

With our flagship *DataInFormationSM* suite of solutions for Image Labeling, Data Annotation, NLP Calibration and Form Data Capture, Liberty Source is proud to serve a wide range of clients across many industry verticals. Our *Solution Briefs* provide a snapshot of the challenges clients bring to us, and the outcomes we achieve for them.

Historic (Liber-based) Real Property Data Capture

Industry	Financial Services / Banking	Data Type	Semi-Structured Text (.pdf)
Project Duration	5 Months	Ongoing?	No

Challenge

A large financial institution needed historic real property information for a home affordability analysis. Originally contained in liber volumes, the pages had since been digitally scanned to .pdf images. The low-resolution images were of poor quality, including wrinkled pages, skewed columns, mis-aligned rows, etc. Given the large volume of data (>50MM characters), manual data entry was not feasible.

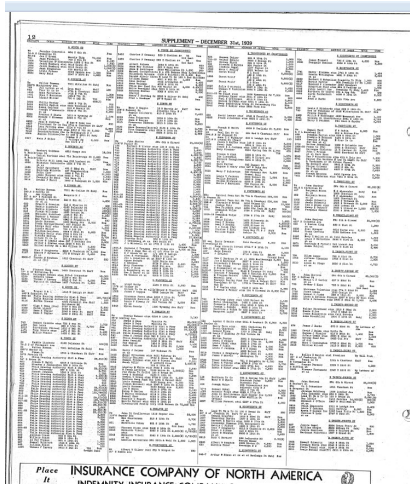
Solution

Our solution included multiple phases: initial staging and preparation of the data, automated scanning and capture, quality assurance checks, and output formatting. Data Associates initially parsed each scanned page image to enable enlarging the image to reduce the density of the data. An AI-driven scanning process captured 50-70% of the contents of each page, augmented by human-in-the-loop review and correction to increase the capture rate to 99%. A two-step quality review verified the final captured information against the original image to bring the capture accuracy to the required 99.5% level. Finally, the captured information was formatted to match the input requirements of the analytic platform.

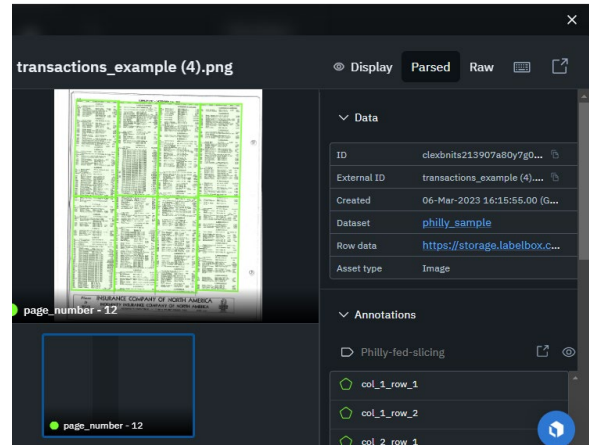
Outcome

The multi-stage intake process, combined with human-augmented automated data capture allowed for a solution that complied with all requirements – including a faster throughput speed than was anticipated.

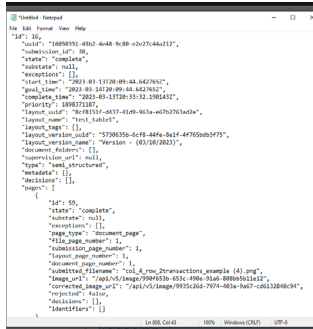
Historic (Liber-based) Real Property Data Capture



Raw Data (scanned document)



Data staged for first-pass (automated) labeling



JSON Output

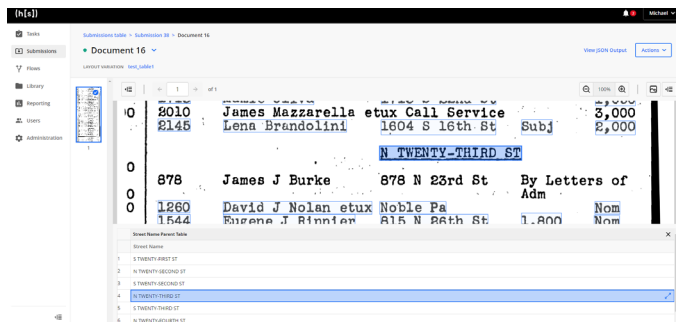
Denotes 'human in-the-loop' process flow



Analyst validates/corrects initial machine annotation



Analyst segments raw data for downstream labeling



RPA tool labels each data element