

# **Distributed Document Imaging: Maximizing Your Investment in Microsoft® Technology**

**Integration with Windows® SharePoint® Services (WSS)  
and Microsoft Office SharePoint Server (MOSS)**

## Contents

Introduction .....	1
Challenges of SharePoint integration .....	1
Distributed Document Imaging in Microsoft IT Infrastructures... ..	2
Flexible implementations.....	2
Real-time integration with SharePoint .....	3
Records management compliance.....	3
Support for SharePoint workflows .....	3
Zero footprint on the SharePoint server... ..	4
Hosted (internet-based) SharePoint servers.....	4
Summary.....	4

## Introduction

Selecting a distributed imaging solution that supports Microsoft SharePoint enables you to store paper-based content with your Microsoft Office and other electronic documents in your SharePoint repository.

This offers several major benefits to organizations already using SharePoint for content management and collaboration:

- Paper-based documents are available immediately to all those with the authorization to view them
- You can implement appropriate security controls to ensure that confidential information is only available to those with the authorization and need to access it
- You can safeguard paper documents by backing them up to an offsite data storage facility
- You can establish and enforce lifecycle policies to eliminate outdated or redundant information from potential discovery during litigation

This eCopy technology brief is one in a series of four briefs that examine the requirements for successfully integrating a distributed document imaging application into your existing Microsoft-focused IT environment. Other technology briefs in this series include:

- Active Directory
- Exchange
- SQL Server and Access

The other three technology briefs in this series can be downloaded from our Web site.

## Challenges of SharePoint integration

The dynamic nature of the SharePoint environment and the ease with which users and administrators can modify the environment present many challenges:

- As users and administrators create new sites and modify existing sites, the client application must adapt in real time to changes in the underlying site structure
- The user's security profile must determine the level of access granted to the various sites, libraries, and folders
- As new content types and metadata (columns) are defined, the client user interface must reflect these changes
- Since metadata is critical to site-wide searching and rules-based workflows, the imaging client must ensure that all documents are stored with the required metadata

## Distributed Document Imaging in Microsoft IT Infrastructures

Distributed document imaging solutions enable knowledge workers to convert paper documents into electronic files.

As a result, these solutions offer significant benefits, including:

- Making paper-based information available throughout the organization
- Speeding up the processing of paper documents while simultaneously reducing the associated costs
- Enabling administrators to apply policies for compliance with records management and security regulations
- Safeguarding paper documents through electronic backup to offsite facilities

To achieve these benefits, the imaging application must be easy to use and must integrate with the applications people already use on a daily basis for communication, collaboration, and document storage. Users must be able to walk up to any scanning device and store, distribute, and share paper documents the same way they handle electronic files at their desktop – by browsing the network, storing files to pre-configured locations, selecting recipients from address lists, and indexing documents for quick retrieval.

For organizations with Microsoft-focused IT infrastructures, this means integrating with the Microsoft technologies and applications already in place, such as Active Directory®, Exchange, SharePoint®, and SQL Server®-based business management applications.

Dynamic integration with back-end servers (domain controllers, Exchange servers, SharePoint servers, etc.) through programmed interfaces ensures that the user interface reflects the latest changes to the underlying applications, directories, and site structures. It also eliminates the need for preconfigured scanning cover sheets that some imaging solutions require. The application interfaces must be sophisticated enough to handle the infinite variety of complex network environments involving multiple domains, multiple forests, outsourced IT management, and internet-hosted services.

## Flexible implementations

For pre-defined workflow scanning, administrators must be able to configure profiles for standard document types, where the destination is fixed, the content type is pre-selected, and metadata is entered with a minimum of keystrokes through use of pre-populated fields and lists of valid options.

Based on the business process requirements and the organization's security environment, the administrator can either bypass the authentication step by including service account credentials in the workflow profile, or can enable runtime authentication to ensure that each document can be traced back to the individual that submitted it.

For ad hoc scanning, where the workflow is not pre-defined, administrators must be able to specify just the site name and allow users to navigate through the site based on their access rights, as they would when uploading documents through Internet Explorer. To simplify navigation, the imaging client must display only those web parts that support document storage or link to other areas of the site. Authentication is essential, since the

user's security profile determines the regions of the site the user can access and the type of access granted (Full Control, Design, Contribute, Read).

## Real-time integration with SharePoint

An ideal imaging solution is flexible enough to support workflow-oriented scanning and ad hoc scanning.

To achieve this level of flexibility, the solution must be tightly integrated with the SharePoint environment, mirroring to the greatest extent possible the client environment available through the SharePoint web browser interface. Document imaging solutions that rely on pre-configured scan templates or intermediary software agents are unable to do this.

Real-time integration is essential due to the dynamic nature of the SharePoint environment. Authentication must determine the user's level of access, and the interface must reflect the latest changes to the site. Additionally, the imaging application must be able to indicate successful delivery by displaying a confirmation message when the document is saved to the specified location.

## Records management compliance

Retaining documents beyond the required retention period creates many potential problems, including increased costs of storage and maintenance, possible exposure to litigation based on outdated documents, and the illegality of holding personal information longer than necessary.

For imaging solutions that support only simple scan-to-file, this presents a problem, since it is virtually impossible to effectively manage content that is scattered randomly across the network. By integrating directly with your organization's content management system, you can effectively leverage existing controls and information lifecycle management systems, thereby mitigating risk of information exposure and discovery liability. Additionally, storing documents with intelligent metadata reduces search-to-deliver cycles during the document discovery process.

## Support for SharePoint workflows

The Windows Workflow Foundation, now integrated into WSS and MOSS, provides powerful tools enabling site administrators to configure workflows to implement business processes.

As new documents arrive on the server, the workflow agent can initiate a process to automatically notify the appropriate individuals by e-mail or through their personal Task List.

Rules-based workflows typically rely upon content types and metadata to determine subsequent actions. For example, when handling expense reports, the report may be routed based on the individual's department and manager. If expenses exceed a preset amount, additional approvals may be required. This type of rules-based processing relies upon accurate indexing that must be enforced at the scanning device when the document is submitted.

## Zero footprint on the SharePoint server

Most IT organizations that support business-critical SharePoint infrastructures do not allow 3rd-party software applications to install modules on the SharePoint server. To do so would be highly risky, since a failure within the 3rd-party module could potentially bring down the entire SharePoint environment.

It is critical that any distributed imaging solution communicate with the server only through the SharePoint Web Services interface, thereby maintaining zero footprint on the server. The network administrator must ensure the application can send HTTP packets to the SharePoint server and is not prevented from doing so by firewalls or proxy servers, but if the application uses standard ports this is typically not an issue.

## Hosted (internet-based) SharePoint servers

The “zero footprint” requirement has a secondary benefit, since it opens up the possibility of working with hosted SharePoint services.

Establishing and supporting a SharePoint infrastructure is impractical and cost-prohibitive for many small and mid-size businesses. One solution is to outsource the SharePoint infrastructure to a hosting service. Even large organizations can benefit from outsourcing Windows SharePoint Services in this way, and outsourcing is predicted to become a significant trend in the years ahead.

Working with an externally-hosted SharePoint service is highly desirable, but presents challenges:

- It is essential that the solution does not require any modules running on the SharePoint server.
- Support for SSL and HTTPS is required to ensure the security of data transferred over the internet.

## Summary

Many Microsoft-focused organizations already use SharePoint for managing electronic content and supporting collaboration. A distributed imaging solution makes it possible to include paper-based information in the same manner.

Such a solution must present a simple navigational interface at the scanning device, while simultaneously providing dynamic integration with the SharePoint server.

Considerations when selecting a distributed imaging solution:

- Does the scanning solution provide real-time integration with the SharePoint server, reflecting changes to the underlying site structure automatically?
- Does it apply authentication, access rights, and security policies with no administrative overhead?
- Is it flexible enough to simultaneously support workflow-oriented scanning and ad hoc scanning?
- Does it support SharePoint workflows by enforcing metadata requirements for rules-based processing?
- Does the solution require any components or modules to reside on the SharePoint server?
- Can it work with hosted SharePoint environments?

The experience speaks for itself™