Turning Documents into Data
- using data capture to improve document processing

Sponsored by

HYLAND
SOFTWARE

aiim ®
The Global Community of
Information Professionals

About the White Paper

As the non-profit association dedicated to nurturing, growing and supporting the user and supplier communities of information management systems and services, AIIM is proud to provide this research at no charge. In this way the entire community can take full advantage of the education thought-leadership and direction provided by our work. Our objective is to present the "wisdom of the crowds" based on our 80,000-strong community.

We are happy to extend free use of the materials in this report to end-user companies and to independent consultants, but not to suppliers of information management products and services, other than Hyland Software and its subsidiaries and partners. Any use of this material must carry the attribution – "© AIIM 2013 www.aiim.org / © Hyland Software 2013 www.hyland.com"

Rather than redistribute a copy of this report to your colleagues, we would prefer that you direct them to www.aiim.org/research for a free download of their own.

Our ability to deliver such high-quality research is made possible by the financial support of our underwriting sponsor, without whom we would have to return to a paid subscription model. For that, we hope you will join us in thanking our underwriter for this support:

HYLAND
SOFTWARE

Hyland Software
28500 Clemens Road
Westlake
Ohio 44145
USA
Email: request-info@onbase.com
Website: www.hyland.com

About AIIM

AIIM has been an advocate and supporter of information professionals for nearly 70 years. The association mission is to ensure that information professionals understand the current and future challenges of managing information assets in an era of social, mobile, cloud and big data. AIIM builds on a strong heritage of research and member service. Today, AIIM is a global, non-profit organization that provides independent research, education and certification programs to information professionals. AIIM represents the entire information management community: practitioners, technology suppliers, integrators and consultants. AIIM runs a series of training programs, including the Capture Certificate course. www.aiim.org/Training/Certificate-Courses/Capture-and-Imaging

About the author

Nick Geddes is a senior analyst within the AIIM Market Intelligence Division. He has over 15 years marketing experience in a variety of industries. Nick is also a Six Sigma Blackbelt with 10 years' experience working on business improvement projects. He has a BSc in Information Technology, an MSc in Marketing and is a member of the CIM in the UK.

© 2014

AIIM
1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910
(+1) 301 587-8202
 www.aiim.org

© 2014

Hyland Software
28500 Clemens Road,
Westlake, Ohio 44145
(+1) 440 788-5000
www.hyland.com

Table of Contents

Introduction

The concepts of scanning and capture are not new, with many organizations actively working to reduce the amount of paper they process. AIIM research[1] shows that 74% of respondents have business improvement campaigns that would benefit from reducing paper in business processes.
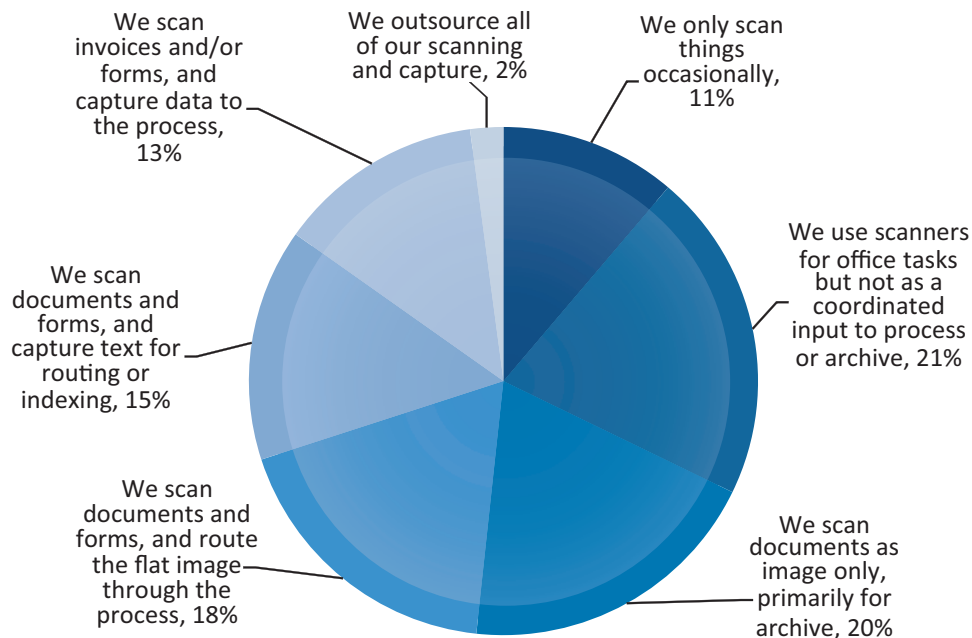
Scanning documents is the first step in the process and brings many business benefits; however, it is capturing information from the page that is the biggest value proposition. Automatically capturing data from the scanned document, validating and interpreting it, eliminates costly process steps, leading to improvements in productivity, customer satisfaction, accuracy of information, and better governance and compliance. Capture complements existing investments in enterprise content management (ECM), ERP and other downstream systems. By eliminating manual processes, capture can provide a solid return on investment (ROI), making a sound business case for change.

In this report, we'll investigate current scanning and capture trends, the benefits organizations realize from this technology and some of the issues they report. We also take a deeper look at what is possible using the latest capture technology to improve business processes.

Moving Toward Automation

As Figure 11 shows, document capture technology is used by most organizations (68%) to get information from paper into a workflow, but with differing degrees of sophistication. For many (20%) the objective is merely to provide electronic images for archive, and in these cases, capture generally takes place at the completion of the process.

Figure 1: How would you describe the level of image capture maturity in your business unit? (N=420)

- We scan invoices and/or forms, and capture data to the process, 13%
- We outsource all of our scanning and capture, 2%
- We only scan things occasionally, 11%
- We use scanners for office tasks but not as a coordinated input to process or archive, 21%
- We scan documents and forms, and capture text for routing or indexing, 15%
- We scan documents and forms, and route the flat image through the process, 18%
- We scan documents as image only, primarily for archive, 20%

Eliminating paper from the process at the earliest opportunity provides more significant benefits such as cost reductions, process flexibility and improved customer response times. Although most organizations recognize that eliminating paper is an important aspect of process improvement, they have generally made little progress. AIIM research[1] shows that 47% have so far made no better than 5% progress toward completely paper-free operations, and 18% haven't even started yet. The slow adoption may be due to budget concerns, but the same research indicated a payback period of 12 months or less for 50% of capture adopters, with two-thirds achieving payback in 18 months or less.

Figure 2: Thinking of all the potential processes where you could drive out paper using scanning and capture, including mobile capture and digital mailroom, how far would you say you are toward that goal? *(N=341)*
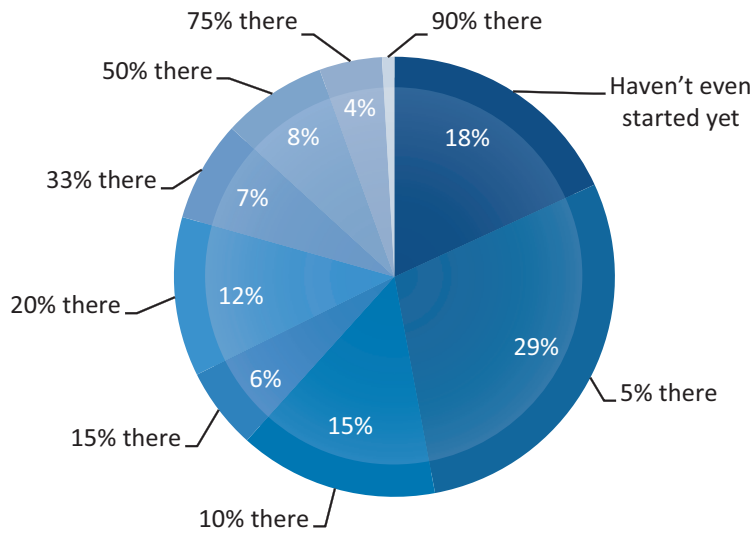

Scanning paper documents and importing existing electronic information (email, PDF, etc.) is important, but it is really only a first step towards getting the most business value from capture.

Recognizing the data on a paper document or form and validating that information is the next stage for many, providing significant process improvements by automating or eliminating process steps. Typically, scanning a document, keying-in data and then routing it through to the appropriate workflow is a manual exercise, with humans needed for data input and classification. When organizations automate or eliminate these manual tasks the process is more efficient, with less potential for error.

The best candidates for automated scanning and capture are organizations that rely heavily on forms and/or process documents. The most popular processes shown in Figure 3 include those used by HR (employee onboarding, timesheets, etc.), finance (accounts payable and receivable) and customer correspondence (help desk). While these forms are the most likely to be automated, other more complicated workflows such as new account applications or case-based processes can also benefit from some level of automation.

Figure 3: Have you made any of the following processes capture-enabled or paper-free?
*(N=408, line length indicates N/A)*
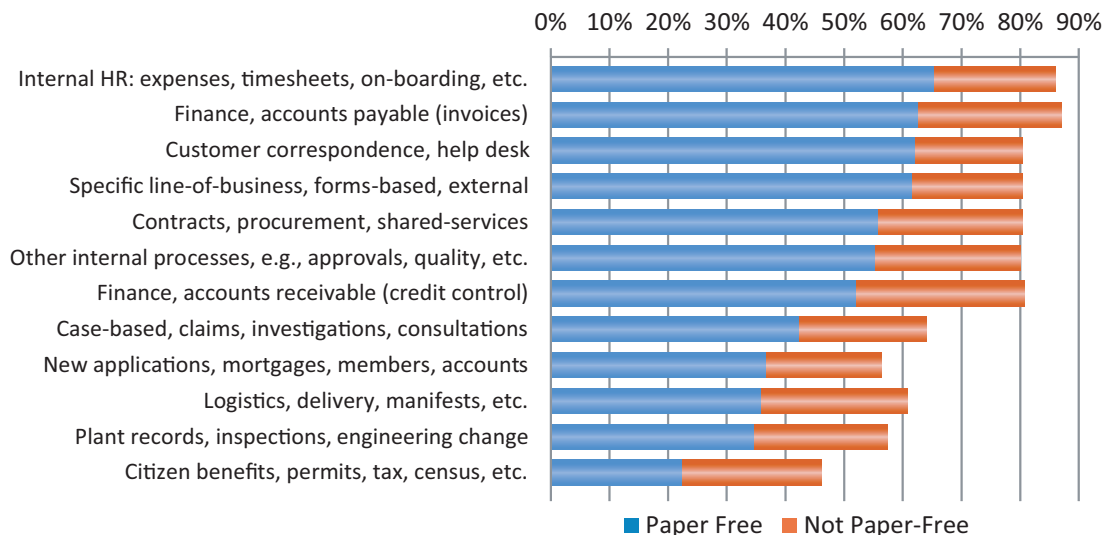
The Automated Process

Automated processing includes several important stages: document or image capture, classification, data capture, validation, and processing. After documents are scanned (or document images imported) they can be automatically classified, using a combination of technologies to maximize the solution accuracy.

During the data capture phase, optical character recognition (OCR) technology recognizes and extracts the data to eliminate the need for manual entry. Data is then validated using customizable checks and balances to ensure accuracy (Figure 4[2] below shows different OCR recognition technologies currently in use). In the conversion phase, the validated data is converted into the appropriate format for the organization's ECM, ERP or other backend system and is exported to it.

Figure 4: Which of the following recognition technologies do you use? (N=280)
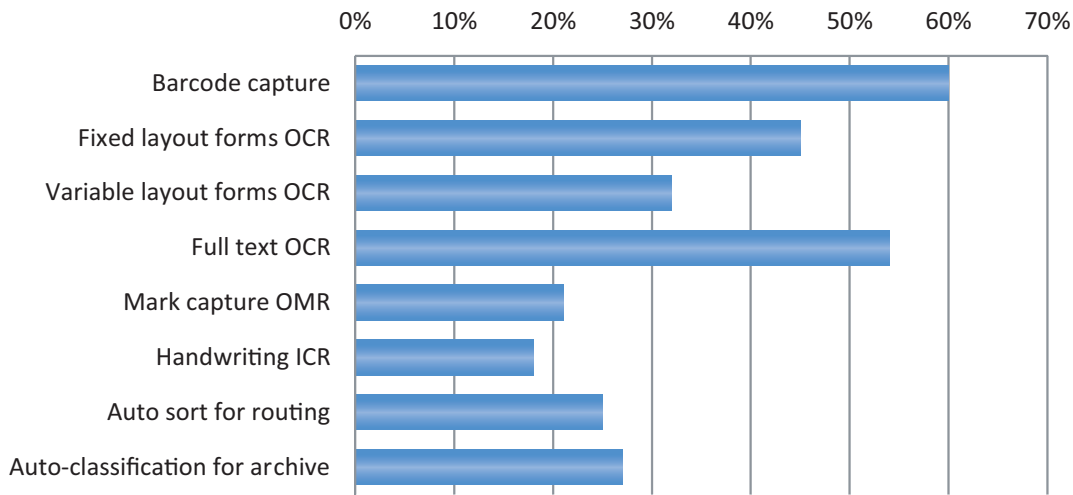
Step 1: Image Capture

The automated process begins with a scanned, imported or e-mailed image of a document, such as a fax, TIFF, or PDF. Documents can be scanned at a central location or from a variety of locations — whichever fits best with a company's existing business model.

The capture system also works with the scanned paper documents from your mailroom as well as all the electronic documents you receive. The software monitors incoming email, automatically identifying each document type and capturing key data elements to sort and route documents and their attachments to the appropriate person or network location. For example, in the case of invoices, vendors may email them as PDF attachments to be automatically routed to the accounts payable (AP) processing queue. Data-feeds from a bank lockbox or scanning service (and the associated images) can also be utilized for organizations that outsource this function.

No matter the method, users view document images within seconds — the information isn't trapped in the mailroom or someone's mailbox, avoiding potential bottlenecks. With instant access to information, time spent previously searching or waiting for information can be spent on more value-added tasks.

Step 2: Classification & Data Capture

Document images can also be automatically identified, sorted and routed to the appropriate person, department, or network location, by applying a variety of classification techniques. This includes utilizing the physical attributes of a document image, matching to a library of images or even "reading" the document to determine its content. For example, documents less than four inches in height can be determined to be checks, invoices can be matched to known vendor logo images, or the data on the document itself can identify it, such as the word "invoice." This ability to classify incoming documents reduces the need for clerical or mailroom staff to manually sort the documents before processing.

Document recognition and classification also eliminates the need to use patch pages, removing another hidden cost. Patch pages are a common document preparation technique that involves an operator physically inserting a blank page wherever document separation is needed. A classification solution easily eliminates this step because it identifies where separation is needed and routes document sets accordingly.

Step 3: Validation & Processing

In the early years of data capture, templates - roadmaps with "structured zones" for locating static data on a document - were the only way to automatically capture data from document images. Fortunately, that is no longer the case. Intelligent data capture technology processes unstructured or semi-structured documents where the location of data on the document constantly varies. For example, in AP departments each vendor will have its own document layout, but this can easily be processed with full data extraction, including line item details.

Instead of creating a unique template for each vendor, the data is automatically located regardless of where it is on the image by combining a variety of methodologies, including searching for defined data labels. Today's technology automatically locates and captures data, such as invoice number, total due, line item details, terms, due date, etc. Advanced systems can also remember where the data was located for the next time that vendor's invoice appears.

Once located, the software solution applies the appropriate recognition technology to automatically capture the data. To increase accuracy during this phase, most solutions are easily customizable to alert the software to expect a certain type of data in a specific location. To ensure consistency, items like a date can be set to be indexed in a user-defined format such as MM/DD/YYYY, no matter how the date is formatted on the document image.

In addition, automatically captured data can be validated against the user's business rules to ensure data accuracy. Most user-friendly solutions allow hundreds of rules to be applied. This ensures the user's organization does not need to change the way it does business to accommodate an automated accounting solution — instead, the automated solution is easily customized to match existing processes.

The most effective solutions also include the ability to look up data already contained in a company's existing databases, such as finance systems, ERPs and existing ECM systems. Using this detailed database information to auto-populate and verify data fields pertaining to a document - or to flag problems for review - creates a more complete and accurate data record for use downstream. This increase in data accuracy also eliminates the bottlenecks previously caused by missing or inaccurate data.

The accurate, verified data is then transferred to the backend system, either on-the-fly or scheduled to occur at an optimal time according to an organization's business practices and resources. A direct integration picks up the captured data and transfers it into the ECM, ERP or other backend system. At this point, a process workflow can be triggered, using the scanned document or form alongside the pre-populated data. With data quickly available, users make faster and more reliable decisions. For example, invoices can be instantly routed for approval after verification, allowing them to be paid quickly.

Benefits of Automation

Organizations must continually look for ways to reduce costs by leveraging technology designed to integrate and augment their existing systems rather than perpetuating costly business practices or investing in new system infrastructures. By replacing manual processes with an automated data capture solution, companies experience a number of quantifiable and qualitative benefits, including:

- **Reduced manual labor costs**
  Automated document processing can eradicate up to 75% of the often hidden labor costs associated with performing manual data entry and other manual processes. Along with avoiding the planned costs of paying employees to perform the tedious task of keying in data from incoming documents and forms, companies utilizing an automated processing solution no longer face backlogs of incoming documents. And, because a potential front-end bottleneck is removed, the need for expensive, unplanned overtime or hiring additional staff is eliminated.

- **Increased data accuracy**
  Manual processing invites the potential for human error. Inaccurate data entry means process staff need to manually correct errors before it wreaks havoc downstream. Dealing with these errors takes employees away from performing more value-added tasks such as servicing customers. By applying custom business rules, an automated data capture solution validates captured data without any human intervention. When these rules are applied, accuracy rates can reach upwards of 99%.

- **Lower per-document processing cost**
  Reduced labor expenses, improved data accuracy and faster turnaround times allow organizations to process more documents in a shorter period of time and at a lower cost without requiring additional staff. This means companies that invest in an automated data capture solution quickly achieve a positive ROI and realize a lower per-document processing cost. This results in significant savings that positively contribute to a company's overall fiscal health by boosting gross profit.

- **Speeds access to critical data**
  Automated data capture speeds the flow of data and images into a user's ECM, ERP or other backend system. Automated verification of data also allows discrepancies to be caught immediately. This means accurate data enters the workflow faster - in hours instead of days - making it readily available for downstream business processes.

- **Saves time and supports compliance with automatic indexing**
  With an automated solution, data fields can be selected to automatically serve as index fields in the ECM system, eliminating the need for staff to manually key in this information. These metadata fields can then trigger records management and retention rules, ensuring compliance with industry regulations, while also improving search and discovery.

- **Faster response to customers**
  When asked, those responding to AIIM research[1] confirm a significant improvement in response times can be achieved using scanning and capture technology. As we see in Figure 5, with an average improvement of 4-times, a typical customer response of four days will be reduced to one day, a huge improvement, especially when used as a differentiator in a competitive business environment.

Figure 5: Taking the widest view, by how much do you think driving paper out of the process using scanning and capture, including digital mailroom and mobile capture, improves or would improve the speed of response of your organization to customers, suppliers, citizens or staff? (N=341)
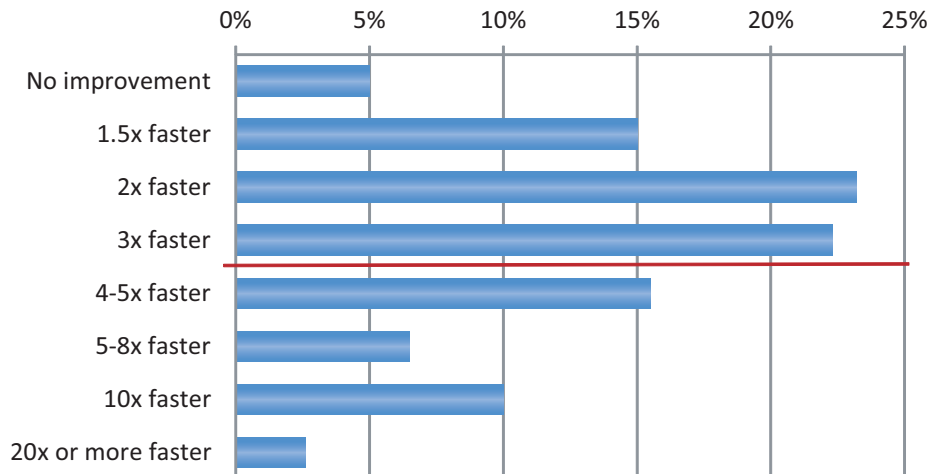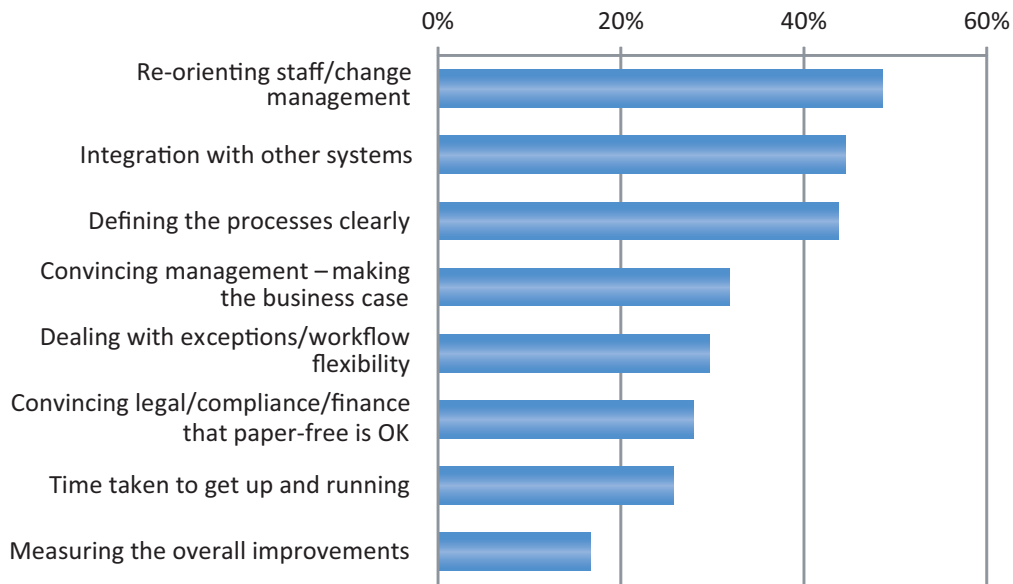
Addressing the Difficulties

As with any process change there are potential difficulties that can be encountered. Organizations should be aware of these challenges so they can include this information in their improvement project plans. AIIM research[1] shows (Figure 6) that re-orienting staff and managing change is the biggest issue. Involvement of key users, communication, and appropriate training are the best ways to resolve this.

To address concerns related to internal business issues, the project documentation should include an in-depth business case (the supplier can provide input here with cost and demonstrable real-world savings) as well as plans for change management, process definition and dealing with exceptions. Designing a new process with enough flexibility to handle exceptions will be one of the key challenges, but vital for a successful transition to an automated workflow.

Integration with other systems is the second most common difficulty encountered; however, this is an area where the supplier should provide clear answers and show experience from similar installations. If off-the-shelf ERP and ECM systems are being used, this will be more straightforward than with in-house developed technology. Providing industry references from comparable organizations can help alleviate concerns when making the business case. These industry references can also relieve legal and compliance fears, as well as helping to understand real-world improvement gains.

Figure 6: What were the main difficulties you encountered in these projects?
(Max THREE) *(N=276)*

Difficulty	Percentage
Re-orienting staff/change management	~49%
Integration with other systems	~44%
Defining the processes clearly	~43%
Convincing management – making the business case	~32%
Dealing with exceptions/workflow flexibility	~29%
Convincing legal/compliance/finance that paper-free is OK	~28%
Time taken to get up and running	~25%
Measuring the overall improvements	~18%

Conclusion

As organizations strive to streamline business process, scanning and capture technology is increasingly used in a variety of business areas. While removing paper is undoubtedly bringing process improvement, taking the data from the page and using it to automate or eliminate process steps is the next phase organizations should move toward as it not only improves productivity and speeds access to information, it provides calculable ROI and continuous cost savings well into the future. The use of automation can leverage the large investments made in both scanning and capture technology as well as in backend systems such as ECM, ERP and CRM for further business performance gains.

Recommendations

■ Review your current situation with paper, scanning and capture. Look at current processes and identify those where paper can be removed.

■ Look at the steps that can be eliminated if information is electronically captured from the document – whether it arrives on paper, PDF or email.

■ Speak with vendors about available technology and how it has been implemented in similar organizations within your industry. Discuss your concerns, and ask them to provide real-world examples of how these have been overcome.

■ Identify and initiate a pilot program, integrating scanning and capture into one of your processes to prove the benefits and ROI for your organization.

■ Always plan to roll out scanning and capture to multiple processes rather than as a single-point solution. Equipment costs and integration with other core systems can then be amortized over several business areas.

References

1 AIIM Industry Watch, July 2013, "Winning the Paper Wars – capture the content and mobilize the process troops"
   www.aiim.org/research

2 AIIM Industry Watch, February 2012, "The Paper Free Office – dream or reality?"
   www.aiim.org/research

UNDERWRITTEN BY

HYLAND®
SOFTWARE

For more than 20 years, Hyland Software has helped our more than 11,500 lifetime customers by providing real-world solutions to everyday business challenges. That dedication is why Hyland realizes double-digit growth, and why 98 percent of our customer base continues to renew its annual maintenance. Our customers see the ongoing value of partnering with Hyland and continue to work with us year after year.

Hyland's enterprise content management (ECM) solution, OnBase, is one of the most flexible and comprehensive ECM products on the market today. OnBase empowers users to grow their solutions as needs change and business evolves. It is tailored for departments, but comprehensive for the enterprise, designed to give you what you need today and evolve with you over time.

AnyDoc, a Hyland Software solution, provides award-winning document, data capture and classification solutions. Thousands of companies worldwide rely on AnyDoc to accurately capture and verify data from both structured and unstructured documents to dramatically reduce the cost and inefficiency of manual data entry.

For more information, please contact request-info@onbase.com or visit Hyland.com.

aiim®
The Global Community of Information Professionals

AIIM (www.aiim.org) is the global community of information professionals. We provide the education, research and certification that information professionals need to manage and share information assets in an era of mobile, social, cloud and big data.

Founded in 1943, AIIM builds on a strong heritage of research and member service. Today, AIIM is a global, non-profit organization that provides independent research, education and certification programs to information professionals. AIIM represents the entire information management community, with programs and content for practitioners, technology suppliers, integrators and consultants.

© 2014

AIIM
1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910
301.587.8202
www.aiim.org

AIIM Europe
The IT Centre, Lowesmoor Wharf
Worcester, WR1 2RR, UK
+44 (0)1905 727600
www.aiim.eu