PARASCRIPT®

AUTOMATION STRATEGIES FOR AMDOCS

Intelligent Automation in Document-Intensive Processes

Takeaways

> ▸ *While only 13% of businesses are realizing the full impact of digital investments, for service providers like Amdocs that adopt advanced intelligent capture, manual data entry and verification becomes nearly obsolete.*
>
> ▸ *Applying three keys to automation—data accessibility, process repeatability and process value—and a unique project ranking framework facilitate identifying the quick wins and more challenging projects for Amdocs process automation.*

Summary

Our two-dimensional framework approach identifies quick wins that shock-proof Amdocs' client operations. From the attributes of the process to the nature of document-based information, many areas require a controlled approach to ensure that intelligent automation meets project expectations. This paper identifies specific processes for quick wins to improve efficiency, reduce costs and insulate operations from unforeseen disruptions due to economic, infrastructure and geopolitical events.

What's in this paper?

Automation Strategies

Are Amdocs document processing services keeping up with industry efforts?

Only 13% of businesses are realizing the full impact of digital investments because adopting automation is a complex and risky process, especially when there are alternatives based on manual labor that have been simpler to implement and manage. It doesn't matter if the process is within a client organization or operated by a sophisticated outsourcer like Amdocs. When Intelligent Automation is applied to document-intensive processes, many aspects of a project impact the level of performance and ultimately its success.

From process attributes to the nature of the document-based information used to support workflows, many areas require a controlled approach to ensure that intelligent automation meets project expectations. The first step is to apply a proven framework to identify and select processes that benefit most from automation relatively quickly with a higher success rate. These processes represent quick wins for service providers to improve efficiency, reduce costs and insulate operations from unforeseen disruptions due to economic, infrastructure and geopolitical events.

Process Framework - Identifying Quick-Wins

When it comes to Amdocs business processes, including those that involve staff and document-based information, they have different levels of criticality. Processes can be categorized by a number of factors including whether they are "front office" vs. "back office," whether they support revenue directly or are administrative, and how controllable they are in terms of standardization or tasks and output. Through many engagements across industries, Parascript found that three attributes of processes affect project success more than others. So, we created a framework to aid with identifying and ranking processes. Below the key attributes are listed along with the variables that contribute to a process with high automation potential vs. "nice-to-have" automation.

Keys to Automation	Best	Maybe	Little ROI	Forget	Next Best	Risky
Data Accessibility	High	High	High	Low	Low	Low
Process Repeatability	High	High	Low	Low	High	Low
Process Value	High	High	Low	Low	High	High

Keys to Automation

Data Accessibility

When it comes to intelligent automation, the focus is on the precision of task configuration and output. Data is essential to identifying how to configure and measure systems. If machine learning is used, data is critical because algorithms—especially deep learning neural networks--are data hungry. Some processes may already have structured data stores. Others won't, especially if they involve document-oriented information. It is important to select processes that offer all the data required or be prepared to address data accessibility first.

Process Repeatability

Industry analyst firms often recommend examining processes based on a complexity scale. The most straightforward way to do this is to rank processes according to the level of repeatability and level of exception handling involved. Starting first with highly repeatable processes with little chance of exception or variability ensure quick wins. Although since Amdocs has fewer of those, it is best to focus on the middle range processes with limited variability.

Process Value

Processes must also be prioritized based on business value. Some are significant, such as the ability to process more orders more efficiently. Others have limited value such as internal technology provisioning. The qualities of each provide a rough guideline as to which of the 5 approaches to take. For highly repeatable, lower value processes, a raise-and-replace is recommended. While for high value, highly complex processes, an incremental approach or inside-out approach probably makes the most sense.

The Role of Machine Learning

With everyone talking about AI and machine learning, it is important to understand how it is applied in intelligent document processing. Machine learning is good at repeatable processes that can involve some level of variance. For higher variance, relying on pure machine learning-based automation may be difficult. Put another way, machine learning is best suited to the "input" side.

According to Horses for Sources (HfS), machine learning is most effective when focused on automating the side that provides improved access to data in its "enriched" form. This enables Amdocs to focus higher value staff hours on the "output," which entails a higher level of scrutiny on the "meaning" of the data and how to react to it. It is very beneficial to include this attribute as part of the both the "repeatability" and "value" equations.

Framework for Ranking Projects by Technology Domain

Another dimension further identifies and ranks the best processes to apply automation based on technology domain. Here the focus is on the attributes of the document-based information and the types of tasks involved to arrive at three levels of complexity. Using this approach, it is possible to segment operations involving documents into three categories: Easy, Moderate, and Difficult as outlined below. This framework takes into consideration tasks that fall within three technology areas: classification, data extraction and validation.

Complexity	Classification Tasks	Data Extraction Tasks	Validation Tasks
Easy	Single-page documents or page-level classification	▪ Mark and signature detection, ▪ Barcode Recognition ▪ Structured English Text ▪ Structured, unconstrained, and low-variance or pattern-controlled handprint	▪ Page number/order ▪ Signature Presence ▪ Date Validation ▪ Field Completeness ▪ Validate with 3rd party system.
Moderate	Document-level classification with separation	▪ Semi-structured text ▪ Structured handprint recognition, constrained(boxes and combs) ▪ Multi-lingual text ▪ Higher-variance Handwriting – structured ▪ Unstructured handwriting transcription ▪ Unstructured handwriting – word-spotting	▪ Field value consistency for document
Difficult	Package classification	▪ Unstructured Text (any language) ▪ Semi-structured handwriting	▪ Signature correctness ▪ Package completeness ▪ Package field-value consistency

Project Rankings: Easy Projects

The proverbial low-hanging fruit consist of the automation needs that fall within the Easy category. Within this easy category, we further break-out each functional area with example tasks.

Classification. For classification, the task is focused on identifying documents on a page level vs. a document level. Page-level classification dispenses with the additional complexity associated with multi-page document classification including identifying the relationship between pages (e.g., first, second...last) of a document and identifying boundaries of one document with another (often referred to as "document separation"). By removing these additional functions and the automation tasks that support them, the overall complexity of the classification task is reduced. Any projects—no matter how many documents are involved—where the majority of documents are single pages fit into the easy project category.

Data Extraction. For data extraction, we focus on the most mature field-level tasks. Starting with detection-oriented tasks such as checkboxes, signatures and then moving to barcodes, these tasks are some of the most common and practical ways to deal with document-oriented data. They are also some of the most precise, easy-to-configure functions. When it comes to data field extraction, structured forms where the data is always located in the same place are less challenging. By not having to manage data location, a significant amount of complexity is removed.

For the data within the fields, we focus on English text and limited amounts of handwriting. While handwriting recognition may appear complex, when we can deal with limited data sets such as numeric amounts or pattern-controlled fields (such as dates and social security numbers), the amount of configuration is fairly low and the results—even with limited effort—can be quite good.

Data Validation. A system can be configured to automatically evaluate the results of classification or extraction and then *verify* its accuracy and/or transform it. Within the Easy category, simpler validation tasks are the focus such as taking a date and verifying that if falls within a certain range (sometimes dates must not be in the past or future depending on the application), scanning for page numbers, identifying fields in structured forms that are blank, and performing rudimentary signature presence.

Another often simple validation capability is to check the output of the system against third-party data stores (i.e., data stored in a separate system of record). For example, a customer request form might collect an account number and verify key customer data stored in a database by looking-up the account number. Generally, third-party data stores are a good source of validation data.

Project Rankings: Moderate Projects

Moderate complexity projects are a step beyond Easy projects when it comes to configuration and maintenance of an intelligent capture system. These projects often represent higher costs when it comes to the manual effort required to support access to document-based information and therefore automation typically results in larger positive impact to the process itself over projects in the Easy category.

Classification. For instance, classification now focuses on document-level which often entails time-consuming and error prone activities involved with reviewing and assigning multi-page documents into a certain class. Reviews often take longer due to the need to view more than one page. Additionally, many document-level classification tasks involve the need to review and separate large multi-page files into discrete documents which is very time-consuming. Automation of these tasks is more-complex since the document classifiers must be configured to take into consideration the relationship of each page to another and identify which pages belong to specific documents. Additional steps must be taken to split files into individual documents. All of these sub-tasks require extensive analysis and testing.

Data Extraction. For data extraction, moderately complex document layouts where data is not consistently located in the same place requires more time to configure and test. Often referred to as "semi-structured" documents, these represent multiple tasks that must be configured to reliably locate specific data and extract it. We also introduce more complex handwriting extraction. Individual characters handwritten into separate boxes or sections is more difficult to extract than handwriting in open spaces. This is because the separators pre-printed on the form interfere significantly with the handwriting recognizers. Handwritten data with a greater variety of potential values such as names, addresses and other alpha-numeric data can also be included.

These cases are moderately difficult because of the potential for errors with the expansion of the potential values unless the right mix of rules or other mechanisms such as dictionaries are available. Unstructured handwriting applications such as the need to transcribe a paragraph or more of information or the need to locate specific terms anywhere on a page are also part of this category. Use cases include extracting comments on a form, preexisting conditions on a medical form, or identifying words like "urgent" on a document that support the ability to escalate a complaint or claim.

Data Validation. Complexity when it comes to validation is largely in the area of taking multiple values of a page or multi-page document and comparing one against the other to ensure that values and/or formats are consistent.

Project Rankings: Difficult Projects

The third category of project complexity deals with the processes that involve many different documents within a case and require support of complex extraction and validation scenarios. Processes that involve complex document automation are often high-value processes that require staff that are trained on specific subject matter that the process supports. As such, even automation of as low as 30% can create significant efficiencies and cost reductions.

Classification. For classification tasks, the automation for difficult projects involves a series of documents where there is a need identify and organize a full package. Sometimes the documents are out-of-order, incomplete or missing altogether.

Data Extraction. For extraction, the use cases become more complex as the data is no longer formatted in a structured form or presented as tag-value pairs. Instead, it is unstructured prose. The ability to parse unstructured text to locate specific information is complex as it requires much more time for analysis of the text to discover the range of variations in terms of both layout and language. Examples include agreements, contracts and unstructured financial-oriented documents such as mortgage notes.

Data Validation. For validation, again the complexity is increased due to the need to perform cross-validations of data stored within many documents, dealing with differences in value, format and position on the page. One use case for a complex project is automation applied to a home mortgage origination or auditing process. Here, potentially 50-60 different documents must be classified, separated into individual files with data extracted from certain documents. And then, the data must be validated and verified to ensure that it is internally consistent.

Amdocs and the Pursuit of Automation

There is no question that organizations worldwide are aggressively pursuing automation projects to benefit from increased efficiency, lower costs and greater adaptability. However, the complexity of automation technologies often hampers adoption to the levels needed and certainly increases project risks. The good news for Amdocs is that examining potential projects through the lenses of process and document information attributes, enables qualitative and quantitative project ranking by likelihood of success and ease/speed of implementation.

Through careful planning, using Parascript's two-dimensional framework, Amdocs benefits from not only rapid deployment of intelligent automation for document-based processes, but also from higher-precision configurations that support key operational needs.

PARASCRIPT®

Contact **Us**

✉ info@parascript.com

🌐 parascript.com

📞 (888) 225-0169

INTELLIGENT CAPTURE

Parascript Intelligent Capture *is* intelligent automation. Smart Learning puts the power of data scientists in your hands to easily create and deploy document classification and data extraction processes for any document or data type including handwriting. More than 100 billion documents for financial services, government organizations and the healthcare industry are analyzed annually by Parascript software. Visit us at www.parascript.com.