How to Improve **Mortgage Processing** with Document Automation

The Ultimate Guide

PARASCRIPT®

How to Improve **Mortgage Processing** with Document Automation

The Ultimate Guide

PARASCRIPT®

Let’s Dive In

Initial loan approval is faster and easer than ever, thanks to online processes. But while much has been done to dramatically improve the customer experience with online portals that allow simple submission of required information including document capture via smart phones, the hard part comes next.

Loan file review is the most critical stage of the loan process. A single mortgage loan file can contain over 100 diverse loan forms and documents.

This is an incredibly labor-intensive process prone to human and software errors. If anything goes wrong at this stage, the rest of your business is affected.

To automate the process, some companies have deployed OCR (Optical Character Recognition) or RPA (Robotic Process Automation) software, but the results have not been acceptable.

It can take months or even years to train the software before it is useful for production. Even then, there are still far too many exceptions requiring human intervention and that defeats the purpose of RPA.

Recent advancements in the application of AI and Machine Learning technology have led to a new breed of “smart” OCR software called Cognitive Capture.

This eBook, addresses the challenges faced by legacy OCR and RPA and then focuses on how Cognitive Capture Software (CCS) addresses and overcomes these challenges.

PARASCRIPT®

The Challenges

The shift away from paper documentation toward a complete digital loan origination will still take years for the mortgage industry. This is because the industry does not use a common platform shared by all the players involved in the process to avoid duplication and repetitive tasks. This failure of standardization results in gaps between aggregators' point of sale platforms and mortgage lender loan origination system platforms because information is not often seamlessly transferred between these systems. As a result, any information passed on, such as the presence of a document type and its contents, must be revalidated and confirmed.

A single mortgage loan file can contain over 100 diverse loan forms and documents. Federal agencies, states, counties, and cities have different forms and regulations with an endless variety of file formats. Applicants cannot be counted on to submit high-quality scanned or phone camera snapped documents. Poor quality source documents are the curse of every automated system. Loan files often contain images that are unreadable by most systems. Applicants also forget to file documents.

The process steps involve opening the file, reading the file contents, separating one document from the next, and verifying the documents. Once that’s all done, then data must be extracted from the documents for input into the loan origination system (LOS). It is an incredibly labor-intensive process prone to human and software errors. If anything goes wrong at this stage, the rest of your process is affected.

Every loan file must be thoroughly reviewed, every document in it must be sorted, verified, and data must be extracted correctly and matched to a database. One poor quality or missing document, or one incorrect piece of data, can stop the loan dead in its tracks until humans resolve the issue. When refinancing or taking out a new mortgage to buy a house, processing delays can result in anxiety, stress, and extra cost. As a mortgage processor or service provider, providing a speedy and precise customer experience is critical for three fundamental business reasons:

1. **Faster processing means faster capital realization for your company.**
2. **Delays mean you risk losing your customers to competitors.**
3. **Data errors can risk fines and sanctions for compliance violations.**

PARASCRIPT®

To automate your mortgage process, using the best possible document classification technology— better known in the mortgage industry as OCR—leads to significant improvements. If loan file processing were a human body, document classification would be the eyes and the brain. Eyes to read, and a brain to learn and to process the information.

What about OCR?

The document classifiers in use today rely on a 30-year-old Legacy OCR method based on rules and templates. In loan processing, there are at least 800 naming conventions that equate to approximately 250 discreet or specific document types. This is one of the most complex classification challenges in any business process.

The software first must extract text from the loan file, then attempt to sort out the wide diversity of document types, where they begin and end, are there missing pages, etc. Because of its rule-based inflexibility and over-reliance on textual recognition, Legacy OCR is simply not the best solution for a complex file stream such as this. The makers of Legacy OCR try to overcome the inherent limitations by creating file libraries that need constant updates and requiring extensive and expensive professional services to eternally create rule upon rule to catch file variants. But the result is still the same. It falls far short of Straight Through Processing (STP) and requires continuous human assistance to deal with false positives and exceptions the software cannot resolve.

What about RPA?

RPA uses software robots (bots, for short) to automate repetitive data entry tasks such as screen-scraping data from an open application window and entering that data into a database or another business application. The fundamental value of a bot lies in its programmed ability to run completely unattended and never need a break. We call this Straight Through Processing (STP).

Bots rely on predictable and reliable data inputs. Anyone with experience at loan file processing knows how hard it can be to achieve, let alone maintain, true straight through processing of data from unreliable and unpredictable documents. RPA is only as good as the OCR software that is feeding the bots and the RPA vendors are using Legacy OCR as well. The same problems of Legacy OCR apply here.

Addressing the Challenges

Smart OCR investment is a must-do for mortgage lenders

In a recent report, **Infosys** identified the top trends for mortgage processing in the new era. Trend number 5 is accelerating the mortgage loan onboarding process. To address the challenge, Infosys recommended Smart OCR as a must-do investment for mortgage lenders.

Where does one find this newly sophisticated OCR?

Is there an OCR University and advanced degrees so you can identify which is the really smart OCR? Thankfully, Cognitive Capture Software (CCS) is now available that includes smart OCR as part of the solution. But it doesn't stop there. CCS also includes automated document classifiers that rely on advanced machine learning to train and learn faster and with fewer samples.

Machine learning systems can process enormous volumes of data, much more in less time than any human can do. They also collect more information and attributes more precisely and accurately. Machine learning algorithms automate the process of analyzing each sample document to discover what we call features or the attributes of each type of document to automatically create rule sets. Literally hundreds if not thousands of hours of effort required to create the very explicit rules using Legacy OCR can be reduced to five or six hours of machine learning by the computer.

Automated classifiers do require the curation of samples. You need to have an adequate number of samples that are tagged for the system to learn. Legacy OCR classifiers process only textual information. CCS typically includes visual classifiers, which can classify poor quality images or images without predictable text and do so at higher speeds than textual classification.

PARASCRIPT®

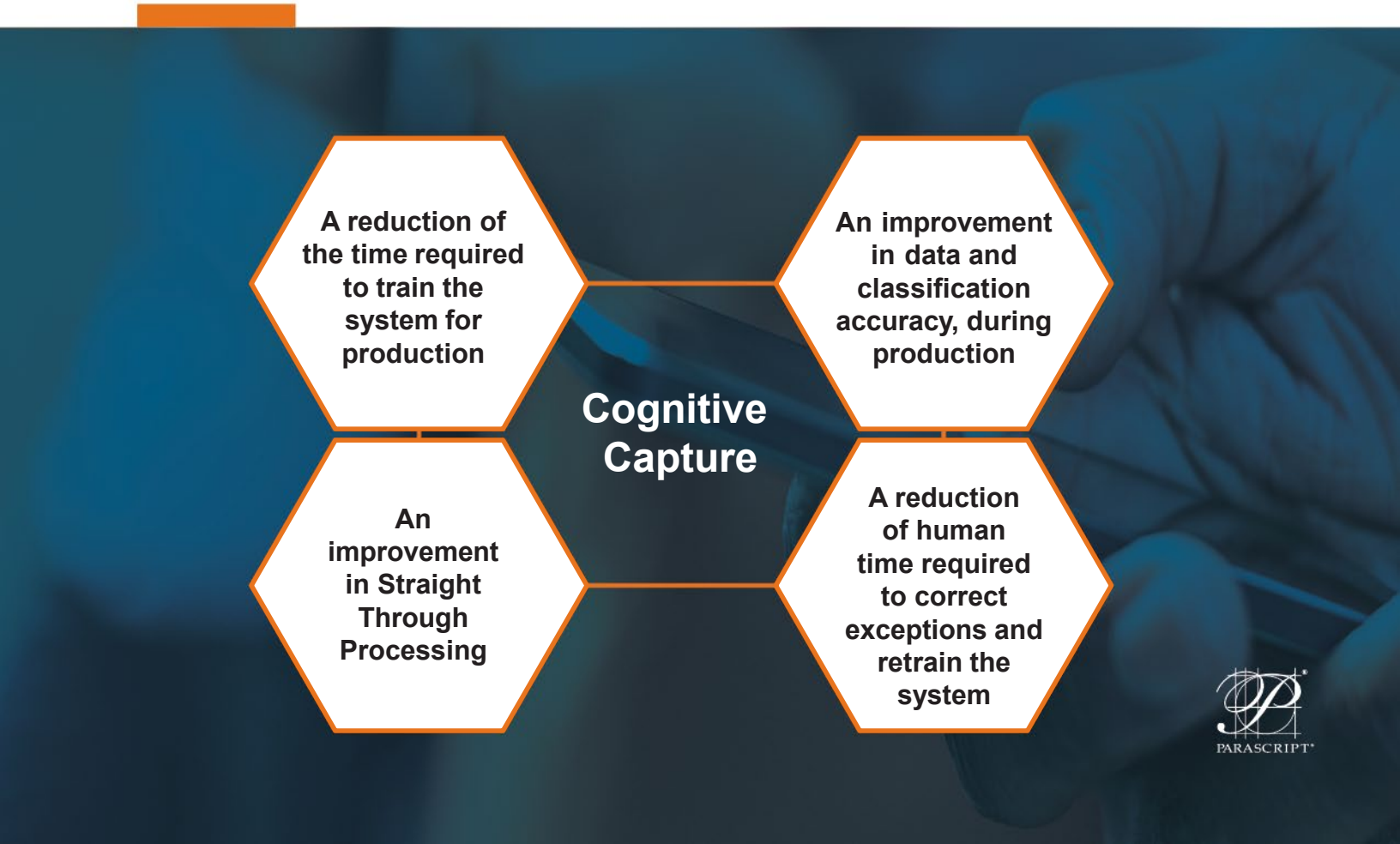
Moving Beyond OCR

Recent advancements in the application of AI and Machine Learning technology have led to an evolution in capture software development and has resulted in a new breed of “Smart” OCR software. Deep Analysis, a leading industry analyst firm, coined the term “Cognitive Capture” for this new software.

Cognitive means “connected with thinking processes” and implies the ability to know and to learn. We believe it aptly describes the quest to harness AI so our computers can learn to think and process that information, with the goal to improve our skills and productivity.

In a nutshell, Cognitive Capture improves document capture processes by learning what works best for capturing documents, then applying that knowledge to future tasks with minimal or no human intervention. To achieve this, Cognitive Capture Software (CCS) uses AI / Machine Learning to enhance the legacy process of scanning documents and extracting data with OCR.

When compared to legacy capture methods, the benefits of CCS can be measured using several key metrics:



Smart vs. Dumb

How to know the difference between Cognitive Capture and Legacy OCR

1

Does it use trainable machine learning (ML)?

Many OCR solutions say they use ML but still depend heavily on Subject Matter Experts (SMEs) to assemble document samples. With potentially hundreds of diverse document types, this is resource-intensive, time-consuming and costly. Look for Cognitive Capture tools with highly advanced ML that can automatically learn a new document class with only a few samples, and augment or even replace SMEs. Don't be misled by vendors who fly the AI flag. AI without advanced ML is not productive.

2

Can it apply multiple document classification techniques?

Look for a Cognitive Capture solution that organizes documents based not only on features and text, but also on imagery and handwritten information on the document including the presence of signatures. Documents with sensitive information such as handwritten social security numbers can be included in your classification workflow and easily identified. Regardless of how the information is presented, classification should be based upon all the available information in the document, not just a select subset of that data.

3

Can it be operated by business users?

Does it provide an accessible and clear-cut user interface for business users? Let's face it: this is a very sophisticated and complex software operation. Look for a solution with a user interface that doesn't require programming skills. Fine-tuning document classes should be as simple as correcting results by dragging and dropping results from one class to another, then re-running the task. A Cognitive Capture solution will hide the complexity without compromising the flexibility.

4

Does it offer high accuracy with low error rates at top speed?

Most classification systems become slower and less accurate in cases such as mortgage automation where several hundred document types are involved. Advertised accuracy rates are a bit like the miles per gallon sticker on a new car. Actual mileage will vary with options, driving conditions, driver's habits and vehicle condition. This must be demonstrable in your proof of concept. A Cognitive Capture solution will improve your results and performance.

PARASCRIPT®

Use Cases in Mortgage Lending

Loan File Processing

A single mortgage loan file can contain over 100 diverse loan forms and documents. Federal agencies, states, counties and cities have different forms and regulations with an endless variety of file formats. The file might also contain insurance, financial statements, appraisals, hazardous material reports, mineral rights, well testing certificates and a myriad of other unstructured data.

Cognitive Capture is the most advanced method available today to separate the loan file into the individual documents within. Once the software knows the document type, it understands the context of the data inside the document, which can be extracted at peak accuracy and speed.

Income verification

(W2s, Paystubs, Bank Statements and Tax forms)

Borrowers are required to submit more documentation than ever before to verify their income and ability to pay for loan closing. There is no standard form for everything, so the LOS must deal with another bewildering variety of forms and layouts. The LOS needs specific data from each document such as:

- Social security number, date of birth, and address;
- Proof of income and employer information within tax documents and paystubs;
- Proof of homeowner's insurance;
- Proof of available funds in a bank account; and
- Details about assets and liabilities, including account numbers, within supporting documents.

For Legacy OCR, automation has been difficult to impossible. Cognitive Capture is designed to deal with the variations.

PARASCRIPT®

Use Cases in Mortgage Lending

TRID Validation

TRID is an acronym for the TILA-RESPA Integrated Disclosure rule. Federal law requires the U.S. mortgage industry to combine the disclosure information that consumers receive when they apply for and close on a mortgage under the Truth in Lending Act (TILA) with the settlement disclosures under the Real Estate Settlement Procedures Act (RESPA).

There are two mandatory disclosure forms: The Loan Estimate and The Closing Disclosure. Both forms are complex and difficult for Legacy OCR to read. During the origination process, typically there are multiple versions of both forms generated. Cognitive Capture can quickly identify the file type and version, then extract every necessary data element from the final versions. This data is passed to the LOS for reconciliation. In cases where the LOS did not generate all versions and some were submitted from external sources, CCS can read the PDF or scanned image and produce the same results.

UCD Audit

A government-sponsored enterprise (GSE) is an organization chartered by the U.S. Congress for the purpose of providing citizens with fair and equitable financial services. The Uniform Closing Dataset (UCD) provides a standardized industry dataset to support the Consumer Financial Protection Bureau's (CFPB) Closing Disclosure. The GSE gathers this data to assess credit risk management and to detect fraud and misrepresentation in the loan process. The UCD also provides data from the loan file to help assess whether the loan meets the GSE's eligibility requirements.

Cognitive Capture is used to read the Closing Disclosure document and look for mismatches in the UCD's XML file. This ability can also be used outside of mortgages on other loan transfers.

Compliance with Data Privacy Regulations

Lenders must comply with a plethora of federal and state regulations. For example, lenders may be required to report on granular data contained in their loan file archive such as:

- List of mortgage documents compiled for a specific customer
- Mortgage applications filed within a certain date range
- Mortgage applications filed within a certain state

When Cognitive Capture is applied to a loan file, the lender can be assured the complete file is searchable with no errors. This allows lenders to accurately and quickly search through loan files and review documents.

PARASCRIPT®

What's Next?

Automating your mortgage loan file review and approval process is the key to success. When the ultimate goal is Straight-Through Processing (STP) with minimal human intervention, you need to upgrade to the latest advanced tools.

How efficiently does your current OCR or capture system process a loan file? How automated is this step? Are you still seeing too many exceptions and human interventions that cause delays? Does your OCR solution continually miss data that is essential for compliance?

Waiting to implement the best possible OCR solution is simply not an option for any mortgage processor.

Why not start a review process today?

PARASCRIPT®

Cognitive Capture: Get Smart Learning

Parascript's Cognitive Capture software uses its patented Smart Learning method to enable true unattended automation for your complex mortgage data. Parascript software analyzes data and the location of data to generate models to locate and reliably extract it. Parascript software constantly measures the results and adapts to those results.

When the system encounters new document types or variances of documents during your production stream, Parascript software runs in the background. It takes that information and uses it to adjust and optimize the systems. The care and upkeep that is typically time consuming and expensive are now automated.

Downtime is expensive. Parascript software contains most of its intelligence and configuration internally. This reduces the dependency on highly specialized and expensive technical resources. This also eliminates another point of potential human error. The system learns as a function of processing and validation in a feedback loop. Due to this feedback loop, the system eliminates errors over time even when new ones are introduced.

Ultimately, what matters is whether your system reliably classifies your documents and extracts the right data. Parascript software automates initial configuration for reliable straight through processing and ongoing adaptability of the system once you put it in into production.

It's quite common to see less than 10% of your overall volume consuming an inordinate amount - sometimes up to 50% - of your workflow resources, handling errors within that process. By reducing the overall processing time per loan, they will hit your ledger as an asset much faster. This gives you the ability to close a significantly higher number of loans in a year.

PARASCRIPT®

How to Improve **Mortgage Processing** with Document Automation

The Ultimate Guide

www.parascript.com
info@parascript.com
(888) 225-0169

PARASCRIPT®