WHITEPAPER

Creating a Flexible Production Document Archive

Background

Most Businesses today recognize the needs and advantages of creating an enterprise wide electronic document archive. Often such projects originate to satisfy legal requirements or as part of a "Go Green" paperless office initiative. Reducing paper use and satisfying legal requirements are obvious and valid reasons to implement an archiving solution; however the advancement of technologies and the subsequent reduction in storage costs have created the opportunity to get significantly more business value through archiving. Archiving can and should be used as a primary way to insure better customer service, giving your employees, vendors and clients new options on how to locate, retrieve and disseminate relevant information. Such information is often, but not limited to, being electronic copies of business forms.

This white paper discusses the best way to architect a document archiving solution that not only satisfies legal requirements and reduces or eliminates paper; it is flexible enough to enhance various business processes throughout the organization. A special emphasis is placed on production document archiving. Production documents include all business forms such as invoices, purchase orders and payroll slips, the original business reports that created them, and any associated documents that are incoming to the business such as orders or delivery receipts. In addition we encourage organizations to look beyond production documents to all other documents and consider how archiving those streamlines business practices and enhance relationships with staff, vendors and clients.

Identifying Business Procedures to be Supported and Enhanced through Archiving

Most companies do not perform formal Business Process management (BPM) evaluations, however they are keenly aware of various functions within their organization that service and support their staff, vendors and clients. Thus it is logical to look at these three groups to identify which functions can be enhanced via a well architected document archive.

Staff related documents that should be considered for archiving include production documents such as payroll slips, reports on vacation and sick time accrued, and monthly sales and commission reports. Additionally there can be value in archiving any documents completed by the employee as a prerequisite for hiring such as non-disclosure agreements and offers of employment. Also, it often makes sense to give staff access to a repository of company forms such as requests for vacation or sick leave. Finally, training documents, company brochures, price lists and company policy documents can be also be made available through an archive. While this is not a complete list, the point is made. Every finalized document that staff touches can be archived. Production document archiving is the catalyst for creating the archive, but savvy businesses expand the archive and begin to think of the archiving system as a document repository. That said, care should be taken not to use the repository as a collaboration tool for developing documents. Only final in-use documents should be archived or the repository will be become bloated with both too many documents and too many document associated business rules.

Vendor associated production documents include purchase orders and their associated delivery receipts, invoices and statements. One may also consider including catalogs, price sheets, terms, and product brochures. By ensuring such documents are well indexed and searchable the purchasing department can make quicker more informed purchases.

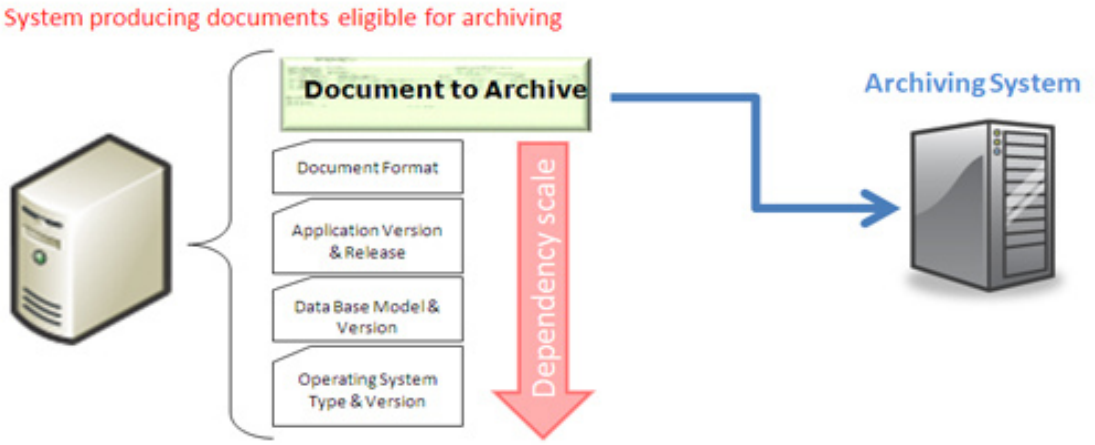
Typical client related production documents include orders, invoices and delivery receipts. In addition many of the vendor associated documents mentioned above would also be appropriate. Your company's catalog, terms, and product brochures for instance. Any document that ships with a product or is made available during presales can be archived. By making these types of documents available indefinitely and searchable customer service and satisfaction are enhanced at very little cost to the organization. A customer that needs a manual for a product purchased five years ago can simply log in, search by model number and instantly download it.

Three Area's to Consider when Designing an Archiving Solution

There are three broad areas to consider when ensuring a production archiving solution is both robust and audit friendly while also flexible enough to streamline business practices and enhance relationships with staff, vendors and clients. First is the basic architecture including the hardware, datastore and workflow between the process that creates the document and the archive. Next is deciding what information about the document (metadata) will be made available and what portions and fields of the document will be indexed and searchable. Third are access considerations meaning who is able to search for which documents, from where and what they can do with the document once they retrieve it.

Key Architectural Considerations for Efficient Flexible Archiving

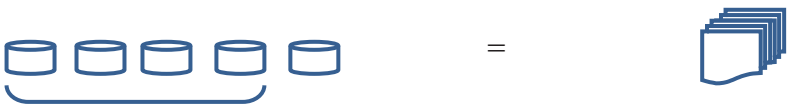
To insure a production archive satisfies legal requirements and is both robust and flexible it is important that the archive be fully independent from original system that generated the documents. The archiving system should to be hosted on a different platform from the ERP, CRM, and/or BPMS that are create the production documents. This insures that the document archive is not dependent on future upgrades or versions of these systems. It frees the organization to change these systems as technologies and business needs evolve and allows access to the archive to be independent from access to the ERP, CRM and BPMS.



In addition the architecture needs to support one or more free readers and have web access. Web based access, usually via an Intranet insures that anyone anywhere can be allowed access to stored documents. PDF and XML are today's most popular formats not requiring a registered user license. Additional formats that may require a licensed reader should also be supported for added flexibility, but should be employed only when alternatives are not easily available, for example Excel to store spreadsheets. It is also important that binary large objects (BLOBS) be supported so that the archive can contain images and perhaps even video. The datastore needs to be fast and storage space inexpensive and easily upgradable. Standard Windows Server technologies fit these requirements whether the machines are physical or virtual. Thankfully, expensive CD archiving systems and cartridge equipment are no longer needed and are not viable solutions for a robust flexible production archive. They simply aren't flexible enough.

Let's consider an entry RAID 5 disk configuration and determine the associated disk hardware archiving cost:

5 disks at 15000 rpm speed and 300 GB capacity, with one disk reserved for Raid5 checksum offer 1.2 Terabytes of total storage space at a cost of about US$1250. This will store approximately twelve million documents if they average 100KBs each. Symtrax's experience with real world clients archiving terabytes of documents confirms that these figures reflect the real world. The per document disk cost of such a system is incredibly only $0.0001.



1.2 TB disk space

12 Millions PDF documents
(averaging 100KB each)

The overall size of the database, how it is organized, optimized and indexed as well as the amount of memory on the server are also all important considerations. For the archive to be considered robust and flexible it is imperative that the system to be able to retrieve documents within at most a couple of seconds. This means that a sizable index that utilizes extensive metadata fields will need to be loaded into memory and that the Intranet bandwidth is sufficient to support a large number of simultaneous users.

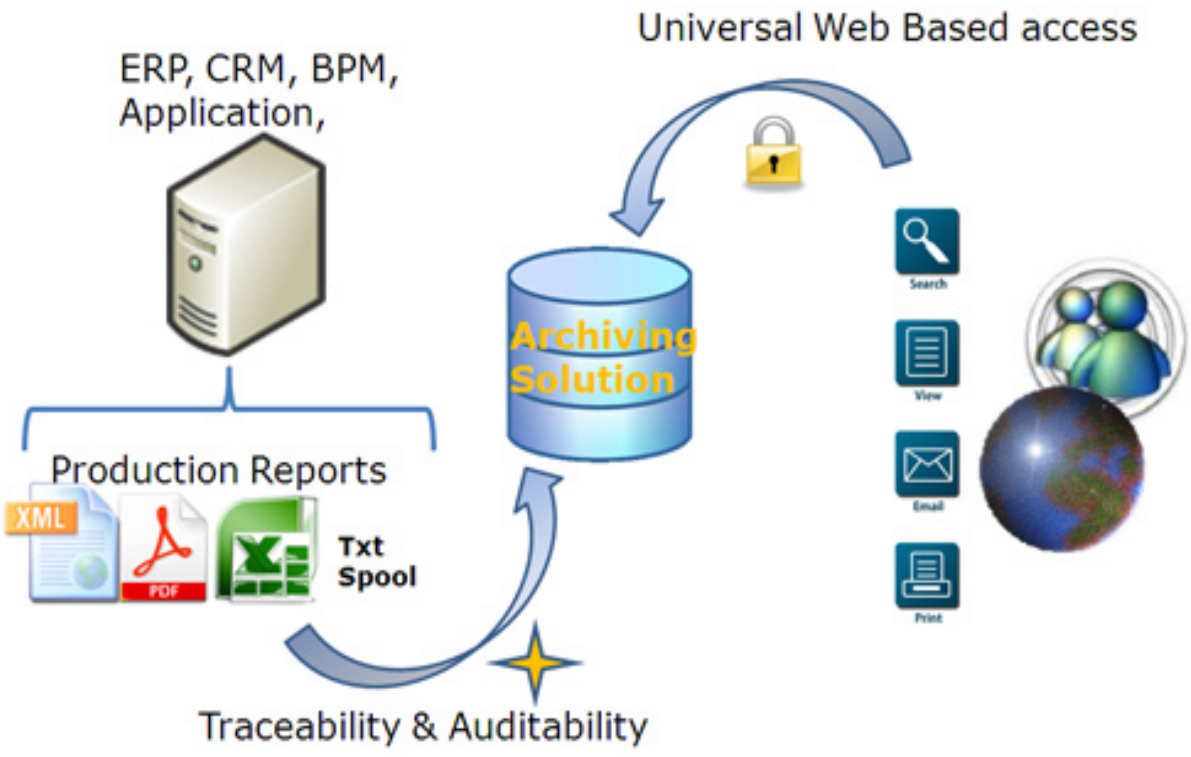
Because of all the variables the only way to insure the above requirements are met is to require an archiving vendor to provide you with benchmarks on their solution and include specifics on the configuration they employed.

Populating the Archive for Optimized Document Retrieval

The front-end of an archiving system is the interface employees and sometimes clients and vendor staff use for document retrieval. This interface needs to be intuitive enough so that new users can find what they need without specific training. It should support pre-defined searches based on the type and characteristics of the specific document being retrieved. This means that the system must allow for a different search model for each type of document and to make it even more flexible, these search models should be "cloneable" so that they can be reused for other similar documents. Millions of people are familiar with the Google search syntax so using similar techniques where possible makes sense. Extensive use of drop down lists particularly dynamic drop down lists is also advised. As with Google, the system should support "wild card searching" because it isn't always possible to anticipate what data a user will decide to search upon.

To support the search capabilities discussed above the system used to upload documents will need to allow for choosing which fields in the document will be indexed. This will greatly reduce retrieval time and help keep the size of the archive from growing too rapidly. In order to support wild card searching while also limiting the size of the database, the uploading process should allow for choosing what part of a document is wild-card searchable. In some cases it will be necessary to make the whole document wild card searchable however for many documents only a small portion of it needs this ability.

Metadata associated with a document needs to be maintained both to support legal and auditing requirements and also to support searching. It may, for example, be useful to search for all documents created by certain process or a certain person or within a specific date range. It is not enough to know when the document was archived; you need to know its creation date, who created it, what process was used to create it, etc. All available metadata should be retained.



Traceability and "Auditability" need to be handled as part of the workflow.

Traceability certifies that the retrieved document is referring to its originating system. It's based on Metadata or properties embedded in document. Traceability requires at least following information:

- System ID or Application ID that generated the document
- User ID, or Batch ID, or Transactional job ID that authored the document
- Date & Time document was generated
- Date & Time document was archived

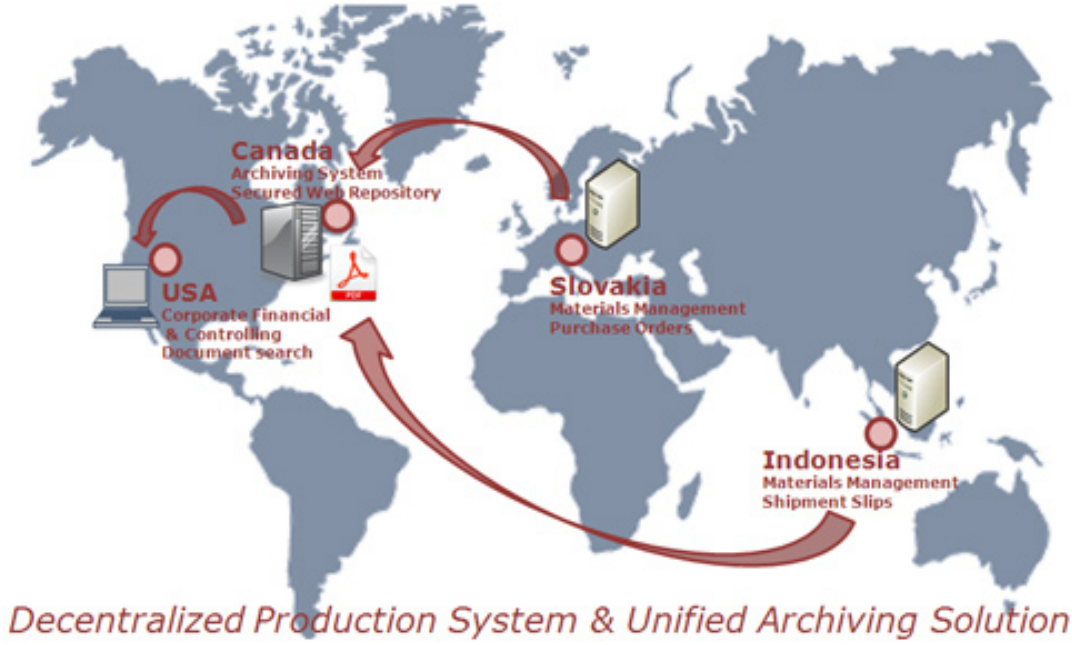
Auditability protects the document and guarantees its integrity. Auditability requires at least following information

- Digitalized signature using a certificate.
- Timestamp
- Reference to the original document

Workflow includes the ability to identify a new production report, to maintain its identity card as metadata or via a properties attachment and uploading it into the archive while logging all operational details and subsequent accesses of it. Production archiving systems require that as soon as a production report is generated, it is burst into individual documents and both it and the associated documents are indexed and archived. Workflow also includes rules on when a document can be deleted from the production archive; some you will only need access to for a few years while others may need to be stored for 10 years or longer. Some documents will always be replaced by subsequent versions while others will never be replaced. Once again we caution about not turning the archive into a full blown workflow management product. Workflow management can be quite useful, but should not be tightly integrated into the archive. The archive is for completed in-use documents, while the bulk of workflow management is best targeted to the process that created the documents and production reports.

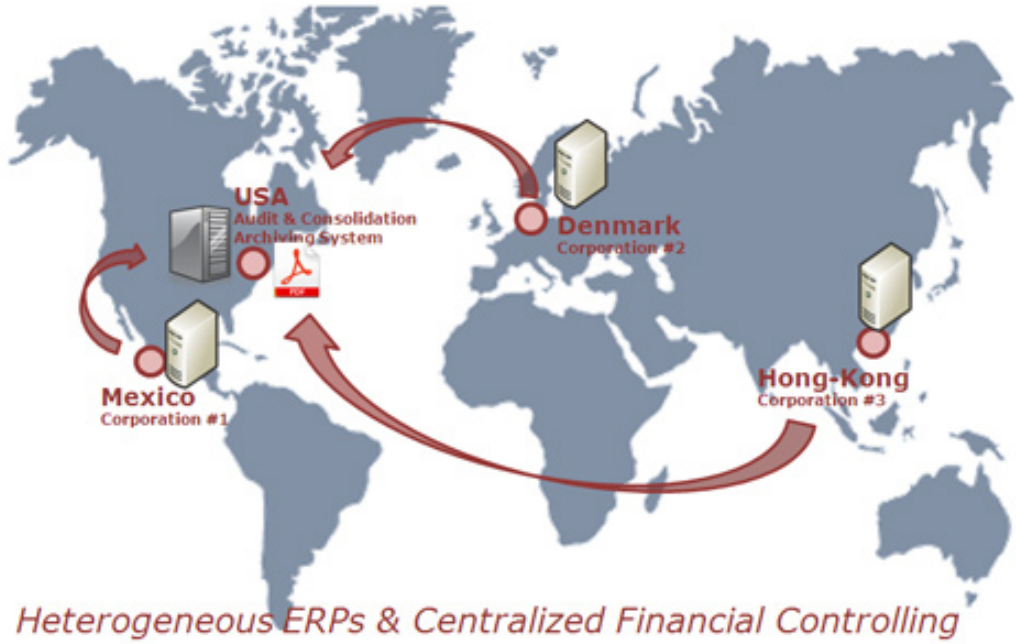
How Accessibility turns an Archive into a Platform that supports Multiple Business Processes

To create a truly flexible document archive that can support multiple business processes across many types of users it is necessary for each user or type of user to have different mechanisms to retrieve documents and selective access to the archive. We have already mentioned that providing a web interface to the archive is essential so that users can gain access from anywhere using an interface they will understand intuitively. This interface will necessarily require a user log-in and password. The system will need to understand, based on the user's credentials, which documents they can have access to and what they are allowed to do with the documents once they are retrieved. Because of varying access rights each type of user may get a different search result even though they searched the same terms in the same search model. In addition, we recommend that some users be able to publish retrieved documents to a secure weblink so that documents can be provided to unregistered users.



Decentralized Production System & Unified Archiving Solution

Registered users will include employees and sometimes select individuals amongst your venders and clients. The security system should be able to assign capabilities based on individual needs as well as group needs. There may even be "super groups" that contain thousands or millions of users with the same rights. Thus, for example, you can maintain a group called clients and everyone in that group will have the same rights.



Heterogeneous ERPs & Centralized Financial Controlling

After it is determined which document search models the user has access to, it needs to be determined what the user will be allowed to do with the document retrieved. Most common are the abilities to view, print and email the document. Additionally some users may have the ability to add notes to a document or even add attachments to the document. This will help support the full lifecycle of the business process associated with the document. For example, with a shipping document, you may need to match the original with a signed version that is returned after delivery. The signed copy can be scanned and attached to the original providing proof of delivery months or years removed from when the delivery occurred. Additional functionalities such as being able to send multiple documents at once and zipping up one or more documents before sending them are also desirable.

Conclusions

Advances in accessibility and decreases in hardware costs make this an opportune time to rethink your archiving strategy. Implementing a flexible production document archive can support many different business processes while also satisfying legal requirements and limiting or eliminating paper. It is not an electronic duplication of a paper storage facility that is only accessed once a year when the auditor arrives. It can be as important to the business as an ERP or CRM, accessed hundreds or thousands of times a day by employees, vendors and clients.

About the Author

Frank Yacano has spent over 20 years in high tech sales and marketing. He is currently the Director of Sales and Marketing for Symtrax, Inc. a company focused on delivering user friendly, fully automated Business Document Management and Business Intelligence Solutions. Prior to Symtrax, Frank served as Director of Business Development for SYWARE, Inc. the maker of Visual CE, a database, reporting and forms development tool for the mobile computing platform Windows CE. Mr. Yacano holds a BS in Engineering from Dartmouth College.

Symtrax USA
(West Coast) Tel: 001 310 216 9536 - Toll free: 800 627 6013 - sales@symtrax.com
(East Coast) Tel: 001 617 728 9910 - Toll free: 866 640 9615 - sales@symtrax.com

Symtrax France:
salesfr@symtrax.fr - Tel: +33 (0) 4 66 04 54 05
Symtrax Europe:
sales@eu-symtrax.com;
NL: +33 4 66 04 54 19 - GE: +33 4 66 04 54 22 - SP: +33 4 66 04 54 23 - EN: +33 4 66 04 54 05
Symtrax UK:
Tel: +44 (0)207 533 6616 - sales@symtrax.co.uk
Symtrax APAC:
Tel: +91 22 4076 5503 - salesapac@symtrax.com