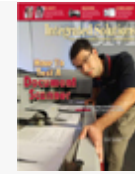




# Integrated Solutions

Thursday, December 11



[Click here for your free subscription](#)

[Home](#) [Magazine](#) [Market Resources](#) [Advertising](#) [Editorial](#) [Solutions Locator](#) [About Us](#) [Tech Centers](#) [Search](#)

[Home](#)

[Back to December 2008](#)



## Forms Processing Solutions Get Smart

The intelligence of capture technology is increasing, paving the way to unstructured forms processing while also providing end users with greater ease of use.

*Integrated Solutions*, December 2008

Written by: [Sarah Howland](#)

Historically, forms processing solutions were designed to meet the document imaging and data capture needs of companies that handled not just large volumes of documents, but large volumes of documents with similar form or structure. The 'magic' of forms processing in these companies was driven by static, template-based imaging software. While this software was a definite improvement over manual data entry, it left plenty of room for improvement. Today, forms processing solutions have escaped the narrow bounds of templates and continue to make strides in unstructured data capture.

### DATA CAPTURE INTELLIGENCE GOES UP, COST GOES DOWN

By far, the greatest advancements in forms processing are being driven by the need to process unstructured documents. Among the most promising improvements is the continuing development of IDC (intelligent data capture) technology. Forms processing solutions based on IDC eliminate the need for predefined templates, rules, or anchors to identify, extract, and classify data. Instead, IDC technology takes a mathematical approach by employing complex algorithms to identify and extract appropriate data during the scanning process.

"Intelligent data capture products essentially imitate the work of the human brain, enabling the software to accurately extract content from documents, both structured and unstructured, without foreknowledge of the layout of the documents or the type of data they contain," says Charles Kaplan, VP of marketing at Brainware. "The ability for these tools to rely on content analysis, rather than templates or keyword recognition, means they can be used to process documents of any type within a company." According to Kaplan, IDC consistently is able to reduce data entry by 80% across an organization.

Another factor driving the development of IDC is the need to address the ease-of-use concerns that are common with earlier forms processing solutions. For example, most traditional forms processing solutions require continual programming expertise to build or revise templates or rules in order to recognize and capture data for each new form encountered. "Making data capture software simpler to configure in the first place and making the solution easier to maintain going forward in terms of adding new document types is the revolution we are going through now," says Richard Brierton, a senior product marketing manager at Kofax. "Vendors are addressing ease of use by actively reducing the need for end users to possess an extreme skill level."

### CAN YOU TRAIN YOUR FORMS PROCESSING SOLUTION?

Dean Tang, CEO of ABBYY USA, agrees that vendors are making strides to reduce the skill level required to deploy and maintain forms processing solutions. "Data capture is still a relatively complicated technology, and not everyone can use it right out of the box," says Tang. "This is why ease of use usually tops the list of important criteria when evaluating forms processing technologies." Recently, this has led

to an emphasis on the creation of data capture software solutions that employ training-based, rather than rules-based or template-based extraction. The introduction of training-based software has been on the rise during the past 12 to 18 months.

Solutions that are training-based are not centered on end user training. In fact, it is quite the opposite. This type of solution is one in which the software itself is trained by the user where, in essence, the software learns by example. Training-based software solutions are deployed with a knowledge base of documents already built in, based on the most common form types a specific company may deal with. For example, in a financial office the software would be preprogrammed to recognize the most common loan application formats. The solution will 'learn' new documents as they are encountered during the scanning and capture process.

Training-based capture tools work by sending an unrecognized document into a queue for review, just as in a traditional validation setting. Either the user or a supervisor will key in the questioned data. The difference is that once the data in question is identified, entered, and validated, the software will 'learn' that document type and recognize it on all future occurrences. Since the need to create a template or set of rules is eliminated, data capture software that uses a training-based model is faster to update and also requires less skill from its end users.

Kofax' Brierton offers a word of caution about training-based tools, however. "It is very important to have proper safeguards in place, especially if you opt for automatic training," he says. Automatic training is where the user performs training on the spot, rather than relying on a supervisor or validation employee. An easy safeguard is to have all documents go through a double training process, where the document or form type must be trained twice before it is 'learned.' The software won't consider a form type as learned until the data is validated the same way in two consecutive encounters. "The chance of people making the same mistake twice, such as misidentifying invoice number as customer number, is very slim, especially if the form is encountered by different individuals," says Brierton.

#### **EFFICIENCY IS STILL THE #1 GOAL OF FORMS PROCESSING**

Of course, forms processing solutions at any level continue to yield many benefits — from reducing paper storage burdens to improving information accessibility. However, the increasing capability to capture data from forms of all types, especially semistructured and unstructured, is driving demand for solutions that can boost the efficiency of entire business processes. This means capturing information at the earliest point possible, classifying that data, and ultimately releasing the data to back end systems.

"More end users are looking for integration with their existing ECM [enterprise content management], ERP [enterprise resource planning], and document management systems," says Samuel Schrage, president of AnyDoc Software. Most vendors are addressing this need by providing solutions capable of transforming captured data into multiple formats to suit the needs of an endless variety of line-of-business applications. "Users are also looking for capture solutions that can feed SharePoint with images, indices, and data," says Schrage. The widespread adoption of SharePoint in companies of all sizes is no secret. Recent reports have the collaboration tool listed at more than 100 million licenses sold. Vendors are paying close attention to this trend, as the number of capture solutions that support integration with SharePoint continues to climb.



If you are looking for more insight into the world of unstructured forms processing, [click here](#).

Overall, forms processing technologies are going to continue to become more powerful and intelligent. What should be expected is for the user interface of advanced forms processing technologies to become even more simplified, intuitive, and easy to use. As a result, it's not unreasonable to anticipate the integration of these intelligent technologies into more distributed applications. The transition may not happen overnight, but it will happen.

[Back to top](#)

---

If you would like to receive new information about **ECM** as it becomes available, please register your email address: