

Electronic Document Archiving with ASG-Cypress™

A WHITE PAPER

Executive Summary

ASG-Cypress™ is an enterprise-class document archival solution that may be easily implemented in different ways to optimize your key business processes. This document describes the ASG-Cypress archival solution, identifies key criteria that influence configuration, and presents different implementations to deliver the most efficient and cost-effective solution based on your business requirements.

Introduction

The goal of any enterprise-class document archival solution is to provide long-term retention of content in such a way that it cannot be altered, and that access to content is fully secured and auditable. ASG-Cypress has long provided these benefits.

However, the way in which an archiving solution is implemented from beginning to end, can be unique to each organization or customer based on the volume of documents they need to archive, the frequency with which archived documents are accessed, the desired benefits, the tradeoff between storage media cost and speed of document retrieval, and many other factors. For example:

- Some organizations have many departments and employees who need frequent, yet secure, access to a wide variety of archived content, from any location. These customers need fast access to content so that worker productivity is not impeded and business processes do not slow down.
- Some organizations need to archive the entire corporate memory over long periods of time. These customers need a secure, cost-effective, and highly automated solution for retaining documents.
- Other organizations seldom need to access archived content and have very low archive volumes. These customers simply need “document backup” functionality that is low-cost and easy-to-implement.
- Still other organizations must comply with government or industry guidelines concerning electronic document archiving. These regulations may specify the media type, document format, metadata that must be included with each document, and how content is to be accessed. These customers need a solution that will allow them to comply with specific regulations related to document archiving.

Due to its exceptionally flexible and extensible nature, ASG-Cypress can be configured to meet the archiving needs of each of the above situations, and many others.

Whether configured as a stand-alone archiving solution or combined with Hierarchical Storage Management (HSM) or Write Once Read Many (WORM) products, ASG-Cypress provides the degree of backup, security, and access that you need for protecting the data permanence and the evidential weight of your documents – based on your requirements.

Archiving Content in ASG-Cypress™

Virtually all parts of the corporate memory may be archived within the ASG-Cypress DocuVault, extending far beyond conventional archiving products that are limited to simple document and report storage. In a single solution, ASG-Cypress can store:

- Documents and reports
- Any source file
- Scanned documents
- CAD drawings/engineering documents
- X-Rays/MRIs
- Images
- E-mail and e-mail attachments
- Audio and voice files

A Standalone Document Archiving Solution

ASG-Cypress is designed to be the only document content storage solution an organization needs, and it meets the requirements of virtually every department and employee in your enterprise.

If you choose, ASG-Cypress can be used in concert with third-party, inactive storage solutions, such as HSM products. ASG-Cypress incorporates a mechanism that enables content migration, demigration, and retrieval using these third-party systems.

Comply with Government and Industry Regulations

ASG-Cypress enables customers to meet key compliance criteria related to electronic document storage and retention. The section on Legal Archiving, which appears later in this document, expands upon this concept.

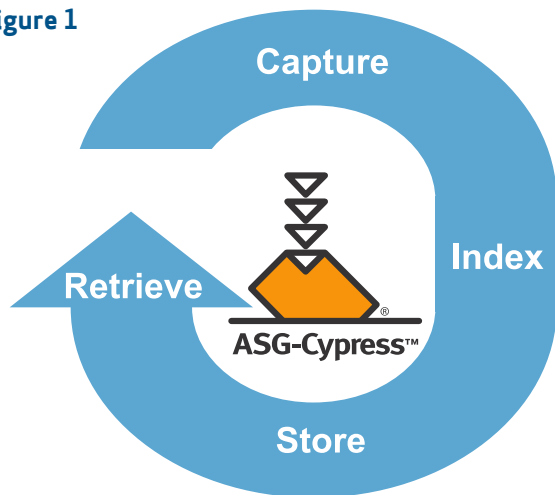
Electronic Document Archiving with ASG-Cypress™

The ASG-Cypress™ Archiving Model

ASG-Cypress is capable of capturing and storing virtually all content created in your organization, and yet may store the content in different ways to best meet your requirements – and those of your colleagues.

The ASG-Cypress archiving model is based on the following functions. Each function is described below.

Figure 1



Document Capture

ASG-Cypress captures content from any system in your organization. ASG-Cypress may capture documents in their source format, or may capture documents in a “common” format in such a way that individual pages may be uniquely secured, retrieved, bundled, and delivered to any enterprise destination with 100% fidelity.

Capturing documents in a common format provides you with far greater value as compared to source file capture, since pertinent pages from any combination of applications may be bundled together to satisfy each customer’s, employee’s, or partner’s requirements. This approach provides content consumers with the exact information they need, regardless of the source, resulting in greater productivity, higher satisfaction, and new business processes and services - all of which increases your organization’s competitive advantage.

Capturing source files is accomplished via ASG-Cypress’ proprietary DDI In technology, by using a FTP function, or by storing documents in specific directories monitored by ASG-Cypress. When detected, ASG-Cypress automatically acquires the file and applies pre-defined retention and security settings.

Capturing documents in a common format is accomplished by “printing” documents from the original authoring application to ASG-Cypress, although the final destination of the document need not be a printer. By acquiring a “print file,” ASG-Cypress captures all information about the document, including text, images, fonts, margins, character position, etc., so that pages may be stored individually and be reproduced with 100% fidelity on any device or system. This also ensures that the original authoring program need not be available for further document viewing.

As illustrated in Figure 2, ASG-Cypress captures documents for any enterprise system and stores them as individual pages in a common format within the ASG-Cypress DocuVault.

Security to all captured content in the ASG-Cypress DocuVault is automatically set at the time the document (or source file) is captured. Access and use permissions can be specified at the file-level, document-level, sub-document level (for jobs containing multiple documents), or page-level. This degree of flexibility and granularity allows customers to easily meet the even the most stringent privacy protection guidelines, but does not require additional labor and document storage space (to store multiple document instances for multiple permission levels).

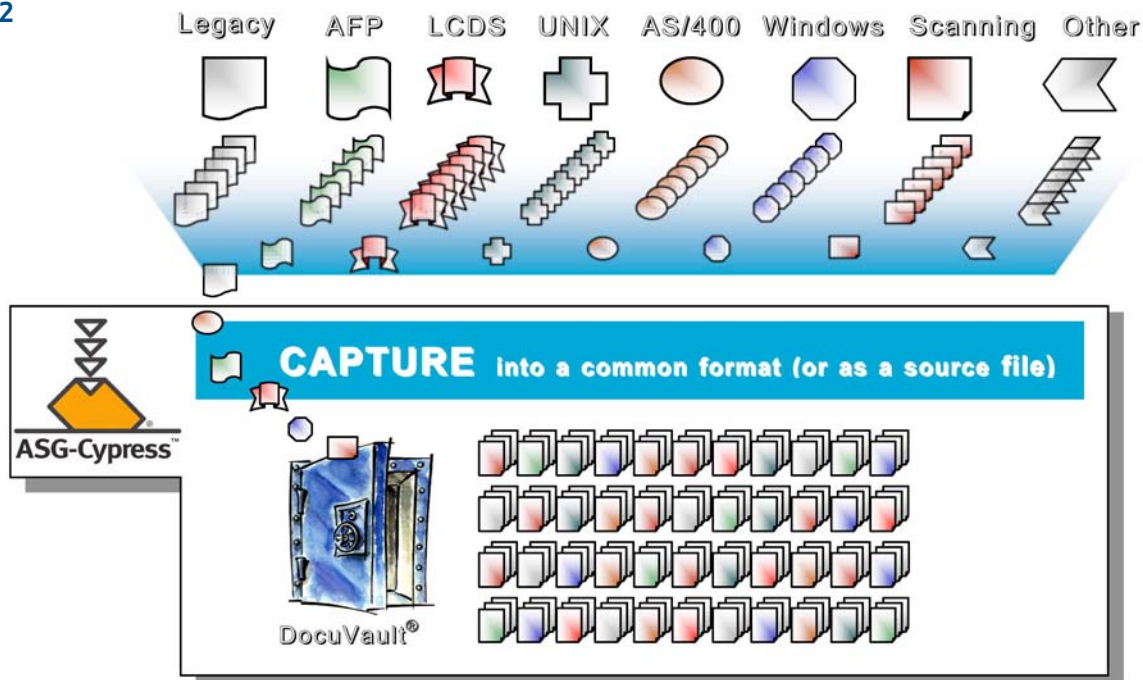
Protecting document authenticity and preventing alteration is provided by the ASG-Cypress Capture process. By definition, documents captured by ASG-Cypress (in a common format) cannot be altered. Captured documents are stored by their constituent parts, with no authoring tool available for modifying the document.

Document Indexing

ASG-Cypress maintains internal system indexes for the metadata related to every job processed by ASG-Cypress, such as creator, document deletion date, timestamp, number of pages, and other information. However, more useful and extensive indexing may be implemented based on the way documents are captured, and the level of indexing you wish to use. Each is described below.

Source documents are indexed based on “job ticket” information. A job ticket is an ASG-Cypress object that stores metadata related to a job, such as creator, sending system, file name, timestamp, and so on. Because ASG-Cypress stores the source file “code,” and does not open source files to parse content, content-based indexing (e.g., full-text index) is not available for source documents.

Figure 2



Documents captured in the common ASG-Cypress format may be indexed using content from the document, in addition to the system indexes and job ticket information discussed above. Specifically, documents and pages may be indexed as follows:

- Full-text index: as documents are captured, every word on every page is indexed within ASG-Cypress. This enables page-level retrieval based on any word (or string) contained in a document.
- Data indexes: as documents are captured, content within regions (areas on a page defined by you) is automatically extracted and indexed. Samples of extracted content include customer number ("CID 4908-03"), division ("Western: Sales"), invoice total ("\$46,987.09"), part description ("No. 9 shipping container"), and so on.
- Ad-hoc indexing: Authorized users may also add index keys to specific pages when viewing documents within the ASG-Cypress client user interface.

Document Storage

ASG-Cypress stores a unique instance of every document it captures, whether documents are captured as source files or in the common ASG-Cypress format.

Source documents are stored in the document's native format (e.g., .doc, .xls, .psd, .mpg, and others). Source files are not modifiable when

in the ASG-Cypress DocuVault, ensuring the authenticity of the file. Even though source files may be retrieved, edited, and placed back in the ASG-Cypress DocuVault, the original file is untouched and is not overwritten or deleted. The "second" version of the source file is stored as a separate document.

Documents captured in the common ASG-Cypress format are stored page-by-page within the ASG-Cypress DocuVault. Each page is individually indexed and secured, and cannot be modified.

Storing content on secondary disk (DocuVault partitioning): By default, all documents are stored on the ASG-Cypress DocuVault's primary disk. The primary disk has the advantages of high speed, high reliability, and immediate document access to archived content. However, using the primary disk to archive large volumes of documents or to store documents for long periods of time may not be the most cost-effective approach.

To minimize storage costs, and improve the speed of ASG-Cypress DocuVault backups, ASG-Cypress offers the DocuVault partitioning feature. This feature enables content to be stored on less-expensive, secondary disks, such as a RAID solution physically connected to the ASG-Cypress Server. Storing documents on a secondary disk maintains an "active storage" environment where all content is quickly retrievable, but at a reduced cost.

Electronic Document Archiving with ASG-Cypress™

The ASG-Cypress DocuVault partition feature is an extension of the ASG-Cypress DocuVault architecture. Effectively, this feature splits the ASG-Cypress DocuVault into read-only pieces, called “partitions,” that contain the documents you specify (shown below). As documents are moved into partitions and written to disk, the document partition system index is updated to reflect each document’s new location.

Note: ASG-Cypress DocuVault partitions, and the documents they contain, are still an integral part of the ASG-Cypress DocuVault even though they are not stored on the ASG-Cypress DocuVault’s primary disk.

When partitioned documents are selected as a result of an ASG-Cypress DocuVault query, ASG-Cypress transparently and automatically identifies the partition containing the needed document, retrieves the document from the partition, then processes and delivers the document as appropriate.

Storing Content on HSM (Hierarchical Storage Management) Devices

Figure 3 also illustrates how ASG-Cypress DocuVault partitions may be stored in an “inactive storage” environment, such as HSM or other separately-managed storage systems. Whenever possible, we strongly encourage you to store content in the ASG-Cypress DocuVault’s primary or secondary disk storage, as this will eliminate delays to

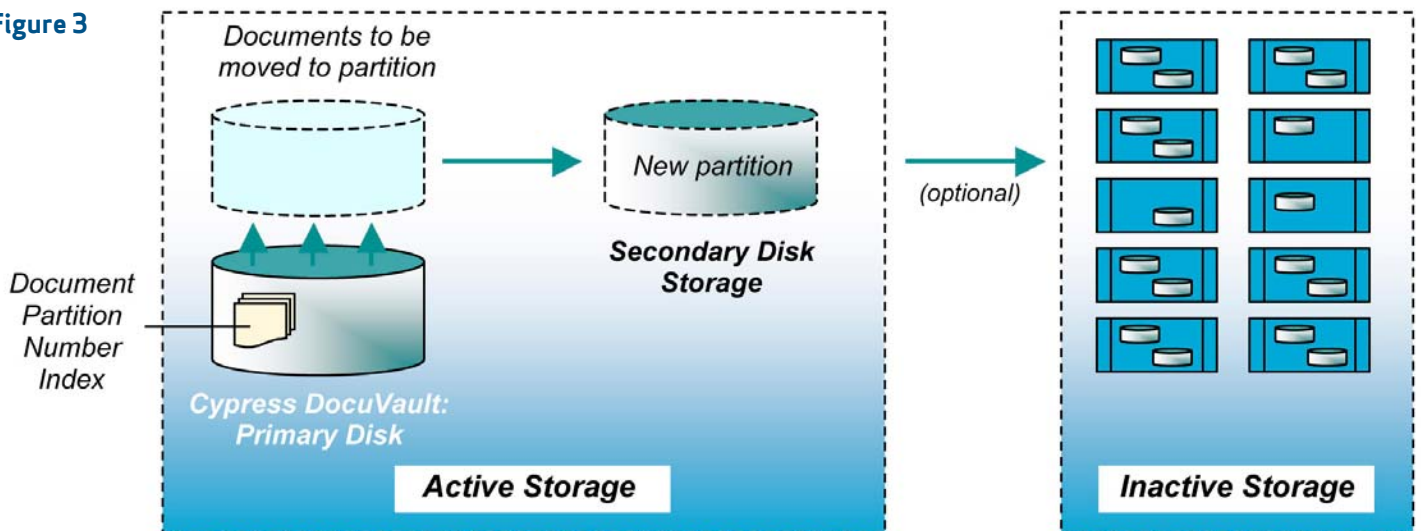
retrieve content and typically costs much less than HSM systems. However, ASG-Cypress can write ASG-Cypress DocuVault partitions to disk, which may then be acquired and migrated by an HSM system to its internal library.

Storing Content on WORM-Type (Write Once Read Many) Devices

Customers who need to store content on WORM-type devices (e.g., CD, DVD) are typically concerned with storing documents over long periods, ensuring a document’s legal admissibility, or complying with government or industry regulations. Since content within the ASG-Cypress DocuVault can be stored indefinitely, is fully secured, and cannot be modified, WORM-type systems may not be needed to meet your requirements.

Customers who are required to employ WORM-type systems are encouraged to keep a “working copy” of archived content that remains live within the ASG-Cypress DocuVault. Keeping a live version of offline content ensures that users can easily and quickly retrieve content as needed, yet comply with regulations that mandate the use of WORM or optical storage. Also, utilizing a “working copy” for everyday use promotes WORM media longevity, as optical media can degrade with excessive reads.

Figure 3



Documents can be written to WORM-type solutions in a variety of formats, such as .pdf, .tiff, or other universal formats, and stored in the following ways:

- Standalone media: Using the ASG-Cypress External Archive Module and the ASG-Cypress Content Processing Facility (CPF), customers can aggregate and export documents stored in ASG-Cypress to stand-alone disk in a universal format. This feature also enables disk-resident source files and directories to be written to external disk. Additionally, ASG-Cypress automatically includes an XML file indexing all the documents on the disk. Currently, the index information listed below may be included in the XML file:
 - Document title
 - Creation timestamp
 - Document creator
 - Number of pages
 - Name of ASG-Cypress DocuVault

Document Retrieval

Customers may search and retrieve all content through ASG-Cypress. Since all indexes are retained at the ASG-Cypress Server, regardless of the location of the actual content (e.g., partition in active storage or inactive storage, PDF file on WORM media, etc.), customers have a full audit trail for all document access, viewing, and printing activities.

Searching and retrieving ASG-Cypress DocuVault-resident documents may be performed from within the ASG-Cypress client user interface, or from a Web browser running ASG-Cypress.Web. ASG-Cypress.Web provides document search, retrieval, and viewing from within a Web browser. Note that ASG-Cypress.Web enables document viewing with 100% fidelity, and does not require the original authoring application.

Based on the format of the retrieved content, viewing is enabled as follows:

- ASG-Cypress-formatted documents (viewed with the ASG-Cypress client interface, or a Web browser using ASG-Cypress.Web)
- Source documents: launched within the source application (e.g., Excel, Word).

Note: retrieving documents stored in partitions, regardless of the partition's location, is identical to retrieving documents on the ASG-Cypress DocuVault's primary disk.

Searching and retrieving documents on stand-alone media (exported via the External Archive Module) is also a viable option. Retrieving exported documents can be achieved in a variety of different ways. However, it is important to realize that searching standalone media for relevant content is a non-productive and time-consuming approach to document retrieval. When archiving documents on standalone media, ASG-Cypress strongly recommends that customers first query ASG-Cypress to identify the target document and media. This is accomplished by inspecting the External Volume Tab, available from each document's properties. This Tab contains the information needed to retrieve a specific document from a specific external volume (disk).

The standalone media simply needs to be inserted into any accepting system, searched using one of the conventions below, and then opened within the appropriate application:

- File Naming Convention: Documents stored on stand-alone media may be written with the document ID embedded in the file name. The XML metadata (discussed below) can be used for retrieval to minimize the need to open files to determine relevance.
- XML file written to standalone media: ASG-Cypress generates an XML file containing a brief index of all document files on the media (including each document's ID number). Users can open the XML file using Windows Explorer (or any other tool capable of displaying XML data), locate the needed document, and then open the document within Adobe Acrobat Reader or other application.

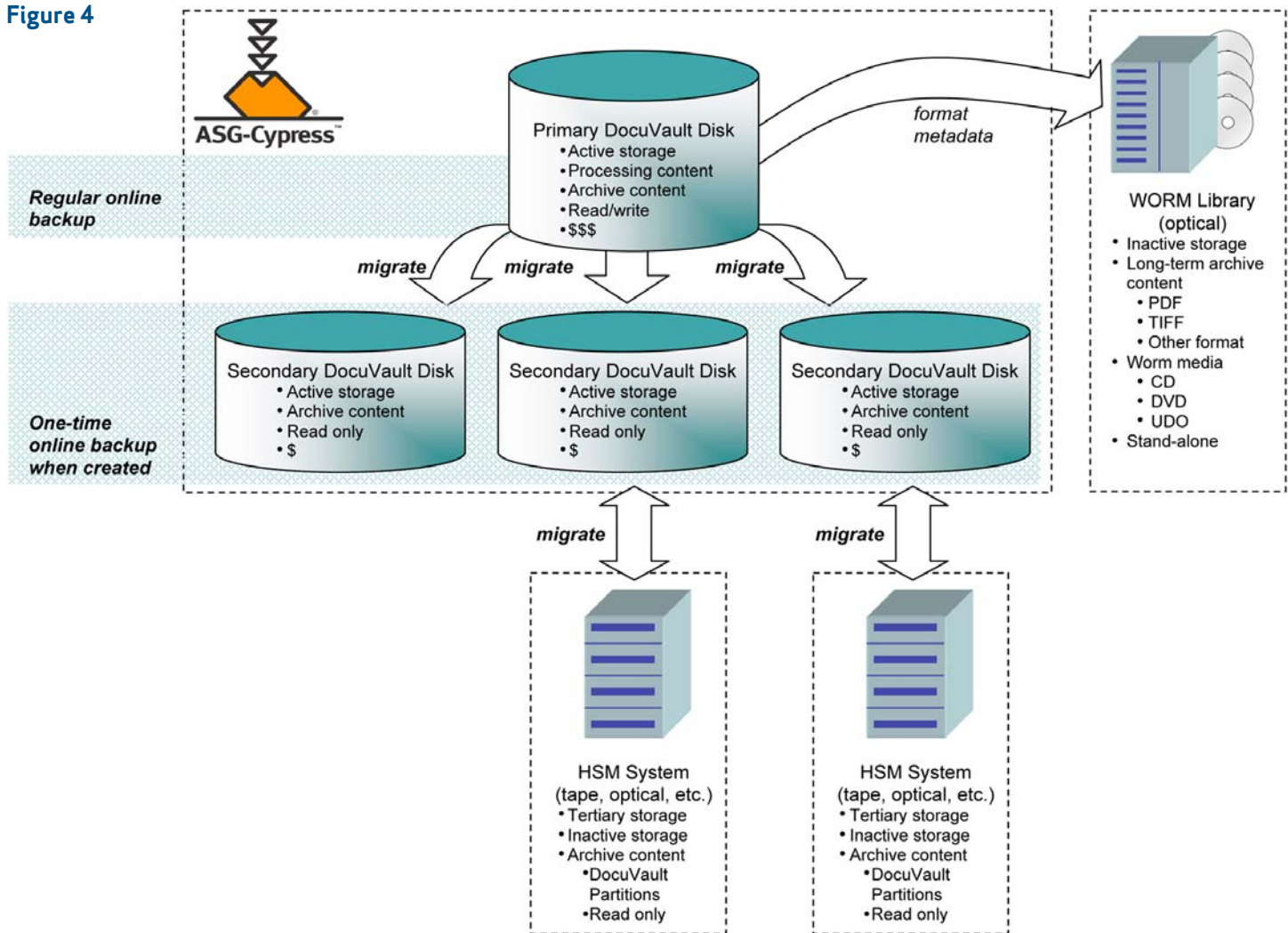
Document access (to documents stored in the ASG-Cypress DocuVault or an integrated offline storage solution) is permitted or denied based on the user's identity and the security established for the document/page at capture time. ASG-Cypress will not return any documents that the user is not permitted to view.

Archiving Architecture

Figures 4 and 5 illustrate the different ways documents may be stored and retrieved using ASG-Cypress.

Electronic Document Archiving with ASG-Cypress™

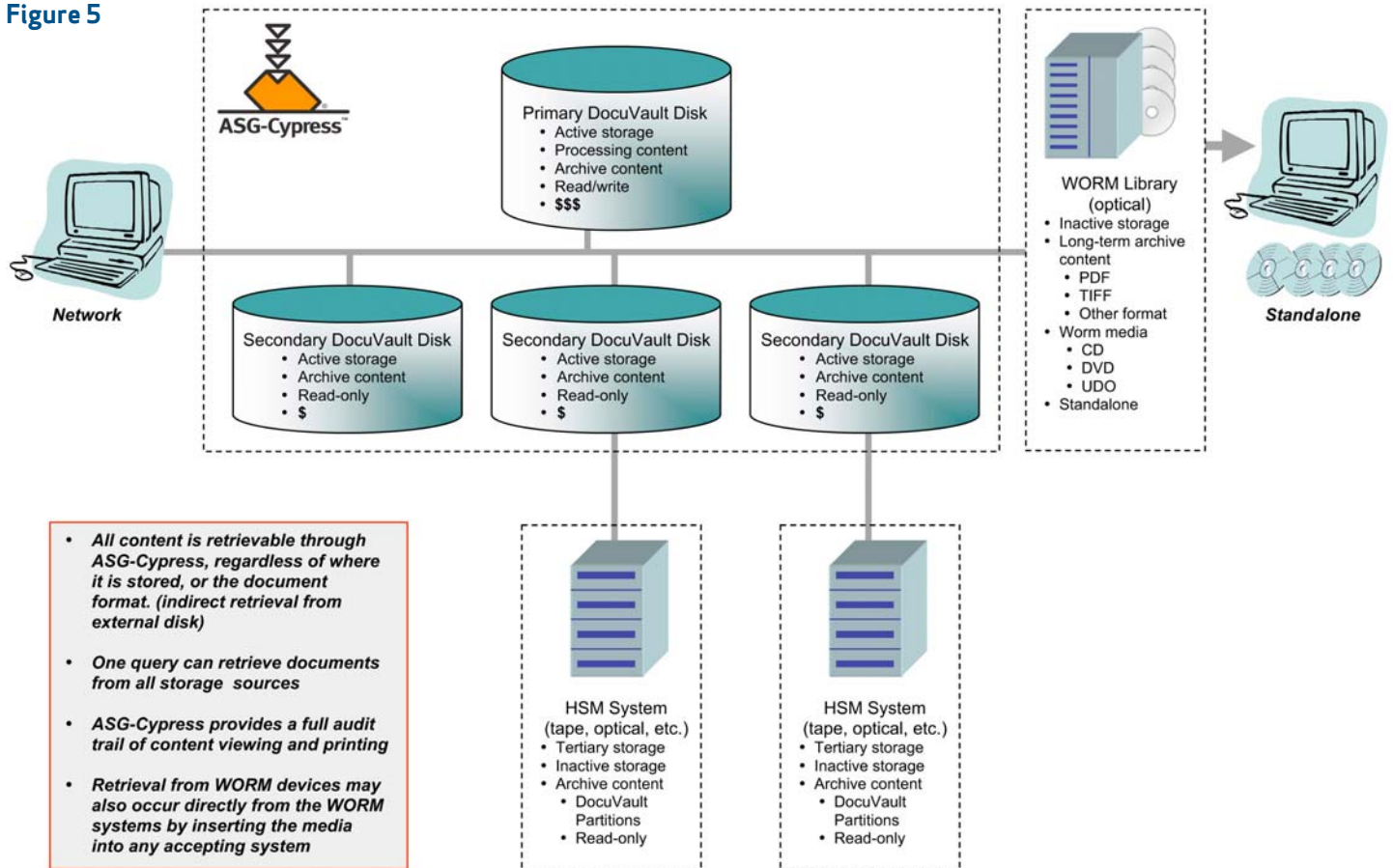
Figure 4



Document Storage

- Store content on the most cost-effective media: ASG-Cypress allows you to store content on the lowest cost media that meets your performance and retrieval speed needs
- Centralized storage environment: Even though documents may be stored on distributed media, all content is fully manageable and retrievable within ASG-Cypress
- Third-party storage system support: ASG-Cypress DocuVault partitions may be migrated to and from HSM systems, and content may be directly written to WORM-type devices in the appropriate format via the External Archive Module
- Faster primary ASG-Cypress DocuVault disk backups: If you partition a ASG-Cypress DocuVault to secondary storage, there is less data to be backed up on the primary ASG-Cypress DocuVault disk
- One-time online backups for partitioned content: Because ASG-Cypress DocuVault partitions are read-only, they only need to be backed-up when created

Figure 5



Document Retrieval

- All content is retrievable through ASG-Cypress, whether stored on the primary ASG-Cypress DocuVault disk, secondary ASG-Cypress DocuVault disk, a tertiary storage system such as HSM, or a WORM-type system
- One query can retrieve documents from all sources (Documents exported via the External Archive Module are not retrieved through ASG-Cypress directly from external disk. Following recommended best practices, documents may be retrieved within ASG-Cypress, and then inspected to identify the external volume (disk) and file information.)
- Queries are exceptionally efficient due to ASG-Cypress' patented technology
- ASG-Cypress provides a full audit trail for any document retrieved through ASG-Cypress
- Documents can be retrieved from a WORM disk inserted into any PC that can read the disk and file format (e.g., .pdf)

Electronic Document Archiving with ASG-Cypress™

Configure the ASG-Cypress™ Archiving Solution

ASG-Cypress allows documents to be stored in different ways, each satisfying different customer needs with regard to speed, cost, authenticity, data permanence, and so on.

While ASG-Cypress is the central component of an archiving strategy, you have choices with regard to centralized or distributed content storage, storage media, and the use of third-party storage devices.

Your particular needs will dictate how you implement an archiving strategy using ASG-Cypress. The approach you choose should be based on the following factors:

- The quantity of documents you need to archive (archiving volume)
- The frequency of end-user access to archived content
- Impact on user productivity
- Cost
- Document retention time
- Legal/regulatory compliance requirements

Archive Volume

The number and size of the documents you archive will influence your decision to store content on secondary disk storage or third-party systems. By default, all content is stored on the ASG-Cypress Server's primary disk. However, this disk can be expensive due to the high speed and high reliability requirements for supporting all aspects of a dynamic content delivery environment. If you plan to archive a large number of documents, consider secondary disk storage, which can significantly reduce document storage costs.

Frequency of Retrieval

Recently-published content, or content relevant to a broad percentage of your user population, should be maintained in active storage, specifically, on the ASG-Cypress DocuVault primary disk. This will maintain overall productivity as users avoid delays waiting for documents to be retrieved from inactive storage. However, content that is seldom retrieved, or is only retained for backup purposes, should be stored to a secondary disk.

User Productivity

Storing documents in any "active" storage media ensures that user productivity is not impacted due to document retrieval delays. Customers who need to store content on inactive storage systems may experience moderate delays in accessing content, depending on the system. Customers using "slower" systems are encouraged to store only infrequently-accessed content, or consider using these systems as backup systems while leaving all content in "active storage" for fast retrieval.

Cost

Storage costs can be easily controlled by storing content on lower-cost disks. Additionally, customers with low archiving volumes or short document retention requirements may continue to use the ASG-Cypress DocuVault's primary disk for the life of the ASG-Cypress Server. The cost for third-party storage solutions are typically eliminated by ASG-Cypress' robust and comprehensive archiving capabilities, and may be needed only to satisfy very specific legal archiving requirements.

Document Retention Time

ASG-Cypress can store content as long – or short – a timeframe as desired. There is no practical limit to the length of time that content may be stored. ASG-Cypress' default media is magnetic disk. Using RAID solutions, magnetic disks may be replaced as needed while still safeguarding content. In contrast, most WORM-type media have a life expectancy of approximately 10 years. Additionally, the life expectancy of WORM-type media, such as optical disk, is directly related to how heavily it is used. Content stored on WORM media longer than 10 years may need to be migrated onto newer disks.

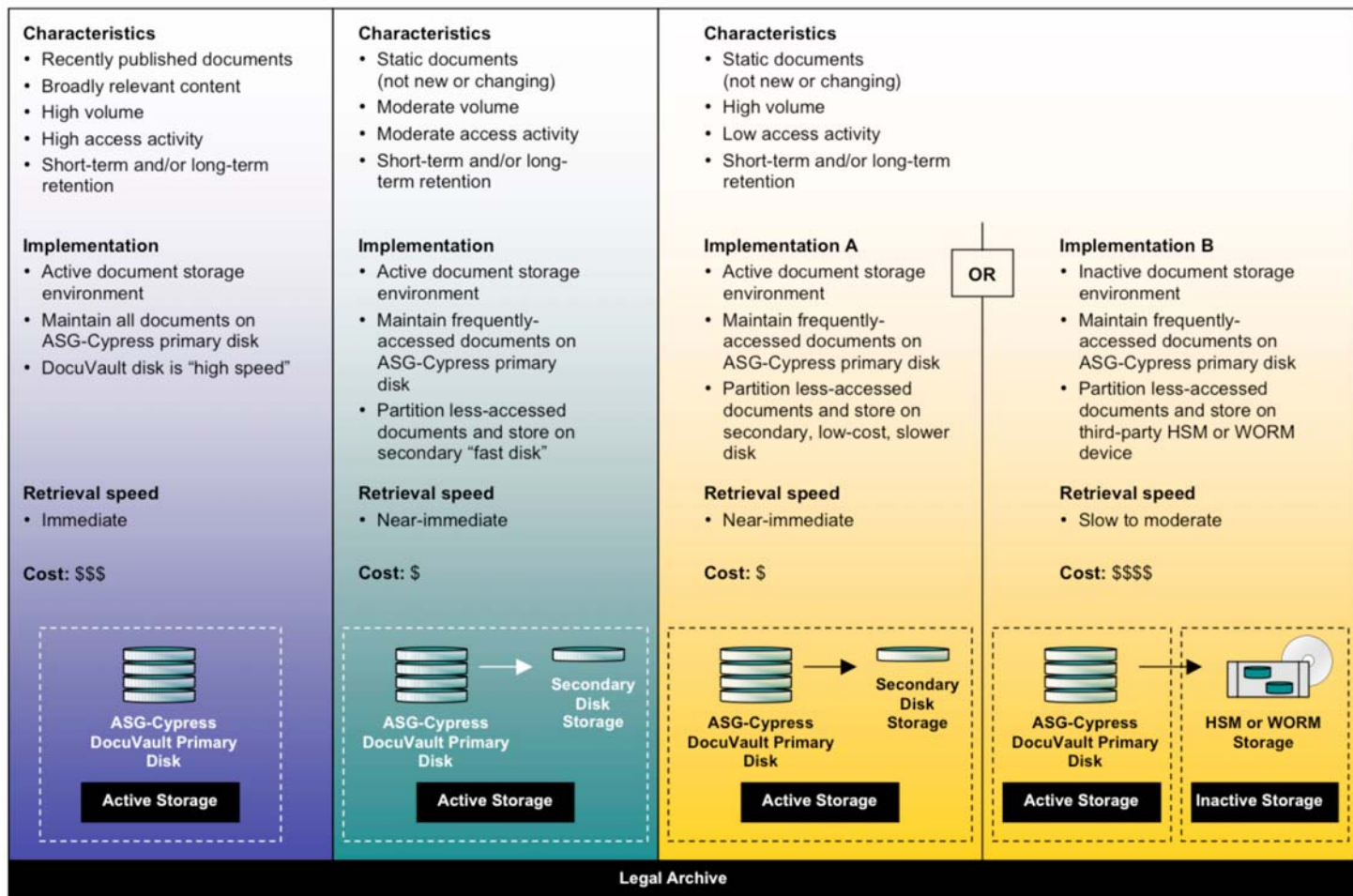
Legal/Regulatory Compliance

ASG-Cypress helps organizations comply with standards and regulations that govern how electronic documents are to be retained (and deleted). Additional information on this specific topic is highlighted later in this white paper.

Sample Configurations

Based on the above-listed factors, customers typically configure ASG-Cypress as demonstrated in Figure 6. Note: All archiving alternatives can be configured to assist in complying with government and industry regulations concerning document archiving and retention.

Figure 6



Recommendation

Virtually all organizations will gain the most benefit from retaining electronic documents in an active storage environment (e.g., disk) so that documents are instantly retrievable and have controlled and auditable access through ASG-Cypress.

As a standalone archiving system, ASG-Cypress provides the following benefits:

- Authenticity
- Security
- Audit ability
- Long-term retention
- Easy and fast retrieval.

Third-party storage systems should only be employed in cases in which a specific storage media technology (other than magnetic disk) is mandated.

Creating a "Legal Archive"

Virtually every government or industry regulation concerning the storage of electronic documents specifies that the following minimum criteria must be met:

- Documents must be stored for a specified length of time
- Documents must be stored in such a way that they cannot be modified
- Documents must be retrievable based on metadata supplied with the documents

Electronic Document Archiving with ASG-Cypress™

Additionally, other criteria that are commonly encountered in different countries and industries also include:

- Documents must be stored on non-modifiable (WORM-type) media to ensure authenticity and eliminate the possibility of alteration
- Documents must be stored in a standard, long-lived file format such as .pdf or .tiff
- An audit trail must be provided to verify document access, archival dates, etc.

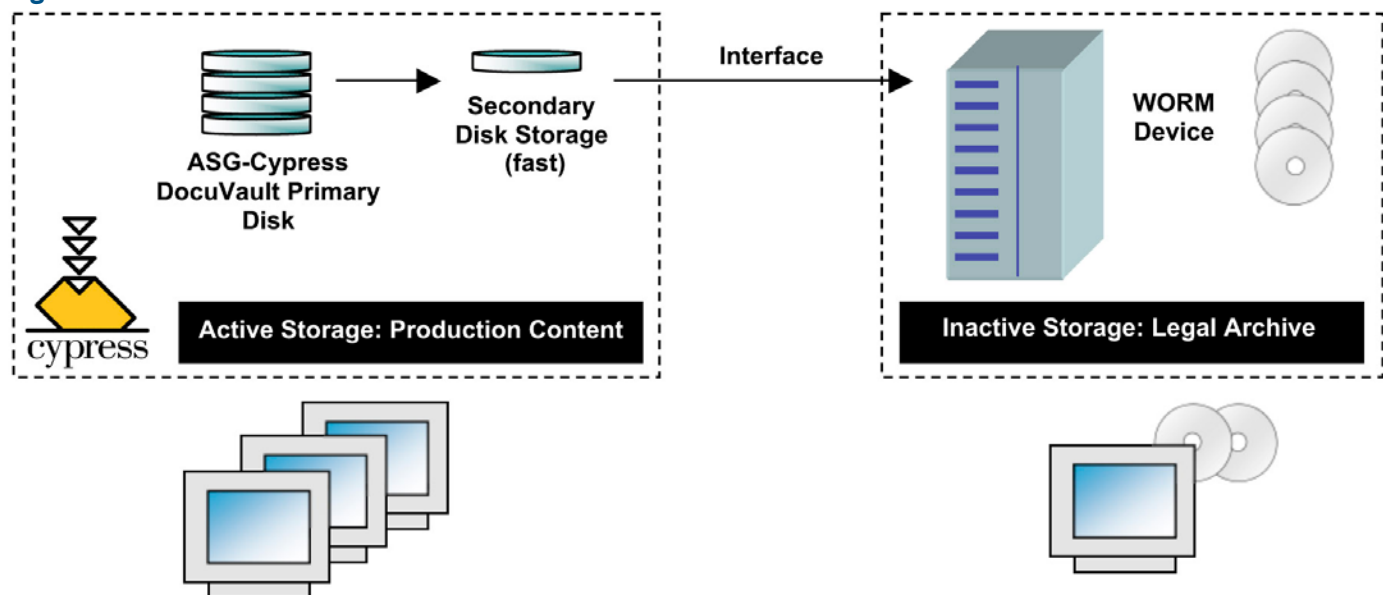
Still other criteria must be met to ensure that an archiving solution facilitates each organization's business processes, is easily administrated, and minimizes labor costs. For example:

- Documents may be selected for archiving based on a combination of document information/attributes, such as length of retention, report type, customer number, etc.

- Document archiving processes should be highly automated to reduce the number of manual procedures to be established and enforced (e.g., automatically control document access based on user ID, automatically set document retention and deletion based on report type – correspondence, invoices/POs, employee data, etc.
- Automatically create a metadata file listing all documents contained on a particular media
- Easily and quickly change document retention dates for archived content, should retention requirements change

ASG-Cypress provides customers with the flexibility to satisfy these and other archiving requirements – not only as a standalone solution – but in a manner that is most efficient and cost-effective for each customer.

Figure 7



The interface between the WORM device and ASG-Cypress provides customers with a single interface for all content, whether accessing production content in active storage, or indirectly from a legal archive (WORM disk).

Note: All document viewing and printing activities are recorded and auditable when documents are retrieved directly using ASG-Cypress.

Content written to WORM disks in a universal format (e.g., PDF) may be directly inserted into a PC and viewed with the appropriate viewing program, e.g., Adobe Acrobat Reader. Search and retrieval is enabled using an ASG-Cypress-created metadata file indexing all documents on a particular disk.

Production Document Content Management Within a Legal Archiving Framework

When using ASG-Cypress as a standalone solution to satisfy regulatory compliance requirements, content-driven business processes continue to benefit from ASG-Cypress without impeding user productivity.

However, certain regulations mandate using specific technologies that require an explicit implementation (e.g., WORM-type media). In many cases, using the required implementation as a production information management tool (in addition to a legal archive tool), may undermine business efficiency.

For example, document retention regulations in EMEA necessitate the use of specific WORM technologies as a way to guarantee document authenticity and prevent document modifications. However, this class of storage device (inactive storage) may be moderately to significantly slower than disk (active storage) when retrieving documents. Slow document retrieval times can lengthen business processes, significantly decrease worker productivity, and reduce an organization's competitive advantage.

For these organizations, ASG-Cypress provides a solution that is analogous to traditional paper-based work processes; the "original" document is stored away in a file cabinet and only used for legal purposes, while a copy is made for "everyday" use. Specifically, the ASG-Cypress External Archive Module can write all required documents to WORM disk in a universal format (e.g., .pdf, .tiff) for legal storage. In addition, however, these documents can continue to be stored as live, everyday-use documents within the ASG-Cypress

DocuVault (e.g., a working copy) so that information remains instantly accessible to users. When authorized users query ASG-Cypress for documents, users view the "everyday" document stored in the ASG-Cypress DocuVault. If the "legal" copy is required, users may inspect the properties of the viewed document and record the external volume and file ID. The corresponding document may then be manually retrieved from stand-alone disk. This approach ensures that employees, partners, and customers can continue to access documents quickly from any location, while complying with legal archiving requirements.

Figure 7 illustrates how ASG-Cypress may be used in tandem with WORM-type devices to enable legal archiving compliance without impacting business efficiency.

Conclusion

ASG-Cypress is designed to be the only archiving solution an organization needs. ASG-Cypress provides customers with the ability to securely archive virtually all content within an enterprise, create new business processes and services, and comply with emerging government and industry regulations concerning document retention.

The key to configuring the most effective archiving solution for your organization is to establish a comprehensive understanding of your archiving requirements – what content, for how long, how much, and accessed by whom. Based on your specific business requirements, ASG-Cypress can be configured to provide the lowest-cost, highest-function archiving solution available.

Content Management

Solutions by 

About Content Management ASG offers Content Management solutions that enable any enterprise to completely control the access and distribution of information throughout the organization. A common architecture allows information to flow seamlessly among disparate information technologies. This unique capability gives users at the strategic, management, and operations levels access to information as needed, without limitations imposed by multiple platforms, operating systems, file systems, applications, and devices.

About ASG | www.asg.com Since 1986, ASG has been using cooperative business practices and more than 200 leading software solutions to help companies around the world overcome everyday business challenges. ASG is headquartered in Naples, Florida, with offices serving the Americas, Europe, Middle East, Africa, and Asia Pacific.