SMONIK
Systems LLC

- **CASE STUDY: PDF Data Extraction – Private Wealth Manager**

SITUATION

In 2018 Smonik was introduced to a service provider who provides staff augmentation to a large mid-western Consulting/Audit/Tax/Wealth Management firm. Many of this firm's wealth management clients held positions in hedge funds and/or private equity funds which made the process of data collection and data input a challenge. In addition, the wealth manager also received broker activity statements in pdf format from which the service provider was required to manually extract data and input to the internal accounting system; Advent Portfolio Exchange (APX). The service provider was looking for a way to automate the data collection and data entry function for these types of investments.

CHALLENGE

The challenge for the service provider was how to automate the input of data from hedge fund and private equity fund investment activity, as well as transaction data taken from broker statements; collectively known as "manuals". Typically, private investment data is distributed to investors in pdf format. Monthly or quarterly statements, capital call notices, distribution notices, etc. are usually created as pdf documents and then sent to the investor either as an attached file in an email, or saved to a website where the investor is able to login, find the file, and then download it. The challenge for investors in private funds is that the data included in the pdf document is not in a format where investors can integrate it with internal systems. The only way for investors to retrieve the data from these pdf documents is to have people read the pdf document and type the data into another system. Private investment fund documents are rarely structured the same; meaning the location of the data is inconsistent.

Firms also face a similar challenge when receiving pdf statements from brokers. Extracting data from these broker statements can be very challenging as the statements can have complex structures and require significant amounts of data cleansing and translation. Commercially available pdf data extraction tools struggle with this type of document.

SOLUTION

Smonik and the service provider agreed to conduct a proof-of-concept using samples of this specific wealth manager's private investment documents and broker statements. After signing an NDA, the service provider sent samples of capital statements, capital call notices, distribution notices and broker statements to Smonik for processing. The POC was a success and the service provider, in conjunction with the wealth manager, agreed to sign a contract with Smonik.

The first step was to configure data templates aligned with the standard data elements to be extracted. Three data templates were agreed and created: a trade template; a position template; a pricing template. Each template is also the source for eventual output files to be integrated with APX.

The next step was to configure the mappings for each type of document to be processed. Smonik worked with the service provider to review the documents and decide which document types required a separate mapping. From over 5,000 documents processed in the first year, approximately 400 – 500 per month, 31 mappings had to be created. The mapping process is a one-time effort, which only requires maintenance if the source document structure is significantly changed.

Once the templates and mappings were configured the service provider began processing documents using the Smonik tools. Output files are then generated, in the specific format required by APX, which create trade transaction, position and pricing files.

BENEFITS

The obvious benefit of employing the Smonik PDF Data Extraction tool is staff **efficiency**. What was once a completely manual process is now automated, significantly improving productivity. The secondary benefit, although arguably more important than the efficiency, is the **accuracy** of the data. Moving to an automated data extraction and input process eliminates the typing errors and other "human" inaccuracies inherent in a manual process. Smonik continues to work with the service provider to configure the most complex document types with a goal of automating 100% of the "manuals".