Market Leading Technology Stack Delivers on Promise of Advanced Capture

Open Text breaks down barriers that have traditionally help back adoption
By Ralph Gammon, Editor and Publisher of *the Document Imaging Report*

Document capture has a proven ROI. Whether it's through reduced manual labor, faster transaction processing times, improved accuracy, better visibility into operations, or a number of other factors, in a recent study by AIIM[1], 84% of respondents reported achieving ROIs from their document imaging projects within 18 months or less, with 59% achieving them within 12 months. AIIM reported that these figures were among the highest ROI rates ever reported—"for capture projects, or any other ECM investments."
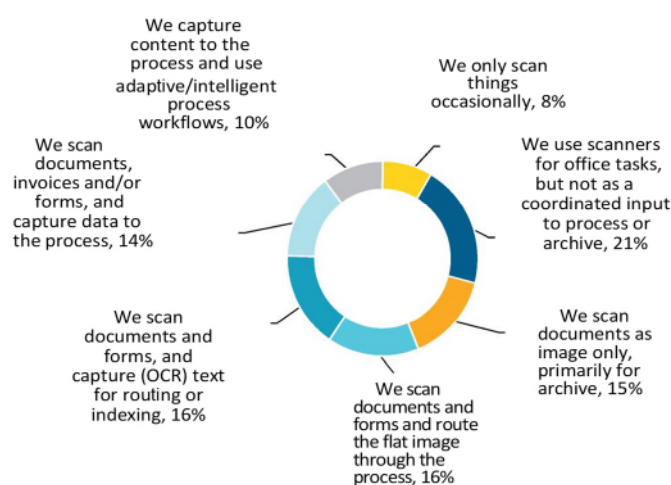
This almost certainly has something to do with the recent introduction of advanced capture methodologies into the market. While automated capture of data from structured documents like medical claims and tax forms, has been in relatively widespread use for close to 20 years, advanced capture like auto-classification of documents and automated extraction from unstructured documents— such as contracts and other complex forms—has only recently started to show up in end user implementations. And some of the results have been impressive!

In 2015, Ameritas Life Insurance Corp., a Lincoln, NE-based business with more than $2 billion in annual revenue and more than $25 billion in assets under management, implemented advanced capture to replace a legacy system. Through this implementation, Ameritas was able to increase its auto-classification of incoming documents by 55%. By being able to auto-classify 15 million documents a year compared to 9 million previously, Ameritas was able to close one of its document scanning centers and absorb additional workload without increasing staff. ([Link to complete case study](.)

The Minnesota Dept. of Revenue has had similar success utilizing advanced recognition to help it define more than 800 document types related to business tax returns. The technology is used to determine if a return is complete, and then, depending on the contents, route the return to the appropriate workflow queue. This application was part of a larger technology implementation that produced more than $1 million in cost savings. ([Click here for the case study](.)

Despite examples of impressive savings and ROIs, the numbers from AIIM's 2016 report entitled "[Paper Free: Are we there yet?]()"[2] indicate that advanced capture is still underutilized in the market. Only 10% of users surveyed for the report indicated that they had implemented the top level of capture technology described in the survey: "in conjunction with adaptive, intelligent processes," and 60% seemed to be using no automated capture at all.

We capture content to the process and use adaptive/intelligent process workflows, 10%

We only scan things occasionally, 8%

We scan documents, invoices and/or forms, and capture data to the process, 14%

We use scanners for office tasks, but not as a coordinated input to process or archive, 21%

We scan documents and forms, and capture (OCR) text for routing or indexing, 16%

We scan documents as image only, primarily for archive, 15%

We scan documents and forms and route the flat image through the process, 16%

Despite these obvious benefits, why are adoption rates of advanced capture so low? Well, it's certainly not because users don't recognize the benefits. In that same 2016 AIIM report, 56% of respondents indicated that one of their immediate priorities for improving their capture systems was "to automate manual processes with document classification"—which was by far the top answer. Coming in second at 30% was "update our capture systems with better technology," which could also mean introducing advanced capabilities.

So, while a majority of users recognize its benefits, adoption rates of advanced capture lag behind. There are a number of likely reasons for this:

- Advanced capture hasn't yet reached mainstream adoption, so many users remain hesitant to implement it—preferring to stick with the herd mentality.
- Introducing advanced capture typically involves adding another moving part to an already complex capture operation.
- Even though the ROI can be substantial, it can still be expensive and time consuming to implement advanced capture.

Putting the Pieces in Place for Success

Open Text is working to break down these barriers. In the past 10 years, the market-leading ECM ISV has made two major capture acquisitions. In 2008, it bought Captaris, a company best known for its fax server technology, but which had, a few years prior, acquired a leading German-based document capture software developer. More recently, Open Text acquired the Enterprise Content Division (ECD) of EMC, so it now owns the market leading Captiva capture software.

When combined with Open Text's internal capture initiatives, the ISV now has a full suite of offerings that can address everything from tools for creating third-party capture applications, to advanced recognition, to market leading invoice capture for SAP environments. "As business is digitally transformed, intelligent advanced capture technology becomes increasingly important," comments Mike Spang VP of research at Harvey Spencer Associates, a leading capture market analyst firm. "Information management companies like OpenText have recognized this market trend and have added advanced capture to their solution suite."

Being able to offer this wide breadth of capture and information management software should make it easier to implement advanced capture than it historically has been. Let's take a look at how Open Text addresses the barriers to entry for advanced capture that we've identified.

What is advanced capture?

We should start out by listing features of basic capture, which include scanning, batch management, image processing, and OCR. Capture from fixed/structured forms can also be included in this category.

Today's advanced capture technology builds on this and primarily focuses on two areas: auto-classification and extraction from forms that have varying structures—in other words the data that needs to be extracted can appear in different places. Let's take a look at each of these functions and how they can benefit users:

Auto-classification: Document preparation is one of the hidden costs of document imaging. Studies have estimated that it can make up more than a third of the cost of a document capture operation.[3] Preparation can involve tasks like staple and paper clip removal and damage repair, which can't be automated. But, it can also include document sorting and batching—and this is where auto-classification can be applied. Instead of having a person manually separating documents, dividing them into stacks, and inserting separator sheets—image and text recognition algorithms can be used to automatically perform these tasks. As a result, auto-classification can drastically reduce document prep time and eliminate bottlenecks that prevent today's high-speed scanners from being utilized to their full capacity.

Auto-extraction from forms with varying structures: Historically, applying automated data capture successfully has been as much about forms design as it is about OCR. But, what about forms that an organization doesn't have any control over? These can include invoices, contracts, explanations of benefits, and other types of documents that contain valuable data. Because of their varying structure, traditional template-based OCR applications are very difficult, if not impossible to utilize on these types of documents. As a result, they have to be manually processed, which, due their complexity, can require expensive skilled labor.

Auto-extraction applied as part of an advanced capture application can combine techniques like keyword matching, natural language processing, and semantic understanding to enable automated data extraction from these variably structured documents. Business rules can also be incorporated, as well as techniques like machine learning and artificial intelligence (which can also be utilized in auto-classification.) It might not automate 100% of the capture, but, like with any data capture operation,

confidence threshold levels can be built in, and steps like manual verification and database look-ups can be implemented, to increase and ensure accuracy. The goal is not to eliminate manual interaction with the documents but to reduce it and therefore increase efficiencies.

A Look at Open Text's Stack

As we mentioned, Open Text has a wide breadth of software to enable a range of capture applications from simple to advanced. Let's take a look at this portfolio:

Capture Platforms

- Captiva Enterprise Capture: A market leading enterprise-level capture application. It's designed to handle high volumes and can run across multiple departments and manage complex capture workflows. It's also tightly integrated with multiple back-end ECM systems, including Documentum. Through a Captiva toolkit, it can also be extended to incorporate mobile capture.
- OpenText Capture Center: A workgroup to enterprise-level platform that can manage batches, as well as apply automated capture through integration with the Open Text Capture Recognition Engine. Capture Center is tightly integrated with Open Text's Content Server and Process Suite software.
- LEAP Snap: A cloud-based capture platform first introduced by ECD in 2016. Targeted at lower-volume, distributed applications, it's designed to enable knowledge workers to create and execute their own automated data capture workflows.
- RightFax: Market leading fax server application that can be used to capture faxes as images and submit them straight to a workflow without ever being printed.

Advanced Capture

- Captiva Advanced Recognition: This is the Open Text's auto-classification and unstructured document extraction technology. In addition to including the functionality we described above, it features automated learning capabilities. This means that the software's accuracy improves over time as it incorporates feedback from the QA and validation processes. Other features like handprint recognition and check capture can also be incorporated through add-on modules.
- Vendor Invoice Management (VIM): A mature invoice capture application that is best known for its tight integration with SAP. SAP resells an OEM version of VIM. In addition to automated capture from invoices, VIM includes ERP embedded workflow for managing approval
- OpenText Business Center for SAP Solutions: A capture solution for onboarding any type of document into SAP environments. It features pre-configured processes for managing sales orders, purchase orders, delivery notes, and HR documents. Like VIM, the Business Center runs within SAP environments, including the SAP cloud, and has been tested and approved by SAP.

Capture Building Blocks

- PixTools and ISIS drivers: PixTools is an SDK for scanning, viewing, and image processing that has been used to build several leading commercial scanning applications. It is a mature and proven platform that leverages the ISIS driver protocol to connect with scanners. ISIS, which is maintained by Open Text, is the preferred scanner driver in many high-volume capture implementations.
- Open Text Capture Recognition Engine (Formerly known as RecoStar): An OCR/ICR engine that is used in many leading automated data capture applications. It can be used to capture both machine- and hand-printed characters in all major languages. It also has features like image

preprocessing, layout recognition, full-page and zonal recognition, and forms recognition and handling. It is a very mature engine that was part of the Captaris acquisition.

- Open Text Document Reader (DOKuStar)**:** A leading toolkit for automating document classification and extraction for indexing. It features Adaptive Read Technology, which enables it to learn the location of data fields through the submission of examples. Self-learning technology can be used to complement the SDK's rules-based classification. This SDK was also part of the Captaris acquisition.
- **Open Text Capture Full Page Reader:** Utilizing the Capture Recognition Engine, it can recognize text on imaged documents to make them searchable. It can output full-text searchable PDFs as well as plain text or XML.
- Mobile Capture SDK: A set of tools for embedding capture within mobile apps, such as those that might be offered by banks to their customers. It offers automated image clean-up and batch submission for integrating mobile capture with production applications. It features a very customizable UI.

Breaking down barriers to entry

Let's take a look at how Open Text's capture portfolio can be utilized to address those barriers to entry to advanced capture that we brought up earlier

- **It hasn't reached mainstream adoption:** Open Text is no start-up when it comes to the capture market, and, even though the market is lagging when it comes to advanced capture adoption, Open Text still has a significant number of advanced capture users, including a large number of VIM customers through its partnership with SAP. So, while the market at large may be in the early adopter stage, Open Text's customer base, and its support team, are more advanced than most.

- **Introducing advanced capture typically involves adding more moving parts to an already complex operation:** As we've discussed, Open Text has a lot of pieces in its capture portfolio. Because it owns these pieces, Open Text's engineers are able to work internally to integrate them before bringing advanced capture solutions to market. The Captaris capture technology has already been fully integrated into Open Text's suite, and the engineering team is currently working on tighter integration with the ECD technology. Open Text can provide everything from the capture tools, to the batch management, to the OCR, to the advanced recognition—this should ensure that it all runs smoothly together.

- **It can be expensive and time consuming to implement advanced capture:** Once again, Open Text's wide breath of technologies comes into play for overcoming this barrier. Part of the expense of advanced capture has to do with the professional services typically required to integrate it with other systems, like basic capture and ECM. Open Text owns extensive technology in both areas, which makes out-of-the-box integration more likely than when working with products from multiple vendors. Open Text's wide ECM portfolio also creates the potential for a more comprehensive technology purchase from a single vendor, one in which advanced capture can be bundled. As far as set up and configuration of advanced capture, Open Text continues to focus on making its technology easier to implement through development in areas like machine learning and AI.

Conclusion

Open Text has invested a lot, and continues to invest more, in capture technology. The ISV realizes that its customers are not getting the most they can out of their ECM implementations without taking advantage of advanced capture features like automated classification and extraction from all types of documents. Open Text's wide breadth of technology and continued aggressive investment in R&D ($281.7 million in its fiscal 2017) and acquisitions should ensure that its advanced capture will continue to move forward faster than the market rate. And Open Text's investment in forward thinking initiatives like cloud capture and cognitive computing (its recently announced Magellan initiative) will help future proof its entire ECM platform.

Concluded HSA's Spang, "OpenText Capture Center and Captiva Enterprise Capture, when combined with Open Text's advanced recognition technology, are well positioned to support wider information management solutions. We believe that these core capture capabilities married with broad market reach will place OpenText in a formidable position to challenge for the market leadership position in capture."

And when considering emerging technologies such as advanced capture, with the potential of a high reward, but also carrying some risk, the backing of a stable, growing market leader like Open Text might prove to be key to a successful implementation.

1. *Paper-Free Progress: measuring outcomes* ©AIIM 2015
2. *Industry Watch: PaperFree: Are we there yet?* ©AIIM 2016
3. http://www.docmanage.com/magazine/backissues/8-4/costscan.htm.