

Imaging with SharePoint and Microsoft Office

Microsoft Office products have out-of-the-box integration for SharePoint there are still some missing parts for a true Document Imaging solution

A KnowledgeLake, Inc. White Paper

Published: December 2007



Introduction

Since the release of SharePoint Portal Server 2001, users have realized that by using the storage capabilities of SharePoint Products and Technologies, documents can be stored, archived and used for collaboration. SharePoint was also a great way for users to get their feet wet with ECM since it was being delivered as part of the platform their organization already had in place or would likely implement in the near future.

In late 2003, Microsoft released Windows SharePoint Services 2.0 and SharePoint Portal Server 2003. These products dropped the Web Storage System, which is used by Exchange, and changed to a SQL Server platform. The new software had many basic Document Imaging features such as Meta data storage, custom views, check-in/check-out, version history and in the case of SharePoint Portal Server, full text OCR indexing. This instantly made SharePoint a scalable, enterprise capable platform. This also instantly gave Microsoft, when combined with Microsoft Office Document Imaging (MODI), a story in the Document Imaging marketplace. Microsoft users almost immediately started saving documents, images, forms and anything else they could find into SharePoint Document Libraries.

In early 2007 Microsoft publically released the entire Office 2007 line, which included a rewrite of most SharePoint components. The new architecture offered better scalability and the base components for a true enterprise ECM system. The 2007 platform added many new features to improve Document Management functionality like workflow, major/minor versioning, indices for columns and content types. In turn, the new architecture gave the base platform even more reasons to be used as an Imaging System. The new packaging is similar to the previous in that Windows SharePoint Services comes free with Windows Server and the upgrade for many more capabilities is to the Microsoft Office SharePoint Server.

However, even though most Microsoft Office products have out-of-the-box integration for SharePoint there are still some missing parts for a true Document Imaging solution. What Microsoft left out of the product is covered by a larger and stronger ISV community than any other vendor.

The Summary of the Parts

Successful Document Imaging solutions require many components working together to accomplish a goal. Whether you're creating electronic medical records, using workflow with insurance applications or scanning lunch receipts there are a basic set of document imaging components typically required. These components are typically defined as the ability to capture, store, preserve and deliver documents to end users. What follows is an overview of the necessities and how Microsoft, developers and the ISV community, along with KnowledgeLake resolve the problems using SharePoint and Microsoft Office.

Capture

Low Volume (possibly remote) Image Capture

In the last several versions, Microsoft Office has included an application called Microsoft Office Document Imaging (MODI). With MODI you can scan paper documents and convert them to TIFF images and save them on your computer's hard disk or a network location. Since SharePoint Document

Libraries can be viewed through windows explorer it is possible to use MODI to scan directly into SharePoint.

Although using MODI sounds simple enough, there are some major caveats. For instance, MODI can OCR a TIFF file, but Microsoft removed the ability to full-text search TIFF files in SharePoint 2007. Another problem is the process of adding the column properties. Document Imaging depends on Meta data for searching and although the properties can be added later, it's hardly efficient. SharePoint also has a large set of APIs available for saving documents and Meta data to SharePoint. Included are web services, HTTP posting, WebDav and a direct API that can only be executed on the servers. These allow any developer to build a scanning solution specifically for SharePoint.

Although many ISVs offer Capture solutions that communicate with SharePoint, KnowledgeLake is the only company dedicated to offering a complete capture solution for the platform. KnowledgeLake has three different but powerful methods for low volume and/or remote capture. The first and the method I use on a daily basis is the combination of a Fujitsu ScanSnap combined with KnowledgeLake Connect, the second is KnowledgeLake Branch Capture Server and the third is a web-based solution built directly into the SharePoint user interface.

High Volume Batch Image Capture

Microsoft does not provide any type of solution that can be used for high volume capture. Like many repository vendors, Microsoft depends on third party vendors such as KnowledgeLake to provide this type of solution.

Store/Preserve

SharePoint has always had the ability to store documents. 2003 added the ability to easily version and store custom Meta data. With 2007, Microsoft added better organization using Content Types as well as major /minor versioning. Although there have been many articles written denying it, SharePoint can store millions of document in a single document library. KnowledgeLake has worked a number of these implementations and one specifically that has over 10 million images in a single library. Microsoft Office SharePoint Server also offers Records Repositories, which are aimed at permanently storing and preserving the document.

Many third party vendors allow users to move data into their databases and repositories, but if you don't already own one of these proprietary, expensive systems you should seriously evaluate whether it's necessary.

Search

Full Text Searching

Microsoft is one of the major Search vendors and does a great job of providing "high relevance" search solutions. SharePoint combines full text data from the stored document along with Meta data to make the document available through what has become known as "Goggle Style" searching from one single text box.

Meta Data Searching

Microsoft handles Meta data “exact relevance” searching by allowing data to be stored in the columns of content types and document libraries. Although there is no great out-of-the-box interface for using this data, it is very easy to construct queries that developers can use to build applications. KnowledgeLake also offers a web part that works inside SharePoint to make Meta data searching very simple.

Delivery

Delivering documents to end users typically depends on both routing and viewing of these documents.

Workflow

Microsoft added a key component for Document Imaging with SharePoint 2007 by adding workflow to both Windows SharePoint Services and Microsoft Office SharePoint Server. There are many third party workflow vendors in the market today, many which specifically address SharePoint. You should however try SharePoint workflow first to see if it meets your needs.

Viewing/Editing

Microsoft handles the viewing and editing of documents as good as anyone, but although they provide TIFF Image Viewing with the MODI application mentioned earlier, any type of edits, including annotations will require a manual step to upload the TIFF back to SharePoint. The other downside is that users must wait for the entire TIFF file to download before seeing any pages.

PDF viewing can almost be considered a standard with dozens of free PDF readers available. Many other ISVs are building viewers that can view any type of document or image. KnowledgeLake does provide the ability, however, to view PDF and TIFF images with no client install necessary, directly inside the browser and without waiting for every page to download.

Conclusion

Both Microsoft Windows SharePoint Services and Microsoft Office SharePoint Server offer a platform capable of Document Imaging. In the case of SharePoint, it’s possible this platform is already in place within your organization. Like other platforms however, third party ISVs or developers may be required to meet 100% of you needs.

CORPORATE OFFICE:

KnowledgeLake, Inc.
3 City Place Drive, Suite 700
Saint Louis, Missouri 63141
Phone: 314.898.0500
Fax: 314.898.0501
Toll Free: 888.898.0555
www.knowledgelake.com
info@knowledgelake.com

This White Paper is for informational purposes only. KNOWLEDGELAKE MAKES NO WARRANTIES, EXPRESS OR IMPLIED, AS TO THE INFORMATION IN THIS DOCUMENT.

© 2007 KnowledgeLake, Inc.. All rights reserved

The names of actual companies and products mentioned herein may be the trademarks of their respective owners.