



## Loan document automation at a top-five Originator

*Paradatec for Mortgage* provides a unique approach for companies wishing to reduce the manual labor costs and increase the accuracy levels associated with classifying and capturing data from loan documents.

The financial services industry is challenged with managing large volumes of documents with varying layouts containing immense amounts of data – part of which is highly critical with regard to compliance. The traditional manual process for classifying and keying data from these documents is time consuming, error prone, and costly due to the sheer volume and complexity of the documents. In an industry where standardizing forms is not possible due to their varying systems and points of origination, an acceptable automation solution must be able to cope with this variability.

General Information	
Project Name	Mortgage Loan Audit Automated Document Recognition (ADR) and Data Extraction
Systems Integrator	Paradatec, Inc.
Products	PROSAR-AIDA, PROKEY, Paradatec Statistics, Paradatec Workflow
Publications	Case Study – Top Five Originator (PDF)



## Client

### Top-Five Originator

This bank is one of the largest in the United States. It is a leading lender offering a range of quality home loans, including government and conventional. These loans are provided through multiple channels.

## Challenge

The mortgage lending industry presents a number of unique challenges for classifying and extracting data from key documents. This is due in part to the large volumes of disparate document variations found in most loan files.

- A typical incoming mortgage loan file may contain 250 to 600+ pages of various size documents, comprising more than 250 potential document types. Older loans files may grow to well over 1000 pages.
- Manually sorting each set of loan documents is a labor intensive and error prone effort, typically requiring the addition of document separator pages if the file is to be scanned.
- Due to the sheer labor effort required, the typical level of detailed document sorting possible with a manual approach is very "coarse". In other words, only the most critical documents and document groups are classified rather than attempting to identify all specific document types. An example of this limitation might be a manual grouping of a series of specific documents into a "Credit Documents Group" rather than breaking these out specifically by document types such as bank statements, credit reports, and brokerage statements.
- To compete in this extremely competitive market segment, organizations are looking for ways to reduce costs and streamline their processes.

In addition to the challenges described above, this top five originator was looking for a solution to help automate the laborious task of providing data for a number of audit-centric applications. These ad-hoc projects commonly had tight timelines and included wide ranges of loans, and millions of pages to be audited.

## Project Description

At the start of the Project, this top five originator had in place a sophisticated document capture infrastructure feeding a well-known enterprise content management system. What was missing from this infrastructure was an advanced recognition module that could deal with the document variations expected in an organization serving borrowers across the nation. The Paradatec solution provided a seamless interface to this capture infrastructure. This greatly simplified the implementation by allowing the existing interfaces to both front-end scanning and back-end image storage to be largely unaffected by the addition of the recognition technology.

Prior to the installation of the Paradatec components a large team would manually classify incoming documents into a moderately broad range of categories or Document Groups. Once these documents had reached the enterprise content management system, a team of underwriters would review, manually enter data, and process the loan.

Limitations of this approach included:

- Heavy reliance on the skills of the people manually classifying documents and extracting data. Error rates varied from operator to operator. Thus, a loss of a skilled operator for any reason had a negative impact.
- Time is of the essence in any mortgage processing environment. Using a human-centric approach meant that



processing times were proportional to staff availability at any given moment.

- People tend to be more expensive than computers and software.
- Regulatory bodies as well as this originator would have preferred a greater granularity in the way documents were classified. This need was however outweighed by the complexity and difficulty presented when attempting to teach and maintain a group of individuals in how to classify documents among over 250 possible choices.

The Paradatec system was selected after an exhaustive evaluation process. A competing solution was initially selected. However, after months of tests it was determined that the Paradatec system had a number of capabilities that surpassed other solutions previously tested or reviewed:

- Paradatec was by far the fastest technology available to read / OCR mortgage documents. Pre-production technical due diligence empirically showed a system that was capable of processing approximately 1 million images per day on a single twelve-core server.
- The Paradatec system was able to use one set of rules to process and recognize all document variations. Because of the extremely large number of documents (and variations of each) which this top five originator encounters, they required the flexibility offered by a non-template-based ADR (Automated Document Classification) and data extraction solution.
- *Paradatec for Mortgage* offered pre-built mortgage logic which “understands” the vast majority of the document types and variations that were required to be recognized. This feature of the Paradatec offering allowed this originator to rapidly implement an ADR and data extraction solution for their specific needs.

The initial focus was to implement an ADR solution that supported more than 250 different document types and potentially hundreds of variations of each document type. The vast majority of the pages in a loan are now identified automatically with no human intervention. The remaining exceptions are presented to operators who either accept the first choice page type or choose an alternative. The Paradatec system is able to narrow down the page types that are lexically possible based on the text on the page. Because of this, in most cases, the operator can choose from a list of no more than five alternate page types. This reduces errors and review time in the verification process.

Upon production implementation of the ADR solution, the focus shifted to automatic data extraction. A list of more than 1500 fields was identified for the first implementation phase of data extraction. Both this project and the ADR work that preceded it were initially implemented in one of the originator’s major channels in order to ensure a wide variety of document sources and variations.

Today both of the projects described above are in production. The amount of manual labor previously required for these tasks has been reduced significantly. Error rates are lower than the human processes that preceded implementation. The end to end processing time has been vastly reduced due to the fact that much of the human labor has now been replaced by lightning fast computer CPU cycles. Additionally, this top five originator has implemented sophisticated downstream mortgage lending business rules to take advantage of the valuable data generated by the Paradatec system.

Paradatec’s solution is an advanced and unique OCR recognition technology. It utilizes neural networks technology and artificial intelligence and is able to read structured, semi-structured, and unstructured documents. It then makes ‘decisions’ about document characteristics in much the same way as a human being does, only many times faster and without human intervention. Paradatec takes a very different approach from other OCR forms processing technologies in that it is a truly template-free design, allowing the system to easily cope with the varying layouts of each document. In performance terms, Paradatec is capable of processing thousands of documents per hour with a single processor. It provides even further scalability by offering seamless



support for the latest in multi-core processor technologies and multi-server configurations.

This top five originator, like any other mortgage lender, is subject to a variety of time-sensitive requests such as internal audits. These audits require that specific data be tabulated from each loan file and reported to the appropriate entity. In some cases, the volume of loans included in these audits can reach into the tens of thousands, with a very limited response timeframe. With the Paradatec system now in production it is possible for this organization to be more agile than in the past. New data fields can be configured and tested in a few hours and a million images can now be interrogated for salient data overnight.

Additional capabilities leveraged successfully at this customer include:

- Verification provides a list of likely document types to further increase speed of verifying exceptions.
- Ability to customize how documents are handled based on the type of process to be conducted (e.g. origination, servicing, audit, etc.).
- Ability to quickly recognize additional document types using the Paradatec automated learning facility.
- Database lookups and business rule logic checks to ensure the highest degree of data accuracy.
- No scripting interface, with easily configurable rules to modify customer's highly sophisticated ADR and data extraction processes.
- Ability to add processor cores (including new servers) to the environment in a matter of minutes to quickly scale and meet tight deadlines or increased staffing demands.

## Outcome

The project was successfully implemented and released to production on-time. As a result of this experience with both the Paradatec staff and the Paradatec solution, this customer is prepared to act as a reference on behalf of Paradatec. Prospective clients are encouraged to take advantage of this opportunity.

Per **Neil Fraser, Director, of US Operations**, "To be chosen by such a high-profile client for a project of this size was a vote of confidence for Paradatec and our technology. I would encourage other similarly placed clients to reach out to Paradatec to setup a one day blind test. In just a day it is possible to see just what this technology can do, right out of the box."

## Further Information

Please contact us:

Paradatec, Inc.

Phone: 513-583-1382

Fax: 614-334-3666

[info@paradatec.com](mailto:info@paradatec.com)

[www.paradatec.com](http://www.paradatec.com)