

# PDF DATA EXTRACTION

WHAT YOU NEED TO KNOW

Data extracted from PDFs can be applied in many applications:

- ARTIFICIAL INTELLIGENCE
- BUSINESS PROCESS AUTOMATION
- eDISCOVERY
- MACHINE LEARNING
- DECISION LOGIC APPLICATIONS
- DATA SCIENCE AND ANALYTICS

Important factors to consider for complete data extraction:

## Ensure that the solution has advanced table extraction capabilities

Varied table formats are challenging for extraction tools to effectively process. Check to confirm that your solution can achieve accurate data correlation from tables for accurate text flow reconstruction.

## Multi-lingual support is critical with increasing globalization

PDF documents that contain multiple languages are a challenge for many tools to process, make sure your solution can tackle documents containing multiple languages leveraging complex extraction capabilities.

## Make sure that full OCR integration is supported

Choose a solution that can offer unique dual processing of text and images to ensure that both are addressed separately. Implement accurate OCR image processing while maintaining text within PDFs as pure text output for optimal data extraction.

## Implement an API or server tool with flexible HTML, XML and EPUB output options

Your needs are unique. PDF Alchemist is available as a scriptable server tool which provides seamless integration into existing workflows, and as an SDK, ideal for SaaS and OEM implementations. Output choices provide flexibility to meet your specific requirements.

Start your free PDF Alchemist trial today!

Subscribe to our monthly newsletter for more tips to effectively manage and work with PDFs.

**Datalogics**

datalogics.com | 312.853.8200