



## **Automating PDF Data Extraction Using Robotics and AI**

Investment management firms, funds of funds, family offices, RIAs and fund administrators alike receive a deluge of financial documents such as investment reports, broker statements and K-1 forms through unstructured digital documents such as PDFs on a regular basis. These documents are packed with important data, often buried in an enormous amount of standard compliance jargon, disclaimers and small print. Emerging robotics and AI tools that automate the extraction and processing of data from PDFs can greatly improve the operational efficiency and agility of financial firms.

It is not uncommon for these institutions to receive PDF financial documents of over 200 pages from each one of their brokers, custodians, or portfolio managers at the end of each month or at irregular intervals. At the same time, clients and investors with external accounts and assets held elsewhere may also choose to share their holdings by similar means. Opening such a document, and quickly and efficiently identifying relevant information is a daunting task that many choose to forgo to their own detriment.

Not only do these documents come in different forms, such as PDF, inline emails, and text attachments, but they also come at different times and at different frequencies. Making things even more challenging, the data that they include are often not exchange-traded securities with common symbology. Assets can be referenced by name, ticker or CUSIP within the same document, while positions and balances may be calculated by trade date in one section and by settlement date in another.

Financial institutions have long grappled with extracting data out of electronic documents and most of them already have processes in place to address the problem. Some have relied on OCR extraction, some on purely manual methods, and others on a combination of both. These solutions have not produced satisfactory results, as they tend to be slow, error-prone, expensive, and ultimately unable to adjust to

unpredictable format changes and sometimes even intentional obfuscation by the institutions that generate them.

## Difficulties in Data Extraction from PDFs

The process of extracting data from financial documents in PDF format presents a number of problems for investment managers that automation can solve.

- **Complex:** Employees have to collect data from various sections of the same document, transform and normalize it before they copy it out to an internal system or spreadsheet.
- **Unreliable:** Getting data from PDF documents is a tedious, manual process that is typically assigned to low-level employees, whose work is rarely cross-checked or proof-read. As such, it is prone to errors, omissions, and misunderstandings.
- **Time-consuming:** In a business where minutes and seconds matter, data extraction is slow and inefficient. It may not be a problem if you are wrestling with a half dozen documents a month, but it quickly becomes unmanageable if you have 20, 50 or hundreds.
- **Expensive:** Employing a battalion of resources to manually extract information from PDF documents is labor-intensive and thus costly, with the expense mounting as the complexity of each document increases.
- **Incomplete:** Because this slow, complex, and error-prone process is so costly, many companies opt to capture only a fraction of the data contained in the PDF documents they receive, potentially leaving game changing insights locked in the documents.
- **Insecure:** In an attempt to reduce cost, manual data extraction is often outsourced, putting extraordinarily sensitive information into the hands of outsiders. The security risk is tremendous.

These difficulties aside, solving the problem of PDF data extraction is well worth the challenge.

## Generating Agility Alpha

The ability to quickly, reliably, and securely extract data from broker statements, tax forms, and other financial documents presents an opportunity for investment managers to reduce costs, create differentiated services for their investors, and gather more assets under management. Extracted data can be quickly integrated with other internal and external data and fed into state-of-the-art portfolio management systems and risk platforms that help investment managers get high-level insights into their investment strategies and better navigate today's tumultuous markets.

Digital data extraction technology using robotics and AI, such as Accordia has developed for its clients, as a standalone service, or an integral part of a cutting-edge, cloud-native portfolio management service, makes unscrambling and extraction of usable data from electronic documents fast, error-free, secure, and

affordable. By accurately and efficiently capturing these external data sources, investment managers can reduce costs, enhance their agility and speed of decision making and gain an edge over their competitors. That's Agility Alpha.