This eBook provides an overview of cognitive document information and what role it plays in Robotic Process Automation (RPA). The process map for RPA and cognitive document information reveals where advanced data extraction fits within an RPA platform, reaching far beyond OCR software capabilities.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRECURSOR TO RPA</td>
<td>3</td>
</tr>
<tr>
<td>ENTER ROBOTIC PROCESS AUTOMATION</td>
<td>4</td>
</tr>
<tr>
<td>WHERE RPA AND AI INTERSECT</td>
<td>6</td>
</tr>
<tr>
<td>EXTENDING RPA WITH MACHINE LEARNING-BASED ANALYSIS</td>
<td>8</td>
</tr>
<tr>
<td>DATA EXTRACTION AND RPA</td>
<td>10</td>
</tr>
<tr>
<td>CUSTOMER ONBOARDING RPA EXAMPLE</td>
<td>12</td>
</tr>
<tr>
<td>PROCESS MAP FOR RPA AND DATA EXTRACTION</td>
<td>24</td>
</tr>
<tr>
<td>FULFILLING RPA’S PROMISE</td>
<td>26</td>
</tr>
</tbody>
</table>
Precursor to RPA

Process automation goes back to the early 20th century, but let’s focus on the 1990s when commercial software-based automation for human-work processes first started to emerge. Before that, most automation was the result of custom-developed, proprietary applications targeted at back-end processes such as what Sabre Technologies provided in calculating airline fares and fare matching/booking automation.

“Expert systems”—a form of AI—emerged in the ‘90s that allowed a person to create encoded rulesets to automate specific tasks. Anyone familiar with the auto-responder in the Outlook email application has encountered work automation using rules. You create a rule based upon certain conditions that tells the application when to automatically send a response email. This type of user-based automation emerged as features of existing applications.

Fast forward to the early 2000s, when companies started to look to process automation for easier ways to integrate data from one application to another, especially with more services offered from websites. Software solutions emerged to screen scrape web-based and other GUI-based data to do things like automate filling-in forms or to aggregate account data from many different sites or applications. In addition to data integration, other variants using a similar approach included automated software testing (e.g., for websites) such as simulation of users executing different commands.

What separated these newer automation solutions was that they were purpose-built, commercial solutions with user-friendly interfaces that did not require programming. You simply create a configuration by visually selecting the data from one application GUI and “mapping” it into another.
Enter
Robotic Process Automation

Around this same time, software vendors focused on providing an underlying automation platform offering a generic way of applying these data integration methods with a workflow engine to automate an entire process. These tools gained popularity in the IT world where efforts involve a lot of rules-based, repetitive tasks. Take, for instance, provisioning a new employee with a computer desktop and software.

Prior to these automation platforms, an IT staff member needed to manually install the necessary software on a desktop and configure the network, email and other centralized systems with the new employee’s information and access levels. Most of these tasks are rote and very rules-oriented. With automation tools, these type of processes were accomplished without any manual effort.
Where RPA and AI Intersect

RPA is founded on one of the earliest examples of expert systems. This form of AI simply encodes the knowledge of a subject matter expert (SME) for a given task into a set of rules that can be executed by a computer-based system. The more specific and known the given task, the more applicable RPA is in successfully automating it.

However, to date RPA software has not been implemented with automation of complex processes. Most “bots”—a term used to identify specific automation work against given rulesets on structured processes—work with simpler, rote tasks. More advanced capabilities that monitor operations and learn from mistakes moves into the realm of AI referred to as “machine learning.”

The most mature application of machine learning for RPA is the advance of “learn by example” into the bot configuration process, where in the background a system can monitor a given SME’s actions for different situations.

This monitoring manifests itself as a “learned” set of rules that can be implemented and is more flexible than traditional rules-based approaches.

When it comes to automation of complex processes where there is complex data involved, use of simpler rules-based instructions and learn-by-example is insufficient. In order to achieve true unattended automation with processes, machine learning techniques must be applied not only to how tasks are defined and updated, but also to the effort of understanding what tasks are completed accurately vs. what tasks need to be reviewed.
PHASE I: RPA THE DATA SUPPLIER

The most practical way to apply Cognitive AI to RPA implementations is to treat RPA as a **supplier of a large set of operational data.**

PHASE II: LEARNING SYSTEMS

This data is fed into a **learning system** that then identifies patterns and other “data features,” which can be used to optimize existing bots that work on specific tasks.

PHASE III: CONSTANT FEEDBACK

Ongoing learning has significant value since you cannot ask operational staff to **constantly report their every input and output**, but you can require a computer system to provide this level of detailed reporting.

PHASE IV: MULTIPLE PROCESSES

This ability presents an opportunity to **apply machine learning-based AI to many different processes.**
Enter Robotic Process Automation

Data Trails
One use of machine learning is to evaluate the data trail created by production RPA activities in order to enable adaptive capabilities.

For instance, in an RPA “bot” that automates the provisioning of a new email account, if the bot encounters an exception where the new employee’s name is already being used, a new rule can be created automatically based upon analysis of this data.

Data Extraction as RPA
Data extraction can be a key component of RPA. For instance, a configuration is created to use OCR to extract data from a given document and then the configuration or “bot” is put into production. However, RPA bots are meant to automate a chain of actions and not one single task such as data entry. The other major difference is most bots deal with repeatable and well-defined tasks where precision is taken for granted, and there are few exceptions.

RPA as Data Consumer
RPA is increasingly being extended to other processes that involve the need for data stored in documents. These processes are more complex than the easy-to-define, highly repeatable tasks. Here RPA becomes a consumer of this document-based data in order to conduct its work. This is where most RPA vendors see value with cognitive document automation.

Real-Value Cognitive Automation & RPA
Real value lies beyond simply adding OCR.

Rather, it is the ability to embed a flexible and adaptive document automation capability so that the overall system (RPA + document automation) can gradually learn and improve.

This is where Parascript, and not an OCR SDK or other page-based OCR software comes into play.
Cognitive Document Automation & RPA

Parascript provides cognitive digital workers that use machine learning to automate both initial configuration and ongoing production for true unattended document automation.

A bot that needs access to document-based data invokes a Parascript routine that identifies and separates a stream of documents.

Parascript software then extracts the data and provides specific information to the bot so that it can continue on with its process.
Customer Onboarding

COGNITIVE RPA USE CASE

The customer onboarding process at a financial institution is one use case. This process involves extensive coordination between internal staff reviewing required documentation and the customer supplying it. Many perfunctory tasks are involved in this activity and ideal for RPA:

1. Verification that all required documentation is submitted
2. Review of data within documentation against defined policies (e.g., income and employment)
3. Know-your-customer checks including verification of authenticity of ID (e.g., address verification and ID verification)
# Customer Onboarding
## COGNITIVE RPA Use Case

<table>
<thead>
<tr>
<th>Activity</th>
<th>Description</th>
<th>RPA</th>
<th>AI Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Document ID</strong></td>
<td>Customer submits required documentation in-person or through submitting via smartphone app.</td>
<td>RPA, using document classification, can identify and sort documents, providing notification of missing documentation to the bank staff.</td>
<td>RPA can be trained via a learn-by-example approach, but cannot handle unknown documents that are submitted, which are treated as exceptions. These exceptions can be automatically identified and routed through the training system to update the set of known documents.</td>
</tr>
<tr>
<td><strong>Customer Data Entry</strong></td>
<td>Bank staff enters data from supplied documents into the banking system.</td>
<td>RPA, using automated data extraction via OCR, locates and populates data into the software.</td>
<td>While RPA can benefit from OCR, location of this data will vary with exceptions handled by bank staff via manual data entry. The data not found, can use values entered by the staff to update rulesets automatically so that data located on similar documents can be automated the next time.</td>
</tr>
<tr>
<td><strong>ID verification</strong></td>
<td>Bank staff view supplied identification.</td>
<td>RPA, using image analytics can identify presence of watermarks, microprint or other security features. Name, ID number and address can be extracted and input into credit check system.</td>
<td>Just as with the previous two examples, AI benefits RPA when exceptions are encountered. These exceptions can be submitted to the learning system in order to learn new security features.</td>
</tr>
<tr>
<td><strong>Blacklist check</strong></td>
<td>Bank staff reviews the name and compare against company-published KYC blacklists.</td>
<td>RPA, using OCR, can extract names and match database of KYC entries.</td>
<td>Names incorrectly located or extracted can be identified as exceptions and added to a database to improve OCR and reduce the likelihood of future errors.</td>
</tr>
</tbody>
</table>
The main benefit of applying AI-based data extraction to RPA systems is to reduce the potential for exceptions, which require costly and slower manual processing.

This exception reducing workflow takes advantage of RPA’s inherent ability to collect key discreet data in the background on successes and failures which can create a closed-loop system through an automated update of existing rulesets using machine learning-based AI.
Fulfilling the Promise
Robotic Process Automation

RPA platforms promise to bring major change across many industries. They’ve matured to the degree where successful automation is possible without a large investment of staff time and capital.

The RPA leaders are adopting a machine learning approach that benefits from easier-to-use, more flexible and adaptive logic in place of its more brittle rule-based AI cousin using expert systems.

Cognitive Document Automation extends support for these complex workflows by enabling these systems to also incorporate new information gathered from unstructured documents in unattended mode, increasing the amount of straight-through processing.
ABOUT PARASCRIP

Parascript Smart Learning-based Digital Workers provide the highest levels of unattended document automation with the lowest complexity and risk.

Single Workflow
Combine scanned and electronic document support in the same workflow so single stream, straight-through processing.

Advanced Automation
Provides automated configuration of classification and data extraction for your varied document types and formats.

Any document. Any data.
Process structured, semi-structured and unstructured documents—including handwritten data—from mobile, fax, fileshare, email, scanners and more.

Machine learning
Get the highest accuracy results using Parascript software powered by machine learning, which automatically refines and improves performance.
Contact Us Today to Learn More

www.parascript.com
☎ 888.225.0169
✉ info@parascript.com