InfoTrends
A Questex Company
Focused | Responsive | Credible | Visionary

October 23, 2008 **Analysis**

Input*Accel* 6: Intelligent Capture for the Enterprise

Authors
Randy Dazo
Omri Duek

Published by
Dynamic Content Software Strategies Consulting Service

Abstract

EMC Captiva continues to build upon its core strengths for enterprise document and data capture, including enhanced support for centralized and distributed capture environments, significant throughput performance and high availability improvements, and ease of development. Input*Accel* 6 also responds to market demands for lower ownership costs, better business usability, and entirely new ways to integrate intelligent capture within enterprise business applications, including providing capture services within a Services-Oriented Architecture (SOA).

For More Information

If you would like to order extra copies of this report, receive permission to use any part of the report, or be informed of upcoming market updates, reports, and related projects, please e-mail us at info@infotrends.com.

Headquarters:
97 Libbey Industrial Parkway
Suite 300
Weymouth, MA 02189
United States
+1 781 616 2100
info@infotrends.com

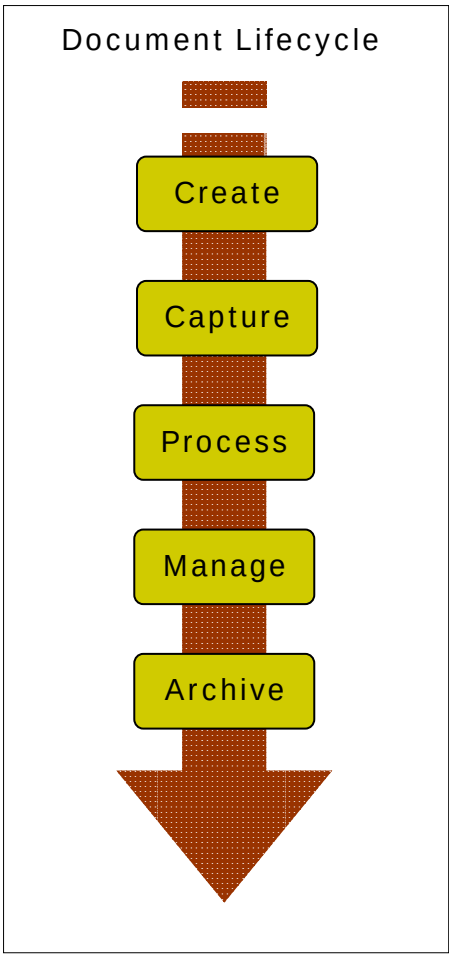
Europe:
3rd Floor, Sceptre House
7-9 Castle Street
Luton, Bedfordshire
United Kingdom, LU1 3AJ
+44 1582 400120
euro.info@infotrends.com

Asia:
Hiroo Office Building
1-3-18 Hiroo, Shibuya-ku
Tokyo 150-0012
Japan
+81 3 5475 2663
info@infotrends.co.jp

Table of Contents

Introduction

InfoTrends' research reflects that while 58% of enterprise documents[1] will be sent via e-mail and 41% will be exposed through the Web, almost an equal number will be printed (51%). That said, organizations must invest in a "hybrid" strategy for ingesting content, one that recognizes the enduring importance of paper while delivering on the cost-saving promises of electronic document capture and management.



Cost pressures are driving organizations to capture the information included in documents, not just the documents themselves. By exposing this information to other business systems, significant cost savings can be gained through process integration, especially for high-volume capture scenarios that involve documents ranging from highly structured forms to unstructured documents with free-form text.

Effective enterprise capture plays a critical role in enabling the automation (or optimization) of subsequent steps in the document lifecycle, broadly depicted in the image to the right. Information extracted in the "Capture" stage—whether from paper or electronic channels—informs the processing and routing of documents in the "Process" stage, the security and privacy needs of the "Manage" phase, and (ultimately) the compliant retention policy required in the "Archive" phase.

[1] Refers to documents with multi-channel delivery requirements.

Introducing Input*Accel* 6

EMC's Captiva product group has historically recognized that capture technologies interact with a much broader ecosystem, including capture hardware devices, content management technologies, and other back office business systems, such as Business Process Management Systems (BPMS), Enterprise Resource Planning (ERP), and Customer Relationship Management (CRM). For larger organizations, this ecosystem often manifests as a disparate environment—agile business requires capture technologies that can scale for centralized high-volume scanning centers while providing reach across highly distributed departmental document scanners, senders, and multifunctional peripherals (MFPs).

These enterprise environments also require capture to scale and to be able to flexibly integrate with the IT infrastructure—capture "services" should be made available for Services-Oriented Architectures, enabling developers to create and integrate more efficient and innovative capture solutions that span business-specific processes and needs. This architecture approach, increasingly being adopted by enterprises, is core to the deployment of integrated enterprise solutions that span capture, BPMS, ERP, and other enterprise systems.

EMC Captiva's Input*Accel*, the platform to the Captiva intelligent capture solution, offers strong support for environments such as these. To that end, batch capture capabilities are matched by EMC's multi-server technology, which load-balances processing needs across available servers and workstations, delivering scalability for high-volume capture needs. For distributed capture environments, integration with hundreds of distributed MFPs is offered by way of a connector to eCopy ShareScan, and the Captiva eInput Web client for remote scanning and indexing addresses the needs for distributed processing. In addition, to easily manage this complex environment, Captiva's Input Management Console (an add-on to the Input*Accel* platform) provides centralized monitoring over the capture infrastructure and process.

Relative to process integration, Intelligent Document Recognition (IDR) powered by EMC Captiva's Dispatcher classifies documents, extracts information, and enables greater automation for downstream process needs. Of course, Input*Accel* is also highly customizable—interface changes, document classification tweaking, cross-system data validation, and pre-defined job types can be tailored to enterprises with many specific needs. The latest edition of Input*Accel* (version 6) builds on these tenets of enterprise scalability, agility, and usability that drive further benefits for organizations.

Simplified Customization Using SOA and .NET

Notable in this EMC Captiva release is broader support for **Services-Oriented Architectures (SOA).** This approach simplifies integration with enterprise applications and systems, enabling organizations to customize and develop capture applications that are standards-based, extensible and reusable, scalable, and platform-neutral. EMC Captiva Web Services exposes Input*Accel* capture processes as services to other enterprise applications and systems, enabling capture request and creating loosely-coupled components that can be linked to business systems and processes. These services simplify development tasks and increase the ability to develop and monitor end-to-end business processes across heterogeneous environments, providing an entirely new approach to how capture is leveraged within other applications and processes.

Additionally, this Input*Accel* release adds **client-side scripting** based on .NET, which enables developers to write custom scripts as Microsoft .NET assemblies. The supported languages include Visual Basic .NET, C#, and other Windows .NET languages. These scripts direct information routing, conduct error handling within a capture process, perform third-party database lookups, modify the behavior of the current capture task, and perform field-level validation and population of data.

Enterprise Capture Ready

A new, Web-based **Administration Console** provides centralized view and control over the capture environment. Key enhancements include allowing administrators to define custom security roles, monitor and review problem batches in a temporary holding area, and perform advanced searches for locating specific batches. The administration of the capture environment has also been enhanced to enable administrators to manage a multi-server environment as if it were a single server. This provides administrators with tools to add, modify, and remove servers; manage page volume licenses across servers; and modify server settings.

This release also adds several performance enhancements such as **multi-threading** to the Input*Accel* server, allowing multiple input/output threads to run concurrently across multi-core processors. By taking advantage of today's multi-processor servers, the Input*Accel* server is better able to manage requests from the client modules, resulting in performance gains. Other Input*Accel* server enhancements that will benefit performance include the ability to process and synchronize modified batches between the client and server, and a separate, work-in-progress database that allows administration functions to be handled independently of the Input*Accel* server. The latter prevents core Input*Accel* processing jobs from being affected by an administrator request.

Failover support for Input*Accel* now includes **Active/Active clustering** technology. With Active/Active cluster support, existing cluster nodes can be used as a failover option, allowing a second virtual cluster to be triggered on the same server, ensuring that no server is idle during normal operation or while waiting for failover recovery.

Easier to Use and Deploy Capture

EMC has invested in substantial client improvements for the version 6 release. The re-designed distributed **scan and index interfaces** make it easy for administrators and users to customize these common clients to meet usability requirements, streamlining the user experience. For example, a workspace layout enables users to resize, move, add, and delete task panels. This gives developers of the capture processes tremendous flexibility to how the user interfaces are displayed for each type of user and process; users will find that a personalized interface, tailored to the business process at hand, maximizes productivity gains. Finally, IT administrators will find that these high-volume scanning and indexing applications are Web-deployable and can be easily delivered across the entire enterprise, leveraging Microsoft ClickOnce technology for highly simplified installation and updating of client configurations. This simple method of deployment and maintaining updates means that capture can easily be distributed and maintained throughout an enterprise organization.

Visibility and Access to Critical Information

Reporting is another key capability that has been added to Input*Accel*. **Integrated system reporting** supplies document audit trails as well as document and data capture statistics—detailed processing information such as interactions between modules and individual pages, module performance metrics such as page recognition times, module processing times and metrics associated with automated processes, and operator metrics (including operator performance information such as characters typed, fields indexed, and pages processed). Custom reports can also be designed with Business Objects' Crystal Reports.

Coping with a Paper Heritage

The reality for most organizations is that a substantial number of documents persist in paper form today, including faxed contracts, invoices, financial applications, and claims. Moreover, as a result of less-than-adequate information management practices, many so-called "electronic" business processes have retained "paper" attributes.

In many electronic processes, documents and their information are rarely well-structured or classified—e-mails, office documents, and scanned pages in particular are the most notorious offenders. In addition, knowledge workers are often required to hunt down, syndicate, and make decisions based on information located in distributed systems and repositories. As a result, not only is productivity and process performance hindered, but performance visibility is limited to the audit trails created by knowledge workers.

Capture vendors are increasingly responding to these needs in the market, recognizing that effective, intelligent capture enables greater innovation in subsequent steps of the document lifecycle. First and foremost, Intelligent Document Recognition (IDR) and Optical Character Recognition (OCR) technologies provide the document classification and information extraction capabilities required for driving an electronic business process. These capabilities eliminate the need for manual pre-sorting and keying of information, thus reducing costs and accelerating a paper process. Secondly, these vendors are providing better options for capturing these documents in a distributed manner. Support for remote offices, MFPs, network scanners, and other capture devices enables knowledge workers to contribute documents to a business process without having to go through a batch back-office process that often includes mailing documents to a central scanning facility. Third, integration between third-party systems provides a one-stop interface for knowledge workers (or automated rules engines) to make business decisions. End-to-end business process solutions can be developed using loosely coupled, highly integrated services with a services-oriented architecture. Information population or validation across several data sources can be automated by leveraging capture scripting tools. Finally, audit trails and performance metrics are managed by capture technologies and integrated ECM (enterprise content management) platforms, giving line-of-business managers a complete picture of the document and information flow in a process. These audit trails and metrics also provide the basis for maintaining process consistency and compliance.

In short, an increasing need for process automation and optimization is manifesting capture requirements both upstream and downstream in the document lifecycle.

Advanced Capture Solutions: Integrating with a Business Process

Although the specific capabilities required will vary by case, business process optimization initiatives are driving several key capture requirements. Foremost, as noted in the introduction, capture of information must include both paper-based and electronic documents—including e-mail, e-forms (Web documents), and their respective attachments.

Another critical requirement for a capture platform is its ability to address environments ranging from centralized, high-volume scenarios to highly distributed scanning. For example, many business processes are initiated from remote branch offices whose paper-intensive content can be effectively captured at the point of entry. Distributed capture cuts the time and costs required in such a process when compared with centralized facilities to which documents must be mailed. Furthermore, distributed processing can often involve the indexing and validation of documents by an off-shore subsidiary or partner, which necessitates a capture platform that can support the unique requirements of this type of environment.

Beyond "hybrid" and distributed capture needs, other key capabilities for optimization or automation initiatives include validation of information against client parameters or third-party system data, intelligent workflow initialization such as routing documents and notifying downstream process participants, and integration with third-party business systems for storage or further processing of captured documents. As organizations continue moving toward services-oriented architectures for deployment of solutions that span enterprise processes, the availability of reusable and standardized capture Web services will be crucial.

Sample Use-Case: Claims Processing

In the insurance industry, claims processing is a fairly common procedure that requires obtaining information from a customer, validating that information, making a claim decision, and initiating a claim payment or denial workflow. For high-value, high-risk, or complex claims or customers, additional steps may be required.

In the following Table, consider how the classification of documents and extraction of document information in the capture phase empowers automation in subsequent stages. Moreover, consider how capture technologies, integrated with the broader IT infrastructure, can drive the "Process" stage of the lifecycle by applying business rules to captured content.

Table 1: Claim Processing (Simplified) Workflow

Create / Capture	✓ Customer submits claim via Web, mail, or e-mail ✓ Print document is sent via fax or scanned by local branch or high-volume scanning center ✓ Electronic and print claims are ingested by capture solution: o Claims are automatically classified and information is extracted using intelligent document recognition (IDR) and OCR, respectively o Claims are indexed by claim type, customer type, claim value, or other business parameters o Customer and claim information is validated against enterprise databases (and possibly, external agents) ✓ The resulting "well-structured" document is stored natively or sent to a downstream repository such as ECM or ERP
Process	✓ Based on extracted information, claims are subject to business process rules and… o Approved o Denied o Escalated for further investigation by claim specialists ✓ A decision initiates a tangential business process, such as claim disbursement or the mailing of a denial explanation
Manage	✓ Following approval/denial, claim-related documents must be stored, managed, and monitored for compliance ✓ Claims must be searchable and accessible by (among others): o Customer service representatives responding to inquiries o Auditors and accountants monitoring claim performance o Risk management personnel o Third-party IT systems, such as business intelligence (BI) ✓ Payment information and supplementary data may be added to the claim and customer files
Archive	✓ Documents and supplementary information are retained securely according to records retention policies.

Without intelligent capture capabilities for identifying and extracting document information, costly labor resources are required for pre-sorting and managing the routing of documents as well as keying, verifying, and validating a document's data. In an optimized process, these resources are only called upon for exceptional cases or where document and data recognition accuracy is statistically questionable. Notably, best-of-breed capture products will provide capabilities for classification and data extraction beyond structured forms, including unstructured office documents, e-mails, and other content types. By automating the capture and processing of documents, greater efficiency can be achieved and scaled. What's more, increased process accuracy can be achieved due to reduced human intervention.

Modern capture technologies also provide the integration required for automating business decisions and content flow throughout the document lifecycle. Information captured from documents, for example, can be used by rules engines to automate the "Process" phase. When knowledge workers are required to handle exception cases or "Manage" documents, information can be easily searched and syndicated from capture, ECM, and other enterprise repositories, leveraging the data extracted in the capture phase. Once again, the availability of capture Web services for developing these highly integrated business solutions should be underscored.

Scenarios such as these dominate business processes across several verticals. The following Table provides a non-exhaustive sample of documents that can benefit from the process integration discussed above.

Table 2: Sample Document-Centric Processes

Financial Services	✓ Loan applications, remittance processing
Insurance	✓ Claims processing
Government	✓ Tax documents, case management
Healthcare	✓ Electronic health records (EHRs)
Engineering	✓ Contract management
Manufacturing	✓ Order management
Retail	✓ Customer service correspondences
Legal	✓ Case management

There are other, non-industry-specific processes that can highly benefit from process integration as well, including accounts payable/receivable and human resources.

The New Standard in Enterprise Capture

How does Input*Accel* 6 respond to these modern, process-centric capture requirements? Primarily, EMC has continued investing in native enhancements to its products, while acknowledging the importance of creating an open platform that is extremely flexible in nature.

Enabling Enterprise Capture

When it comes to processing business-critical information in an enterprise capture environment, organizations want a solution that is scalable, adaptive, and proven. The latest version of Input*Accel* builds upon a proven capture platform which for years has played a central role in supporting the day-to-day processes that organizations depend on to run their businesses. Input*Accel* 6 adds to this technology platform by introducing enhancements around performance and high availability, while making it easier to manage and use capture applications throughout the enterprise.

Relative to the applications described above (such as claims processing), the Captiva Input*Accel* 6 platform provides the capabilities to manage and support the unique business requirements that are typically associated with these critical processes—whether they involve the classification of thousands of different document types, the processing of millions of pages per day, the extraction of critical information from these documents, or the application of complex business rules to this data based on client parameters or integrated external business systems, rules engines, and databases. Captiva Input*Accel* Web services provide the reusable components required for agilely developing these solutions within a Services-Oriented Architecture.

Distributed Capture and Processing

Historically, paper-based processes that originated from remote locations have been highly inefficient, with most organizations relying on centralized scanning centers that required costly and inefficient postage. With the preponderance of distributed capture devices and sophisticated capture platforms such as Input*Accel*, however, organizations can take an enterprise approach to capture while responding to remote scanning and distributed processing requirements.

Input*Accel* continues to provide a platform for a distributed capture and processing environments that is flexible enough for a wide gamut of organizational requirements. The Captiva eInput Web client, fully integrated with Input*Accel*, provides distributed scanning and indexing capabilities through a thin-client, even including features for offline use. In the claims processing example described, the Web client allows agents to capture claims and supplementary content (such as notes or images) from remote or mobile work locations. The captured information can be indexed, packaged as a document, and delivered to back-office systems for further processing. Organizations who have made investments in MFPs can also take advantage of the latest Captiva Input*Accel* connector for eCopy ShareScan, connecting the capture server to hundreds of MFPs via the eCopy solution. These MFPs become much more than basic scan-enabled devices in this context, enabling the automation of remote paper-based processes that would otherwise require manual processing by workers.

With capture support extending to high-volume centralized scanning centers, low-to-mid volume distributed scanning and indexing offices, ad-hoc scanning and indexing, and office peripherals such as MFPs, EMC Captiva's platform allows an organization to scan individual documents or document batches from the most cost-effective locations while retaining centralized visibility and control over the capture process.

Intelligent Document Recognition

Intelligent document recognition (IDR) is essential to the input phase of capture and the overall success of any capture system. The best IDR technologies deliver a unique combination of classification and data extraction techniques that enable organizations to streamline their information flow, thereby reducing costs, increasing productivity, and enhancing the capture of information.

Captiva Dispatcher version 6 continues to support the direction in which organizations are headed, further reducing costs surrounding manual document preparation, classification, and data entry. The release introduces updates to recognition that should provide customers with improvements in document classification and data extraction, further enabling organizations to drive down costs and improve business processes efficiency.

Dispatcher customers will also see new support from Captiva for capturing, classifying, and extracting data from personal and business checks, including reading Courtesy Amount Recognition (CAR) and Legal Amount Recognition (LAR) fields. This new capability will offer financial institutions and other businesses a solution that can automate the capture and processing of checks typically associated with remittance processing.

InfoTrends' Perspective

In light of the persistence of paper documents and the abundance of unstructured documents and information, organizations must provide enterprise-wide capabilities for ingesting structured forms as well as unstructured content. The reality is that simply digitizing and providing centralized access to these documents is not enough. The meaning of documents must be extracted for use in process optimization and automation initiatives. These projects also require deep integration between capture technologies and downstream IT systems such as ECM. EMC's Transactional Content Management application framework is one offering that provides a robust set of applications—addressing information input, management, output, compliance, and process—that can deliver value throughout the entire document lifecycle.

EMC's Captiva product group has recognized that capture technologies play in a much larger ecosystem that includes Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), and many other enterprise technologies. Continued investments in intelligent document recognition and updated development tools for application integration in Input*Accel* 6 are testaments to these needs. The enhanced capabilities provide the classification, information extraction, validation, and routing required to inject these documents into downstream systems and processes. Moreover, uniformly opening up the capture platform to third-party systems will drive the innovation of a stronger partner, developer, and customer community, even where those participants are not leveraging Documentum (to EMC's chagrin). The availability of Captiva's Web services speaks to EMC's commitment to provide a capture platform that delivers solid functionality today without limiting future innovations or flexibility.

Although version 6 enhances the already-strong scalability and distributed processing features of Input*Accel*, this EMC Captiva release has a lot to offer in the areas of stronger process integration, faster deployment, and increased ease-of-use.