

Helpful Tools

Vendor Questionnaire

**To Assist Your
Organization's Evaluation
of Document Imaging
Systems**

**By: Randy Van Ittersum and Erin Spalding CDIA+
www.disusa.com**

Vendor Questionnaire

to Assist Your Organization's Evaluation of Document Imaging Systems

The following identifies critical questions that you should ask any vendor and provides you with a brief explanation of each question's importance. It does not include every question that you should ask; rather, it focuses on the pertinent topics that every customer should know.

Presented by:

Document Imaging Solutions, Inc.
P.O. Box 517
Spring Lake, Michigan 49456

Phone: 616-847-5055
www.disusa.com

Document Imaging Solutions, Inc.

Copyright © 2005, Document Imaging Solutions, Inc., All Rights Reserved, Duplication Prohibited

The information in this document is subject to change without notice. No part of this document may be reproduced, stored or transmitted electronically or mechanically, for any purpose, without the expressed written permission of Document Imaging Solutions, Inc.

1. What is the price of your document imaging system software?

This quote should include the price of the server license, database license, scanning station license and any workstations that are connected to the network.

2. What are the scanner features and price?

Scanners vary by make and model, as well as by the speed of the automatic document feeder. Does the scanner include a flat bed to scan torn or crumpled documents? At how many pages-per-minute does it feed the documents? What size of documents can it scan? Does it scan single-sided documents (simplex) or double-sided documents (duplex)? Is it a black-and-white or a color scanner?

3. What is the manufacturer's warranty on the scanner?

Does the warranty cover on-site repairs, or are you required to send it into a repair depot? What is the length of the warranty?

4. What is the cost of an extended maintenance agreement on the scanner?

Extended maintenance agreements can range from a technician coming on-site to repair the scanner to having to ship the scanner to a service depot.

5. What is the cost of extended maintenance agreements on the software?

Maintenance agreements normally cover software upgrades and support.

6. How long does it take the average user to become proficient in the operation of your system?

The vendor should demonstrate all of the steps that an operator must know in order to operate and maintain the document imaging system.

7. In which format do you primarily save your scanned images: PDF or TIFF?

PDF

An advantage of the PDF file format is that anyone can view the document using Adobe Acrobat® Reader, which is free software that can be downloaded from the Internet. This feature makes it very easy to share documents within or outside of an organization because everyone can view the document at no additional cost.

The PDF file format has become the de facto file format for the Internet. This is due to its small file size and the ability to begin viewing the document as soon as the first page downloads (other file formats require you to download the entire file before you can begin viewing the document).

A significant advantage of the PDF file format is that you can add document security at document level. Using Adobe Acrobat® and other third party plug-ins allows you to add file security to prevent unauthorized viewing of the document.

In addition, you cannot alter a PDF document without leaving an electronic footprint. This is a major reason why the U.S. Federal Judiciary has mandated its use for electronic filing of case documents. It also gives an organization the freedom to keep the scanned PDF images on a server to be shared throughout the organization, without the fear of document manipulation.

TIFF

Many document imaging companies that use the TIFF file format use proprietary software to view the TIFF images because the TIFF file format can utilize a variety of compression ratios. This can create a problem for your organization if you were to migrate your imaging system to a different vendor in the future. Your new vendor must have the ability to open and view your archived TIFF documents, which is frequently impossible.

Because of the requirement for proprietary viewing software, it may be difficult to share TIFF files with people outside of your organization. It is important to find out the cost of providing viewing software to outside parties that will be accessing and viewing your documents.

Since a TIFF image can be altered without leaving an electronic footprint, it will require you to save the document to an optical disk or a WORM drive as a read-only file. These devices restrict the ability to alter the document.

8. Does your system use a database or an index to search for documents?

Database

Many document imaging systems use a database to search for documents. With this configuration, the relationship between the database and the file structure must be maintained, or you will not be able to retrieve and view your documents. Due to the technical expertise required to manage a database, systems of this nature are usually installed in organizations that have access to an IT staff. Most database systems only establish one database per location in an organization. The advantage of a database is that an organization can access the data within the database and integrate the information into another application. This usually requires programming knowledge on how to customize the program that you wish to integrate with the document imaging system.

There are significant disadvantages to a database system. They are expensive to maintain. They not only require a high level of IT support, but the routine maintenance requires more time than an index. Furthermore, if the document imaging software vendor no longer supports its software, the cost of performing a database-to-database conversion can be extremely expensive (prices can range from \$50,000 to \$10,000,000 to migrate from one database to another).

Indexes

The alternative is an indexing system. A major advantage of an indexing system is that it is dramatically easier to maintain and does not require the level of technical expertise required by a database system. Many vendors provide multiple indexes that can be department-specific to search for documents. Rather than everyone searching a single index as you would search a single database, an organization would set up a different index for each department. Your permission for access to an index for searching is determined by your network administrator.

Indexes are less expensive to initiate and maintain than databases. Many indexes are "closed," which prevents someone from accessing the data stored in the index. This provides a high level of security.

You should consider a document imaging system that uses an index if you do not have access to IT staff, or if your IT department does not have the time or resources to maintain another system.

9. Do your search words embed into the scanned image or do they reside in an index or database which only hyperlinks to the scanned imaging?

Embedded Search Words

Embedding relevant search words into the document is the safest method to label documents. In effect, wherever the document resides, the search criteria to find the documents resides with it; i.e., the documents become the database. Documents embedded with their search words are easily transferred between locations, such as from the server to a laptop computer or a Tablet PC without losing your ability to search for them.

Another advantage of embedding the label is that the documents are web-enabled and can be searched and retrieved from an Internet web site. This feature makes sharing documents over the Internet inexpensive compared to a database system.

Disaster recovery is a simple process when documents are embedded with the search words. One simply copies the documents from the backup file onto the server and builds a new index. Recovery can take place in minutes, instead of days or weeks as it would with a database system.

Probably the most important aspect of embedding the search words into the document is that it protects your investment into a document imaging system. If the company no longer supports their software, hundreds of search engines such as Microsoft, Adobe and Google can retrieve these documents. In a database system that only links the document to the database, you would incur the costly process of doing a database-to-database conversion.

Non-embedded Search Words

Most document imaging systems do not embed search words into their documents. Their process saves the search word into a database, which in turn hyperlinks to the location of the document. It is imperative that the relationship of the file structure and the database be maintained; otherwise one will be unable to find the documents. Essentially, the database only “knows” the location of a document; if that document is moved from the specified location, the database will be unable to locate it.

These systems allow you to easily retrieve documents but you sacrifice the portability features of an embedded label in the document. If you move the document from one file folder to another, you can no longer retrieve the document using the search engine. You would have to manually open each document to determine what it is.

Disaster recovery is more difficult a process because you must maintain not only the file structure but the directory structure as well. For example, if the documents resided on the “E” directory under the old system, it must also reside on the “E” directory of the new system. This sound easy, but when you have documents ranging over multiple directories, it can be a significant task to recover a document imaging system of this type.

10. What is the training time for the operator and IT personnel to be fully proficient using the document imaging system?

Hidden Costs

The hidden costs of system implementation can be substantial, and are often the culprit behind delays and failures of adoption. These costs can range from training time to integration woes. Tangential costs include post-deployment training, integration, and the repurposing, firing, and hiring of personnel associated with the system. The cost of training is an ongoing responsibility; there is the training you need long after deployment for new people or those who change positions. What effect will it have on your document imaging system if you lose your IT personnel and have to teach someone else the system? What about ongoing training and technical support?

11. Does the system integrate into your current network?

Does the new system require your network administrator to maintain an additional set of passwords and administer the system in addition to the organization’s network? This would be required in a database environment where you have a single database sharing all the organization’s documents.

The problem with this setup is that it increases the amount of time required by your IT staff to maintain your document imaging system. The advantage of integrating into your existing network is that the network administrator only needs to maintain one system. Passwords and permissions are based upon your organization’s network protocol and does not require any additional time by your IT staff.

12. Does your system OCR the documents and search off of the OCR text?

Full Document OCR

The dirty little secret of document imaging – what the vendors will not tell you – is that the OCR process is more time-consuming and produces less accurate searches than if you label a document. They lead a prospect to believe that by processing the document through the OCR engine that it will not be necessary to manually intervene in the document imaging process. This is a wonderful concept – press a button and everything will be done for you – but it simply is not true.

The problem lies in the fact that no OCR program is 100% accurate. If a company's OCR process is 95% accurate, then 1 out of 20 characters is wrong. The result is that an operator must go through the document and manually correct the suspect characters. Since this process is extremely time-consuming and takes far longer than labeling the document, some vendors propose that you do a Fuzzy search for your documents and not manually correct the document. Unfortunately, this merely postpones the process, because a Fuzzy search will find many irrelevant documents for which the end-user is not looking. The burden is now on the end-user to sort through the search results to find the documents that they need. As long as you have only a few documents, this may not be a problem; however, if you have one million documents and a search result produced 1,000 documents to view, the problem is enormous.

Zone OCR

In order to restrict the area of a document that is recognized by the OCR process, a document imaging system may permit you to set up templates that OCR a “zone” of the document that is defined by the template. Although the process is still not 100% accurate, the amount of effort needed to correct the OCR zoned documents is significantly less than if the system were to OCR the entire document. This process also increases the accuracy of the search because it is limiting the search to just the OCR zones.

This process is excellent for structured documents such as company invoices where the information resides in the same place from one document to the next. However, if you are dealing with unstructured documents – where the relevant information is located in different places from one document to the next – this process becomes extremely awkward. It requires you to set up multiple templates that must be changed every time a new document is introduced to the system.

Freehand Zone OCR

This process can be used on both structured and unstructured documents. It allows you to determine “on-the-fly” which areas of the document you wish to OCR. Again, because the OCR process is not 100% accurate, it may be necessary to correct inaccurate results. Searches tend to be quite accurate because the OCR process captures only a specified amount of relevant information. The advantage of free-hand zone OCR is that it works amazingly well with unstructured documents.