



**Flux7**  
Case Study



**Oil Leader Digitizes Document Library for  
Analytics, Compliance with Amazon Textract**

# Case Study

## Oil Leader Digitizes Document Library for Analytics, Compliance with Amazon Textract



### Profile

This organization has a long and rich history of providing services to the oil and gas industry around the globe. This firm has established itself as a leader by specializing in demanding areas of the industry.

### Challenge

Working with detailed geological data and regulatory compliance requirements, this energy sector firm has over seven million physical documents that it has saved and stored. However, the process to access these documents filed in deep storage requires an employee to get in their car and drive to the warehouse, and physically locate the files. This system is labor and time intensive as files are not always accurately filed or cataloged.

For these reasons and more, the firm sought to digitize its sizeable document holdings. In the process, it wanted to make sure that it would be

1. Easy to archive documents moving forward,
2. Simple for operators to search for and find the desired information, and
3. Straightforward to conduct data analytics on its impressive data store.

**Solution** Working with AWS Consulting Partner Flux7, the company created a working plan to digitize and catalog its vast document library. AWS had recently announced a new solution, Amazon Textract (which was still in preview mode) that Flux7 knew would be ideal for the task. Amazon Textract is a service that using machine learning automatically extracts text and data from scanned documents. Unlike Optical Character Recognition (OCR) solutions, it also identifies the contents of fields in forms and information stored in tables, allowing the company to conduct full data analytics on its data once digitized.

The teams started with a proof of concept, in which several dozen physical documents are scanned. Once scanned, the documents are uploaded to S3. Upon upload, Lambda functions are triggered which in turn launch Textract document parsing. The output is a JSON file which is stored in an existing Elasticsearch cluster which is in turn queried by Kibana, which allows the organization to visualize their Elasticsearch data. In addition to the data being presented to Kibana, URLs for specific documents are presented to users.

Interfacing with the data via Kibana, end users can now create smart search indexes which allow them to quickly and easily find key business data. Moreover, operators can build automated approval workflows and better meet document archival rules for regulatory compliance.

## Oil Leader Digitizes Document Library for Analytics, Compliance with Amazon Textract



### Benefits

The Amazon Textract POC solution exceeded the team's goals, allowing them to quickly digitize, parse and store document pages; every file in the initial POC was scanned and indexed in Kibana in milliseconds. Next steps for the company are to apply these steps en masse. Starting with handwritten and other hard copy documents, a third-party vendor will begin scanning in the vast document library.

As Amazon Textract automatically detects the key elements in a document or data relationships in forms and tables, it is able to extract data within the context it was originally created. With a core set of key parameters, such as revision date, extracted by Textract, operators will soon be able to search by key business parameters across the library, finding everything they need -- all from the quick comfort of their desk.

### LEARN MORE ABOUT FLUX7

As DevOps and AWS experts, Flux7 offers a suite of solutions that help organizations design, build, own and manage IT modernization projects. Focused on architecting and optimizing their clients' AWS infrastructure and training internal IT teams to manage their own infrastructure, Flux7 solutions are rooted in DevOps best practices. Flux7 has delivered hundreds of agile, right-sized projects to satisfied customers across industries, creating a well-architected core from which these business can own and expand their IT modernization.