

The logo for INTENDA, featuring the word in a stylized, blue, sans-serif font. The letters 'I', 'N', 'D', and 'A' have small red dots above them. A thick red horizontal line is positioned below the logo.

INTENDA

PUBLIC SECTOR
CASE STUDY

INNOVATIVE
REAL-WORLD
SOLUTIONS / FOR
TODAY / FOR THE
FUTURE

FRAXSES



**Firstly, we're eliminating the client's
cumbersome copy-and-paste approach
to lifting text from PDFs.**

THE CLIENT

Our client compiles data that is sourced from the public domain and published by government municipalities for public consumption.

The data is lifted from public PDFs, stored in a central database, and served to users in the form of dashboards.

The client needed an entirely different approach to data processing and management, and a solution that would provide best in class processes for PDF processing, data storage and business intelligence.



THE SOLUTION

Our solution for this client entailed two parallel work streams, the first being the conceptualizing of the future state end-to-end process and the other to configure the corresponding technical components this would require, using the fraXses data fabric.

We deployed fraXses on-premises to host and virtualize the final data set, so that content is delivered via the fraXses visualization tool.

The client's original time estimate for completion of their project was two years. We delivered the entire solution in under four months, effectively streamlining two of their major production bottlenecks by implementing two microservices:

- **Firstly**, we're eliminating the client's cumbersome copy-and-paste approach to lifting text from PDFs. Originally, workers would sift through electronic documents and transcribe data, page by page.

Now, business analysts begin by building a metadata library for each PDF.

They identify which pages contain relevant data, record them, and then capture the coordinates of the data in a zonal capture programme. An OCR service can then digest those coordinates and automatically lift the text data from the PDFs.

fraXses can then re-assemble useful data and filter out the noise from the result sets provided.

The client's original time estimate for completion of their project was two years. We delivered the entire solution in under four months...

(CONTINUED)

- **Secondly**, we have developed a classification microservice for the client, which automates enrichment and classification - this would otherwise be a manual process. This enrichment service is built using human-defined keyword matching.

This process adds over 100 unique dimensions to the dataset. These dimensions are later used to customize the user experience in the fraXses visualization tool. Technically, we delivered an on-premises hosting solution. The input to the data fabric platform is an excel database that a business analyst has reviewed for accuracy.

The FlowZ module runs a batch process to dimension the data into its components while virtual data objects rebuild the original structure to create a hosting-friendly data model.



This process adds over 100 unique dimensions to the dataset.

BENEFITS

While the current end-to-end processes are loosely coupled, the future state of these processes entails tight integration with document management software.

Enrichment and classification will move to machine learning on the Basis Tech NLP engine.

Essentially, business analysts will be able to run these procedures from a data workbench in an automated fashion, which would not be possible without the metadata catalogue approach and automation we have provided for this client.



Essentially, business analysts will be able to run these procedures from a data workbench in an automated fashion...

____WE
SEE
DATA
INSIDE
OUT____

THANK YOU
GET IN TOUCH

www.intenda.tech