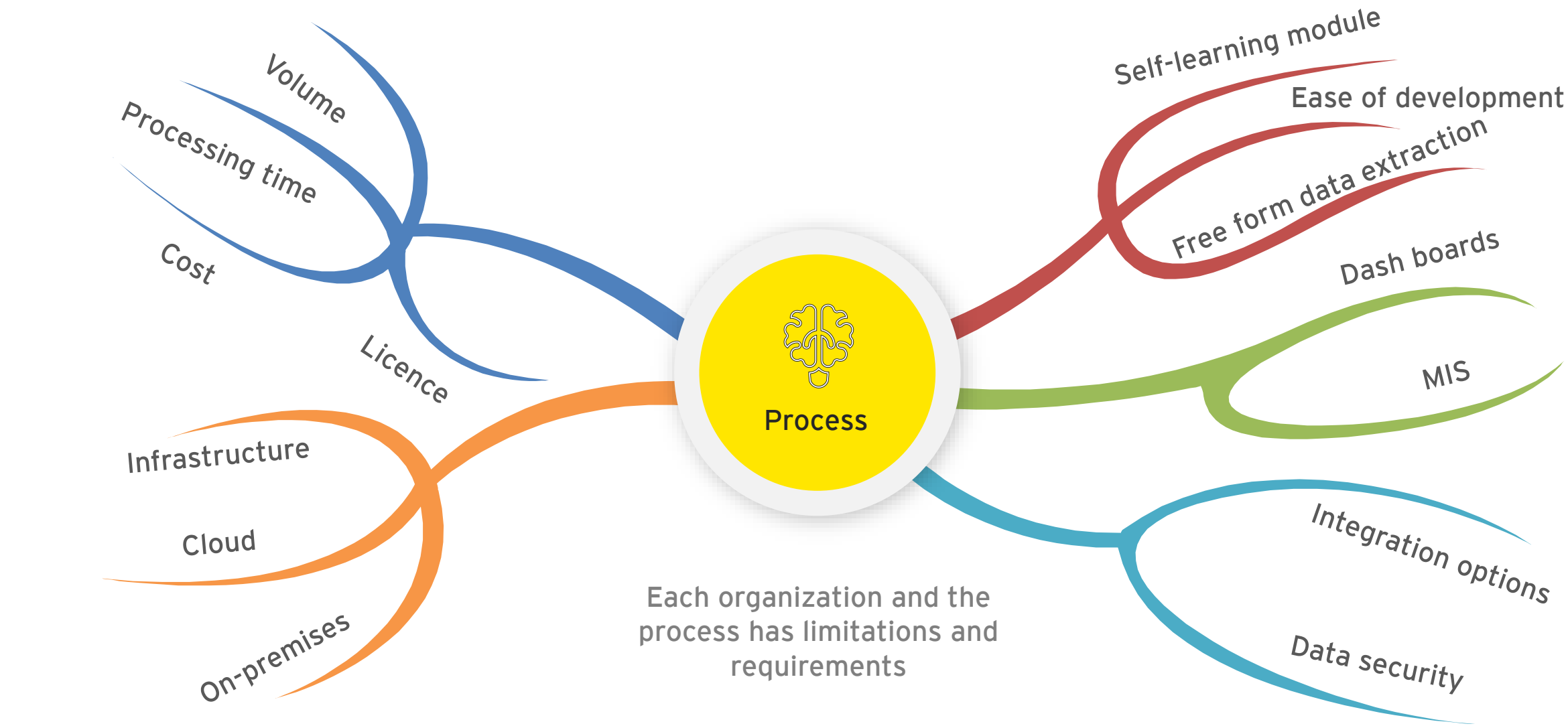


Use-case specific considerations



Each use case for document digitization requires numerous feature consideration

Process specific considerations



Each organization and the process has limitations and requirements

Technology Stack

ICR pre-built platforms

IRIS Document to Knowledge A CANON COMPANY

KOFAX

ANTWORKS Reimagine · Rethink · Recreate

docdigitizer

ABBYY

Xtracta your data engine

AUTOMATION ANYWHERE

pendo systems FUELING THE AI REVOLUTION

Datacap an IBM® Company

SIMPLEKYC

WorkFusion

ess DOCS

LEADTOOLS THE WORLD LEADER IN IMAGING SDKs

Custom ICR toolkit

Open source tools

Tensorflow

Python

Keras

Tesseract OCR

OpenCV

APIs (Computer vision/OCR/Language)

Azure

1. Computer vision API
2. Speech to text API
3. Language Understanding API

aws

1. Amazon Rekognition
2. Amazon Polly
3. Amazon Lex
4. Amazon Textract

Google Cloud

1. AutoML vision
2. Cloud vision API
3. Cloud text to speech API
4. Cloud Natural language API

InDEx Tech Stack

Open source tools

Python

OpenCV

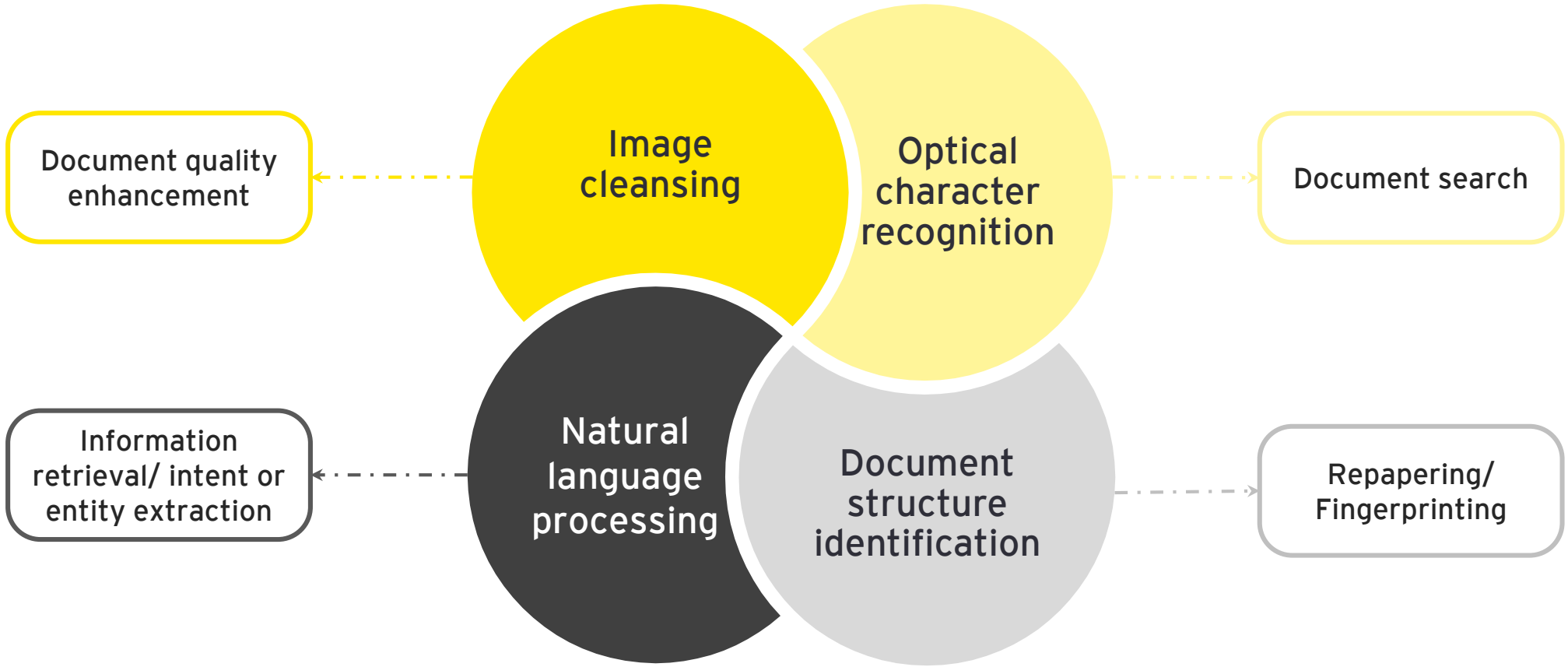
APIs (Computer vision/OCR/Language)

Azure

1. Microsoft Azure Machine Learning services
2. Microsoft Azure Cognitive services
3. Azure Kubernetes services
4. Microsoft API management
5. Microsoft PowerApps

InDEx - Intelligent Data Extraction

To transform un-structured information to structured it is important that we are able to contextualize the information. Thus InDEx leverages machine learning to infer context based on historical information.



Functionalities in ICR

InDEx in action

InDEx automatically extracts data fields from unstructured documents such as scanned contracts.

Input document (image)

The definitions and provisions contained in the 2006 ISDA Definitions, as published by the International Swaps and Derivatives Association, Inc., are incorporated into this Confirmation. In the event of any inconsistency between those definitions and provisions and this Confirmation, this Confirmation will govern.

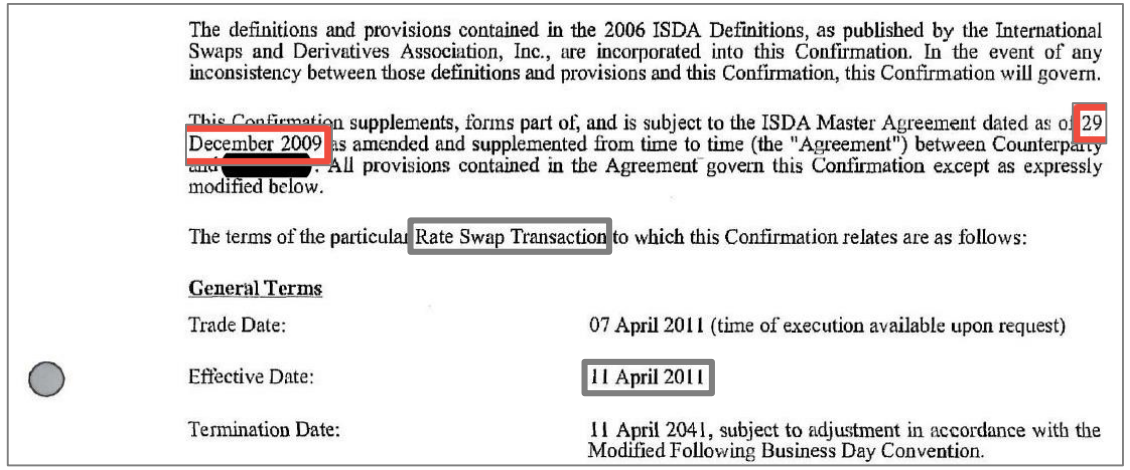
This Confirmation supplements, forms part of, and is subject to the ISDA Master Agreement dated as of 29 December 2009 as amended and supplemented from time to time (the "Agreement") between Counterparty and [redacted]. All provisions contained in the Agreement govern this Confirmation except as expressly modified below.

The terms of the particular Rate Swap Transaction to which this Confirmation relates are as follows:

General Terms

Trade Date:	07 April 2011 (time of execution available upon request)
Effective Date:	11 April 2011
Termination Date:	11 April 2041, subject to adjustment in accordance with the Modified Following Business Day Convention.

Processed document



Required Data Fields

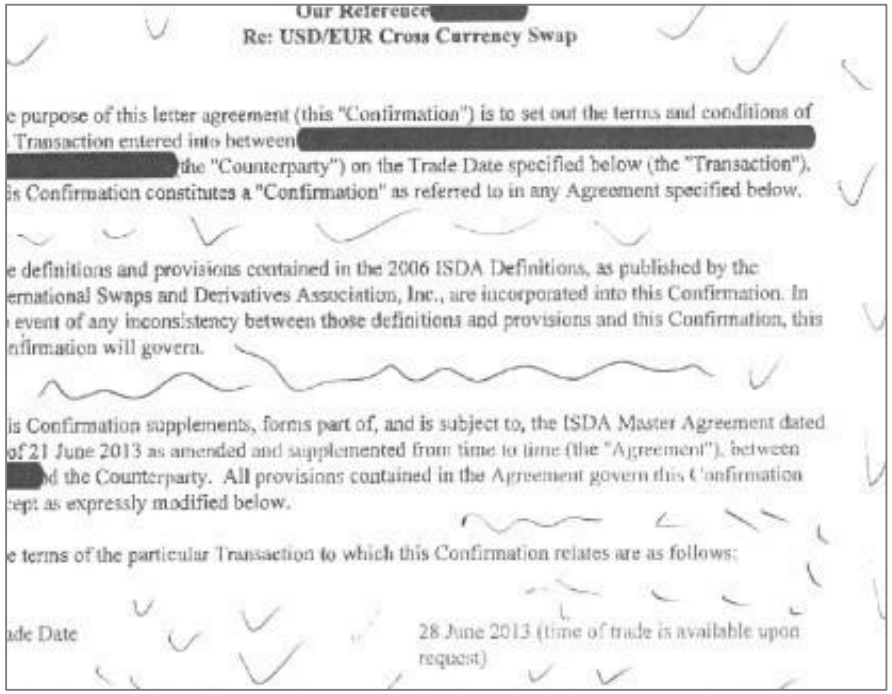
- **Master Agreement version**
- **Type of transaction**
- **Effective date**

Extracted Data Points

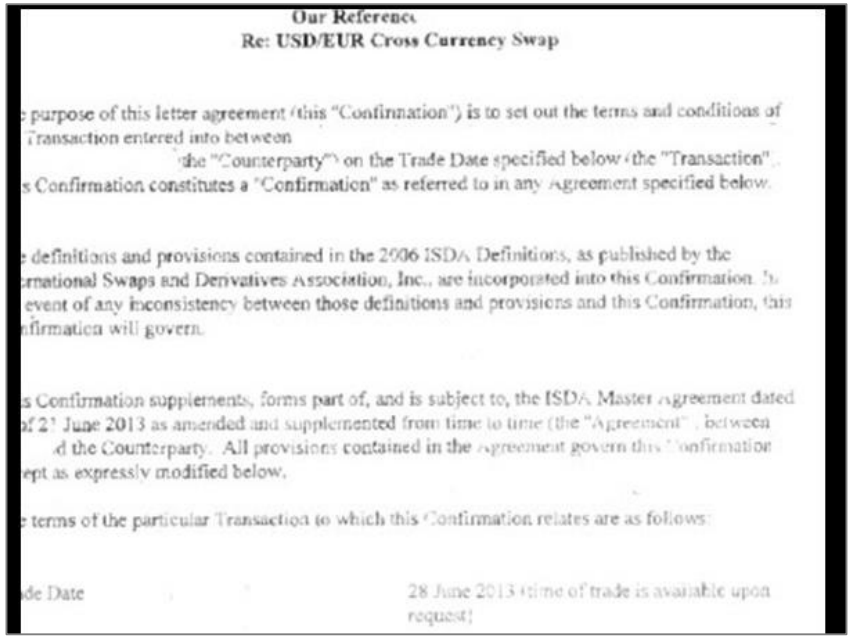
- ✓ **29 December 2009**
- ✓ **Rate Swap Transaction**
- ✓ **11 April 2011**

Image cleansing

In order to improve the success rate of the OCR a set of image processing and cleaning routines are applied. Deep learning techniques such as convolutional neural networks are used to remove noise.



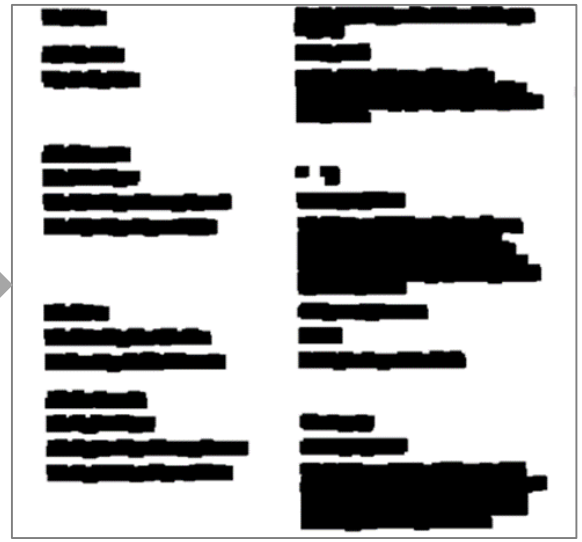
Raw data

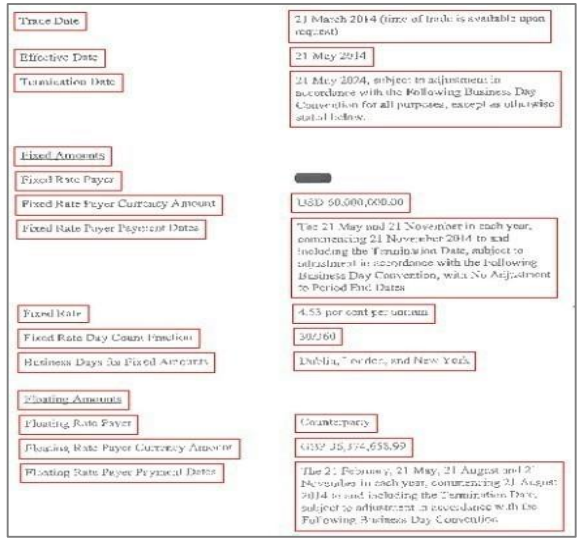


Clean data

Information re-mapping/ fingerprinting

The layout of the document is analysed by algorithms able to locate headers, paragraphs, tables… Dilating techniques are used to identify blocks of information and implicit table structures.





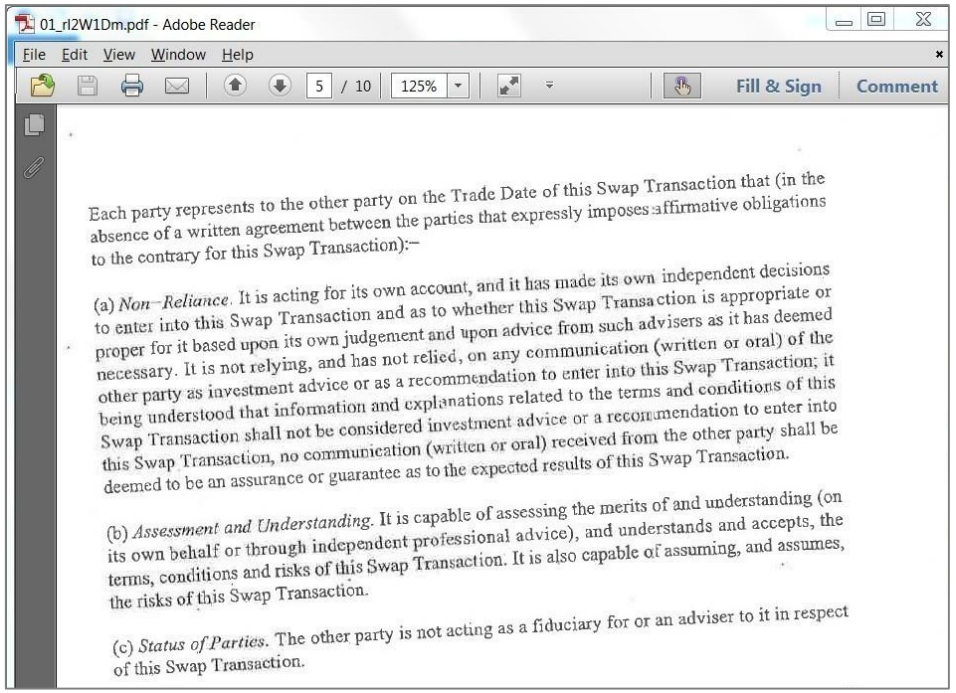
1 Raw data

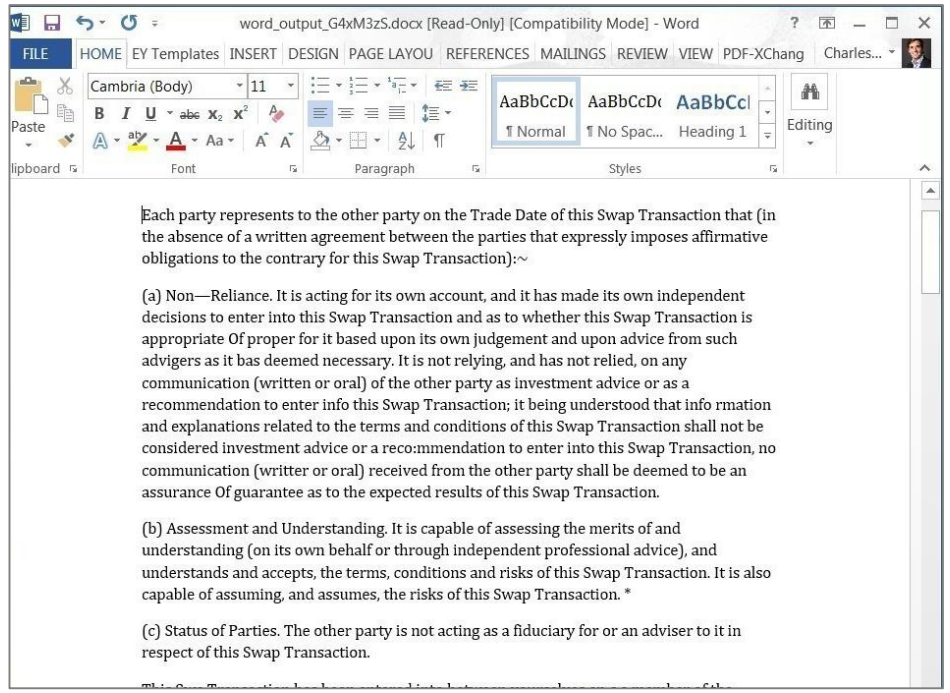
2 Use algorithms to learn format

3 Locate and extract the relevant information

Optical character recognition

OCR engine is applied to each of the identified blocks of information to **transform unstructured data into text.**





Unstructured data

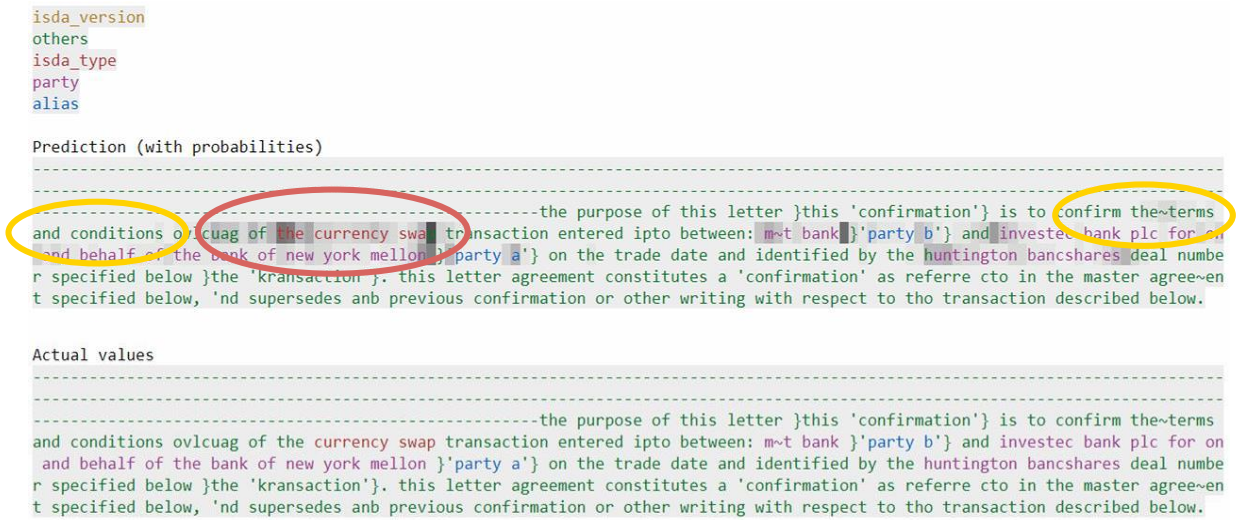
Structured text where further AI can be applied

Natural language processing

Use of Deep Learning techniques such as, using recurrent and convolutional neural networks to tag and classify words based on their context and meaning

1. Tag
2. Classify
3. Extract

Example: identify strings containing the “ISDA Type” information:



Given their meaning and surrounding words, “the currency swap” has a high likelihood of being the “ISDA Type”.

Invoice / Purchase order digitization

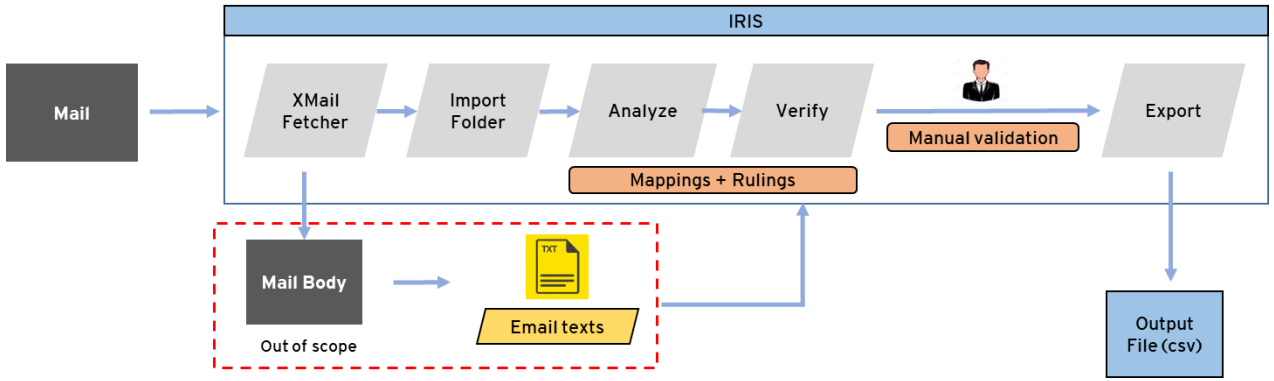
Accounts payable/receivable departments of different organisations receive purchase orders and invoices from multiple vendors. Relevant data is extracted from these documents and fed into the ERP systems. EY has developed a customized solution involving digitization tools and RPA to process these documents and enter the details in the ERP systems thus reducing manual effort, processing time, error and increasing efficiency.

Organizational Challenges
High manual effort and data volume
Manual dependencies and non-traceable processing
Significant time consumption
Prone to human error

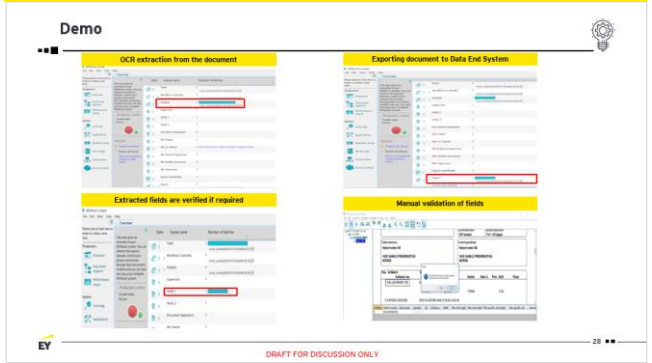
Our Solution
Reduction in manual efforts and processing time
End to end automation with capability to handle large volumes
Scalable and traceable
High accuracy, increased efficiency

Application
Invoice processing
Purchase order processing

Solution Architecture

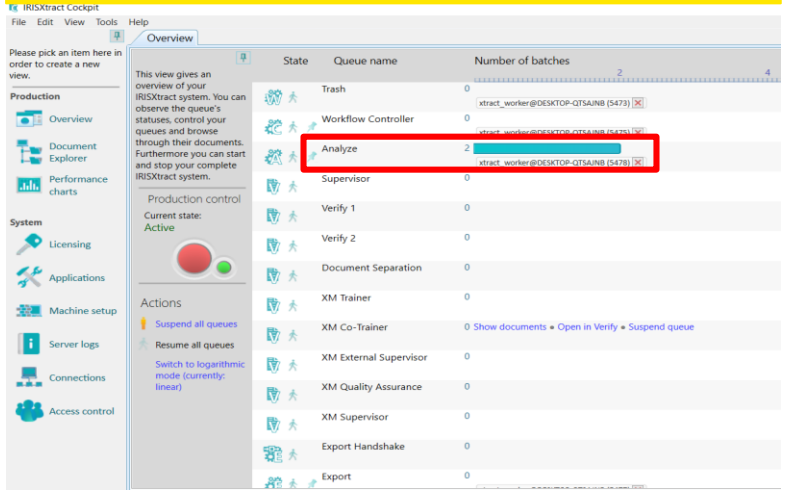


Demo

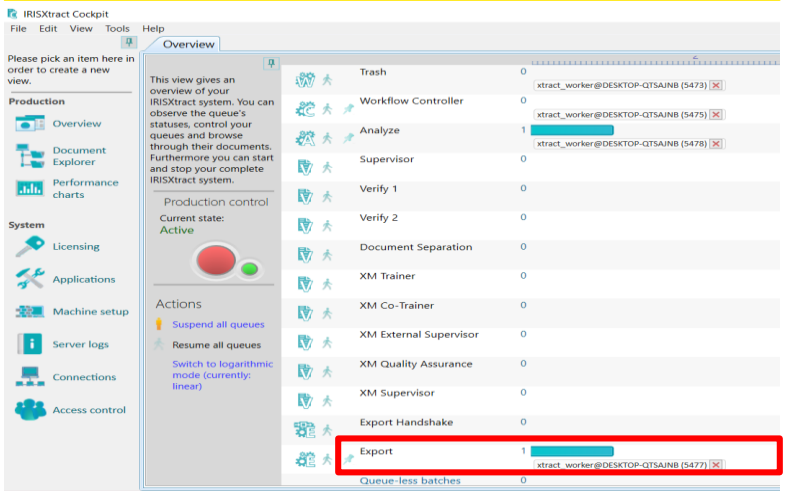


Demo

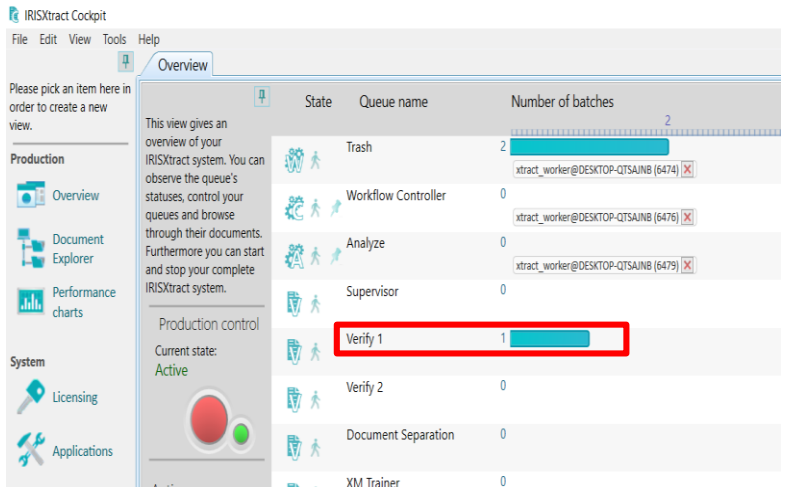
OCR extraction from the document



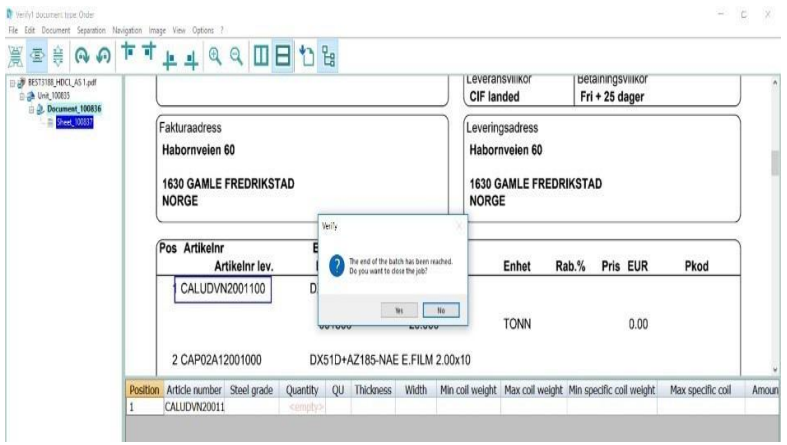
Exporting document to Data End System



Extracted fields are verified if required



Manual validation of fields



Our Credentials

Building a better working world

Business Case – ICR Order Entry

Background

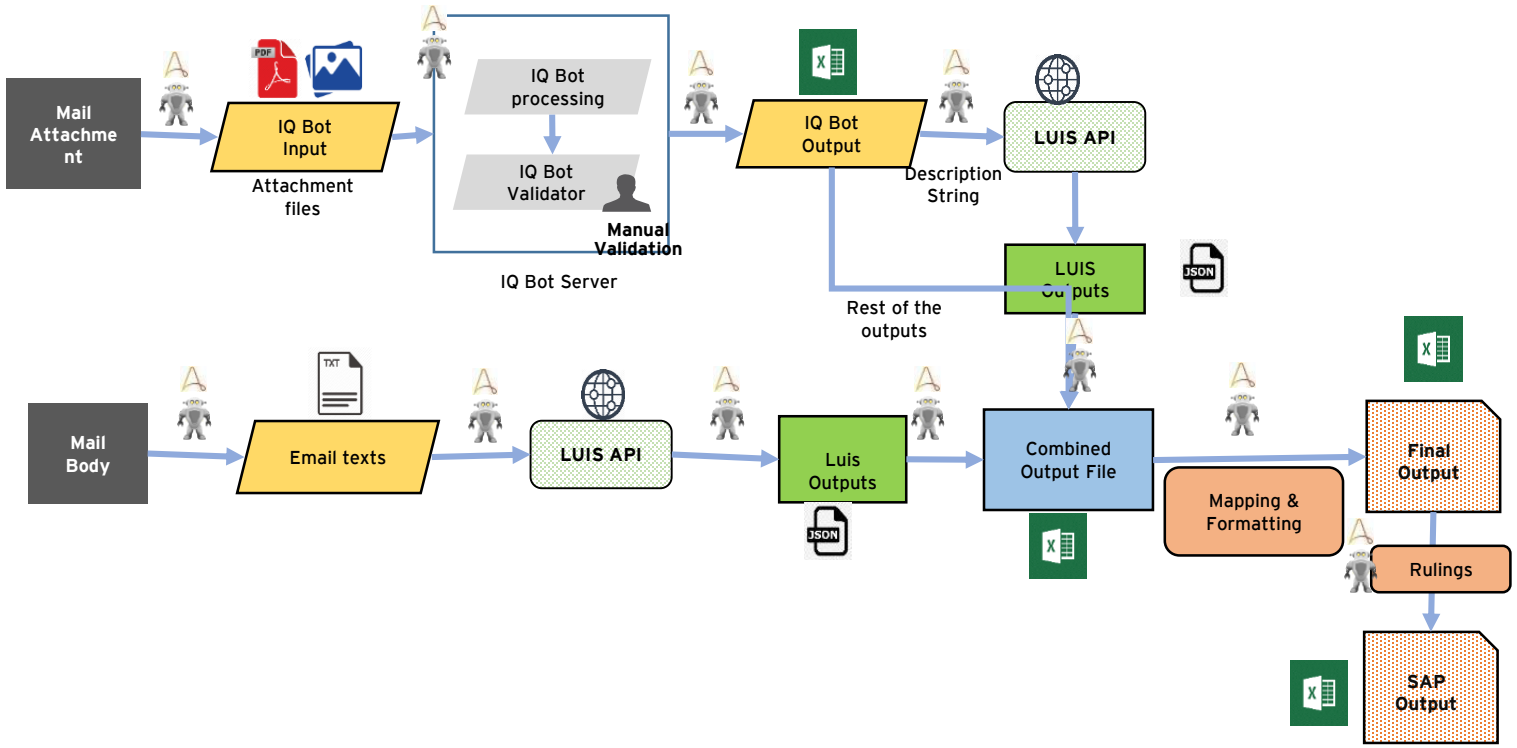
- A multinational steel manufacturing corporation performed the order entry process manually. The process involved manually recording and digitizing purchase orders and invoices with multiple vendors

EY Approach & Solution

- EY a study of the process and shortlisted OCR/ICR platforms to be integrated with the client's system
- An end to end ICR and NLP based solution to be designed for the vendor's PO and Invoice digitization process
- End to end automation integration to digitize PO and Invoice documents and combine the outputs to be fed into the ERP system in place

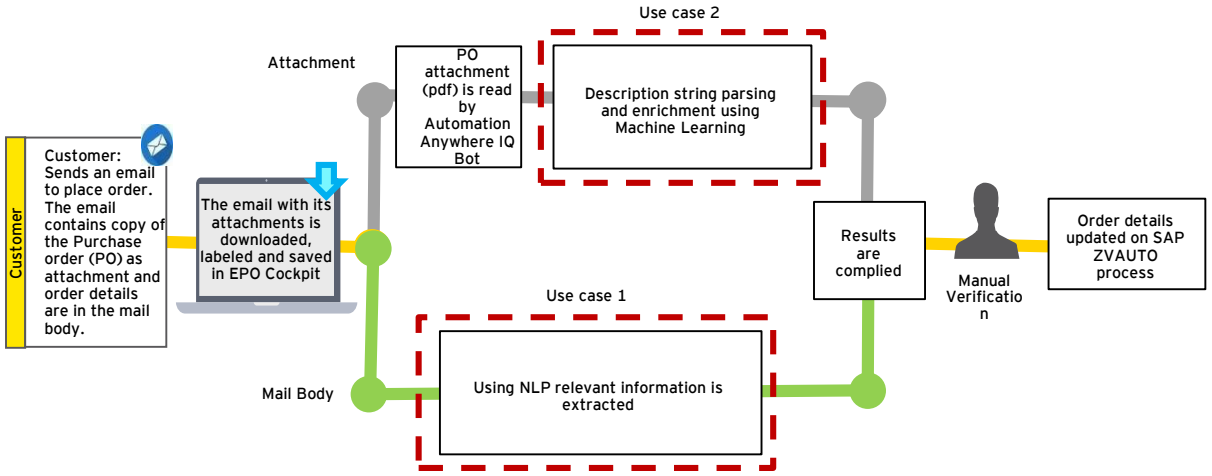
Outcome

- EY conducted extensive feasibility assessment exercises with multiple global vendors
- IRIS Xtract was finalized to conduct OCR/ICR digitization
- Microsoft LUIS was deployed to conduct NLP classification
- End to end automation by deploying Automation Anywhere was achieved
- The solution reduced total SLAs and increased overall efficiency of the process.

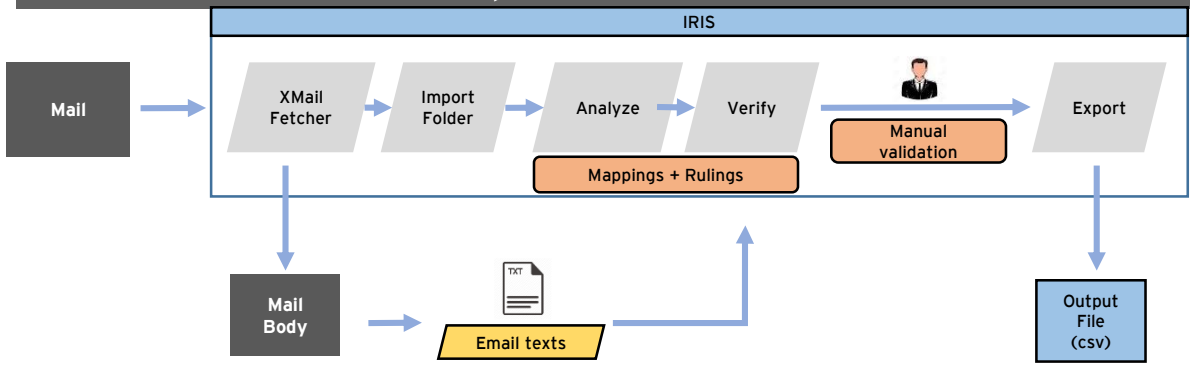


Business Case - ICR Vendor Selection

Current Process / Solution



Proposed Process Flow



Vendor Analysis

Sl	Technology Platform	NLP/ Semantic search	Free form OCR extraction (No template configuration)	API Integration	Ease of configu ration	Customiz able User Interface/ Validation screens	Auto data valida tion rules	Handling multiple PO in single file
1	IRIS Data Capture							
2	EssDocs							
3	Pendo Systems							
4	ABBYY							
5	Infrrd AI							
6	EssDocs							
7	ODYSAI/Doc Digitizer							
8	Expert Systems							
9	Nuance							
10	AA IQ Bot							

Business Case – ICR Vendor Selection

Background

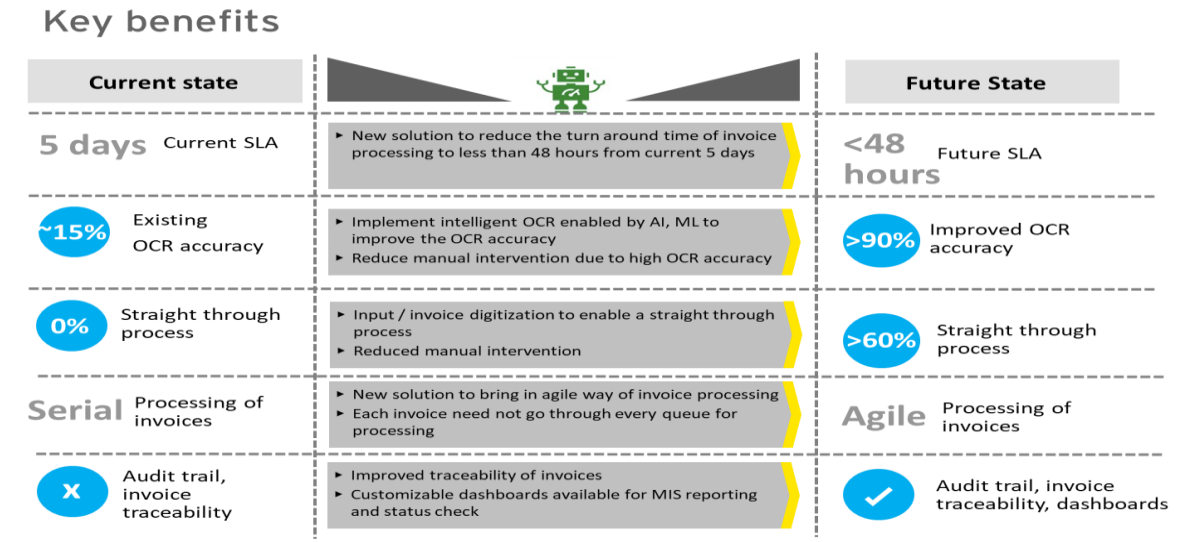
- A multinational investment bank's Accounts Payable process was handled by a vendor in-charge of data enrichment and validation services before further processing in ERP
- Key pain points: Multiple human touchpoints, regional variations, zero straight through processing, primitive reporting, no available audit trail

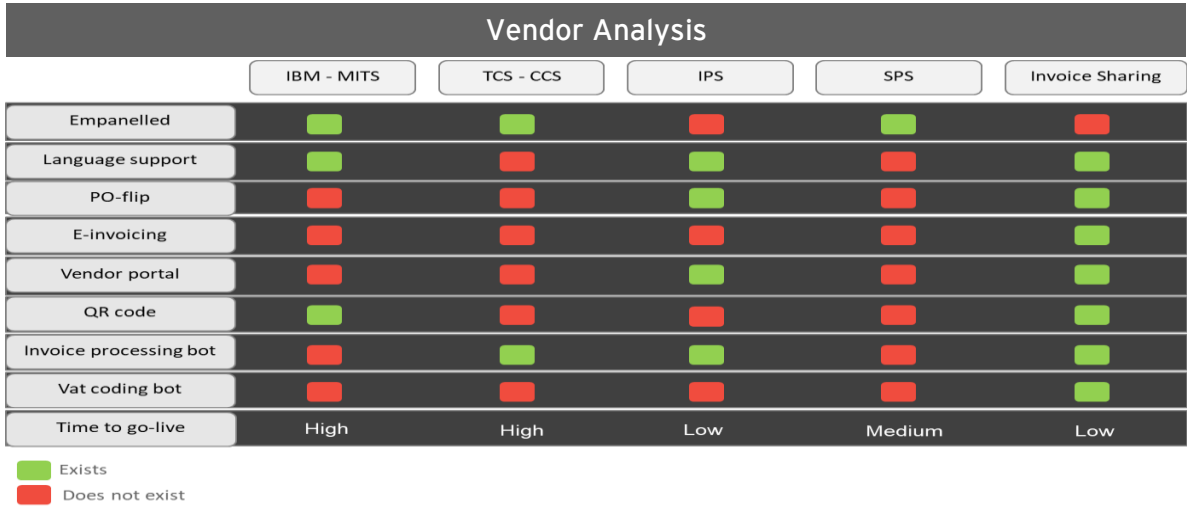
EY Approach & Solution

- EY conducts a study of the process, gathers data around invoice volumes and peak loads, conducts meetings with stakeholders to understand their key issues and challenges
- Best fit analysis to identify evaluation parameters and create an evaluation framework to assess solutions in the market

Outcome

- EY conducted demos, assessed functionalities & captured feedback from stakeholders & users. Built a model around those parameters to assess the solution based on technical , functional, financial and general vendor presence.
- Shortlisted solution based on agreed parameters and scoring was provided
- Suggest top 3 solutions : one primary solution and two alternate solutions & provide detailed implementation roadmap for the selected solution





Business Case – ICR Vendor Selection

Background

- A leading Global investment banking client performs KYC refresh for over 20,000 entities each year. The process involves manual tasks of document collection from various sources, data extraction and data validation.
- The process involves FTEs manually collecting publicly available document from different sources and extracting required data fields and conduct extensive validation through their downstream systems

EY Approach & Solution

- EY conducted a feasibility assessment of data extraction technologies that can be deployed on premises with the ability to extract data using NLP / semantic search.
- EY identified evaluation parameters and measured the technology platforms' ability to integrate with upstream and downstream systems

Outcome

- EY conducted extensive feasibility assessment exercises with multiple global vendors
- Data extraction was found feasible as per our study
- Suggested top 3 platforms: one primary platform and two alternate solutions & provided detailed implementation roadmap for the selected solution
- Due to limited time and training data, EY relied on the platform's robustness in a well trained scenario, first pass accuracy, credentials, references and assessment of the technology.

Vendor Analysis

Functionality Matrix

EY assessed the shortlisted technology platforms specifically for features relevant to GS. Below is a summary of top 5 Technloloy Platforms:

Technloloy Platform	On-Premise Deployment	NLP/ Semantic search	Rule engine	API Integration	Processing Speed	Self learning modules	Ease of on-boarding new document	Training & Support Services	Enterprise Class Management
Pendo Systems									
Expert Systems									
Ikarus									
EY In-house									
IBM Datacap									

Strength definitions

Strong — **Denotes a capability that is;** in a mature state of existence, easily demonstrable.

Intermediate — **Denotes a capability that is;** approaching maturity but with gaps to fully satisfy the client requirement or business need.

Weak — **Denotes a capability that has;** either not been developed or is in its infancy. The capability will quite often be in beta or not a focus for the Technloloy Platform