

Four Challenges in Taking Capture to the Next Level



Positioning capture to meet the next generation of challenges

In a recent AIIM survey we investigated the question of what information capture “leadership” looks like in user organizations. What does information capture look like in leading organizations that want to position this competency not only as a source of *immediate competitive advantage*, but also as a *long-term competency* critical to the coming era of machine learning?

What are the problems that organizations are experiencing with their capture implementations as they consider this evolution? Here 4 key problem areas that surfaced in our survey. We’ll be discussing this survey and its implications in an [AIIM webinar on September 19](#).

1 Multiple document types.

Most organizations are struggling with capture complexity that is driven by the sheer volume of document types that must be managed. 64% of the organizations in our survey are dealing with more than 10 document types. In reality, this probably *understates* the complexity that organizations face at scale. As one survey taker commented about the number document types in their environment: “Myriad, Multiple, Many, Mucho, Massive, Mega.”

New, advanced data extraction solutions move beyond OCR. Using advanced technologies *on top of OCR* is the only way to properly identify the

document type and then accurately locate, extract, and validate the data. The best solutions build on a strong OCR base and then leverage machine learning, content and image pattern recognition, and automated classification to provide a solution that more than satisfies business users’ needs.

2 Uncertainty about processing accuracy.

Data is at the heart of the Digital Revolution. And data *quality* is at the heart of the challenge facing organizations as they attempt to make their data *fit for purpose* and *fit for use*. According to Gartner, at any moment in time, **up to 40% of an enterprise’s data is inaccurate, missing or incomplete**. When leaders at the top [Business Process Outsourcers](#) were asked by [Parascript](#) how they rated the accuracy of their data results from document processing, 10% rated results “very low” and 50% rate their results as “somewhat low.”

65% of organizations do not approach accuracy from the vantage point of statistically predicting the accuracy of the *system*, and instead rely on measuring accuracy from the individual document level by manual inspection of a small production run or a small sample. Measuring and tuning the *capture system itself* – not just automating the processing of individual documents -- must be automated so that the system continues to classify, locate, extract and verify data with great accuracy

over time. This is challenging in a dynamic production environment where documents and images continually change and new types are added to the system.

3 Poor usability.

OCR software is inadequate for businesses that want to use this extracted data to efficiently process transactions, organize their documents for better control and governance, search important documents quickly and easily, access the right data for decision making, and find the content necessary to support business. OCR software supplies text and numbers devoid of context. This data might serve useful for a full-text search. However, as so many businesses have already realized, full-text search is insufficient and fails to provide a basis for knowledge management and information governance.

62 percent of respondents in our survey rate their capture software “very difficult” or “somewhat difficult” to configure. 44% of organizations do not have expertise or staff available to tune accuracy, and thus “out-of-the-box” functionality is important.

4 Information capture is more complicated than imaging.

Capture is often assumed to be synonymous with scanning. The reality is that most organizations need to do far more than just process images. According to AIIM, 42% of organizations will be spending more on inbound workflow automation over the next 12 months.

Information and data are coming into business organizations from all types of devices and in all types of formats. In fact, when you look at the broader spectrum of the Internet-of-Things, information sources are now extending to remotely connected devices that include security systems, health monitors, and more.

Consider the % of organizations that are trying to automatically extract data from the following “non-image” document sources:

Document type	% trying to extract data from this source
PDFs	97%
Active PDF forms	62%
Excel Spreadsheets	76%
Power Point (PPT/PPTX)	53%
Word documents (DOC/DOCEX)	87%
E-forms	70%

31% of organizations say they approach this challenge by “Processing digital documents *in a different system from the one that handles scanned documents.*” The same percentage say, they convert digital documents to images and processes them in the same workflow as they do scanned images. Clearly simplifying, standardizing, and automating this process is key to improving performance.



This tip sheet was sponsored by [Parascript](#). [Parascript](#) automates the extraction of

meaningful, contextual data from image and document-based information to support transactions, information governance, fraud prevention, and business processes. You might also be interested in:

- [Misplaced Data Quality Priorities – OCR Accuracy vs. System Performance](#)
- [5 Reasons Improving Data Quality Should be a Strategic Priority for Outsourcers](#)

We'll be discussing the survey mentioned above and its implications in an [AIIM webinar on September 19](#) at 2 pm Eastern time. Join us!