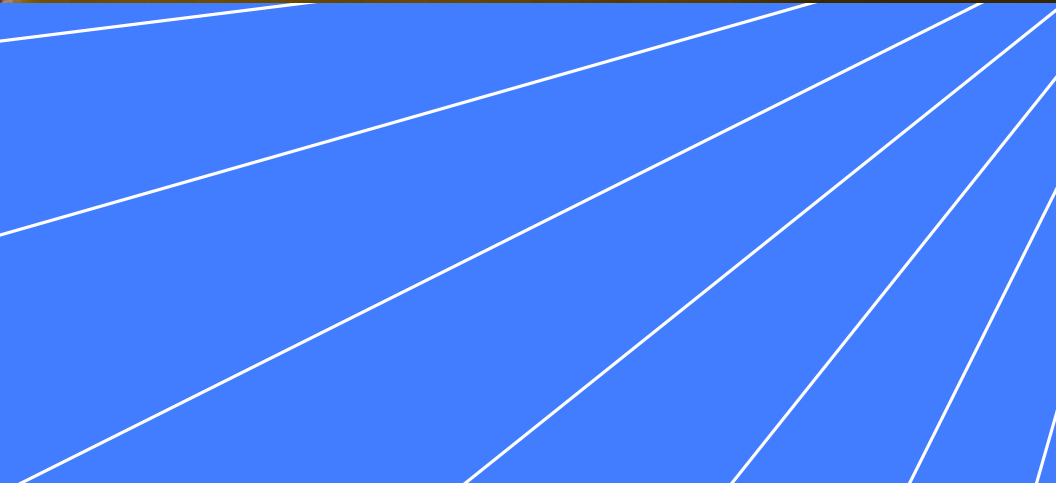


Understanding Intelligent Document Processing

How pairing AI and machine learning with human validation will change the way you work

DataBank IMX



Contents

- 04** Introduction
- 06** So, what is Intelligent Document Processing?
- 12** What content can IDP handle?
- 26** What powers IDP?
- 32** Is IDP just another way to eliminate paper processes?
- 36** Get started with Content Intelligence
- 50** About DataBank
- 56** Sources

Introduction

As the data owner of your organization, you have to make the right technology investments now to keep your team moving forward.

It's a tall order, especially when those investments may require you to make significant internal changes or clean house upfront.

Here are some of the challenges we see every day:

- Valuable data is trapped in physical records or useless PDFs.
- Data resides in siloed, disconnected systems.
- Processes from different sites, offices, or acquisitions are outdated and disjointed.
- The data are incomplete or unreliable. Maybe both.
- Time-consuming, manual indexing leads to errors worsening the problem.
- Your teams are bogged down trying to resolve these issues one by one.

Trusted, accessible data is both the villain and the hero in your story, or else you wouldn't be here.

You need reliable and accessible data to understand impact and ensure your future technology purchases will see dividends.

This guide will prepare you for these conversations with your CEO and other digital allies within your organization to address not only your goals, but a shared vision that will tackle your data difficulties across the board.

We'll look at how Intelligent Document Processing will not only resolve these disparities but also make information more accessible and provide insights that help your teams predict where to focus next.



So, what is Intelligent Document Processing?

IDP DEFINITION AND BENEFITS | IDP VS OTHER DOCUMENT PROCESSING METHODS



So, what is Intelligent Document Processing?

Intelligent Document Processing (IDP) uses modern technology and human validation to turn an organization's content into structured, usable data.

AI is a primary component that uses machine learning to study past data, then quickly extracts and structures essential fields so they're searchable and accessible. The reward is that it's much faster and ultimately requires less human interaction than manual or OCR processing methods.

Benefits of IDP:

- Mass volume automation of document processing
- Enables people to focus on more productive work
- Extraction and structuring of valuable information into intelligible data and insights
- Immediate access to data through your most essential systems
- 98% or higher accuracy with the assistance of human validation




So, what is Intelligent Document Processing?

IDP vs other document processing methods

Here's a comparison of past and present document conversion methods and why IDP is a natural next step for any organization that deals with a ton of paper and/or digital records.

	Dawn of time	70s Disco Fever	Now
Processing method	Manual Human-managed; self-learning	OCR Logic-based workflows with significant human intervention	IDP AI learns from human validation
Input	Unstructured or semi-structured content of any format	Structured content	Unstructured or semi-structured content of any format
Output	Structured, contextual data points from text and images E.g., Understands date of birth = 11- 12-88	Extracted data without context E.g., Recognizes 11-12-88, but doesn't understand this is a birthdate	Structured, contextual data points from text and images E.g., Understands date of birth = 11-12-88
Accuracy	98%+ accuracy	85-90% accuracy	98%+ accuracy
Human intervention	Completely manual	Somewhat manual	Validation only
Cost	High labor, error- prone, slow turnaround	Less labor, moderate turnaround	Low labor, fast turnaround

Now that you have a high-level understanding of IDP and how it compares to other processing methods, let's look at the types of content structures you may have today and how IDP standardizes it.



DataBank IMX

What content can IDP handle

| STRUCTURED CONTENT | SEMI-STRUCTURED CONTENT | UNSTRUCTURED CONTENT |

andle?

What content can IDP handle?

Processing your unstructured, semi-structured, and structured content

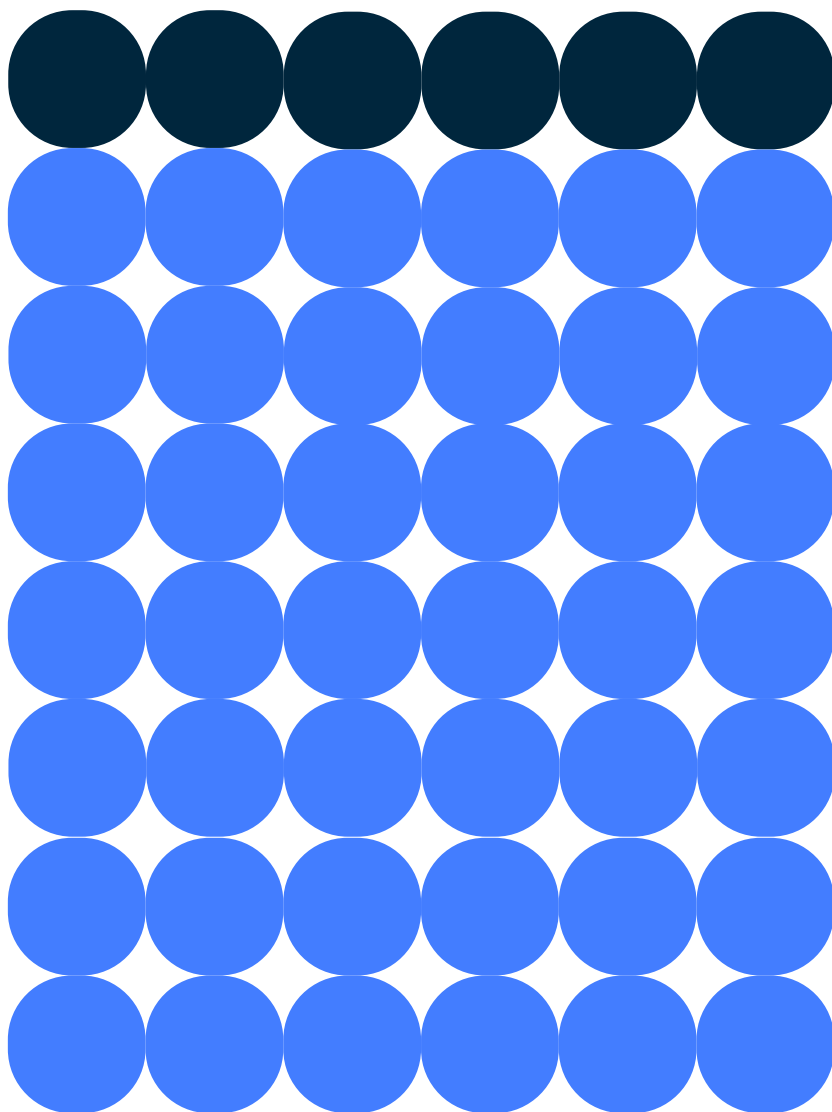
IDP can organize any content you've got - structured, semi-structured, or unstructured - into a synchronized, understandable format that can be distributed to any system(s).

Here are the different types of content and how IDP works with them.

Structured content

Structured content is what we commonly think of as data easily put into a spreadsheet.

It has pre-defined fields like headers with data falling into those categories. It's typically in a fixed format and easy to analyze.



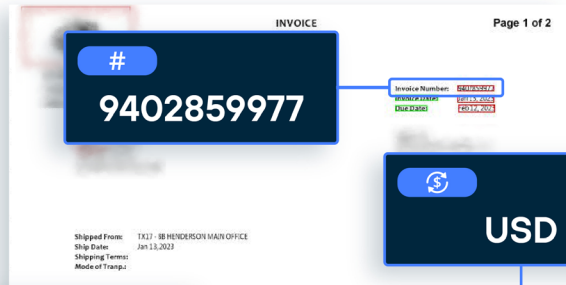
What content can IDP handle?

Real example:

This company has structured invoices. The headers and data are in the same location on every invoice they use.

Through IDP, we feed the machine by indicating what fields are most important, and then it quickly extracts that information.

With AI, it learns from all the documents we serve it over time so data can automatically be extracted and structured.



Buyer:	●	
Supplier:	●	
Debit/Credit	●	Debit
Invoice Number	●	9402859977
Invoice Date	●	01122023
Tax Amount	●	0.00
Total Amount	●	9274.28
Currency	●	USD
POnumber	●	
Account Number	●	
BOLNo	●	
Freight Amount	●	
Terms	●	
Due Date	●	
Vendor Number	●	
Remit Address1	●	
Remit To City	●	
Remit To State	●	
Remit to Zip	●	
Industry Code	●	
Is Statement	●	
Email Address	●	
Company NO	●	

Item Description	Net Qty	UoM	Price	Amount - USD
Freight	24.170	each	33.1600	800.03
jt Surcharges	800.030	each	0.3600	240.01
Freight	24.080	each	33.1600	797.05
jt Surcharges	797.050	each	0.3600	239.12
Freight	23.790	each	33.1600	787.45
jt Surcharges	787.450	each	0.3600	236.24
Freight	24.110	each	33.1600	798.04
jt Surcharges	798.040	each	0.3600	239.41
Freight	24.210	each	33.1600	801.35
jt Surcharges	801.350	each	0.3600	240.41
Freight	23.700	each	33.1600	784.47
jt Surcharges	784.470	each	0.3600	235.34
Freight	24.100	each	33.1600	797.71
jt Surcharges	797.710	each	0.3600	239.31
Freight	23.010	each	33.1600	761.63
jt Surcharges	761.630	each	0.3600	228.49
Freight	24.360	each	33.1600	806.32
jt Surcharges	806.320	each	0.3600	241.90
Total				9274.28

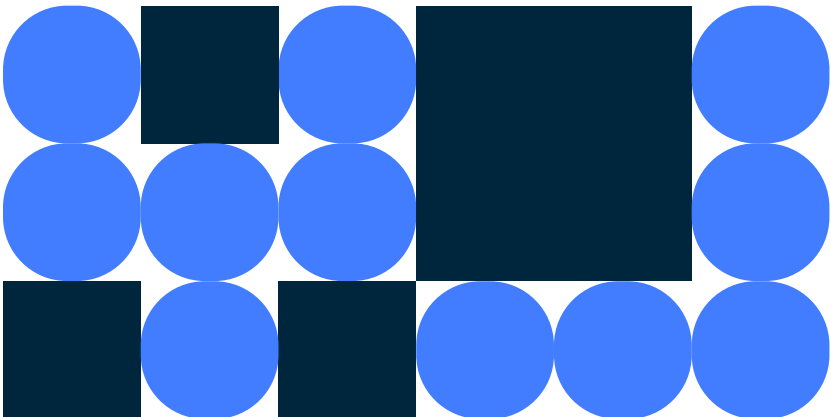
9,274.28

What content can IDP handle?

Semi-structured content

Semi-structured content is what we typically see in today's digital world.

An email may seem unstructured with text and images, but behind the scenes, metadata capture essential data points like header, time stamps, subject, delivery time, from, to, etc.





What content can IDP handle?

A real example:

This organization is structuring its vital records, including this death certificate. Although the actual certificate looks structured, it's only semi-structured.

Why? Yes, there are some handwritten responses, but this organization had many variations of this certificate.

Over the years, forms change. An organization could have 10 versions of the same form in a single year, which would involve many manual human hours to determine which header and field are which.

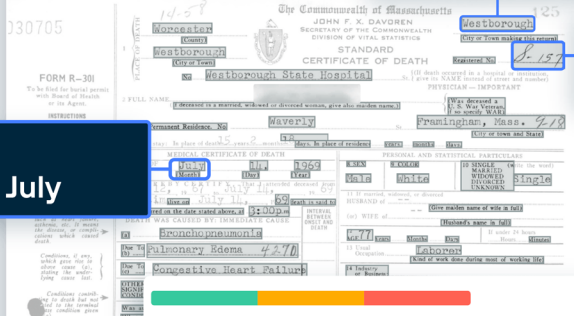
The beauty of IDP is that even with semi-structured content, the machine studies similar documents and then automatically determines those necessary fields on its own. That means all that's left is for humans to check quality.

Document Confidence: 96.916664

Westborough

8-157

July



Form

Registered No .

8-157

City or Town

Westborough

Due (b) To

Pulmonary Edema 4270

8 SEX

Male

(Month)

July

19 MAIDEN NAME OF MOTHER

Westborough

Was autopsy performed?

Yes

15 SSN

123-45-6789

Other significant conditions

Atherosclerosis

Due To (c)

Congestive Heart Failure

What test confirmed diagnosis?

Autopsy

9 COLOR

White

County

Worcester

17 NAME OF FATHER

Westborough

Received & Filed

July 18, 1969

14 Industry or Business

Address

Medical records & questionnaire

Date of issue of permit

July 14, 1969

What content can IDP handle?

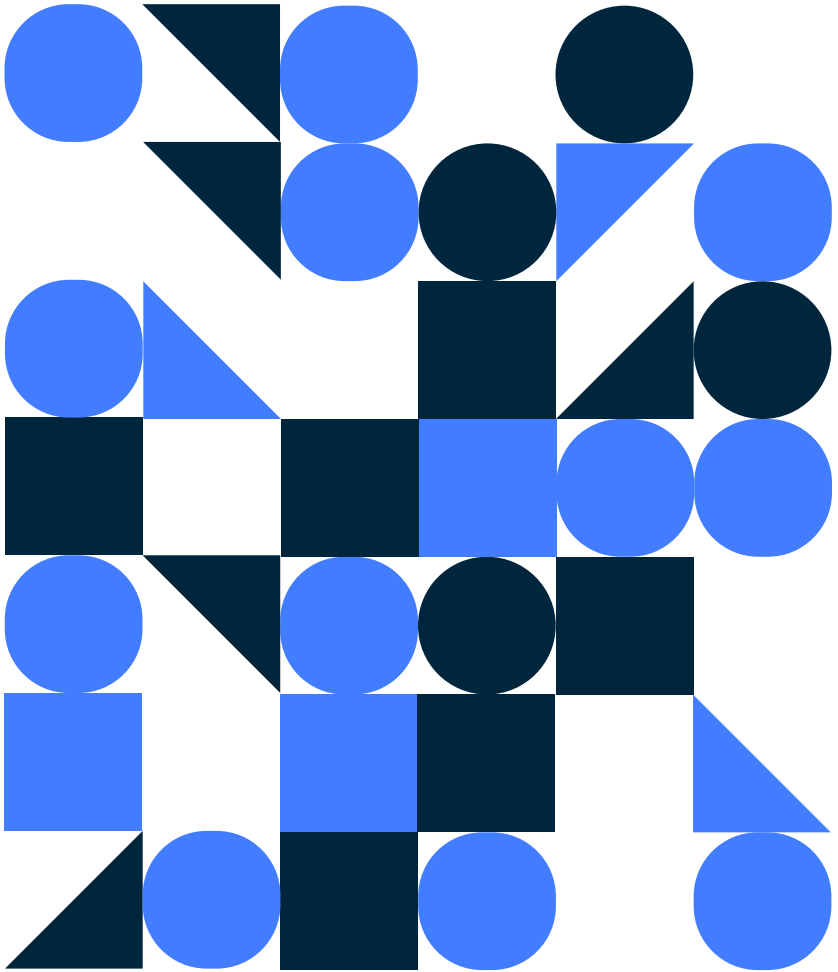
Unstructured content

We often wish we could easily capture and analyze complex information.

A perfect example is a paper form. A customer, patient, or citizen could fill in responses on a form in various places, making capture of that data more difficult. Free-form text, even if typed, or video footage are other good examples of unstructured content.

It can't be analyzed easily.

Most of the world's data is unstructured, so many untapped insights exist.



Unstructured vs. semi-structured vs. structured content

Real example:

This letter was received from a doctor outside of the patient's care network. It had no headers or prominent fields that could be quickly identified.

With IDP, the sections that matter in your unstructured content can be detected, extracted, and arranged into searchable data.


The image displays a medical letter from Advanced Vascular Associates and its structured data representation. The letter includes a logo, company name, website, a date stamp (August 30, 2017), a subject line (Re: Michael J. Collier), and a body of text. The structured data is presented in a table-like format with the following fields:

Date	08/30/2017	✓
Description & Document Type	(Office & Other Facility Documentation)	⊖
Description		!
Comments		✓

Once all your structured information is fed to your most vital systems, the possibilities are endless:

- Automate intake and workflows like invoicing or contract approvals
- Make accurate data accessible to your workforce regardless of location
- Use advanced analytics and business intelligence to power automated insights
- Understand customer, patient, or constituent sentiment and mood to make better predictions
- Support data integration during mergers and acquisitions

Now that you know IDP can quickly extract and structure any content, even images or handwritten forms, let's move on to the people and technology behind the magic.



DataBank IMX

What powers IDP?

AI AND MACHINE LEARNING | FREE YOUR ORGANIZATION OF PAPER

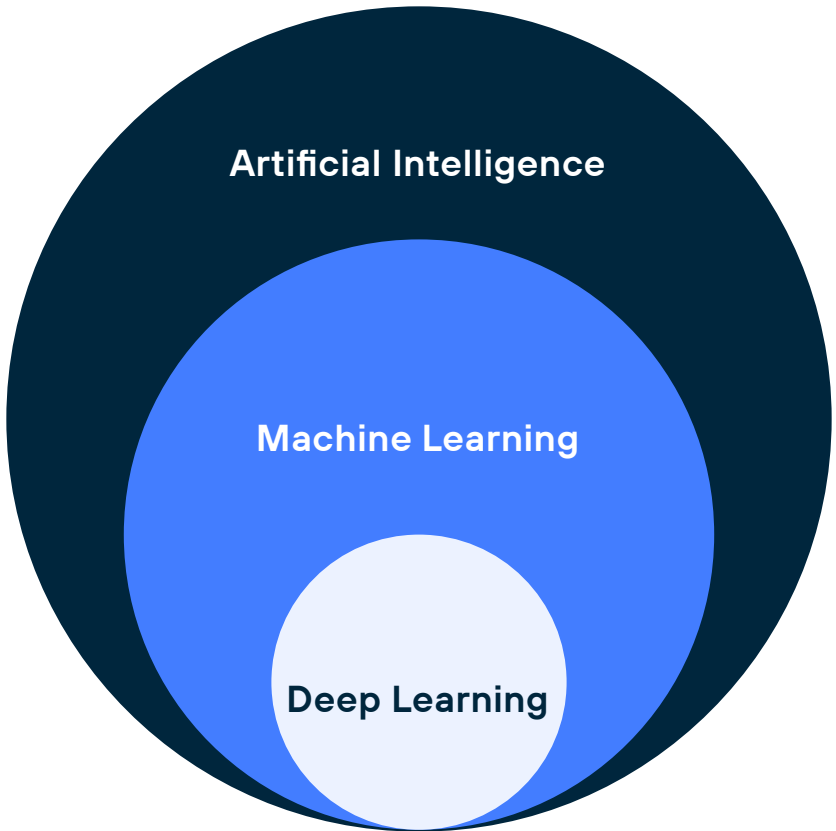


What powers IDP?

AI and machine learning

Artificial intelligence has been a topic of conversation for several decades now. But for most people, our understanding of how it applies to business is limited.

Another element of understanding IDP is knowing how the technology piece works. IDP is not AI alone. That's important to remember. Robots are not taking the place of real people, but you can think of AI as a tool that mimics human intelligence and how we learn. You may want to know only some of what there is to know about AI and machine learning, so here's a short n' sweet version explaining the parts that matter.



What powers IDP?

AI (Artificial Intelligence)

An umbrella term for an application (or algorithm) that helps a computer think and make its own decisions. Examples of AI would be the technology behind a self-driving car or a chatbot application that allows customers to solve a problem online.

The U.S. Postal Service has been an early adopter of AI. How could they not with all that mail? Their retail call centers use an AI virtual agent to assist customers. They have a pilot with the Department of Motor Vehicles (DMV) for intelligent self-service kiosks in post offices. The USPS has a 10-year plan of adopting new technology, like AI, to better serve its customers.

Machine Learning

This is a subset, and the leading edge, of AI that helps the computer learn from data. If you give the machine a large set of data, it will be able to explore and analyze it to help you understand what it means.

This can be done by giving the machine rules to

follow. Once the machine has processed many similar document types, we can rely on it to tell us what “clusters” or data patterns exist.

Through machine learning, AI can analyze any content and output structured information that we can immediately understand to take action.

Where does this leave the humans?

We initially help the machine learn, and then we validate everything. AI is very accurate, however, we still want that additional layer of confidence to know that our information is showing as expected or that what’s being extracted will be beneficial.

We believe only humans can determine that.

DataBank IMX

**Is IDP just another way to
eliminate paper processes**



Is IDP just another way to eliminate paper processes?

The answer is no.

When we talk about utilizing IDP to eliminate paper processes, there's more to it than the immediate and obvious benefits of improving operations, reducing overhead costs, and supporting sustainability initiatives.

All the paper - plus digital - documents, images, and records you own have millions of valuable data waiting to be extracted. The data are begging you to use them.

And in today's business world, where you have to be prepared for any shift, uptick, disruption, or downturn, the data are necessary to predict what's in sight or be ready for what's not.

Here are some examples of what you can do through IDP:

- Automate intake and workflows like invoicing or contract approvals
- Make accurate data accessible to your workforce regardless of location
- Use advanced analytics and business intelligence to power automated insights
- Understand customer, patient, or constituent sentiment and mood to make better predictions
- Support data integration during mergers and acquisitions

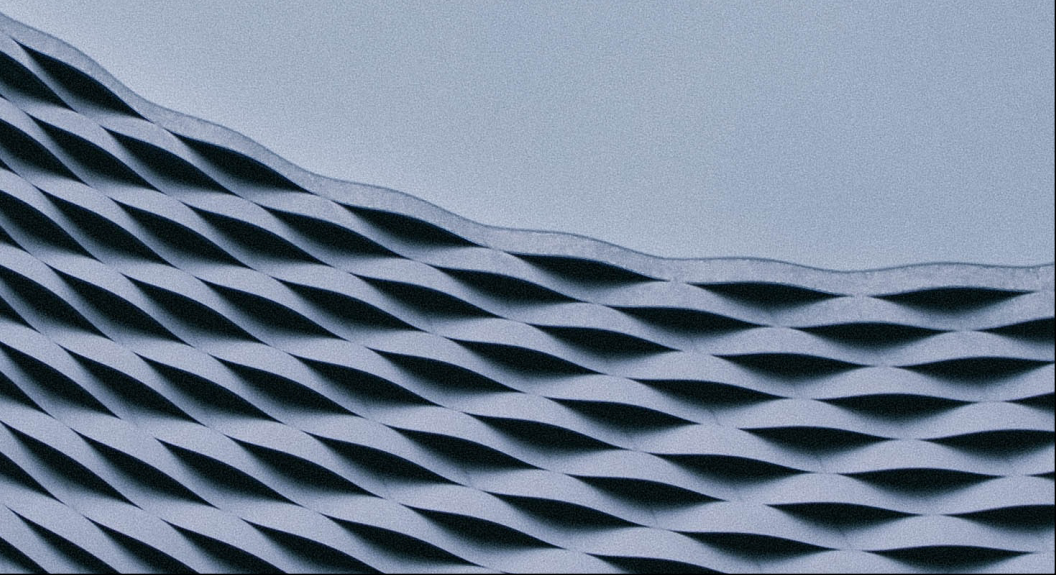
You will need more than simply scanning or digitizing documents to get any of these results.

We designed Content Intelligence, DataBank's IDP solution, to help you unlock that data, make it accessible to the people who need it, and use it to your advantage.

DataBank IMX

Get started with Content Intelligence

BENEFITS | HOW IT WORKS



Get started with Content Intelligence

Turn all your content into trusted data.

Content Intelligence is DataBank's solution that blends data intake, Intelligent Document Processing, and human validation to turn your unstructured content into accessible, usable data to support the outcomes you're seeking.

This is more than just another technology investment. This is a complete, end-to-end solution where we partner with you to reshape how your employees receive and manage information.

Make sense of your content

Turn any digital or physical content into structured, useful information that matters most to you.

Pick up the pace with intelligent processing

Our friends AI and ML do the heavy lifting, reducing manual effort while increasing output.

Trust the data you see with human approval

Modern technology can't replace our value. Information is always validated with human-in-the-loop verification within our secure data management centers.

Gain a partner in the process

We're not a software vendor. We're strategic advisors helping you build a better data foundation and finding opportunities to grow from there.

Benefits



Focus on the important stuff

It's no longer practical to spend hours searching through documents. You have more value-generating projects to work on, and so do your teams.



Transform data management operations

This first step will help you accelerate the flow and management of your most critical information from intake to service delivery.



Share the data wealth

Align people, teams, and departments anywhere by sharing the same information where and when they need it.



Give the people what they want

People are why you exist, so empower your stakeholders - employees, customers, patients, and constituents to make informed decisions during critical moments.



Unlock new insights

With your new structured data, you'll see just how much information you already have on hand, but now you can do something with it.



Move on to the next opportunity

There's always going to be a new business challenge to overcome and a new goal to meet, but your bad data will no longer be a root cause of your pain.

Get started with Content Intelligence

A glance at your Content Intelligence journey

“Digital modernization” is a journey, not a destination. Our offering is intended to meet you where you are. Let’s take this one step at a time.

The work we do is based on the objective you’re looking to achieve. Manage will help you simplify record access. Activate takes it a step further to extract, structure, then activate your data.



01 Define your objective

Our work is based on your goals. **Manage** is about making your information accessible. **Activate** takes it a step further to turn that data into insights that help you make predictions.

02 Discovery

Discuss objectives, file variations, and costs.



04 Content prep

Humans prepare your content to be processed.



06 Structure

Data is extracted and structured based on pre-determine keywords or through AI depending on your objective. Results that don't match your confidence level are sent to a queue to be validated by a human.



08 Activate

Build automated workflows, power reporting, analytics, and insights.



03 File receipt

Physical files are picked up, shipped, and digital assets are delivered to one of our centers.



05 Convert

Files are converted to machine-readable text so rules can be applied.



07 Delivery

Newly structured data is delivered to any system.

Proof of concept

No matter the objective, a proof of concept is delivered before we hit "Go" so you can be sure expectations are being met.

What is Content Intelligence?

Human-in-the-loop validation

We don't believe AI and machine learning can replace human understanding. That's why we've built in human validation to each step of the Content Intelligence process.

A real person verifies that results are as expected and can make real-time adjustments based on your requirements.

Technology isn't perfect, and humans aren't either, but combined, they're a powerful duo.



What is Content Intelligence?

Data Management Centers

With DataBank, IDP takes place in an actual facility with people who have been running data intake and refinement for a long time. Our data management centers are regional and specialized, focusing on the requirements that meet respective industry standards.

The IDP market is flooded with technology providers, but most don't provide data centers that store, process, validate, and handle each document with care on your behalf.



Huntingdon Valley, PA

Primary data management center



Houston, TX

Primary data management center



Chesterton, IN

Primary data management center



Bossier, LA

Satellite center



Canton, MA

Satellite center

What is Content Intelligence?

Information Security

We follow a mature, multi-layered security framework to keep your data safe in transit and at rest.

First layer

Managed endpoint security and network/systems

Second layer

Managed SIEM and threat intelligence

Intrusion detection and prevention

Email protection services

Network behavioral analysis

Advanced persistent threat management

Third Layer

Policies and procedures



ISO 27001 security controls

Information security policies

Asset management

HR security

Incident management



SOC 2 compliance attestation

Logical and physical access controls

Change management

System operations

Risk mitigation

Flexibility to work within our existing or your required framework:

NIST

HIPAA

HITRUST

Other

DataBank IMX

About DataBank

WHY WORK WITH US? | SOLUTIONS



About DataBank

We help people and technology work together to fuel organizational growth.

For mission-oriented leaders who want to better serve stakeholders, DataBank is the data and process improvement partner to help organizations unlock and activate their information.

Why work with us?

In-depth process analysis

We enjoy the journey as much as the destination.

Technology and human ingenuity

We're not a software company but a people company that understands technology's purpose.

Comprehensive solutions

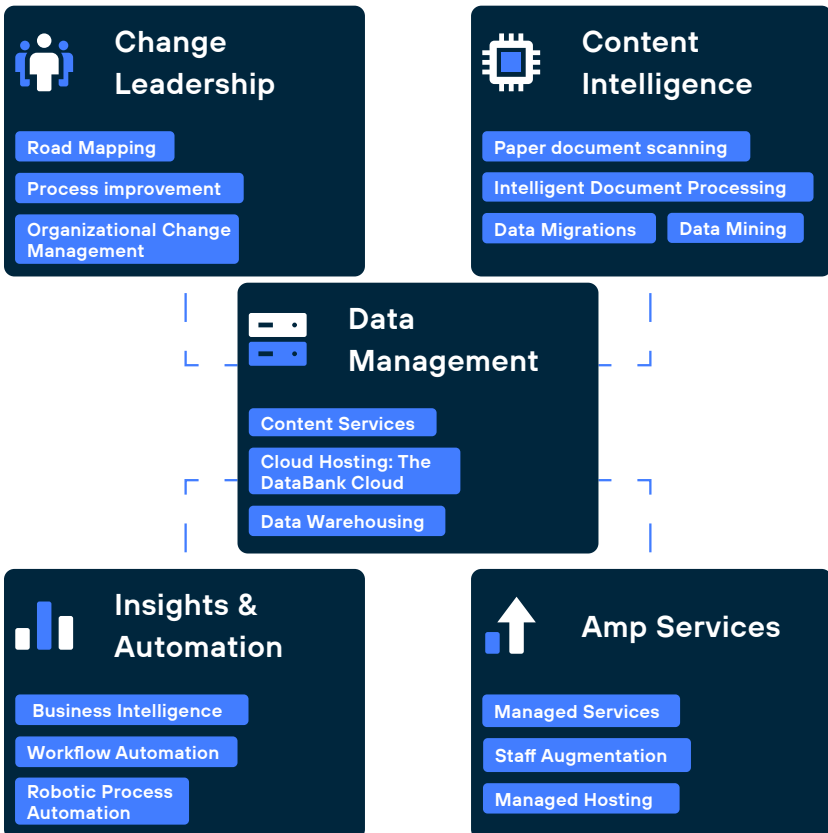
Growth occurs when you consider the impact of every part of a process. Our services and expertise cover it all.

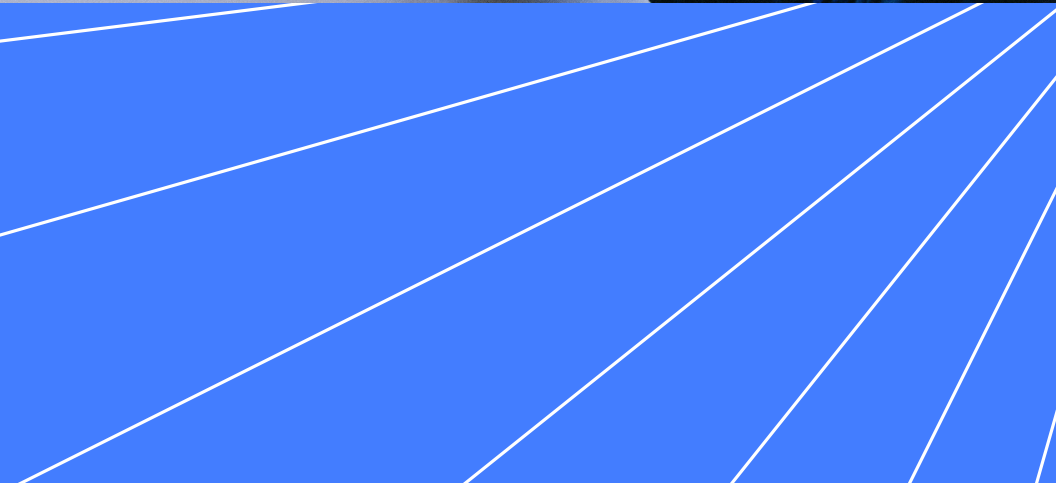
Problem-solving - not temporary fixes

We focus on long-term solutions that eliminate the roadblocks to your success.

End-to-end data and process improvement solutions

Data management is at the center of everything we do. It is surrounded by a suite of services to make the most of your content, data, and technical teams.





Sources

<https://www.oversight.gov/sites/default/files/oig-reports/USPS/RISC-WP-21-007.pdf>

<https://automationhero.ai/blog/guide-to-intelligent-document-processing-2022/>

<https://emtemp.gcom.cloud/ngw/globalassets/en/information-technology/documents/insights/the-it-roadmap-for-data-and-analytics.pdf>

<https://emtemp.gcom.cloud/ngw/globalassets/en/information-technology/documents/2022-cio-agenda-ebook.pdf>





Get started with an expert

<https://www.databankimx.com/connect-with-an-expert/>

