

(Un)intelligent Document Processing





Introduction

- ▶ Disclaimer
- ▶ What are my goals here today?





What this Isn't

- ▶ Explicit vendor comparisons
- ▶ Recommendations
- ▶ Telling you what to do





What this Is

- ▶ Things to think about when approaching IDP
- ▶ Reality vs marketing
- ▶ Reality vs wishful thinking





Data Evolution

Start with Data

What is this?

8675309

Or this?

631021815



Data Evolution

Start with Data

What is this?

8675309

Or this?

631021815

Make It Information

Convert it into something actionable:

"8675309"

Or:

"631021815"



Data Evolution

Start with Data

What is this?

8675309

Or this?

631021815

Make It Information

Convert it into something actionable:

"8675309"

Or:

"631021815"

And then Metadata

Give it context:

Phone Number: 8675309

Or:

City: St. Louis

State: MO

Zip+4: 631021815





Data Evolution

- ▶ IDP evolves Data into Metadata
 - ▶ By turning non-actionable **DATA**
 - ▶ Into actionable **INFORMATION**
 - ▶ And, finally, with context into **METADATA**





What is IDP?

- ▶ AI-powered (isn't everything these days?)
- ▶ Multi-channel capture
- ▶ Full text extraction
- ▶ Content Classification
- ▶ Field Extraction
- ▶ Intelligent processing of fields and context



What is IDP?

- ▶ Machine learning of forms through training sets
 - ▶ Using AI to classify content that defies classification
 - ▶ Falling back on “human-in-the-loop” (eyeballs and brain)
 - ▶ Improving classification and extraction from interactive feedback
- ▶ NOTE: Some IDP solutions lack a HITL UX.



How Can IDP Fail?

- ▶ Poor form design
 - ▶ Poor classification rules
 - ▶ Over-extraction
 - ▶ Overconfidence
-
- ▶ None of these are product failures





How Can IDP Fail?

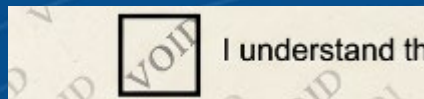
- ▶ Low code / no code solutions
- ▶ Operations and support
- ▶ SaaS vs on-prem



IDP Strengths & Weaknesses

► Extraction

- Is only going to keep getting better as technology/AI improves
- But can still be a victim to
 - Poor form design
 - Poor capture
 - Anti-photocopying watermarks

A photograph of a California Application for Vehicle Registration form. The form is filled out with handwritten information. The title is 'STATE OF CALIFORNIA APPLICATION FOR VEHICLE REGISTRATION'. The form includes sections for Transaction Requirements, Vehicle Information, and Owner Information. The vehicle information includes License Plate No. 8DFN456, VIN 1CHBTW866L234791, Year 2022, Make TOYOTA, and Model CAMRY. The owner information includes Owner Name DAVID L. CHEN and SARAH J. CHEN, and Address 1450 ELM STREET, SACRAMENTO, CA 95816. The form also includes fields for Fees, Odometer Reading (31,450), Signature, and Date (10/21/2023).



IDP Strengths & Weaknesses

- ▶ Validation
- ▶ Who builds the solution?



IDP Strengths & Weaknesses

- ▶ What is your goal for your IDP solution?
 - ▶ Simple mailroom automation
 - ▶ Goal is high accuracy and high confidence extraction of a few fields





IDP Strengths & Weaknesses

▶ Risks

- ▶ Bad form design may obscure folder number or make it hard to reliably extract
- ▶ Name data can be especially problematic for 3+ word names
- ▶ External capture may crop off barcodes or make them unreadable
- ▶ Poorly-defined keyword matching can lead to mis-classifications



IDP Strengths & Weaknesses

- ▶ What is your goal for your IDP solution?
 - ▶ Full extraction of form for downstream automation





IDP Strengths & Weaknesses

▶ Risks

- ▶ Hard (or impossible) to validate some fields in full extraction
- ▶ Meaning lower confidence extraction
- ▶ And some fields may be critical in downstream automation
- ▶ They must come out of IDP with high accuracy
- ▶ More human-in-the-loop (accuracy) vs more automation (process risk)





IDP Strengths & Weaknesses

- ▶ All your forms, all at once
 - ▶ This may look good on paper
 - ▶ But doing so requires that everything is done right
 - ▶ 80% accuracy is a viable *goal* for IDP



IDP Strengths & Weaknesses

- ▶ IDP can be expensive, increasing pressure for ROI
 - ▶ IDP is inherently complex, increasing drag on ROI
 - ▶ These two things are incompatible!
-
- ▶ Too slow, and you may lose funding – it stops being shiny
 - ▶ Too fast, and you risk failing with a complex build
 - ▶ Where is the “Goldilocks” zone?



Playing to IDP's Strengths

- ▶ IDP can be very effective
 - ▶ It is strong at classification if the classification foundation is well designed and built
 - ▶ Classification leads to extraction and extraction can be a victim of poor form design and un-correlated data that defies validation
 - ▶ Critical decision-leading fields **must** be accurate, which can increase drag on efficiency gains from IDP automation



Playing to IDP's Strengths

- ▶ Any new solution will have a learning curve
- ▶ The learning curve, by definition, means initial lower productivity
- ▶ Which will increase friction and dissatisfaction for operations management
- ▶ A quick, high-quality win that achieves high, no-touch automation will offset that and prove the IDP proposition





Playing to IDP's Strengths

- ▶ Select specific scenarios that leverage what IDP does best and build with those to drive early wins
 - ▶ Distinct forms that will classify well
 - ▶ Machine print over handwriting, whenever possible
 - ▶ Limited field extraction, with correlating fields for validation



Playing to IDP's Strengths

- ▶ For example, returned mail of customer statements
 - ▶ Content is of a known format (you sent it out)
 - ▶ Key values are in a known position on the paper
 - ▶ Account number should be in machine print increasing extraction quality
 - ▶ If you can add a barcoded record locator onto the outbound mail, you can bypass all extraction except for the barcode reading





Building with IDP

- ▶ Design, build, maintenance
 - ▶ Outsourcing is faster time to delivery
 - ▶ Insourcing is slower but has legs
 - ▶ Best of both worlds?



Building with IDP

- ▶ Avoid the “kitchen sink” approach
 - ▶ Too much complexity can create a brittle solution
 - ▶ The more you promise, the more you can disappoint
 - ▶ Build for early successes to build momentum
 - ▶ Iterate to more complex content
 - ▶ Highest volume is not necessarily ideal candidate





Building with IDP

▶ Testing IDP

- ▶ Extraction testing – how accurately does it extract information?
- ▶ Process testing
- ▶ The point of testing is not infallibility, but rather confidence



Human-in-the-Loop

- ▶ A goal of IDP is to reduce human work
- ▶ To zero is not realistic or even valuable
- ▶ 100% accuracy out of the mailroom will need HITL

- ▶ Full extraction is really moving the extraction from your service workflow to your mailroom, and then augmenting it with IDP





In Conclusion

- ▶ IDP can be very powerful and very effective
- ▶ Pitfalls are only traps if you don't know they're there



In Conclusion

- ▶ There are good IDP products out there
 - ▶ They can shine if suited to your task – and if your task is suited to them
 - ▶ And if you build incrementally, building customer satisfaction with early successes
 - ▶ Or they can fail if rushed
 - ▶ Or specified poorly
 - ▶ Or built poorly
 - ▶ Or overpromised



In Conclusion

- ▶ Success with IDP depends on how you approach it
- ▶ If you're looking to pick a vendor, I hope you have learned a few things today to help you ask good questions
- ▶ If you already have a vendor, I hope you have learned a few things today to help you navigate to success.



THANK YOU!

