

White Paper

Document Imaging's Dirty Little Secret

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

Database Systems: A Complex Problem

When pitching a document imaging system, a salesperson often presents a database as if it were problem free and easy to maintain. Nothing could be further from the truth. Document imaging databases can be your worst nightmare if you can't afford the support required to maintain them.



Document imaging databases are not the typical databases that most IT people are accustomed to. They are more complex and costly to maintain because they have an added element: a hyperlink to an electronic image located on the server. Furthermore, they are dynamic, with office staff continually opening and closing documents, as well as regularly adding documents to the system. And, if that were not enough, document imaging databases get extremely large, very quickly. These three issues; hyper-linking to an image, the dynamic nature of a document imaging database, and its size, makes a document imaging database time-consuming and expensive to maintain.

The secret that nobody wants to talk about is the magnitude of the hidden costs that are unavoidable with a document imaging database.

Why Not Just Reinstall the Last Backup?

If asked how to repair a database, most IT people will instruct you to reinstall your last good backup and then reenter the lost data from that date forward. The problem is that you cannot

repair a document imaging database using this method because of the dynamic nature of document management. It is impossible to identify those documents that were added into the system during the lost time period. The problem may be further compounded if the paper documents have been destroyed.



Consider a typical scenario: you discover on December 31st that you can't find and retrieve a set of documents that you know were scanned into the system. After further investigation you discover the scanning was done on September 1st. Now the task at hand is to find the missing documents and repair the database?

- You can't reinstall a backup of the system prior to September 1st because you will lose all the documents that were added into the system from September 1st to December 31st. The system must be repaired manually.

This problem repeats itself across the business community daily. The question is how much will it cost to fix the problem? To find the answer, you should ask these probing questions to assess the potential impact on your business. Note: these are questions that should be asked before you purchase a document imaging system, not after you encounter problems.

1. Who will do the manual repair of the database?
2. How much time will it take to make the repairs?
3. How much will it cost to make the repairs?
4. Do you have the expertise in-house to solve the problem?
5. If so, do they have the time to devote to this project and what other projects are delayed during the time they are repairing the database?
6. If not, who will you hire to solve the problem?

How Large is the Problem?

The unfortunate truth is that there are always hidden problems with a document imaging database. You simply don't know that they exist until you are unable to recover a document. When a document imaging system is newly installed, the number of lost documents is so insignificant that the probability that you will discover a problem is slight. But over time, the volume of lost documents will increase until the problem finally becomes noticeable. How soon this happens is dependent upon several factors, such as:

- How many documents are added to the system?
- How many people are accessing the documents?
- How frequently are documents being accessed?

The following table projects a document imaging system that is growing by 10,000 documents a month or approximately 500 documents a day. It illustrates how a small problem can grow into a large problem with a document imaging database whose system is corrupt only 1/10th of 1% of the time:

Year	Total Number of Documents	# of Doc. Lost at 1/10 th of 1%	% of Documents Lost
1	120,000	780	.7%
2	240,000	3,000	1.3%
3	360,000	6,660	1.9%
4	480,000	11,760	2.5%
5	600,000	18,300	3.1%
6	720,000	26,280	3.7%
7	840,000	35,700	4.3%
8	960,000	46,560	4.9%
9	1,080,000	58,860	5.5%
10	1,200,000	72,600	6.1%

Note how the problem of lost documents grows over time. The cost to repair the database can be calculated if we estimate three variables:

1. How many documents will be added to the system monthly?
2. At what rate do the hyperlinks no longer work in your database?
3. When is the problem discovered?



Example: using the table above, let's assume that the problem was discovered after 5 years. Also, assume that you hire an outside database consultant at \$250 an hour who can repair each problem in 60 seconds.

- At a $1/10^{\text{th}}$ of 1% corruption rate it will require 305 hours = \$76,250

The cost to repair a corrupt database is one of the hidden costs that most document imaging companies don't disclose to you. The cost, however, is part of the total cost of ownership and needs to be taken into account in your evaluation of any document imaging system.

Unfortunately, a salesperson will seldom give you a cost estimate upon which to base your cost analysis. Most salespeople know less about document imaging databases than you do. The best way to approach a vendor on uncovering this hidden cost is to negotiate a 5 year contract for them to fix any database problems with their system and use the quoted price to calculate the total cost of ownership.

DIS-Imaging Solves the Problem

As we have pointed out, the cost to repair a corrupt database is the Achilles heel to a database system. At Document Imaging Solutions we have developed a cost effective solution.

Our system creates metadata fields and embeds the key search words into the PDF document. By doing so, we have created a very stable environment where, in effect, our documents become the mechanism to fix a database. This is accomplished by exporting our metadata into a database and creating a new hyperlink to the document.



By scheduling a complete rebuild of a document imaging database using our system, one fully restores the integrity of the system. The database is automatically repaired without the need for manual intervention. This makes it easy for any IT person lacking the special database skills to easily maintain a document imaging database.

Using an Indexing Search Engine Avoids the Problem

Databases are designed to hold data; they were never designed to link to an object such as a word document, an image, etc. As technology has advanced, a better solution to find computer objects on the server has evolved. It is called an Index. An index will capture the information contained in the file name, the body of text, or the metadata, and allow you to search for the document using the captured data.

The advantage of indexes over databases is that they are easily maintained by someone who possesses just basic computer skills. In fact they are so easy to maintain that some companies like Google, Microsoft, Yahoo and Copernic are now providing desktop versions for free.

Unfortunately, an index will also become corrupt over time. But, unlike a database, an index is easily fixed. You simply rebuild it periodically. In the worst case scenario, you can delete an

index and create a new one in its place, a task that can't be accomplished using a database.

As previously stated, our software embeds the key search words into the metadata in the PDF document. Although our system can be used with a database, it has a built-in index, designed to make our system easy to use and maintain.

There is Still Another Dirty Little Secret

Historically, software companies have decided to discontinue supporting their software, which forces its customers to make expensive upgrades or to change vendors. If you are using a database system and this happens to you; what problems will you face in making a smooth transition to the new vendor? You will encounter two serious problems that document imaging companies don't want you to know about.



1. When you change vendors using a database system, you must go through a database-to-database conversion process. Again, the issue you face is one of cost. The costs can be as low as \$10,000 to as high as \$10,000,000. Due to the range of cost, this issue should be addressed before you purchase a system, not afterward.
2. A number of companies use a proprietary compression ratio on the images used with their systems. This allows them to charge a per-seat fee to retrieve and view the images. In these cases you can be charged a fee to move the documents from one system to another system. Again, this issue is one that should be addressed before you purchase a system.

With our document imaging system we use the PDF file format, which is an open sources (non-proprietary) format. Therefore, you avoid any hidden fees to move documents from our system to

that of a different vendor. You also avoid any database-to-database conversion charges if you change vendors, because virtually every indexing system on the market can be used to retrieve documents created with our software.

We would recommend to anyone who is considering purchasing a database system that they negotiate the cost to move from one document imaging system to another, before the time of purchase. It can be disastrous to wait until you want to move your system to negotiate the price. And, when evaluating a document imaging system, this should be included in the total cost of ownership.

Conclusion

We have revealed that there are hidden costs associated with a database system. You can ill afford to ignore these costs. To uncover them you will need to ask a number of probing questions, such as:

- Who will be responsible for repairing a corrupt database?
- How much will it cost?
- Will you use in-house personnel or hire someone from outside your organization?
- How will this impact business while the system is under repair?
- What is the cost to do a database to database conversion if you voluntarily change vendors?
- What is the cost to do a database to database conversion if the vendor no longer supports the software?



There is no doubt that you will have a problem with a document imaging database; the question is who is responsible to fix it and how much will it cost? When document imaging systems entered the business scene, they were geared for mega-size organizations, which had experts on staff to manage the problems that arose. Now that document imaging has moved into the mainstream of American business, the question you must answer is: does your organization possess the expertise needed to manage and maintain a database system?

We are of the opinion that every organization will derive significant benefits from a document imaging system. The next step is choosing the best system for your organization. Should it be a:

- Database system,
- Full-text system, or a
- Metadata system?



