



harvey spencer associates



Intelligent Enterprise Capture: *A Document On-Ramp to SharePoint 2010*

Harvey Spencer Associates Inc.

David Lakness, Research Director

2 Penfield Dr.

East Northport, New York 11731

1-631-368-8393



Executive Summary ii

Metadata-driven Intelligent Enterprise Capture 1

Intelligent Enterprise Capture with SharePoint 2

 Centralized capture 3

 Distributed capture at point of origination 3

 Document classification 3

 Extract, validate, and deliver to SharePoint 3

 Central administration 4

Enterprise collaboration intersects with Intelligent Enterprise Capture 4

From the digital office to capture for business processes 4

Conclusion 4

EMC® Captiva® Intelligent Enterprise Capture Solutions 6

 Captiva – an intelligent enterprise capture solution 6

 Capture flow process – drag and drop design 6

 Document classification and data extraction 6

 Secure processing and reporting 6

 Conclusion 7



Executive Summary

Managing content that is derived from paper continues to be expensive. Industry analysts estimate that as much as 80% of all potentially usable business information originates from paper, or computerized information sources, requiring analysis and interpretation to be recognized for processing. Information capture from this structured and unstructured content is the primary information source for loan origination in banking, claims processing in insurance, invoice processing and new account enrollment across all industry segments.

To succeed in today's digitally driven market; organizations need to streamline their business operations by structuring the flow of business documents and data. In an effort to keep pace with the ever changing market, many companies have begun to adopt Microsoft SharePoint across an enterprise as a place to store documents and collaborate in an effort to improve productivity and access to critical business information. But in order to automate critical business processes, the ability to capture and store paper as digital content in a repository like SharePoint is required, so that the information can be easily retrieved and acted upon.

This white paper explores SharePoint® 2010 and how Intelligent Enterprise Capture provides a platform for capturing all types of documents and data, in a distributed and centralized environment. The paper further explores how Intelligent Enterprise Capture provides a common framework for implementing automatic document type classification, intelligent data extraction, validated accuracy, and delivery into SharePoint 2010.

Intelligent Enterprise Capture defined

Intelligent Enterprise Capture, integrated with SharePoint, enables a standardized approach to deploying document capture across all departments and processes, offering the following automated services:

- Transform paper based documents into digital images at point of entry to the organization.
- Capture, classify, extract data and validate accuracy for content regardless of format or type.
- Automate the straight-through processing of information into workflows and business systems.
- Audit processing of all documents from point of receipt through to archiving.

Intelligent Enterprise Capture is fundamental to making document capture a normal part of the business process. Sophisticated document and data recognition and validation provides the “on-ramp” to automated processing from front-office capture devices such as multifunction peripherals (MFPs) and network scanners, to back-office high volume scanning, fax, and email.



Metadata-driven Intelligent Enterprise Capture

In SharePoint 2010 the management of metadata is core to how documents get stored, searched on, acted upon, and retrieved. Content type, taxonomy, business process, storage assignment, permissions, records declaration, and lifecycle management are all centrally administered through SharePoint Enterprise Metadata Management (EMM). Unique Identifiers (guaranteed unique across the enterprise) are available to be assigned, enabling automated content enrollment.

These new set of capabilities provides consistent metadata and shared taxonomies across multiple SharePoint sites in an enterprise. This is important for intensive metadata processes like capturing data from an invoice which contains critical information to the business workflow process – date, vendor ID, purchase order number, invoice number, invoice amount, line item details, and more. The information contained in an invoice is critical to the accounts payable approval process given the data is required for updating an Enterprise Resource Planning (ERP) system, storing the information in a repository, and initiating a workflow.

One of the key components missing from SharePoint 2010 is the ability to capture and deliver paper-based content into SharePoint. Intelligent Enterprise Capture compliments SharePoint by capturing documents from scanners, fax, multi-function devices (MFD), email, and electronic data-streams. By leveraging the new set of SharePoint Enterprise Metadata Management features, enables organizations to assign captured items a “content type” which defines specific metadata and workflow processes to associate with items. Content type is the primary key for storage policy assignment as items are submitted to the “drop-off-library” for content organizer routing to specific SharePoint folders.

SharePoint 2010 metadata navigation also offers a location independent view of content as keyword or managed term (hierarchically) organized virtual folders. Content can be enrolled in multiple virtual folders through term association, providing logical views of content collections specific to role, function, department, and location. For data extraction that does not map to a managed term or keyword, a keyword may be created which will subsequently be available for tagging. Keywords are easily promoted through the SharePoint Term Store Management Tool, offering a process to provide content based input to expand the managed term vocabulary, perhaps incorporating some cross referencing and semantic tagging across content categories.

These new metadata capabilities provide the basic building blocks for a more powerful set of tools to manage paper based information stored in SharePoint as digital content.



Intelligent Enterprise Capture with SharePoint

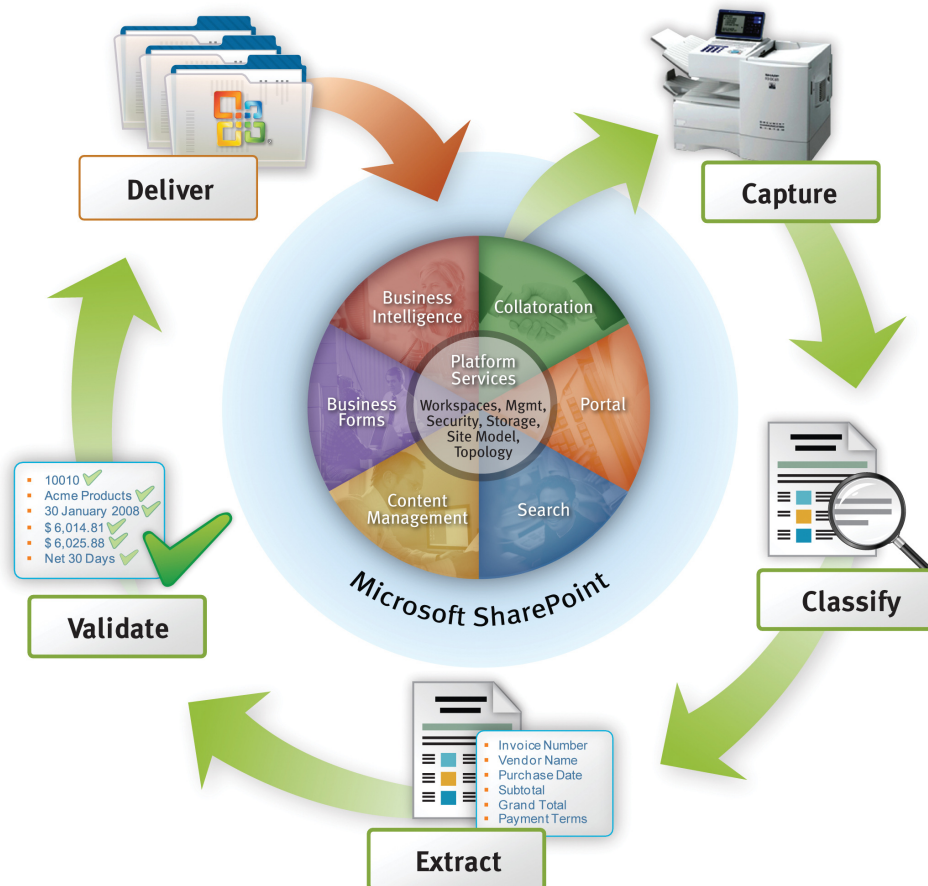


Illustration: Intelligent Enterprise Capture supports the capture of paper documents and electronic files from scanners, multi-function devices, fax, and email, in a centralized and distributed environment.,

Intelligent Enterprise Capture supports both centralized and distributed capture, and, when connected with SharePoint, offers a strategic platform for organizations to automate the capture and delivery of documents from anywhere in the enterprise. Intelligent Enterprise Capture can also provide automated records on-boarding to SharePoint Records Center through the same facilities, enabling staff at all levels to capture and enroll new document records through an automated classification and data accuracy validation process.

Metadata, workflow, lifecycle policy, hierarchy and relationships are defined as content type properties in SharePoint 2010. Content type syndication publishes content type defined policy to multiple SharePoint sites and/or farms across the enterprise, thus establishing consistent policy and structure. This new model for managing content type provides a consistent structure for integrating capture across multiple SharePoint sites and/or farms. Taking full advantage of Intelligent Enterprise Capture with SharePoint is discussed below.



Centralized capture

Centralized capture is a term used to describe the automation of capturing incoming structured and unstructured paper content at dedicated facilities with trained staff and resources provisioned exclusively to automate the capture process for the organization.

Sometimes referred to as mailroom capture, centralized capture enables companies to standardize their unstructured and structured content processing procedures. Corporate governance and compliance policy best practices are often enforced through this centralized approach. Dedicated facilities enable provisioning high throughput production capture devices responding to paper volume and turnaround requirements.

Distributed capture at point of origination

A mix between centralized and distributed capability efficiently addresses organizational needs serving multiple locations, and departments. Intelligent Enterprise Capture enables a distributed capability with the same level of content integrity validation and processing enjoyed by central capture, and straight through processing capability, at the front-office point of origin, offers significant process optimization and latency reduction benefits. The benefits of distributed capture are quite significant to a business, which include eliminating shipping cost, enabling front-office knowledge workers to provide additional context about the documents being captured, and reducing delays in delivering content into back office systems.

By centrally managing and utilizing the same capture platform that is used for centralized processing, administrators are able to control the process of captured content into the SharePoint environment through automated procedures. This ensures consistent delivery of content into one or many SharePoint server sites and/or farms.

Document classification

Intelligent document classification provides a way to automatically recognize the document type, which is the key interface point to delivering information into the right location within SharePoint. Document content type also drives the data extraction rules during the capture process. The data captured from specific document types are then associated with columns in SharePoint and are the primary keys used to retrieve documents later on.

The classification of documents may be configured to incorporate specific business objectives into the categorization and subsequent metadata association process, for the purpose of driving additional business processes within SharePoint. For example, insurance claims are driven by a variety of documents received over time from first notice of loss to settlement. Rich association of otherwise separate content types through the policy number data can form a virtual claim folder for storing essential claim settlement documents in SharePoint. Automated document classification and integration with SharePoint accelerates the claims submission time and improves customer service, by reducing manual paper handling steps like pre-sorting of documents. .

Extract, validate, and deliver to SharePoint

Intelligent Enterprise Capture provides the means for automatically extracting and validating data from documents to associate with SharePoint document properties . Sophisticated recognition capabilities are employed to extract all types of data – type written, handwritten, barcodes, and checkboxes - from all types of documents, including free-form data extraction from less structured documents. Additionally, field level business rules, and external data validation are utilized as well for ensuring a very high degree of accuracy. This type of approach to capturing data helps companies reduce the cost associated with manual data entry and accelerate the delivery of content into SharePoint.



Central administration

Administrator tools enable central governance of the capture environment, including maintaining document capture process flows, security rights, audit and tracking of document processes.

This central oversight provides the necessary means for maintaining strong controls over a capture environment that can span across many departments and processes. By controlling the capture environment centrally, administrators can manage the capture processes that are deployed to remote sites ensuring documents captured and delivered into SharePoint adhere to strict content definition guidelines.

Enterprise collaboration intersects with Intelligent Enterprise Capture

Intelligent Enterprise Capture synchronized with SharePoint EMM, syndicated across the enterprise, brings seamless information recognition and validation to the front office of a business. This front office capability makes document capture a normal part of the business process. Automated processing is enabled from document portals in the “front-office”, such as multifunction peripherals (MFPs), network scanners, fax, email, and electronic data-streams.

From the digital office to capture for business processes

Getting content to the right place is simpler with Intelligent Enterprise Capture for SharePoint. Automated document classification and data extraction allows businesses to organize each individual document into the appropriate library and folder based on its contents, enabling users to find content by topical navigation.

Utilizing Intelligent Enterprise Capture, allows businesses to address both simple and complex business challenges. On the one end, simply getting paper into SharePoint as digital content is possible using the broad set of capabilities that Intelligent Enterprise Capture provides, which is referred to as creating a digital office for managing paper information.

On the other end, paper is a critical part to lots of different business processes – accounts payable, claims, new account applications, and more. To support these types of processes, document classification and data extraction are leveraged to submit content into a library/folder that is associated with a SharePoint workflow. From there SharePoint Business Connectivity Services (BCS) streamline deep integration of captured content, external data and content-enabled applications (CEA)..

Conclusion

SharePoint 2010 offers a comprehensive response to the functionality, reach, extensibility, and deployment requirements of businesses – across people, departments, divisions, partners, internal processes, inter-organizational processes, and customer-facing front-office processes.

Organizations will quickly find that with every department or business process, paper is a critical part. SharePoint 2010 capabilities build upon how content needs to be delivered, accessed, and stored.

Intelligent Enterprise Capture, connected with SharePoint, enables seamless document capture, recognition and validation throughout the enterprise (central & distributed). Documents containing business critical information, from wherever they enter the organization, are delivered to enterprise systems and applications as a normal part of the business process. Intelligent Enterprise Capture provides a consistent platform to effectively deploy departmental and large scale applications, to automate the processing of large volumes and varieties of documents organizations come into contact with on a daily basis.



The latest SharePoint 2010 Enterprise Metadata Management capabilities provide the basic building blocks for more powerful and complex metadata solutions that capture applications require.

Intelligent Enterprise Capture, deployed with SharePoint 2010 will enable organizations to achieve document process transparency from point of capture, guaranteed chain of custody audit, and business process streamlining through Content-Enabled Applications.



EMC® Captiva® Intelligent Enterprise Capture Solutions

EMC is one of the information capture solution providers on the market today. The Captiva solutions are reviewed here as they relate to the topics discussed in this white paper.

The EMC Captiva intelligent enterprise capture solutions are architected as a capture platform that supports centralized high volume batch capture, distributed capture, and ad hoc distributed capture at the desktop or from a remote capture device.

The solution's functionality is segmented into three basic capabilities: scanning/operator interfaces, capture processing, and system management. The modular approach offers a flexible and open intelligent enterprise capture platform.

Captiva – an intelligent enterprise capture solution

EMC Captiva intelligent enterprise capture is a highly configurable solution that captures and identifies documents and automatically extracts and validates data from all types of documents – structured, semi-structured, and unstructured. An open architecture that provides a .NET development environment and supports web services enables document capture to be quickly integrated with line-of-business business applications, existing document/content management systems, ERP systems, and process models.

Within a centralized and/or distributed environment, the capture server can prioritize processing by levels of urgency, which can be defined by the system using a series of rules. Captured content can be stored in several repositories including SharePoint, and automatically forwarded to the appropriate departments or individuals.

Capture flow process – drag and drop design

Capture flow processes control how documents are handled, including scanning, image enhancement, document identification, recognition, validation and export. Captiva provides a simple drag-and-drop process design application that allows users to develop capture processes in a matter of hours, while at the same time providing developers with tools to customize the capture process to fit unique business requirements.

Document classification and data extraction

Intelligent document classification and data capture is where the core intelligence lies within Captiva. Captiva automatically identifies incoming documents as they're captured, routes them based on pre-defined business rules, and extracts business data from a wide variety of document formats.

Validation processing of extracted data within Captiva includes quality control measures ensuring the data captured meets the highest quality standards that businesses set.

Distributed capture

For distributed capture operations, Captiva provides a Web client for scanning, importing, and indexing images and electronic files via a browser. Control of the capture process occurs directly from within the browser, where a remote user captures images using either a desktop scanner or browsing a network folder for images that were captured via a multifunction device, network scanner, or other source.

Secure processing and reporting

Finally, the Captiva intelligent enterprise capture solution provides a secure platform for the capture process and the captured content. Tight audit trails, granular security settings allow administrators to manage the capture process. System reporting and monitoring allow administrators to maintain and fine-tune the capture operations.



Conclusion

The Captiva intelligent enterprise capture solution allows organizations to deploy capture in both a distributed and centralized capture environment as a single platform. It is worth noting that these EMC Captiva products and solutions are also part of the bigger EMC family of Enterprise Content Management products and solutions. On the one hand, EMC views capture as a platform that can be integrated into virtually any other application. On the other hand, EMC views Enterprise Content Management (ECM) as a modular platform that can address many business challenges and of which capture is just one component.

Although it is part of the EMC family of products (which includes EMC Documentum), the Captiva solution has been integrated with SharePoint and other content repositories.

Now intelligent enterprise capture may sound big, but many successful organizations start small, and grow to environments that span across many departments. And with Microsoft SharePoint increasingly becoming a ubiquitous, reliable, and a product used across an organization, the need for a flexible and scalable intelligent enterprise capture solutions will continue to grow.