

Intelligent Document Processing (IDP)

Presented to:

ISKO Singapore

Agenda

- Definitions
- IDP pipeline
- IDP use cases
- IDP demo
- Q&A

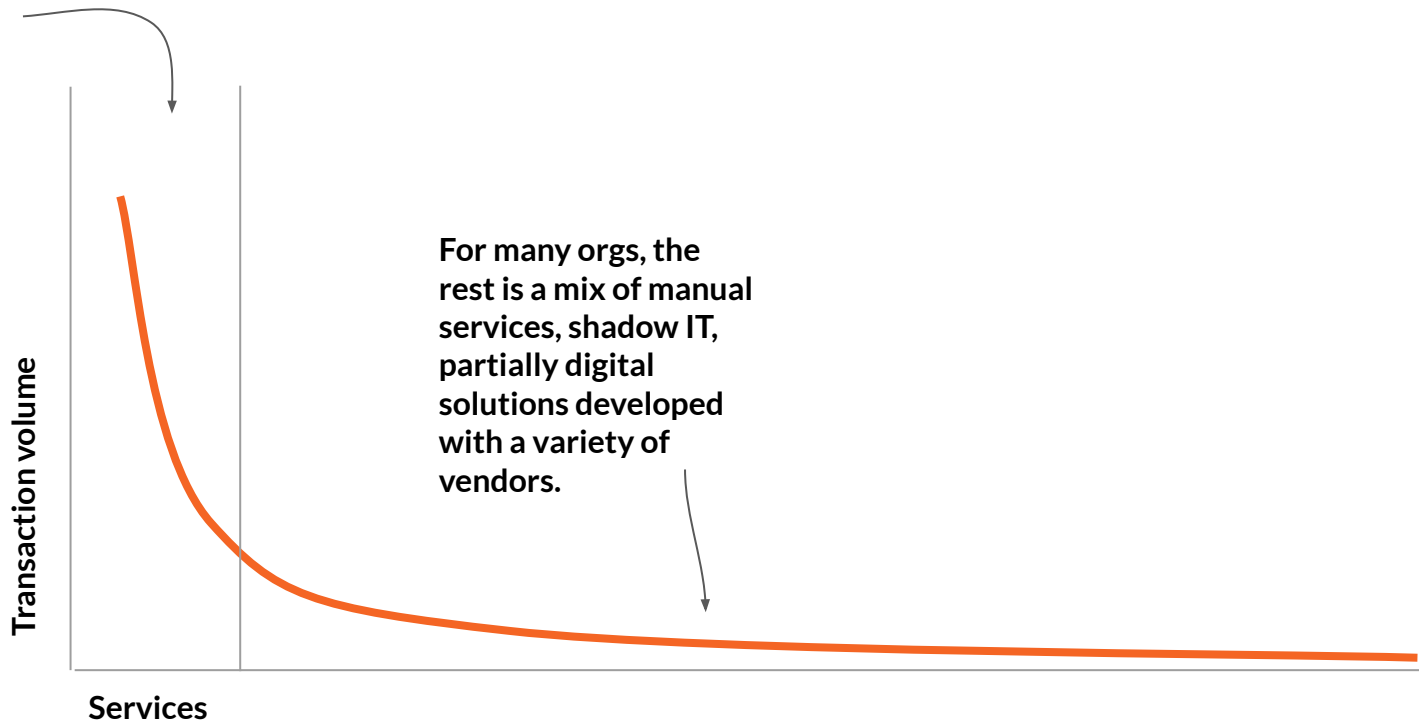


What got us started

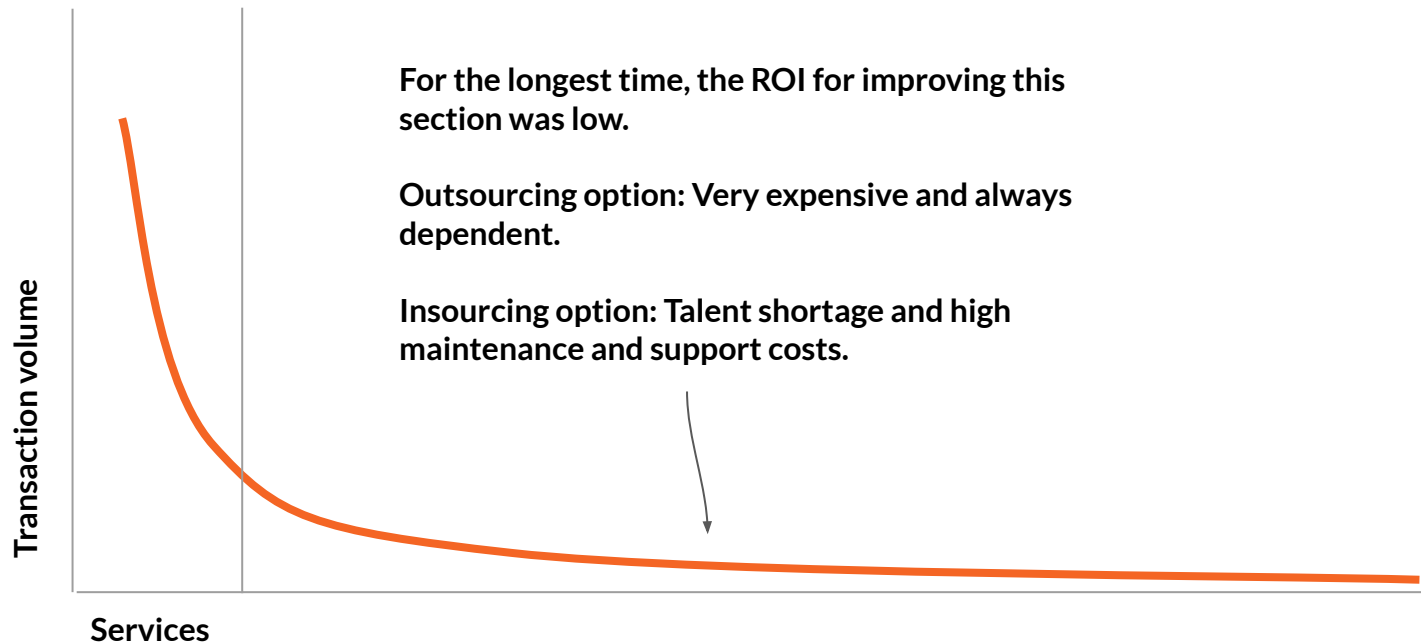


Power law graph of services in an org (long tail)

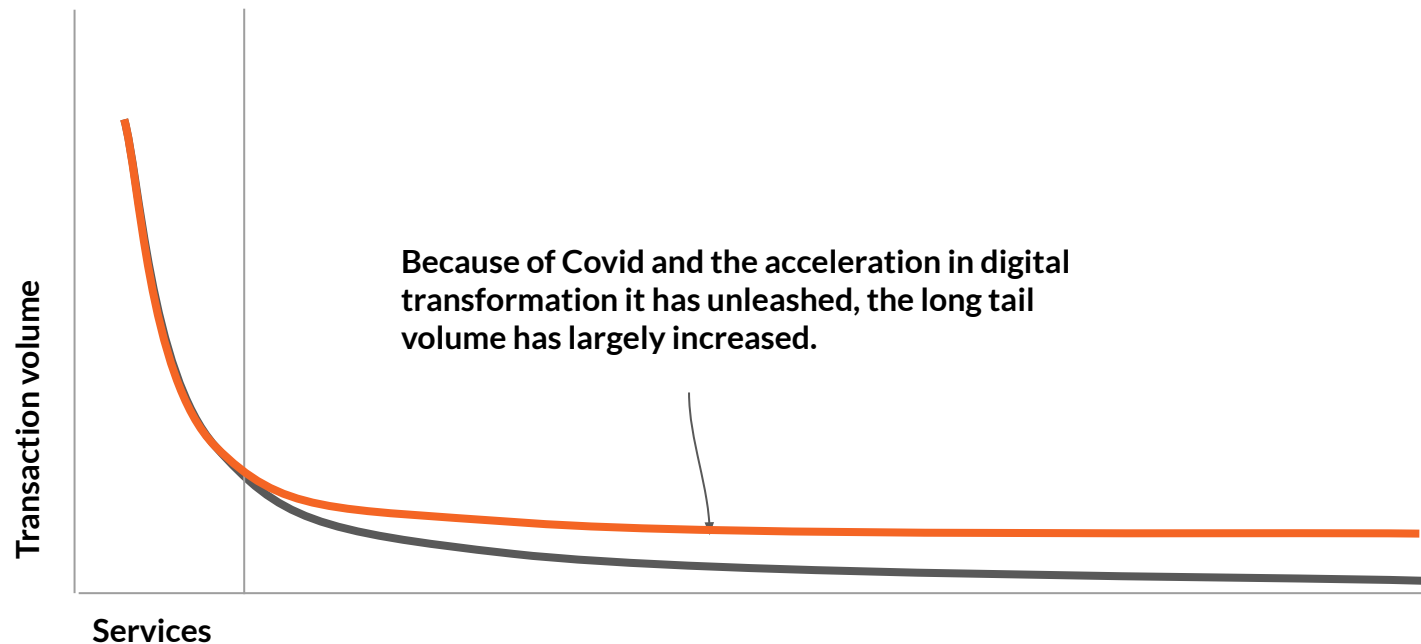
5% - 20%, usually the domain of core IT and big ticket solutions such as ERP, Finance, HR etc.



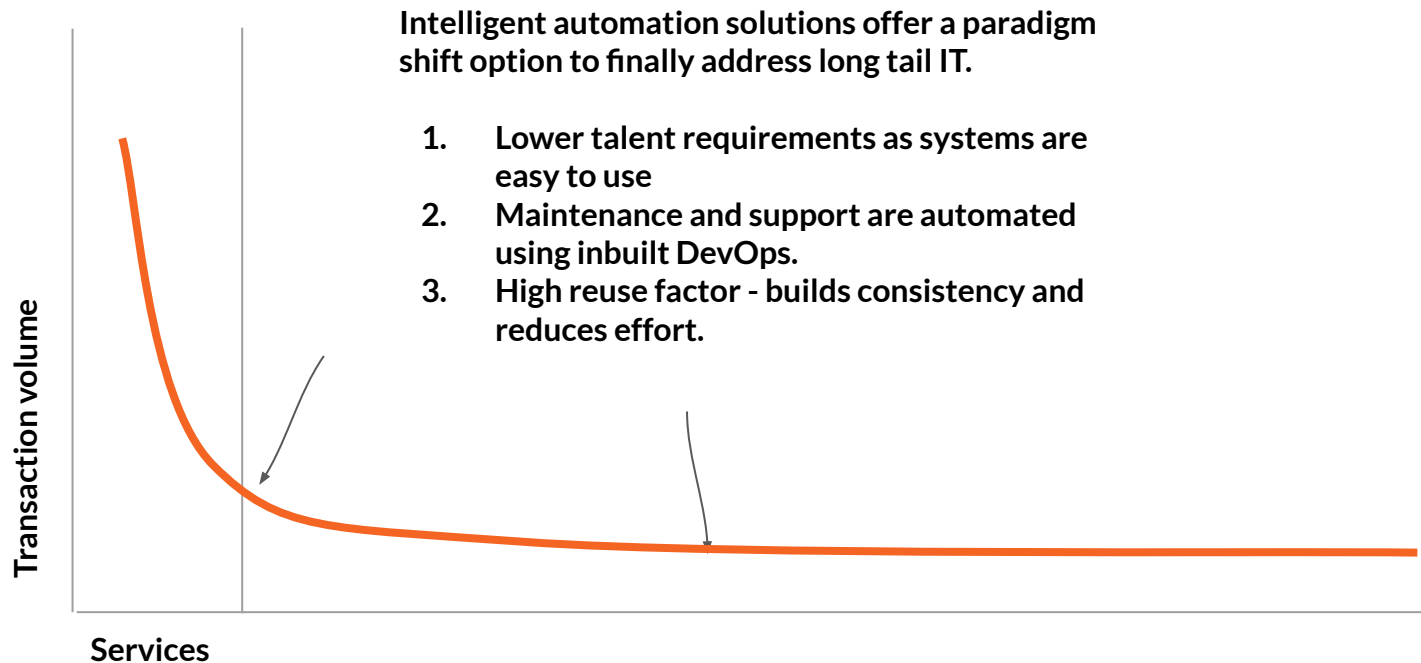
The prevalent challenge



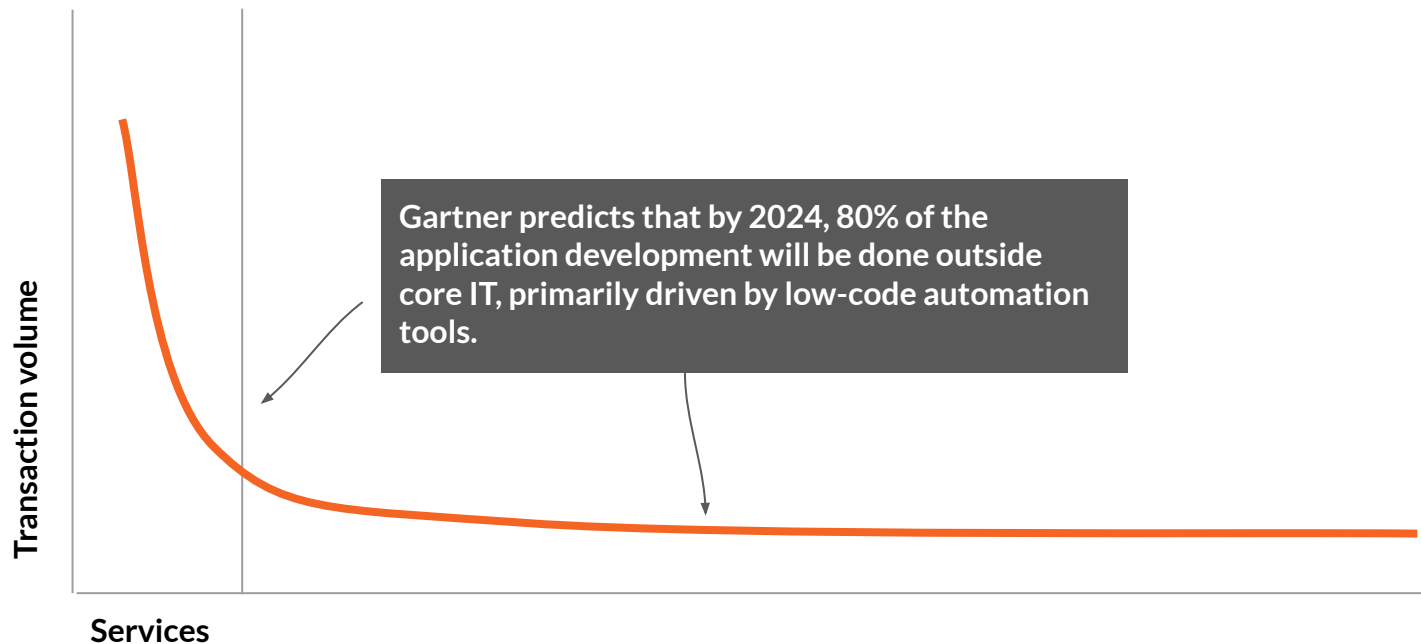
Covid exposes the gaps



The rise of low-code automation



Stats on low-code automation



Definitions



Structured vs Unstructured data

Structured data refers to data that resides in a fixed field within a file or record. Structured data is typically stored in a relational database (RDBMS).

Unstructured data might have an internal structure, but that is not available for external processing.

Typical examples of unstructured data are:

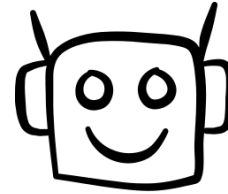
- Rich media
- Text
- Social media posts
- Surveillance imagery
- ...

Unstructured data makes up a whopping 80% or more of all enterprise data, and the percentage keeps growing.



Human workers vs Digital workers

Digital Workers are “software-based labour” that **enhance and augment human work** by combining AI, machine learning, RPA, and analytics (intelligent services) to automate business functions from end to end.



This customer support email is asking for an answer that is available in the knowledge base. Let me send it to them.



No code vs Low-code vs Pro-code

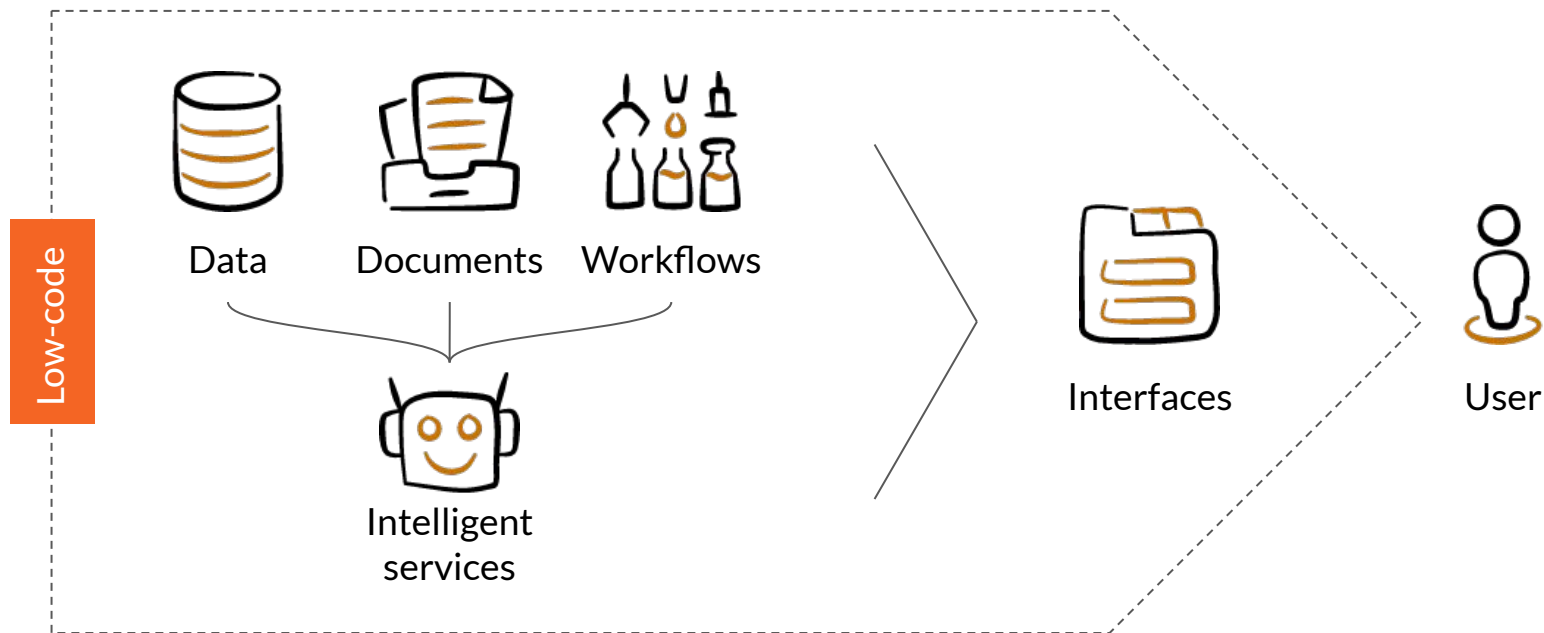
A low-code/no-code development platform is a visual software development environment that allows developers to drag and drop application components, connect them together and create a mobile or web app.

Benefits of low-code

1. Lower barrier to entry, cost and deployment time
2. Speeding up development cycles
3. Lower maintenance burden
4. Increased productivity across teams
5. Faster innovation and development



Intelligent Automation (IA)



Intelligent automation is a **stack of modern technologies** that includes artificial intelligence, machine learning, robotic process automation, low-code development platforms that aim to optimise business process design and delivery with **minimal human intervention**.

For developers



Accelerate the speed of development

Reduce the number of developers required

For business



Increase speed to market

Reduce unnecessary manual intervention

Increase responsiveness to change

For customers

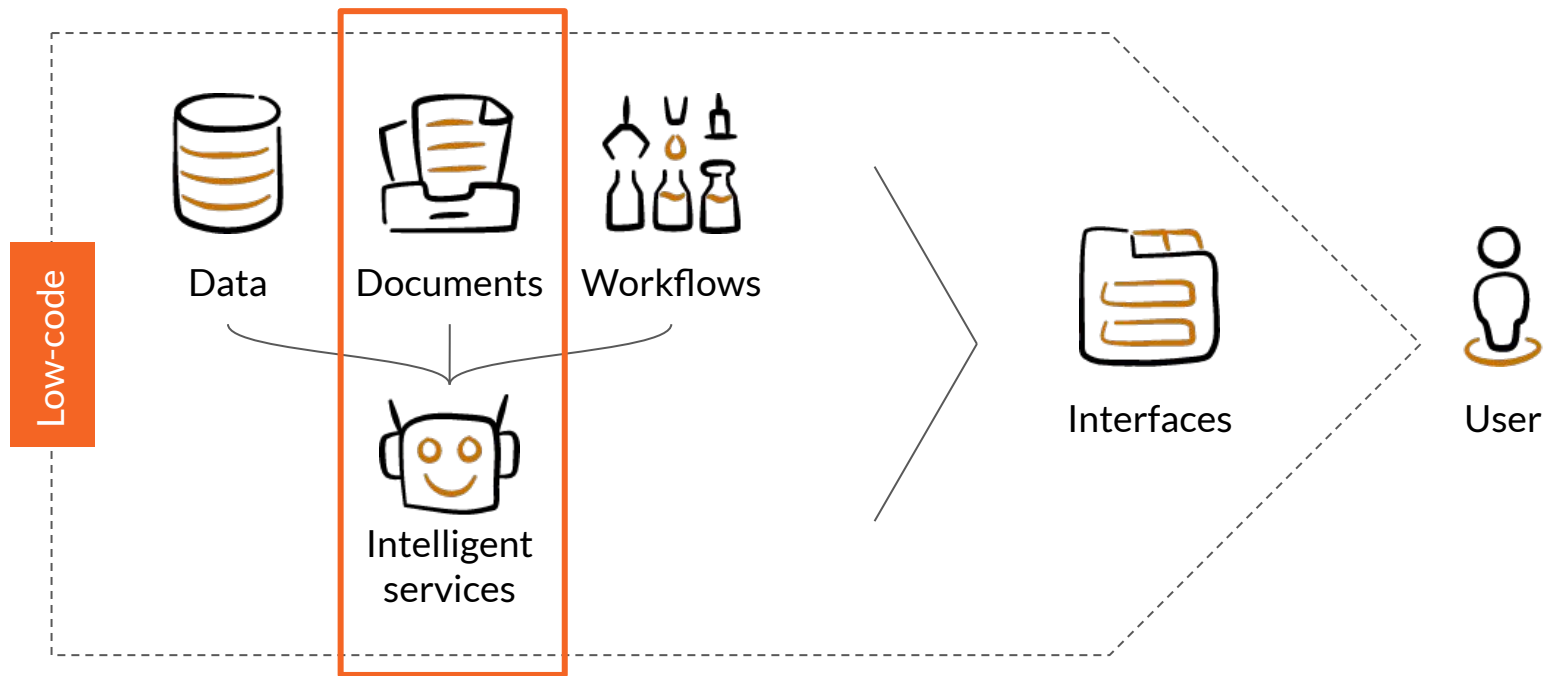


Increase in process efficiency

Increase in user experience



Intelligent Document Processing (IDP)



Intelligent Document Processing (IDP)

Intelligent Document Processing (IDP) is a growing capability of Artificial Intelligence (AI) algorithms to convert unstructured data into structured data.

Robert Smith
Project Architect

Phone: (123) 456 78 99
Email: info@wikiresume.com
Website: www.wikiresume.com
LinkedIn: linkedin.com/wikiresume
Address: 173 7 Marshville Road, Alabama

SUMMARY
Self motivated and goal oriented professional with a diverse professional experience of 10+ years in Architecture design. Maintaining quality control and building positive rapport with consultants, engineers, clients and vendors as a part of construction administration.

SKILLS
AutoCAD, REVIT.

WORK EXPERIENCE
Project Architect
ABC Corporation - January 2011 - December 2013

- Coordinated projects and team assignments with consultants to complete architectural developments.
- Responsible for ensuring that the production of construction documents was congruent with the design intent.
- Resolved complex technical and design issues.
- Gained extensive knowledge of building codes, regulations, and structures.
- Obtained strong construction administration and specification writing experience.
- Expanded comprehension of schematic/land planning and urban design.



First name	Last name	Role	Tel	Email	Skills	...
Robert	Smith	Project Architect	(123) 456 789	info@wikiresume.com	AutoCAD, REVIT	

IDP vs OCR

OCR tech is good at converting data stuck in images to digital text.

IDP tech supplements by taking the digital text and applying Machine Learning algorithms to do some amazing things like grouping the data into categories, picking out names of people, places and organisations, understanding the sentiment of a statement and much more.

But wait a minute, isn't this text wizardry called "Text Analytics" (TA) these days?

Therefore, $IDP = OCR + TA$.

$IDP = OCR + \text{Text Analytics}$



Template-centric vs AI-centric

Can be template-ised

Cannot be template-ised

Structured doc

Acme Corporation		INVOICE	
INVOICE #: 2356		DATE: MAR 20, 2020	
BILL TO:	SHIP TO:	INSTRUCTIONS:	
Rox Roxby 124 Highland Rd NY, NY 10001	Same as recipient	Please pay in full by 4/21/20	
QUANTITY	DESCRIPTION	UNIT PRICE	TOTAL
1	Item 1	\$5.00	\$5.00
3	Item 2	\$10.00	\$30.00
SUBTOTAL			\$35.00
SALES TAX			\$2.10
SHIPPING & HANDLING			\$5.00
TOTAL			\$42.10

Un structured doc

Robert Smith Project Architect	Phone (123) 456 78 99 Email: info@wikiresume.com Website: www.wikiresume.com LinkedIn: linkedin.com/wikiresume Address: 173 7 Marshville Road, Alabama
SUMMARY	
Self motivated and goal oriented professional with a diverse professional experience of 10+ years in Architecture design. Maintaining quality control and building positive rapport with consultants, engineers, clients and vendors as a part of construction administration.	
SKILLS	
AutoCAD, REVIT.	
WORK EXPERIENCE	
Project Architect ABC Corporation - January 2011 - December 2013	
• Coordinated projects and team assignments with consultants to complete architectural developments. • Responsible for ensuring that the production of construction documents was congruent with the design intent. • Resolved complex technical and design issues. • Gained extensive knowledge of building codes, regulations, and structures. • Obtained strong construction administration and specification writing experience. • Expanded comprehension of schematic/land planning and urban design.	

Morgan Menetries Licensed Architect	
Phone 405-658-0043	
E-mail morganmenetries@gmail.com	
LinkedIn linkedin.com/in/morganmenetries	
WWW www.menetries.com	
Oregon State Licensed Architect with 9+ years of experience. Seeking to deliver design excellence for G with budgets over \$10M for LHP at 20% lower cost than the competition. Commended 12 times by management and budget adherence.	
Experience	
2014 - 2019	Architect (Design & Production) Locklin, Horowitz & Phair
• Lead architect in a top 300 architecture firm. Designed 18 \$10M+ jobs from red completion.	



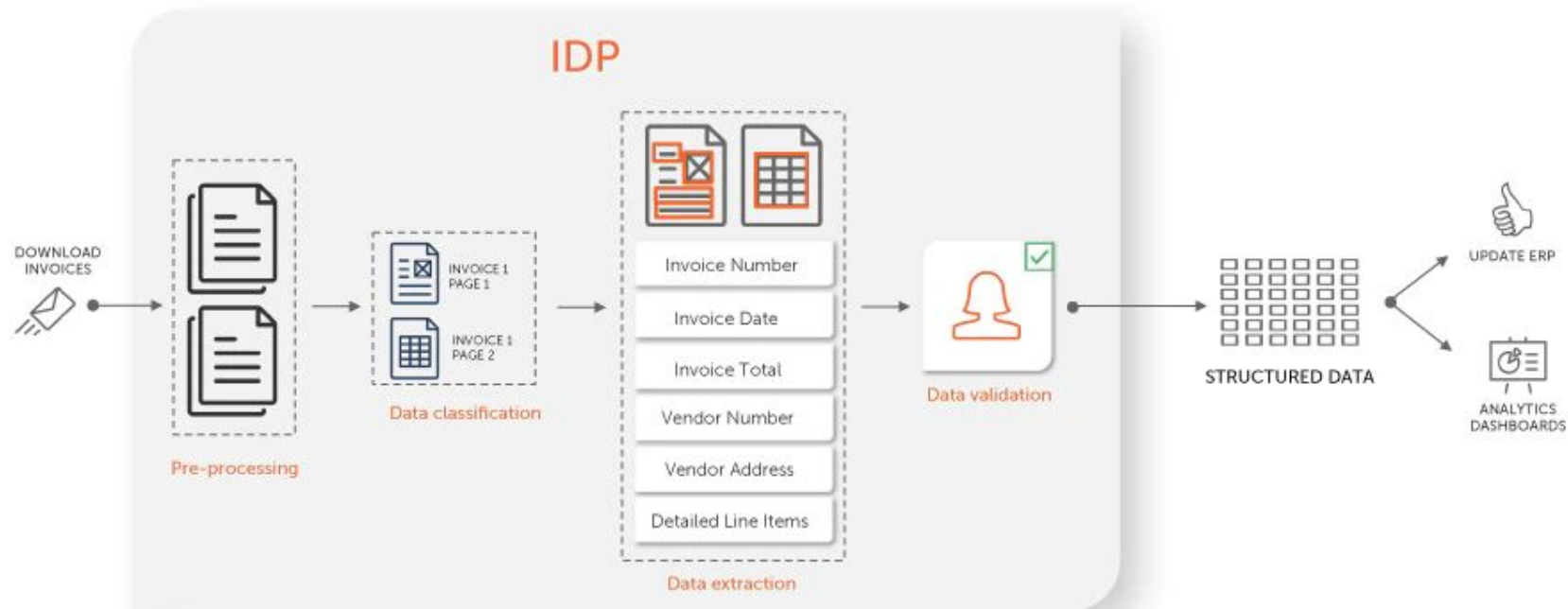
Template-centric vs AI-centric

	Template-centric	AI-centric
Best for	Structured documents	Unstructured documents
Uses	OCR + RPA + Templates	OCR + ML models (transfer learning)
Accuracy dependent on	OCR + document structure	OCR + ML models
Training set requires	1000s of documents	50-200 documents
Tools	Abbyy, UiPath	Indico, Metamaze



IDP pipeline





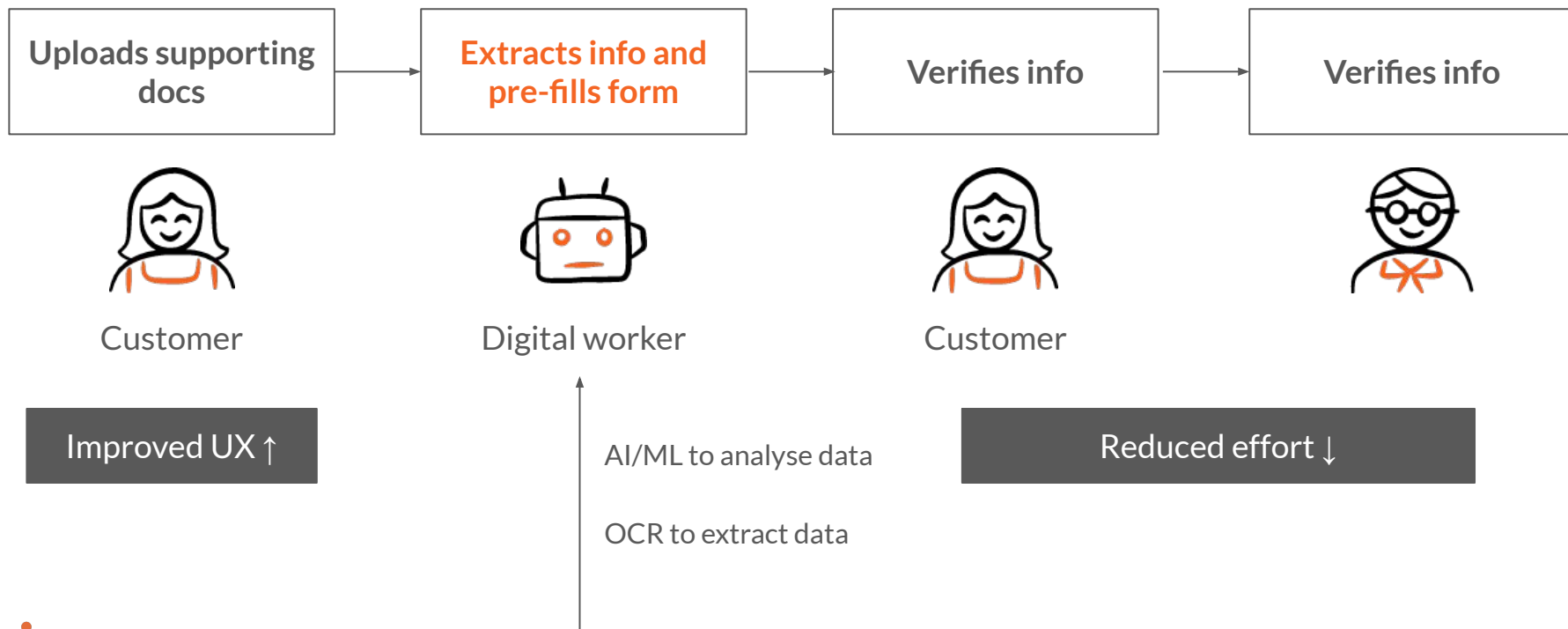
Use cases and benefits



Old school document processing

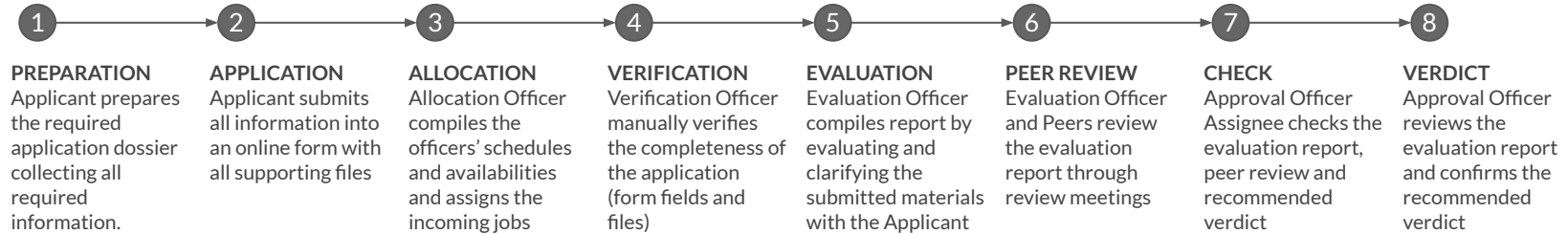


Digital worker processing documents

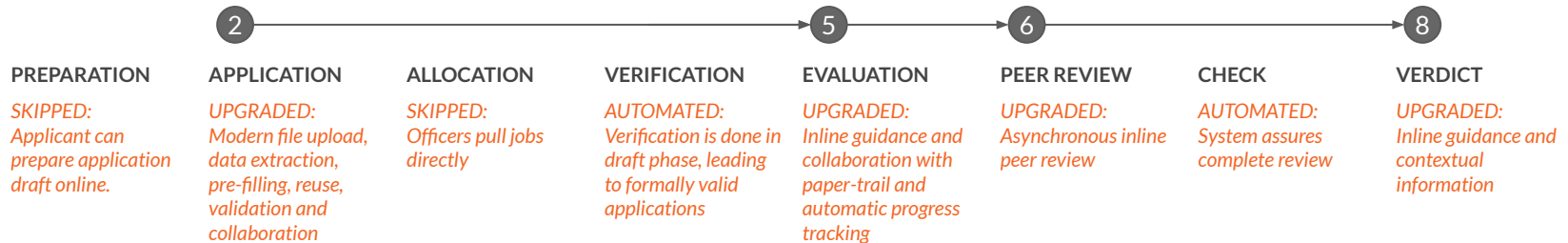


Use case: From 8 steps to 4

AS-IS PROCESS



TO-BE PROCESS



More use cases

Life insurance underwriting

Applying intelligent document processing to life insurance underwriting can help companies dramatically improve the process by largely taking humans out of the equation. With IDP, companies can create models to quickly categorize and extract data from reams of documents.

Customer onboarding

IDP can take the various documents required to onboard a new customer and automatically classify them, extract relevant data and input it into the bank's digital management system. Customers are onboarded more quickly, with increased accuracy, resulting in faster time to revenue for the bank and improved customer satisfaction.

KYC process automation

As part of the commercial banking client onboarding process, these laws require banks to make an effort to verify the identity of customers as well as the risks involved in any business relationship with them.



Benefits of IDP

**Lower human
involvement**

Fewer errors

**Spillover
effects on
modernisation**

**Faster
processing
times**

**Increased
data-driven
insights**



Demo of IDP platform — Indico Data

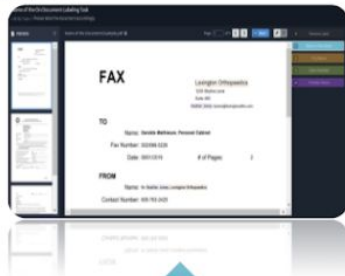


Inbound
Docs



Indico Teach

Subject matter experts label
~200 sample for each
DocType to create custom
model



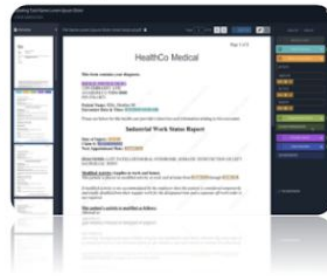
Indico Workflows

Models sequenced together
into a workflow for doc
classification >>extraction
>>validation



Indico Review

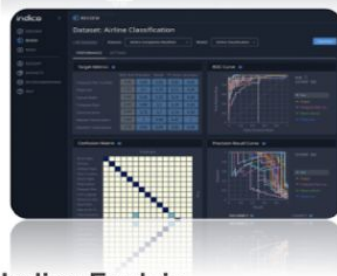
Integrated "human in the
loop" interface for rapid
review and exception
handling



Structured
Output



Improved
Models



Indico Explain

- SME's and Data Science team are able to fully understand model performance and update prod models



Staggered
Loop
Learning

Output - JSON file

```
1 {
2   "submission_id": 18724,
3   "review": {
4     },
5   "file_version": 1,
6   "etl_output": "indico-file:///storage/submission/22/18724/18724_068796df-4c46-4845-bad3-696d19bc4a35_etl_output.json",
7   "results": {
8     "document": {
9       "results": {
10         "Passport Merged - v1 ": [
11           {
12             "start": 146,
13             "end": 147,
14             "label": "Type",
15             "text": "P",
16             "confidence": {
17               "<PAD>": 1.10993114503799e-05,
18               "CCode": 0.00011147817713208497,
19               "DOB": 4.654355507227592e-06,
20               "DOE": 6.626712547586067e-07,
21               "DOI": 1.4365814422490075e-06,
22               "FULLNAME": 7.347647965616488e-07,
23               "FirstName": 7.324117177631706e-06,
24               "Gender": 3.273238689871505e-05,
25               "LastName": 1.7070243529815343e-06,
26               "MRV": 2.7284829684504075e-06,
27               "Nationality": 6.514822530334641e-07,
28               "POB": 2.0398699973611656e-07,
29               "POI": 1.945982575080052e-07,
30               "PassportNo": 4.567677024169825e-06,
31               "Type": 0.999819815158844
32             }
33           },
34           {
35             "start": 148,
36             "end": 151,
37             "label": "CCode",
38             "text": "CHN",
39             "confidence": {
40               "<PAD>": 5.352817424864043e-07,
41               "CCode": 0.9998747110366821,
42               "DOB": 7.310385967684851e-07,
43               "DOE": 2.9362392979237484e-06,
44               "DOI": 1.6601943571004085e-06,
```

Thank you

Maish Nichani
maish@pebbleroad.com