



# White Paper



## Distributed Content Capture

Key Factors to Consider for Distributed Content Capture Deployments



<http://www.nsius.com>

Notable Solutions, Inc.  
600 Jefferson Plaza  
Suite 500  
Rockville, MD 20852  
ph: +1.240.683.8400  
fax: +1.240.683.8420  
[www.nsius.com](http://www.nsius.com)

Document capture technologies handle scanned documents, imaging devices, indexing images, and processing of data into back-end applications. Content capture technology, on the other hand, expands the capture spectrum beyond document scanning to include all other types of corporate content.

Traditional document capture technology primarily concentrates on production imaging, desktop scanning, and ultimately dealing with images and processing of images. The new generation of content capture servers expands the production scanning environment and provides a uniform methodology to deal with content in all possible formats including un-structured, semi-structured and structured data.

Furthermore, defining and managing corporate content enables the organizations to deal with content as a whole, uniformly administer content capture business rules, properly manage corporate business content capture, and effectively control cost.

### Distributed Content Capture

The key to effective and efficient business process is to capture business content as early and as accurate as possible. The actual cost and effective cost of processing information increases as a business process progresses. Factors such as information loss, inaccurate data, and improper delays all add to the business cost of completing a task.

### Early capture

The cost of capturing corporate content increases dramatically as the content flows further through the business process. Upfront and early automation of business corporate processes provides additional benefits by:

- Decreasing the rate of content loss
- Increasing security
- Promoting sharing of knowledge
- Faster content flow
- Decreasing content handling cost
- Centralizing the administration of content capture process

### Distributed nature of content

Most corporate content is distributed within multiple organizations. In larger corporate environments, typically there are multiple content sources, users and content management applications that are geographically dispersed. For these organizations, distributed content capture is the most effective solution. For example, invoice processing for accounts payable of a trucking company requires a distributed content capture application. Since each truck delivers goods to various locations, resulting into delivery tickets and signed receipts, an application that allows the driver to capture and send all information related to the delivery to the central office application would best lend itself to this business process.

An organization can use Fax servers, desktop scanners, network scanners and other on-ramp devices to capture documents. Additional content sources commonly include desktop applications capable of generating electronic files, PDA's, printed report files and other applications generating structured or semi-structured data files.

### Key Factors to Consider for Distributed Content Capture Deployments

A distributed content capture application can be represented in three sub-components:

- **Capture** - All modules used to capture corporate data content.
- **Process** – All components used to reformat, extract information, index, verify, and manipulate the captured content.
- **Route** – All components used to integrate captured content into applications.

### NSi products are designed considering the following key factors

*Capture component key factors are:*

- **Diversity** – ability to capture content from a diverse set of content source (fax, desktop, internet, XML, on-ramp devices, etc.).
- **Intelligence** - ability to interact with intelligent devices such as smart copiers, printers, and scanners.
- **Dexterity** – ability to adapt to new technologies and new content sources.
- **Scalability** – ability to grow as the amount of content grows and the rate of content increases.

*Process components key factors are:*

- **Interlink** - ability to link to all back-end application, and interchange information.
- **Feature rich** – providing a rich set of standard processing engines to enable reformatting into all standard formats.
- **Flexibility** – ability to easily adapt to the changing business needs.
- **Platform independence** – enabling the use of multiple engines to better optimize processes.

*Route component key factors are:*

- Basic transport – integration with the most common transport protocols (e.g., TCP/IP, SMTP, FTP, Directory Store, etc.).
- Document Management – ability to support all commonly store-into document management applications.
- Content management – ability to support common content management, workflow, and collaboration features and various applications.
- Customized integration – ability to support custom built application interfaces.

### **When to use centralized content capture**

Some business applications lend themselves better to centralized content capture solutions. For example, an insurance claims processing center where all claims are sent to a central site is better suited for a centralized capture.

Centralized content capture is more efficient where all content is gathered within a central location. In some instances organizations have used the internet to increase efficiency by converting a centralized content capture center into a decentralized and distributed internet-based capture environment. NSi internet based AutoCapture® provides an effective means of using internet-based capture solutions.

### **Integrated capture process**

A successful deployment of distributed content capture requires proper integration of front-end capture components with back-end business applications.

Using an integrated capture technology, the user experience is non-segmented and enables a uniform presentation of back-end corporate content that makes an easier, more accurate front-end capture possible.

The AutoStore® product provides application integration with on-ramp devices such as HP digital senders. AutoCapture® uses HTML forms and XML messaging for an easy and familiar web-based capture solution.

### **Multi-step capture**

When implementing distributed content capture environments, the captured content may need indexing, verification, or quality review prior to storing the content into back-end applications. The indexing, verification, or quality checking steps might be done locally or anywhere within an enterprise. NSi products are designed with these capabilities in mind. Using AutoStore users anywhere within the enterprise can contribute knowledge to the captured content prior to final submission into the enterprise back-end content management application.

### Distributed capture technique

A diverse and distributed capture technology such as NSi family of products provides many alternatives for capture of information in a distributed environment:

1. Shipping paper documents to central office location
2. Off-line scanning and shipping CDs + index data files
3. Using facsimile to send scanned images
4. Using File Transfer Protocol to transfer files over the IP connection to back-end server application
5. Using SMTP to send images and data to SMTP Gateway services for capture and routing into back-end applications
6. Using on-ramp devices with TCP/IP connectivity capabilities to scan documents and index data. These devices are smart enough to present application forms to users and collect index data fields
7. Use integrated scanning, indexing, and routing applications that are capable of using XML and proper transport protocols to integrate scan devices directly with applications.
8. Using XML/SOAP based applications designed using MS .NET framework such as NSi AutoCapture to capture any electronic file into back-end applications.

### NSi Content Capture Architecture

NSi family of content capture products provide all necessary components for an efficient and effective implementation of centralized or distributed capture deployment. NSi enterprise content capture server, AutoStore®, AutoCapture® internet-based content capture application, and Refero<sup>2</sup> Digital Workflow Ticketing (SMARTickets) are designed to provide a full spectrum of features necessary to properly install and deploy enterprise wide content capture solutions.

### Enterprise Content Capture Solution

Expanding organizational needs into enterprise content capture requires special capabilities and features. An enterprise content capture solution must not only be scalable but also be capable of providing more complex features for broader range of enterprise level requirements. This section highlights some of the features for the more complex enterprise deployments.

### Scalable Content Capture Architecture

The content capture solution should be able to scale up and accommodate for additional local capture stations, incremental review/indexing stations, and additional processing servers. Additionally the server architecture should be to perform distributed content capture, and distributed content processing. Load distribution and messaging to properly handle load transfer with customizable queuing algorithms are required for larger size deployments. All server components should be able to transfer data and files into other servers to distribute the load and take advantage of unused resources.

### Central Administration

Ability to centrally control capture stations, user permissions, indexing forms, validation rules, and business process flows are key elements within an enterprise deployment. New York City Human resources has deployed AutoStore servers with over 700 network scanning devices and using AutoStore administrative components has been able to successfully synchronize all device attributes with internal applications, thereby dramatically reducing administrative resource requirements.

### Multi-Process Support

Multi-Organizational content capture requires multi-process support. Requirements for the multi-process support goes beyond supporting multiple threads within a process. It should include the ability of the components within a process to:

- Encapsulate files and data fields and properly distribute information to various servers.
- Ability to use an open and define interface to link to other applications
- Ability to exchange information and files with external applications at any point within a process
- Ability to create multiple chained processes to enable the content capture to flow-thru
- Create an easy to use, easy to administer interface to create, modify, and maintain all interfaces

## Workflow automation

Create an ability to make content capture workflow decisions based on data provided by user, devices, or extracted from content. Capture workflow design interface must both be flexible and intuitive to use. Enterprise deployment often require programmable interface design to allow for customizable workflow modules.

## Desktop Capture Support

Desktop scanning workstations software provide a powerful method for capturing various type of paper documents. For enterprise deployment also look beyond just scanning capabilities on your desktop. Key elements to consider on desktop are:

- Ability to poll and capture for any type of files not just scanned images.
- Flexible form design capabilities.
- Back-end application integration for validation and indexing assistance.
- User permission control to limit access to various features or workflows
- Modularized design to allow cost control in license distribution
- Desktop processing capabilities
- Central management desktop capture features and capabilities
- Management of usage rates and submission volumes
- Flexible programmable filters

A true enterprise deployable desktop content capture application must provide a flexible and expandable architecture through full support for industry standards such as XML, SOAP, Web Services, program-mable VB components and an open API.

## Network Scanner Support

Network scanners are leading the imaging and printing market. Virtually all manufacturers have been quick in manufacturing and deploying intelligent network imaging devices. The speed of acceptance of these devices into IT corporate marketplace has been tremendous and they are fast becoming an integrated part of enterprise device deployment. Your content capture solution must seamlessly integrate with these devices and take advantage of their unique capabilities. No matter what content capture applications you are planning to deploy with your network scanning devices, make sure you can:

1. Establish two way communications between your content capture servers and the devices.
2. Reduce, not increase, your administrative tasks required for maintaining these devices
3. Take advantage of all indexing and scanning features available within these devices
4. When dealing with MFPs, use printing as well as scanning sub-devices to capture content such as printed pcl report files, etc.
5. Manage these devices centrally

## Internet Capture Interface

A large number of companies have successfully used internet to connect information sources together. An internet capture application enables an organization to capture content from anywhere within its enterprise. Your content capture interface must include a no-footprint or small footprint internet-based content capture component. Make sure you have central control over user permissions, forms, and capabilities. Use industry standard Internet Web servers and allow for any number of deployment around the enterprise. Capability to centrally control the content capture processes is the key to expanding content capture to all internet enabled desktops.

## Device to Application Synchronization for Enterprise Deployment

When deployed, network devices are smart enough to present the user with flexible forms. It is essential to deploy applications that are capable of synchronizing and presenting application level data elements on these network devices. Such synchronizations allow users to perform more accurate and reliable content capture and indexing tasks thereby increasing efficiency and reducing cost.

## Indexing and verification workstation

When content capture is not performed by a knowledge worker, often the business process requires a validation or indexing station to be introduced within the automated business workflow to insure proper content is captured. Look for the following features within your content capture technology:

1. Flexible index form design toolkits
2. Ability to integrate forms with back-end applications
3. Ability to customize the queuing, routing, and workflow of jobs by each index station user
4. Open API for customization of indexing capabilities.
5. Indexing station inter-connectability and support for chaining indexing processes is also a required capability.

## Flexible Design

All content capture components must be build on open platform supporting most recent open interfaces including the latest in programming languages, messaging formats, and interface access protocols. Custom built workflow components allow the indexing station to be designed into your unique business application requirements.

## Large Partner Network for Support and Services

Above all features, are a reliable and qualified network of partners to provide local level support and services. An enterprise deployment would require an enterprise level of support. Enterprise content capture is the backbone of your corporate content. A worldwide organization like Hewlett Packard, Canon, Xerox, Sharp, Kyocera and Ikon can provide the right level of service and support at any location worldwide.

## Summary

Scalable enterprise deployment of content capture technology requires certain unique features. Table below provides a list of these features and cross section to NSi content capture family of products:

| Component/Feature                                      | AutoStore | AutoCapture<br>Internet<br>Capture<br>Application |
|--|-----------|---|
| Scalable architecture                                  | ✓         | ✓   |
| Multi Process  | ✓         |   |
| Indexing and Queuing                                   | ✓         | ✓   |
| Internet Capture                                       | ✓         | ✓   |
| Security (secured capture, Filtering, etc.)            | ✓         | ✓   |
| Flexible workflow                                      | ✓         | ✓   |
| Easy to Administer                                     | ✓         | ✓   |
| Supporting Open Programming Interface                  | ✓         | ✓   |
| Modular design and licensing                           | ✓         |   |
| Central Administration                                 | ✓         | ✓   |
| Network device support                                 | ✓         |   |
| Open Transport Gateway support (SMTP, POP3, FTP, etc.) | ✓         |   |
| A large network of partners for support and services   | ✓         | ✓   |

Table 1 - Summary of Enterprise Content Capture Capabilities



**Notable Solutions, Inc.**

---

**Corporate Headquarters**

600 Jefferson Plaza; Suite 500

Rockville, Md 20852

ph: +1.240.683.8400

Fax: +1.240.683.8420

---

**Western U.S / Canada**

ph: +1.415.333.3615

fax: +1.240.683.8420

---

**European Office**

Gebr.-Wahl-Str. 21

35619 Braunfels, Germany

ph: +49 6442 962 349

fax: +49 6442 962 359

<http://www.nsius.com>

Powered By **AutoStore™**