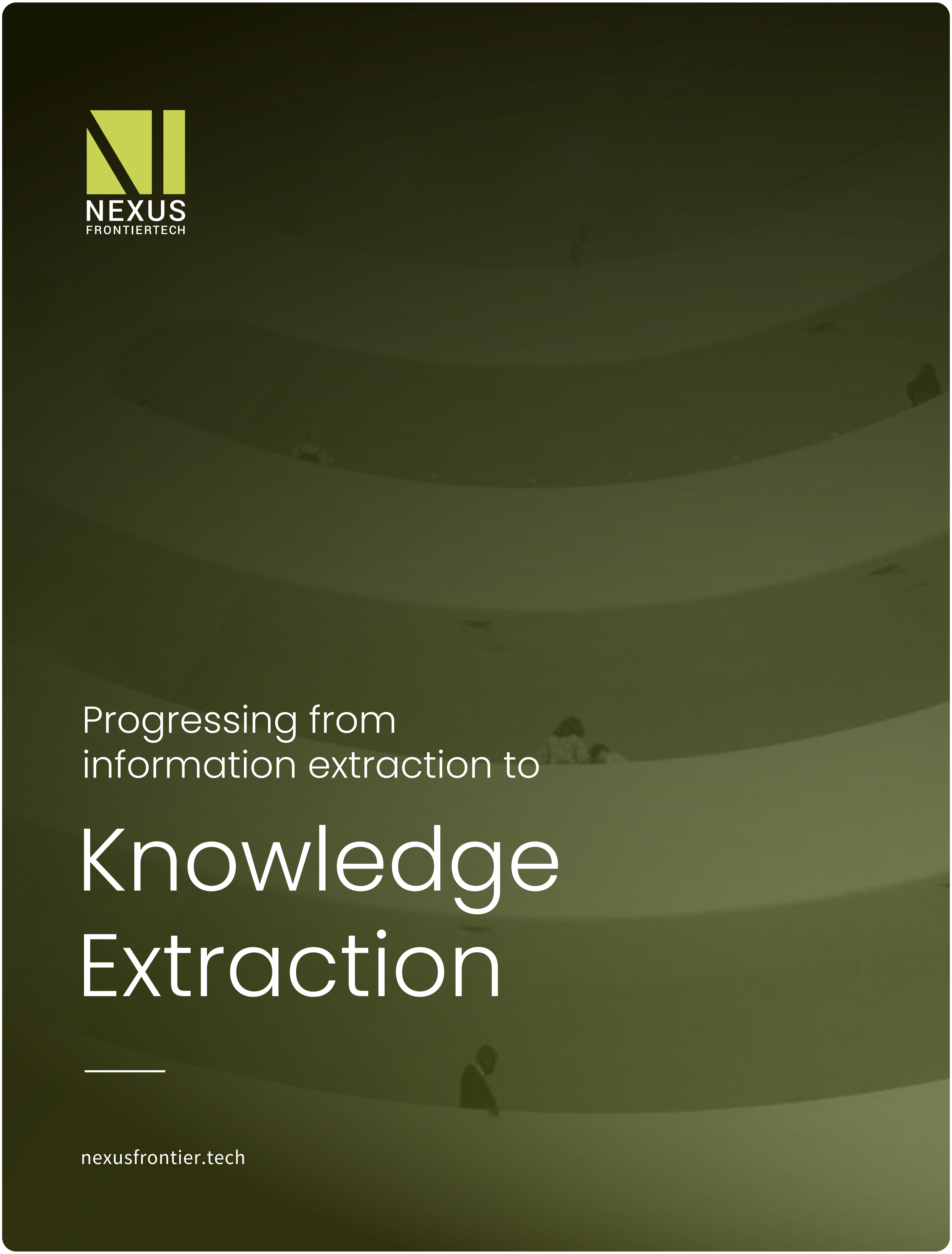


NEXUS
FRONTIERTECH

Progressing from
information extraction to

# Knowledge Extraction

nexusfrontier.tech

# Executive Summary

*While data is increasingly available for consumption, organisations often fail to transform it into useful knowledge to drive real business improvements.*

Part I of this paper will discuss some challenges and techniques for extracting data from unstructured formats. Following this, Part II will address why merely extracting data is insufficient and how it is instead necessary to extract precise data for specific business needs, to be integrated into existing flows. Finally, Part III will explore some practical case studies of applying these techniques to maximise the value of data and provide intelligent insights.

by **Nexus FrontierTech**

PART 1

# Data extraction

Data, when used properly, is a valuable asset for companies to stay competitive and to provide intelligent insights into internal operations and external trends. It is increasingly used as a tool to gather intelligent insights to improve business processes, leading to better decision-making.

While more and more data is becoming available every day, organisations are failing to maximise its potential. A survey by NewVantage Partners found that whereas 56.5% of participating organisations are driving innovation with data, only 26.5% reported achieving the goal of becoming data-driven and only 19.3% had established a data culture.

| | |
|---|---|
| 56.5% | *participating organisations driving innovation with data* |
| 26.5% | *achieving goal of becoming data-driven* |
| 19.3% | *established data culture* |

**Part I of this paper will discuss the common reasons for this and some data extraction techniques to make data more usable.**

## Diverse and unstructured data

Companies often fail to maximise the value of data due to its diversity in terms of sources, quality, and format. As larger volumes are created and made available from different channels internally and externally, it is becoming more challenging to centralise and provide a holistic overview of the necessary data. This makes data processes prone to duplicate efforts and miscommunication.

Another obstacle is the fact that the majority of organisational data is still in unstructured format. While structured data is standardised according to pre-defined formats and hence is easy to search, extract, and analyse, unstructured data is more difficult to handle since it lacks organisation.

This type of data, such as images and scanned documents, is traditionally unusable by machines without additional processing steps. It can take a lot of manual labour to standardise, sort, and clean this data which is time-consuming and costly.

## Intelligent Document Processing

Fortunately, technology to convert unstructured data into structured formats is advancing. This form of technology is referred to as Intelligent Document Processing (IDP). The value of IDP is being recognised across many sectors, and is increasing in adoption. In fact, it is predicted that the Intelligent Document Processing market will grow to $3.7 billion in 2026.

### EXAMPLES OF TECHNOLOGIES USED IN IDP INCLUDE:

**Optical Character Recognition (OCR)**
to convert images of text into searchable text.

**Intelligent Character Recognition (ICR)**
to convert handwriting into searchable text.

**Computer Vision**
to extract information from images, such as floor plan data.

**Natural Language Processing (NLP)**
to perform sentiment analysis and pattern recognition on free text inputs.

Companies are increasingly migrating to digitised forms of information capture, which removes the need to retroactively structure data.

However, using data extraction tools to unlock large volumes of structured data is insufficient if it does not translate into salient knowledge. In fact, data extraction is only the first piece of the puzzle.

To realise the full value of data, it is necessary to use targeted IDP models, consider the broader application of automation, and integrate solutions into existing workflows. The next section will explore these concepts.

PART 2

## How to implement & integrate Intelligent Document Processing

# Targeted Extraction

Data must be extracted precisely and purposefully to drive value. IDP solutions can seek to solve specific business problems, which necessitates identifying the right opportunities for data extraction.

This is particularly important since different document types will require different models and technologies. There is a varying range of techniques available for different extraction types and this technology needs to be carefully selected, adjusted, and combined to apply to the specific scenarios to optimise accuracy and usability.
A single IDP model will be unlikely to extract from all types of documents across a business.
For example, data can come in complex forms, such as tables and diagrams, which require specialised models.

As a result, it is necessary to analyse the operations within a company to detect and prioritise use cases for IDP. Solutions which are scalable across business units are often the best contenders for maximising the potential of such technology.

**Some key questions to ask during this investigation include:**

- Is there a solid business case for applying IDP to this use case in terms of volume and cost savings?
- How many different document groups exist within the use case?
- How many extraction points are there within each document?

- How complex is the extraction?
(E.g. are extraction points always found in the same location? Are extraction indicators consistent or widely?)
- What variations in quality and format exist for each document group?
- What technologies will be required to extract from the documents?

What extraction accuracy is necessary for a viable solution? Is this technically feasible?

# Integrating IDP

IDP is not a standalone technology but rather the initial step in maximising the value of data. Consequently, it is important to consider how IDP models will work within existing systems. However, it can be difficult to integrate AI models into the larger infrastructure of a business. Additionally, data will need to be stored and managed.

The right technology architecture and technical expertise are critical. It will be necessary to have a well-designed production environment which is dependable, flexible, and scalable. This requires not only the right hardware and software but also a competent team of AIOps professionals, with the expertise to build, integrate, test, release, deploy, and manage the system to transform the results of AI models into actionable user insights (Tse et al.)

# Data extraction

## AS PART OF A BROADER STRATEGY

Extracting data is only one part of a comprehensive document processing strategy, which should comprise the following steps:

**STEP 1:**

Capturing document inputs into the system, usually from multiple channels of data and different document formats.

**STEP 2:**

Classifying and indexing documents into appropriate groupings, usually via a machine learning (ML) model.

**STEP 3:**

Extracting data by:

a. converting unstructured sources into a machine usable format and
b. selecting the relevant and valuable data points with which processes can be automated and intelligent insights can be gained.

**STEP 4:**

Validating extracted data to inform future improvements and model retraining.

**STEP 5:**

Integrating extracted data into the appropriate workflows automatically.

An automated process should be in place to automatically transfer data down the pipeline from one process to the next. IDP will make targets available for analysis and visualisation, for example through interactive dashboards. This can help to inform decision-making as well as to uncover previously undiscovered trends.

PART 3

## Knowledge extraction in action

This section will explore some Nexus FrontierTech case studies wherein IDP was used to extract knowledge for tangible business benefits.

## 01 Case study: Licence application processing

Nexus co-created an AI-powered chatbot with a global market regulator to streamline licence application processing in financial services.

### CONTEXT

The regulator's licence applications were information-heavy and often involved missing or invalid inputs. Previously, applications were manually processed through manual extraction and transferral of data to other workflows for future processing. This was inefficient and led to long turnaround times.

▶ **IDP SOLUTION:** As well as creating document processing models, we designed an integrated intelligent chatbot to guide the applicants through the process. Reviewers are then automatically presented with the extracted and compiled information for their assessment.

By considering the broader workflow of the process instead of data extraction alone, we created a solution that extracted useful information and integrated it into an intuitive and user-friendly workflow. The solution allowed for real-time feedback and interaction, reducing information gaps and turnaround times leading to a better user experience. Not only was this faster and more cost-efficient, but human operators could spend their time dealing with more nuanced decision-making.

Fund Report
Market Report
Individual client
Institutional client

Are you an individual or institutional client?

Institutional client

Company name AA-AAA

Name of fund BBB

Investment date 27/10/2021

Valuation currency $$$

# 02 Financial Spreading

Nexus helped a top 20 global bank to automate financial spreading - the process of capturing, spreading, and analysing financial data to calculate credit scores.

## CONTEXT

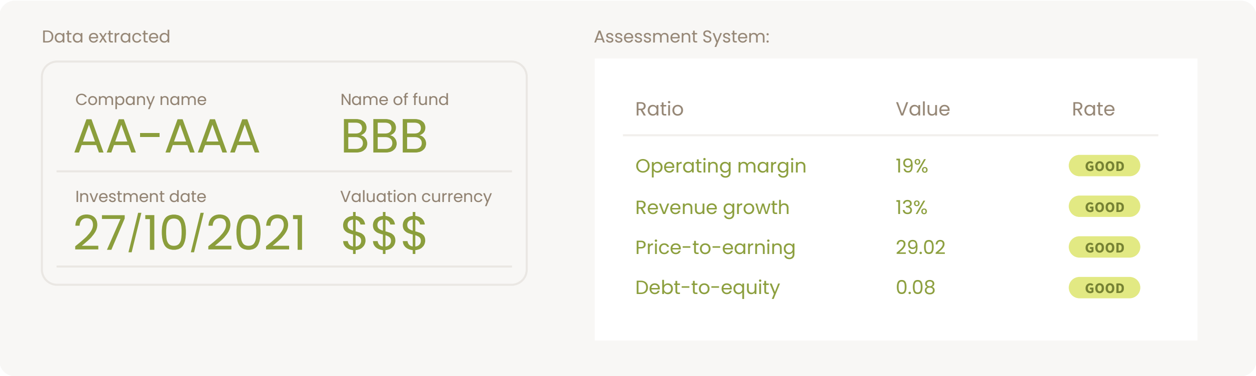
Credit analysts at the bank manually input and spread data from financial statements and transferred it onto the bank's credit scoring model. The previous process was manual and slow and prone to errors and inconsistency.
This was especially due to the lack of universal classification and interpretation of data.

▶ **IDP SOLUTION:** We designed a Financial Spreading solution to retrieve scanned financial statements from the bank's internal database. Following this, IDP models extract and validate data from the documentation according to the spreading rules based on the client's credit scoring model.

This solution required integration with existing workflows and systems; for this to be possible, we had to carefully consider its interactivity with existing systems and workflows. Importantly, the credit analysts are still a key part of the process to review unclear or missing data, although their manual data extraction responsibilities are greatly reduced.

## 1 SCANNED FINANCIAL STATEMENTS RETRIEVED FROM BANK'S INTERNAL DATABASE

## 2 DATA IS AUTOMATICALLY EXTRACTED AND VALIDATED



Data extracted

| | |
|---|---|
| Company name<br>AA-AAA | Name of fund<br>BBB |
| Investment date<br>27/10/2021 | Valuation currency<br>$$$ |

Assessment System:

| Ratio | Value | Rate |
|---|---|---|
| Operating margin | 19% | GOOD |
| Revenue growth | 13% | GOOD |
| Price-to-earning | 29.02 | GOOD |
| Debt-to-equity | 0.08 | GOOD |

## 03 Sales quality monitoring

Nexus developed an automated solution to drive compliance and reduce costs for a large global bank monitoring financial product sales quality.

CONTEXT

A team of 120 reviewers were tasked with checking financial products sales at this bank. However, the review process was lengthy, requiring 180 data points to be checked across 10 document types. This meant it was only possible to check 10-15% of cases around 2-3 weeks after the sale leading to compliance gaps and potential oversights.

▶ **IDP SOLUTION:** We developed a series of AI models to extract the relevant data from different document types to close compliance gaps and to check all cases within an hour after the point of sale. These included several unstructured document groups, such as scanned bank statements.

Due to the wide variety of document types and qualities involved in this project, it was necessary to design different models to perform the extraction tasks. These targeted extraction models were able to reach a high sustainable accuracy (around 80% at early stages, reaching up to 90-95% over time in production).

Bank Stateme
Market Report
Company name
AA-AAA
Name of fund
BBB
Investment date
27/10/2021
Valuation currency
$$$

## Conclusion

Innovative solutions are key for minimising manual data processing and to offer intelligent insights. This involves leveraging the appropriate IDP capabilities that are available. However, merely having access to and extracting a lot of data does not maximise the potential of data solutions. Organisations have to go a step further to transform data into knowledge through targeted solutions that can be automatically integrated into organisational workflows. In this way, the value of data initiatives can be maximised.

## REFERENCES

**Markets and Markets (2021)** *Intelligent Document Processing Market by Component (Solutions, Services), Deployment Mode (Cloud, On-Premises), Organization Size, Technology, Vertical (BFSI, Government, Healthcare and Life Sciences), and Region - Global Forecast to 2026.* Available at: https://www.marketsandmarkets.com/Market-Reports/intelligent-document-processing-market-195513136.html#:~:text=%5B271%20Pages%20Report%5D%20The%20global,36.8%25%20during%20the%20forecast%20period (Accessed 20 June 2022).

**NewVantage Partners (2022)** *Data and AI Leadership Executive Survey 2022. Available at: https://c6abb8db-514c-4f5b-b5a1-fc710f1e464e.filesusr.com/ugd/e5361a_2f859f3457f24cff9b2f8a2bf54f82b7.pdf (Accessed 20 June 2022).*

**Tse et al. (2020)** *'The Dumb Reason Your AI Project Will Fail', Harvard Business Review. Available at: https://hbr.org/2020/06/the-dumb-reason-your-ai-project-will-fail (Accessed 20 June 2022).*

# Ready to kickstart your Intelligent Automation Journey?

Book a free consultation with a data assessment and demo to accelerate your decision-making processes.

Schedule a Meeting

**United Kingdom**
Corporate & Sales

**Japan**
R&D & Sales

✉ info@nexusfrontier.tech

**Singapore**
Tech, Sales & Marketing

**Middle East**
Corporate & Sales

www.nexusfrontier.tech

**Vietnam**
Execution & Delivery

## About Nexus FrontierTech

Nexus FrontierTech accelerates decision-making processes by enabling modular automation solutions on our proprietary  AI engine. We bring visibility, traceability and usability to enterprise data in real-time, empowering the financial services industry to efficiently develop structured processes for compliance, risk management, and innovation purposes.