

Why Archiving Matters

How to Manage Your Enterprise Content with Open Text Archive Server

January 2010

Abstract

Faced with today's climate of strict compliance and demands for maximizing ROI, archiving matters to your organization more than ever. Using a purpose-built solution such as Open Text Archive Server, part of the Open Text ECM Suite, allows for quick and easy retrieval, sharing, forwarding, and reuse of content. With Open Text Archive Server, your organization can electronically archive its content in ways that meet regulatory requirements. Large volumes of fixed content can be stored and retrieved efficiently, and users have access to the content they need in the correct business context.

This whitepaper explains why, with the need for greater efficiency and compliance, archiving really does matter to your business. It introduces Open Text Archive Server, and it describes the features that make Archive Server a world-class archiving solution.



Contents

Why Archiving Matters	3
Strategic archiving	4
How archiving is different from backup	4
The three cornerstones of modern archiving	5
Customer examples.....	7
Why Should You Use Open Text Archive Server?	8
Connect: Enable efficient standard processes.....	8
Comply: Control all content and reduce risks.....	12
Cut costs: Low TCO at highest document volumes	17
Functionality Deep Dive	22
What is Open Text Archive Server?	22
Why is an archive server important in an ECM environment?.....	22
Architectural overview of Archive Server	23
Document pipeline: The conveyor belt into the archive system	25
Usage in heterogeneous environments	29
Simple to integrate and connect	30
Archiving from customer solutions	37
High-volume management.....	37
Document lifecycle management services.....	40
Authorization and authentication.....	42
Secure data transport	43
Digital signatures	44
Secure, long-term archiving and data integrity.....	44
Timestamps	44
ArchiSig: Signature renewal for long-term digital signature	45
Auditing: Long-term traceability	46
Encryption of the stored data	46
Always available through various backup scenarios.....	46
Storage and resources.....	48
Conclusion.....	56



Why Archiving Matters

Faced with today’s climate of strict compliance and demands for maximizing ROI, archiving matters to your organization more than ever. Companies inundated with content—scanned documents, email, files from file systems or Web pages, data files from host, SAP®, Microsoft®, and other applications—must deal with it every day in ways that meet all regulatory requirements and instill organizational trust amongst shareholders and customers. The solution lies in Enterprise Content Management (ECM), a system for effectively managing information across the enterprise so that all employees and selected partners globally can access what they need and when and how they need it. A key component of ECM is archiving, which plays a central role in preserving content in a cost effective way so that it’s available when needed.

Open Text Archive Server, which is included as part of the Open Text ECM Suite, allows for quick and easy retrieval, sharing, forwarding, and reuse of content. With Open Text Archive Server, your organization can electronically archive its collective memory permanently and in auditable form. And your content is put into context—linking the documents with your unique business processes. Large volumes of fixed content for international applications such as call centers, and for external customer-facing applications such as Web-enabled bill presentment, are stored and retrieved using Open Text Archive Server.

Figure 1:
ECM Services Architecture



Strategic archiving

Archiving has become strategic and mission critical to organizations because it can help achieve the following goals:

- Archiving speeds up business processes, such as accounts payable, claims processing, self-service scenarios, or supply chain processes, even those involving external partners.
- Archiving supports your company's compliance efforts. It supports the discovery of information for litigation, and allows you to abide by new regulations designed to ensure the integrity of financial control and reporting.
- Archiving forms the basis for records management (automating the management of record archiving and retention policies).
- Archiving saves costs in the IT department, such as for archiving transactional data from business applications, legacy system decommissioning, consolidation of archiving landscape, and file system archiving.
- Archiving preserves your company's knowledge and makes accessible such vital items as construction drawings and drug development documents. Archiving also enables review of previous work when starting new projects.
- Archiving enables your company to handle the challenges of modern communication (e.g., email archiving, email retention, access to historical messages).

How archiving is different from backup

An overview of content archiving is not complete without a brief explanation of the difference between *archiving* and *backup*. A common misperception is that archiving is just another word for backup, when in fact they serve different purposes:

- **Backup:** The function of backup is to create a duplicate copy of primary data in order to protect that data against loss due to hardware failures, user errors, or data corruption. With backup, a copy of production data is stored in a low-cost format such as tape and often warehoused offsite. Retrieval of historical data from tape backup is a slow, tedious process. Typically backup covers only a limited period of time, for example one to three years.
- **Archiving:** By contrast, archiving is a systematic, intentionally designed process of securely storing valuable content in an



unalterable and tamperproof form over a long period of time. Long-term storage can be just 7 to 10 years but might well extend to 30 or 90 years and more. The archiving solution secures the accessibility and readability of content during the entire lifecycle and, through replication and distribution, protects it from loss in the event of disaster. The archiving is done in such a way that data can be searched and specific content found and accessed quickly when needed. One reason for retrieving archived data is to meet audit requests or a legal discovery request in the event of pending litigation or the threat of litigation.

The three cornerstones of modern archiving

Connect: Enable efficient standard processes

Archiving has never been just a matter of storing data without creating access to it. Back in the days of the paper trail, organizations received invoices and other hardcopy documents, manually transported them along the process, and then archived them at the end of the journey.

However, in today's digital world, that workflow is reversed—and vastly improved because of it. Paper documents get archived right from the beginning and are immediately brought back to the process as electronic documents. Significant savings in time and money are the result. For example, the verification of a manual invoice can take a company three to four weeks to process. With an electronic archive, that process is shortened to three to four days! Organizations reduce 50 percent of their costs per transaction and benefit from vendor cash discounts.

The benefit of geographically independent access pushes efficiency as well. Electronic access to documents is critical for setting up shared-service centers. These centers centralize business administration functions, such as vendor invoice management or HR functions, and lead to productivity improvements and cost reductions.

Comply: Control all content and reduce risks

Archive Server can reduce risks in many ways. It protects content from disasters (such as flooding and fire) that can bring a business to its knees. For example, Hurricane Floyd destroyed all the paper



files of a leading company in precious metals and materials technology with 9,000 employees. Fortunately, every one of their technical drawings was available in Open Text Archive Server, so the natural catastrophe never interrupted production. Archive Server also reduces the corporate risks of being non-compliant with regulations such as SOX, SEC 17a, or CFR 11 compliance. And the solution can be used to implement records management and handle document retention cycles.

Records management legislation governs the content of information in question but not how it is stored, communicated, or conveyed. That means all forms of electronic communication, such as email and instant messaging, are covered. However, due to the sheer volume of email, and the unprecedented rate at which it's being adopted by organizations, email represents a leading information and risk management problem.

Failing to archive email information can lead to major losses. For example, a major tobacco manufacturer was fined \$2.75 million U.S. in August 2004 because key executives there did not comply with a court order to retain emails relevant to pending litigation.

Retention management is another aspect of compliance. Deleting information has become as important as archiving information. Applications such as SAP or Open Text Records Management add retention schedules to archived information. Together with these applications, Open Text Archive Server ensures that information is archived as long as needed, as well as controlled, deleted, and physically destroyed where permitted by legal regulations.

Cut Costs: Low TCO at highest document volumes

Archive Server also stores historic transactional data coming from enterprise applications such as legacy systems or SAP systems. With SAP best practice for data archiving, you can manage the growth of your SAP database while securely storing the archived data on the Archive Server. This means that companies can reduce their Total Cost of Ownership (TCO) by reducing the size of the SAP database and thus administration effort and hardware consumption. TCO can also be reduced through the consolidation of legacy systems into a leading application like SAP. And Archive Server is embedded in the Open Text ECM Suite, making it versatile to use at the lowest TCO.



Customer examples

Here are just a few of the high-volume document management customers of Open Text who have benefited from using Archive Server:

Customer	Highlights
Large international bank	<ul style="list-style-type: none">• Scans 120,000 documents / 300,000 images a day• Manages more than 16 TB of data
German mail company	<ul style="list-style-type: none">• 50,000 new scanned documents a day• 80,000 EDI documents a day• About 10 GB of new documents each year
Large U.S. Retailer	<ul style="list-style-type: none">• 18.2 million incoming invoiced each year—70,000 per day• 8 new 4,6 GB WORMS per day• 2.7 TB of new documents each year
Large German automotive supplier	<ul style="list-style-type: none">• 3.1 million incoming invoices each year, 12,000 per day• 150 million pages archived in the HR department• 250,000 pages of pay slips every month



Why Should You Use Open Text Archive Server?

Archive Server is a scalable and integrated service for archiving all of your enterprise content. That content is archived on a secure document repository, giving you the peace of mind that all documents are safely stored for years, yet are still instantly available when needed.

Connect: Enable efficient standard processes

The work on standard processes, such as those found within administration departments, requires precision along a fixed business process, as well as across functional departments or regional sites. Legal regulations may also require that processes within such departments are fully traceable and auditable.

Many employees involved in standard processes spend 80 percent of their time retrieving, inputting, and sending information, then waiting for follow-up information. Working on standardized business processes is characterized by excess paperwork that mirrors inefficiencies, errors caused by processing bottlenecks, media interruptions, delays, duplications, and inaccuracies.

Typical scenarios for such standard business processes include the following:

- Accounts Payable processing means dealing with the inevitable volume of related paperwork. Because manual processes are resource and time-intensive, they increase costs, create employee inaccuracies, and decrease efficiency.
- Accounts Receivable processing revolves around payment collection. The longer a payment remains outstanding, the less likely it will be paid in full. Employees in Accounts Receivable departments need supporting information from multiple sources to reconcile differences between customers and invoices. They must review account information generated throughout the organization before they can effectively field customer questions, troubleshoot problems, and facilitate timely collection. Lacking instant access to customer documents translates into inefficient and delayed collection, increased bad debt, and erosion of bottom-line revenue.



- Human Resources departments need to store employee records for many years for active and retired employees. Paper-based records require huge efforts for manual processing, and many employee records are incomplete or faulty. Paper-based storage, access, and manual routing result in high costs, long processing times, and decisions based on incomplete or faulty information. The average annual cost of manually handling one employee record is \$15 to \$30 US; retrieving a misfiled document can cost as much as \$120 US.
- Order processing requires fast and standardized reaction on incoming orders. A market differentiator for companies is the speed of the order-to-delivery process. Manually handling order documents, such as incoming faxes, reduces the speed for this critical process.
- Key functions within the insurance industry include claims processing, underwriting and sales, all of which are standard processes accompanied by a huge volume of documents. Manually managing these documents translates into inefficient core processes.
- During the sales process, failing to have the complete customer folder available leads to fewer opportunities to leverage cross-selling.
- Self-service scenarios could save the company much of the cost of sales if customers were provided access to such information as complete delivery and invoice history. In the case of a pharmaceutical wholesale company, the use of self service reduced customer inquiries substantially.
- Other standard business processes that inherently deal with manual paper processing include Contract Management, Quality Management, Customer Complaint Management, and Product Lifecycle Management.

The Solution

Archive Server, a core component of the Open Text ECM Suite, provides services for effectively taking in content, integrating content into leading applications, and functionalities for securing and auditing content and its access. All these basic services are mandatory for making business processes faster. Handling paper documents electronically speeds up processes by giving users a way to instantly access any business document—no matter how, where, or when!



For example, Archive Server is integrated in leading applications like SAP or Groupware (Microsoft Exchange®, Lotus Notes), stand-alone or other applications—ideal for back-office processes. Archive Server also allows for geographically independent access, in the case of larger companies. And distributed companies benefit since Archive Server enables online, parallel access 24 hours a day, 7 days a week.

- Standard business processes with a high throughput of documents profit the most from Archive Server. Faster Accounts Payable processing ensures vendor discounts and good vendor relations. Automatic invoice capturing reduces transaction costs by half (typically \$4.50 US per transaction).
- Faster payment collection in Accounts Receivable means that invoice disputes can be clarified faster. This reduces Days Sales Outstanding and decreases bad debts, so the need for bank credits is decreased.
- Employee records in HR are stored securely, completely, and without the need of a physical storage place. Administration processes like applicant processing runs faster and nearly eliminates the manual workload for HR employees. Self-service scenarios for employees reduce costs and administration efforts (such as SAP ESS).
- Incoming order processing works quickly and along a standardized workflow. This guarantees fast and standardized reaction on incoming orders, such as ensuring same-day delivery.
- Manual paper processing in claims processing can be automated with a high-volume document management system. Employees get the complete customer file in seconds instead of days.
- Because of implementing a complete electronic customer folder, it's simple to set up self-service scenarios where customers can look into their own documents for resolving simple enquiries.

Functionalities

Integrated into standard processes

Many standard processes are covered by Enterprise Applications like SAP. Open Text ECM solutions integrate into these applications and speed up processing through instant access to all relevant documents in the context of the business transaction in the leading



ERP/CRM system. Details on how Archive Server integrates with leading business applications are described in the last section of this white paper, covering the following:

Functionality Deep Dive

Integration with the Open Text ECM Suite	Integration with business applications
Integration with groupware systems	Archiving from Customer Solutions

Enterprise-wide deployments

Standard processes do not stop at the border of a company’s site; they run across the whole enterprise globally or involve partners. Many companies set up shared service centers, such as for Accounts Payable processing. Such centers require that the underlying business document technology can be accessed across many business sites and even countries. Also 24-hour access is mandatory when business takes place worldwide.

Details on how Archive Server can be deployed globally are described in the last section of this white paper:

Functionality Deep Dive

Build for enterprise-wide deployments	Scalability and ability to distribute
Based on defined standards	Vendor with a complete solution
Strong focus on compliance and security	Always available through various backup scenarios

High-volume management

The fastest ROI results from using ECM in mission-critical business processes, which is also where masses of transactions take place. Archive Server is prepared for this by supporting priority document volumes.

For example, the core processes of a retail company are buying and selling. Some customers receive as many as 60,000 invoices per day and generate over 8 TB of data every quarter. In the insurance industry, millions of documents need to be stored for many years and still be instantly accessible, such as during claims processing.



Details on how Archive Server supports high-volume management are described in the last section of this white paper:

Functionality Deep Dive

High-volume management	Access speed optimization
Optimized data handling	Caching

Comply: Control all content and reduce risks

Today's business environment is more complex and regulated than ever and for good reason. Corporate issues involving fraudulent accounting, misconduct, and data quality issues frequently dominate news headlines.

CEOs and Boards of Directors are under public scrutiny and, as a result, regulatory requirements have emerged to address these issues using commonly accepted principals of corporate governance.

Compliance may be defined as conforming to a rule. Types of compliance rules include the following:

- Government legislation and regulations
- Industry standards
- Internal company policy and procedures

Typical regulatory bodies include government agencies created to enforce legislation, industry standard bodies, and corporate directives. Compliance with the rules can be required or voluntary.

As a result of these rules, companies across all industries face an increasing need to make compliance an integrated part of their document management processes.

Consider the following questions:

- Does your organization store regulated documents, such as accounting documents, policies, standard operating procedures, and even email communications, in a document management system?
- Are you appropriately storing electronic documents that may contain evidence for a future dispute?



- Can you be certain that electronic documents will meet admissibility standards and requirements in the event of litigation or a regulatory audit?
- Are you following established and proven best practices for electronic documents retention?

Accepted guidelines and best practices on electronic document handling define how electronic documents should be handled to ensure evidential weight both in court and when under regulatory audit.

The Solution

Archive Server provides a host of functionalities to address compliance requirements associated with electronic document retention for many regulations, including SEC, Sarbanes-Oxley, FDA, GOB, and GDPdU. Most regulations place strict requirements on corporations to manage content not only through its active lifecycle but also to retain it for many years or decades after its use. Archive Server thoroughly addresses these regulations.

Functionalities

Long-term readability and accessibility

Archive Server stores content on a medium that will be accessible throughout the required retention period. Archive Server can store content in any format. (Nevertheless, we recommend archiving standard formats, such as TIFF or PDF, rather than proprietary formats). Archive Server also supports the transfer to alternate storage media if required.

Details on how Archive Server ensures long-term readability and accessibility are described in the last part of this white paper:

Functionality Deep Dive

Retention management	Storage reorganization
----------------------	------------------------

Archive Server mainly stores content on a medium that is unalterable. Not only is tamper-proof storage ensured but several functions prove that content has not been changed. Details on how Open Text Archive Server addresses long-term readability and accessibility are described in the last part of this white paper:



Functionality Deep Dive

Secure, long-term archiving and data integrity	Secure data transport
Digital signatures	Timestamps

Retention handling

Retention handling defines and handles retention of documents and data on the basis of a corporate-wide policy. Archive Server provides retention handling functionality that allows a leading application, such as SAP Applications or Open Text Email Archiving for Microsoft® Exchange, to define and manage the lifecycle of archived documents and data.

Retention handling must address a complex environment. Data can be actively referenced by different sources, and it may be archived for governance reasons. Data should be disposed of intelligently once it has lost its business value.

Retention handling must address this complex environment to ensure that business values and business risks are managed concurrently. Once a document’s active processing phase is complete, its classification determines the rest of its lifecycle. At this point, the technologies that automate records retention and destruction come into play.

How long should the records be saved, and when can they be safely destroyed? Companies often have written policies on document retention. They should define and document policies for records management and ensure that the policies are implemented and maintained at all levels in the organization.

One would think that since almost all documents are now electronic, control and access would be a snap. Sadly, that is not the case. Electronic records exist in many different locations, both on-site and remotely. Employees are accessing and storing records electronically at home and even on handheld computers. Document retention policies are difficult to create and even more difficult to enforce.

For example, when one of the largest US software vendors was fighting its anti-trust case with the US Department of Justice, the prosecution was able to bring forth emails that had been circulated between employees as evidence of anti-competitive business



methods. Had the company been more diligent in enforcing its records retention policies, those emails might have been legally destroyed.

Details on how Archive Server handles retentions are described in the last section of this white paper:

Functionality Deep Dive

Retention management

Controlled deletion

Each type of document has its retention periods. After expiration of the retention period, content must be deleted. Combining Archive Server with the optional Records Management module ensures that content is deleted. In addition, associated content, such as meta data and indexes, is securely deleted. Details about how Archive Server integrates with Records Management are described in the last section of this white paper:

Functionality Deep Dive

Integration into Open Text Records Management

Prevention from unauthorized access

Another aspect of security is protection from unauthorized access. For example, when Archive Server is used for offering hosting services, it's necessary that the hosting company (or its customers) cannot access any foreign content. Data encryption helps protect privacy and content.

Archive Server ensures that content is not accessed by unauthorized or inappropriate individuals.

Other security issues involve protection from unauthorized access during transmission of content via networks and protection against re-usage of URLs. The technologies secKey and SSL help to protect from these risks.

Details on how Archive Server prevents unauthorized access are described in the last section of this white paper:

Functionality Deep Dive



Authorization and authentication	Encryption of the stored data
Secure data transport	

Audit of all activities

Archive Server can retain detailed records of all activities performed on content stored in the archive, including the date and time. Details on how Archive Server facilitates audits are described in the last section of this white paper:

Functionality Deep Dive

Logging	Auditing – long-term traceability
---------	-----------------------------------

Long-term storage

Long-term storage surely is the basis for compliance of ECM systems. However, companies also benefit from long-term storage from the standpoint of process efficiency. Insurance contracts, technical drawings or outgoing customer correspondence (such as utility invoices) often need to be accessed during customer complaint management or repairs. Fast access to archived documents helps in solving business problems or repairs faster.

Details on how Archive Server provides long-term storage are described in the last section of this white paper:

Functionality Deep Dive

Secure, long-term archiving and data integrity	Storage resource management
Powerful management services	Volume migration

Digital signatures and timestamps

Digital signatures have two aspects of usage: fulfilling compliance (authenticity of content, securing evidence) and speeding up standard business processes by emulating personal signatures.

In some industries, it is necessary that documents are signed in order to complete a process step. If such processes have a high document throughput, it makes sense to replace manual paper processing with electronic document processing. Archive Server can store personal signatures together with business documents.



Details on how Archive Server addresses the compliance issue of digital signatures are described in the last section of this white paper:

Functionality Deep Dive

Digital signatures

ArchiSig – Signature renewal for long-term digital signature

Cut costs: Low TCO at highest document volumes

When companies choose a document solution, the driver is usually a specific problem to solve. We call it “point solution.” In many cases, it is the accounts payable process. Over time, when such a document solution proved its value, other departments get interested in this technology as well—for example the Sales department to create a central customer folder with all incoming and outgoing customer and project documents. Legal regulations also force companies to store critical content, such as emails with relevance to business transactions. In other words, one customer could run several point solutions from different vendors.

Each of the point solutions needs to be implemented, connected, and administrated. Hardware needs to be purchased to support all the different applications. Know-how must be maintained for all the point solutions. If companies added up all the costs for their point solutions, they’d discover that it makes sense to reduce TCO. Many different point solutions cost more than a single, scalable ECM. As successful ECM implementations grow rapidly, some factors to consider are

- The system must be highly scalable in volume and regional presence.
- Flexibility in changing hardware technology or vendor is required in order to adapt to most recent reliable technology.
- An ECM system integrates into a heterogeneous environment (e.g., databases, operating systems, business applications, scanning-, storage-hardware).
- Accounting: Who uses the system to what extend—especially when the ECM services are running centralized at an IT service provider?



The Solution

Archive Server is a part of the Open Text ECM suite, which allows companies to be flexible in using this infrastructure—either for a point solution for a specific department or as the basis for many solutions running on the ECM backbone. Existing point solutions from other vendors are simple to migrate to Archive Server.

This reduces TCO because of reduced administration, know-how, and hardware costs.

Part of a whole ECM Suite

Archive Server is part of the Open Text ECM Suite. All Open Text solutions are based on this repository when it comes to the archiving of content. Customers profit from an integrated solutions suite that fits to all ECM-relevant needs.

Functionalities

The needs for archiving can differ. Details on reasons can be found in the last part of this white paper:

Functionality Deep Dive

Simple to integrate and connect	Document Lifecycle Management Services
Strong focus on compliance and security	

Scalability

Archive Server provides archiving and storage management capabilities for all applications that plug into the Open Text ECM Suite framework. This large-scale integration enables you to save costs by using the same archiving framework and capabilities for all of your enterprise content. Even if you originally deploy an Open Text ECM Suite-based solution for the purpose of archiving email and attachments, you can seamlessly extend that solution to quickly and cost effectively archive SAP content and any other type of enterprise content.

True enterprise scalability means that organizations can extend a system in any dimension—whether by geographic distribution,



number of users, or volume of content. Archive Server scales in each of these dimensions.

Details on how Archive Server provides true enterprise scalability are described in the last part of this white paper:

Functionality Deep Dive

Usage in heterogeneous environments	Scalability and ability to distribute
Caching	Access speed optimization

Accounting

If Archive Server runs at an IT service provider, powerful accounting functionalities make sure that the usage of the system can be billed to the right departments in a transparent and fair manner.

Accounting is required and used to reflect the usage of documents and scenarios. Application Server providers, as well as outsourced IT departments of large companies, need statistics about accessed content and billing. Through this, document storage and document retrieval can be charged.

In addition to the accounting data, usage statistics related to performance monitoring can be gathered and used to optimize the system performance.

Details on how Archive Server provides powerful accounting functionalities are described in the last part of this white paper:

Functionality Deep Dive

Built-in Accounting functionality

Optimized storage management

Archive Server virtualizes storage and accessibility, which increases flexibility in storage management by using your choice and combination of storage hardware. Typically, the lifetime of business documents exceeds the lifetime of storage hardware. Compatibility with all major storage providers ensures that companies can seamlessly migrate content to alternate storage media in the future.

Details on how Archive Server optimizes storage management are described in the last part of this white paper:



Functionality Deep Dive

Hardware abstraction	Storage reorganization
Optimized data handling	Logical archives

Content protection and availability

If highly critical or worldwide processes rely on content provided by the Archive Server, it is essential to provide access to the content 24 hours a day, 7 days a week. For these requirements, the Archive Server supports high-availability deployments.

Archive Server supports replication and distribution scenarios, so that data sets can be kept redundantly for additional safekeeping. For example, to help safeguard against the risks of physical disasters and environmental instability, redundant data sets can be stored in multiple physical locations.

Archive Server doesn't just archive content: it also affects how that content is organized. By retaining information about the hierarchy of data, Archive Server can rebuild not just the content itself but also the structure of the information store. The administration interface facilitates disaster recovery processes where administrators can reconstruct Archive Server from storage media.

Details on how Archive Server provides content protection and availability are described in the last part of this white paper:

Functionality Deep Dive

High availability	Remote standby
Disaster recovery	Backup

IT landscape and know-how

Archive Server solution was designed for use in heterogeneous IT landscapes. It supports a wide range of operating systems, databases, and storage hardware, including the following:

- Operating systems:
 - Microsoft Windows Server, Sun Solaris, HP HP-UX, IBM AIX, Novel SUSE Linux, Red Hat Linux
- Database systems:
 - Oracle, Oracle RAC, Microsoft SQL Server



- Storage hardware:
 - Hard disk Write Once media (NAS, CAS)
 - HSM Systems
 - Optical media in jukeboxes

Administration and monitoring

Central administration and monitoring of the server and storage functionalities simplifies the lives of administrators. The Administration Server of Archive Server is used to managing and configuring the system components. The entire archiving system can be managed locally or remotely via the user-friendly administration client.

The Server Monitor checks the availability of system resources and monitors the activity of the individual archive components. It is used proactively to quickly detect problems and pinpoint the source of any errors. The Server Monitor client can also be used remotely via a Web-based client.

Details on how Archive Server simplifies the work of administrators are described in the last part of this white paper:

Functionality Deep Dive

Administration server	Server monitoring
-----------------------	-------------------



Functionality Deep Dive

What is Open Text Archive Server?

Open Text Archive Server is a core component of the Open Text ECM Suite and constitutes the foundation for enterprise-wide ECM solutions.



Archive Server enables storage, management, and retrieval of archived data and documents.

Open Text offers customers several connectors to expand the functionality of Archive Server. These connectors allow you to manage business documents in different applications and to link them to the business processes. For example, with Open Text Suite *for SAP Solutions*, users can access all data and documents they need to process a business transaction in the SAP business suite. Furthermore, Archive Server provides general server interfaces for integrating new or customer-specific applications.

Why is an archive server important in an ECM environment?

Information technology and its environment are prone to change in storage technology, organizational structures, consolidated applications, compliance rules, etc.



If we look back 10 years at how archive systems have been structured, and compare that with an up-to-date ECM system, we realize that the complexity has increased, especially in terms of applications using archive systems and available storage systems. Some storage technologies are not used anymore for long-term archiving, such as Microfiche and CD.

Archive Server provides extensive functionality for data migration and hardware abstraction in order to give corporations the required flexibility in their hardware strategy for decades.

In parallel, together with the increasing significance of electronic documents, new legal requirements have emerged and compliance rules need to be enforced. Archive Server offers technology for implementing retention management and electronic signatures.

A leading application is one that generates archived documents (such as print lists in SAP) or with whose business objects the archived documents are linked (e.g., inbound documents in SAP). SAP, Open Text Content Server, Microsoft Exchange, Lotus Notes, and Microsoft SharePoint can be linked as leading applications.

Companies are increasingly aware that leading applications change. However, the documents referenced by those leading applications may have to be kept over long periods of time—sometimes even 20 years. In order to stay independent from leading applications, customers choose to archive sufficient metadata with the documents. In case that a leading application is being discontinued, the metadata ensures an easy migration path.

Architectural overview of Archive Server

Archive Server is the central unit, providing much of the document management system functionality, the storage capability for documents and data, and the central archiving functionality.

Archive Server comprises multiple services and processes, amongst which the Storage Manager, the Document Service, and the Administration Server are the most important ones. The Storage Manager is responsible for storing documents and data, whereas the document management functionality, the storage of metadata and other properties, and the entire communication is done by the Document Service. Client applications “talk” to the Document Service. (In the following, these two sub-services are referred to as Archive Server.)



Depending on the business process, the document type and the storage media, Archive Server uses different techniques to store and access documents. This guarantees optimal data and storage resource management. Mass data, which is not changed anymore, can be stored as ISO images. The Storage Manager provides access to ISO images within a physical or virtual jukebox. Content that is prone to change and has an individual lifecycle will be stored as single file.

More complex Open Text ECM Suite implementations may consist of several Archive Servers, for example, to reduce access time in large—possibly worldwide—networks or to improve reliability by mirroring an entire Archive Server. If an Archive Server acts as a mirroring system of another server, it is called a Replication Server. Additional Cache Servers complement these servers to a complete, worldwide storage landscape.

Archive Server incorporates the following components for storing, managing, and retrieving your data and documents:

- Document Service - Controls the storage and retrieval of the individual components.
- Storage Manager (STORM) - Controls the storage devices.
- Document Pipeline - Used to transport and process the data and documents to be archived. (The Document Pipeline is optional.)
- Cache Server - Speeds up the access to the archived documents. The Cache Server is optional and used in ECM environments, mostly with worldwide, distributed departments, and low network bandwidth. The Document Service itself contains a service to cache content from slow media like WORMs.
- Administration Server - Allows the Administrator to create and maintain logical archives, physical devices, etc.

In addition, Archive Server supports COLD (Computer Output on Laser Disks) and archives COLD and spool data from host systems. The Document Pipeline controls data processing and archiving.



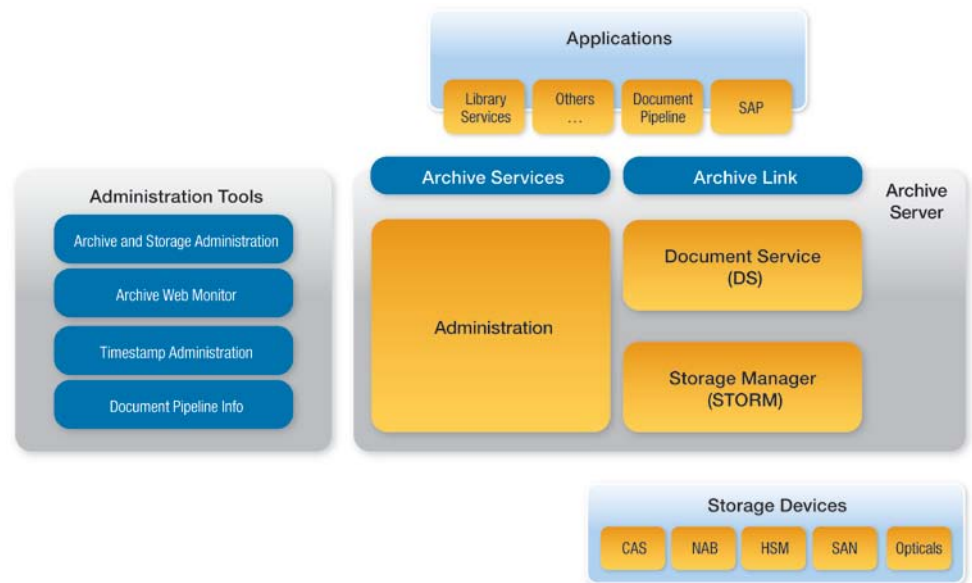


Figure 5: Open Text Archive Server Architecture

Document Pipeline: The conveyor belt into the archive system

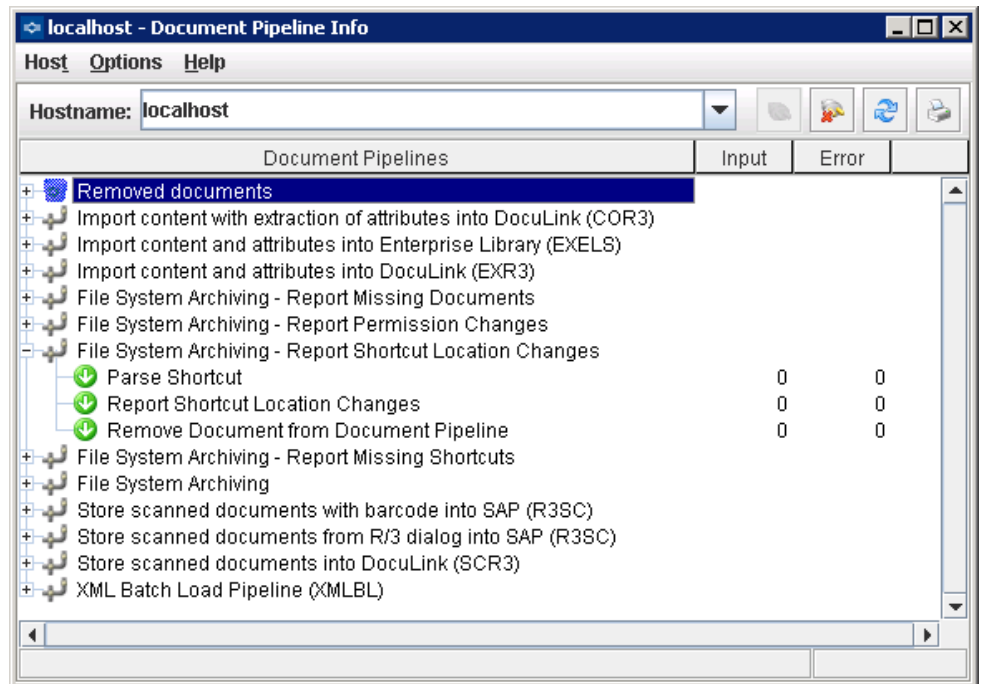
A Document Pipeline is the basic component in almost all document processing software and is used, for instance, to transfer documents to a storage system or another application while performing certain additional tasks. Speaking figuratively, a Document Pipeline is the conveyor belt that transfers the documents through the software.

Individual tools (called DocTools) retrieve the documents from the conveyer belt, process them one by one, and then return them to be processed by the next tool. The last tool in the pipeline generally removes the document from the conveyer belt. Depending on the configuration, Document Pipelines can contain various different DocTools to implement all different kinds of document processing, and further tools can be added as required.

An application called “Document Pipeline Info” displays the status of all document pipelines and their DocTools. In the picture below you see the status of the document pipeline “Import content and attributes into DocuLink, which adds documents to the Content Server application known as Open Text DocuLink for Content Server. Currently no documents are being processed and none are in the Error queue, indicated by the zero in the columns on the right-hand side.



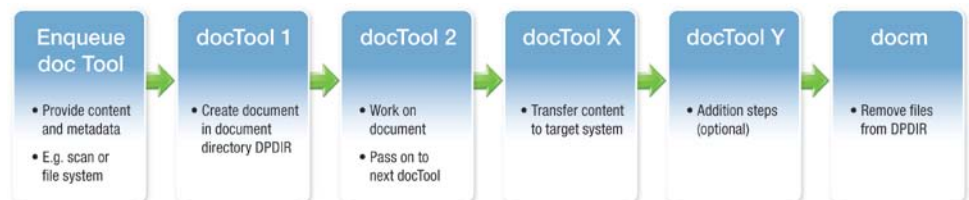
Figure 6: See the status of everything in the Document Pipeline



A scenario in which the document pipeline plays a central role is “batch import of print lists or document lists with form overlay and attribute extraction.” On its way through the specifically configured document pipeline, each document has its attributes extracted. In the next step, a form is assigned to document lists. After the form has been stored by Archive Server, the document list is stored together with a link to the form. Finally attributes are stored with the leading application. When users retrieve a document from a document list, it will be displayed together with the assigned form and dramatically improve usability.

Document Pipelines are available for all major target systems: SAP, TCP, Content Server, Enterprise Library, File System Archiving.

Figure 7: Document Pipeline for batch import with attribute extraction



High-volume filing

An important principle for all Document Pipelines is that processing is always transactional. That means the processing status of the document is always defined: either it has been processed by a specific DocTool or not, and no documents can get lost. If for any reason the Document Pipeline is aborted or processing is cancelled at any time, the document is considered to be unprocessed by the last active DocTool. The current status is retained at all times. Therefore, when the Document Pipeline is started again, processing can continue at precisely the same step the document was at when the program was aborted. This re-entrance provides the security required for high-volume filing.

Some customers archive as many as one million documents within 10 hours. The high-volume filing capability of Archive Server allows large migration projects to be conducted. At a large German bank, the migration of a decommissioned archive system included 160 million documents, 1,000 online and 1,800 offline partitions, which make a total of 2.8 TB online data and 1.8 TB offline data.

Based on defined standards

Archive Server uses established standards to help protect your investment, running on various Windows Server versions and all major Unix versions (including Linux). The archive database can use an Oracle database or Microsoft's SQL Server. It also supports hardware platforms from the leading storage vendors (e.g., NetApp, EMC, HDS, SUN, HP, etc.).

Archive Server stores any content, regardless of its format. Storage of some forms of content is trimmed to optimize the use of storage space or document access, e.g., outgoing invoices that may be numerous but very small. Open Text Content Server based applications come with a set of clients for imaging and displaying documents.

These clients support existing imaging standards such as TIFF, JPEG, and PDF, as well as SAP formats such as OTF, ALF, and ADK. All the desktop applications and the different Windows clients use the Open Document Management API (ODMA) to communicate with the archive system. The ODMA interface allows for seamlessly integrating most applications with the business document system.



Vendor with a complete solution

Leading analysts say that maximum-access protection is achieved when the protection mechanisms of the archive system, the Document Management application and the viewer are configured to work together. In a complete solution from one vendor, this integrity is already built in from the beginning.

One vendor, Open Text, offers a complete ECM suite that is integrated in all components. This reduces TCO and enhances stability and security of the system.

Build for enterprise-wide deployments

Archive Server is built for enterprise-wide deployments. This, in turn, means Archive Server has

- Strong capabilities in the sense of scalability in document volumes.
- Strong capabilities in the ability to distribute the system to all business regions.
- Flexibility to run the system on existing databases and operation systems.
- Flexibility to connect the system to existing or new storage hardware.

Scalability and ability to distribute

Archive Server client/server architecture provides versatile options for configuring, scaling, and distributing the archive system. For example, it is equally possible to install multiple Archive Servers to form a large, distributed archive system as it is to manage multiple logical archives on a single Archive Server. In addition, no matter how large the distribution, it is possible to centralize system administration and management.

A logical archive may represent an individual content lifecycle with individual storing properties like compression, single-instance archiving, and the possibility to purge the content after retention has expired. The pools attached to a logical archive may represent different storage media that accommodate the storage requirements of individual content. This allows for implementing customer-specific storage hierarchies and content lifecycle management.



For instance, two logical archives are created, one to store contracts and another one for personal signatures. Personal signatures need to be accessed very fast, thus the logical archive should be attached to an HD. Contracts need to be stored on a save medium, which ensures they cannot be manipulated. Thus, the logical archive for contracts may be attached to a device with WORM support. Furthermore, retention periods may be different for individual document types. By assigning and explicitly naming logical archives or pools to individual fiscal years, the administrator is given an immediate overview on retentions.

Archive Server can adapt to changing business needs flexibly and cost effectively. Archive Server scales both vertically by adding additional worker processes and horizontally by load balancing. As the number of users grows, it is possible to connect new clients to Archive Server or to install additional Archive Servers or Cache Servers.

Usage in heterogeneous environments

Support of operating systems and database systems

Archive Server is designed for use in heterogeneous IT landscapes and runs under Windows Server operating systems, all major Unix versions, and hybrid Windows Server / Unix environments.

Support of storage systems

Archive Server virtualizes the storage layer and, because of this, Archive Server was designed to support a wide range of different storage technologies and storage vendors. Archive Server supports the following:

- Hard disk Write Once media with WORM feature and retention handling
- HSM systems
- Optical media in jukeboxes
- Cloud storage

For details, please see the Archive Server Storage Platform Release Notes.



Simple to integrate and connect

Integration into the Open Text ECM Suite

Integration into Open Text Content Server

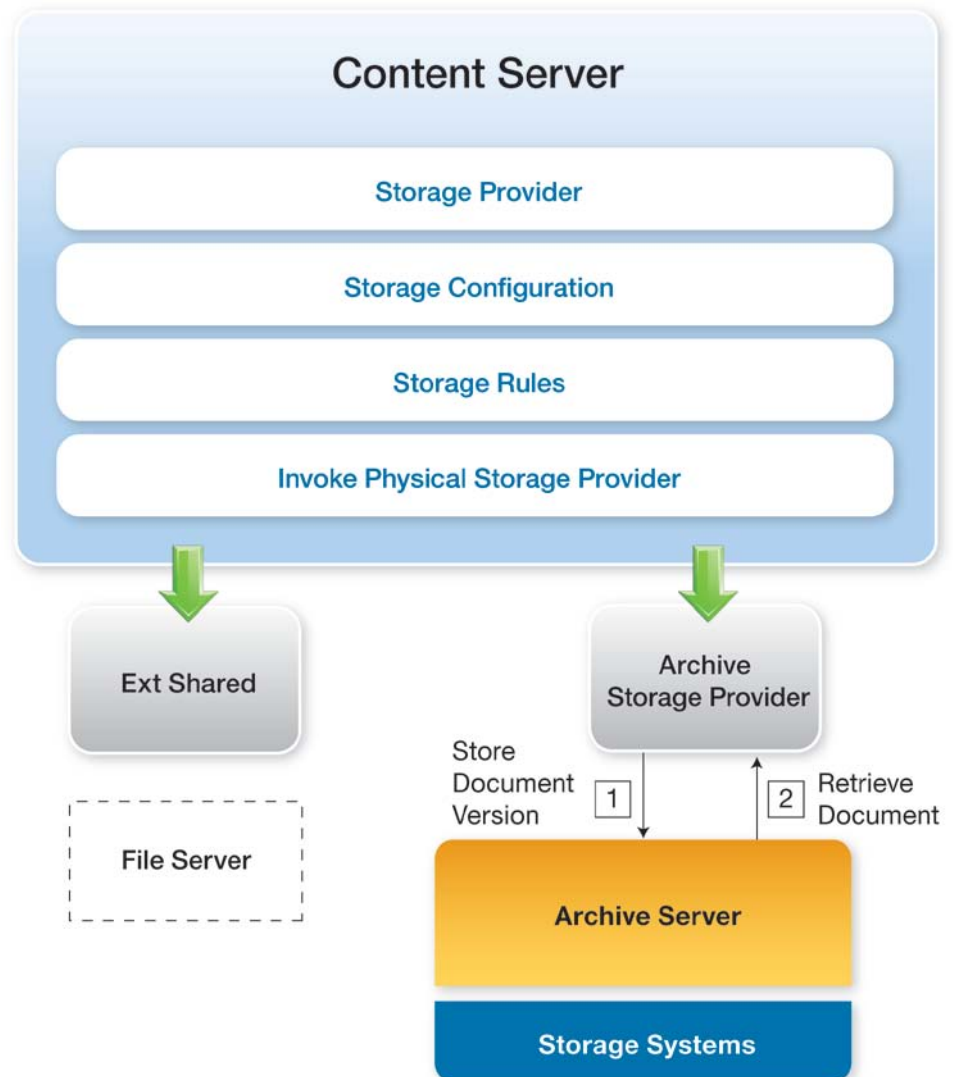
Open Text Content Server is the leading collaboration and content management software for global organizations that brings together people, processes, and content. The information managed within the Content Server can be safely stored with Archive Server. For this purpose, a software option called Open Text Archiving for Content Server has been developed.

Open Text Archiving for Content Server adds the full capabilities of Archive Server—including compliance—to the Content Server. Organizations can deploy a robust solution for managing content throughout the entire ECM lifecycle—from creation through publication to archival and eventual deletion.

The process is completely transparent to end users. The user creates document versions within the Content Server. The document itself is stored on Archive Server, from where it is quickly and reliably accessible. No user interaction is required to store documents on Archive Server. Based on rules, configured by an administrator, the system decides which storage provider is used upon document creation. Multiple logical storage providers can be configured, each related to a logical archive. The logical archives in turn refer to specific storage locations. While storing documents with Archive Server, they are automatically full-text indexed.



Figure 10:
Integration of Archive Server
into Open Text
Content Server



Integration into Open Text Transactional Content Processing

Open Text Transactional Content Processing is a document management system that supports structured and transactional DMS and workflow requirements such as the following functions in a high-volume, often business-critical environment:

- Capture (high-volume imaging, indexing, reports and print lists, faxes, office documents)
- Process (file/search/retrieve folders and documents quickly and easily in a specific business context or as part of a structured business process with BPM)
- Preserve (manage document lifecycle and retention periods)



- Archive (long-term secure storage on multiple hardware)

Documents are managed based on a strong and flexible metadata model.

Transactional Content Processing provides the tools to quickly and easily build customer-specific business applications such as customer folders in insurance, banking, utility, or a patient folder in healthcare. Production Document Management solutions quite often work closely together with leading applications such as SAP, CRM, or host systems. The various interfaces allow customers to build the integrations and thereby support the end user's daily business. The tight integration with Transactional Content Processing allows building document-centric process solutions to improve business efficiency and achieve faster response times. It leverages Archive Server as a highly scalable and secure repository for business critical data and documents and is designed for the complete range of business content and its lifecycle management.

Integration with Enterprise Library and Records Management

Records management is the practice of both retaining and destroying records, enabling organizations to

- Ensure that all information is retained for at least as long as it must be retained,
- Ensure that discovery requests and audits can be performed in an efficient and cost-effective manner (i.e., information can be reliably retrieved), and
- Ensure that information is destroyed on a consistent basis.

In managing the lifecycle of email, hard-copy documents, file boxes, and more, a company can provide litigation support, identify vital records, automate, and administer its corporate retention program efficiently. The organization can also apply descriptive metadata, ensuring integrity of business-critical knowledge and reducing risk due to audit, regulatory compliance, and litigation. By integrating Archive Server, you can ensure compliance and implement your corporate retention program through all layers of the application down to the hardware components.

The lifecycle of records can be managed transparently by Enterprise Library. It offers a holistic approach to manage the document lifecycle determined by the Records Management Classification and associated Record Series Identifier(s). Storage



Rules and Storage Tiers propagate the associated retention period down to the Archive Server.

General Desktop API

The ODMA enables the integration of standard document applications with Archive Server. Files can be archived in their original format from any ODMA-compatible application—such as Office. The “Save” menu item lets users save documents directly to the archive, without requiring special macros. In addition, users can print files via ODMA from any desktop application and store them in the archive in the long-term TIFF format.

The Imaging DesktopLink module, which is part of the Imaging package, can archive documents from any ODMA-compatible desktop application and integrate them, like with SAP business transactions, for example.

Integration into business applications

Integration into SAP NetWeaver

The integration between Archive Server and SAP is based on and certified for various standard SAP interfaces:

- SAP ArchiveLink™ Interface
- SAP HTTP Content Server Interface
- SAP ILM WebDAV Interface (together with other components of *Enterprise Library*)
- SAP Solution Manger Ready

The SAP ArchiveLink interface—developed in 1992 by SAP and IXOS, an Open Text company—is the most important communication interface between SAP and an external archive system. This standard SAP component allows for linking documents that Archive Server manages with SAP business processes and provides retrieval through SAP transactions.

The SAP HTTP Content Server Interface is the successor of the ArchiveLink interface and allows for connection to the SAP Knowledge Provider, which is used for SAP PLM and SAP DMS, for example.



The SAP ILM WebDAV interface is the successor of the SAP WebDAV XML Data Archiving Interface. The ILM WebDAV interface is used to manage the complete lifecycle of archived SAP data. Enterprise Library together with the Archive Server enforces the retention periods and holds, which are transmitted by SAP for the data archiving files and also for the attached documents.

The Archive Server is certified to be Solution Manager Ready. Even more so, it integrates in the SAP support infrastructures at a customer, which is based on Solution Manager. The Archive Server allows monitoring and root cause analysis via SAP Solution Manager Diagnostics. The Archive Server as well as the entire Open Text Enterprise Library can be deployed on SAP NetWeaver CE as Web Application Server.

All these integrations into standard SAP interfaces allow customers to leverage the document functionality of SAP in each and every SAP module. Also through the usage of these standard interfaces, Archive Server can be rapidly connected to SAP.

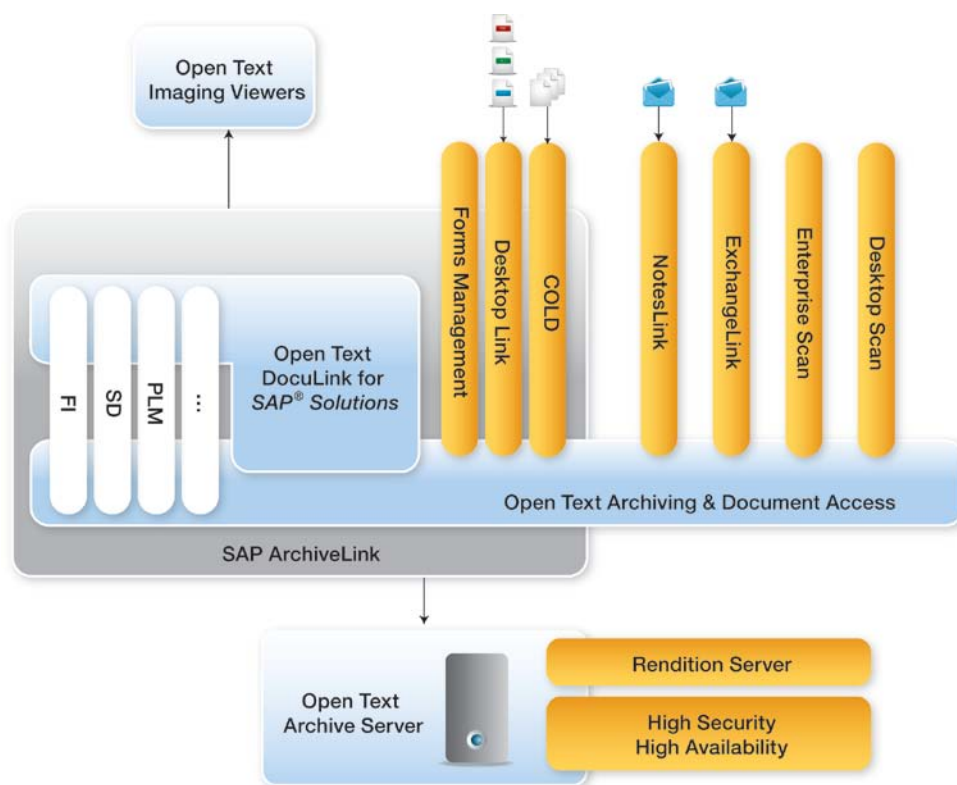
The add-on product, Open Text ECM Suite *for SAP® Solutions*, uses these interfaces to manage and archive all kinds of SAP documents, such as ArchiveLink and SAP Knowledge Provider (KPro) and including the following:

- Outgoing SAP documents (documents that were created by the SAP system, such as purchase orders, invoices, reminder letters, delivery notes)
- Incoming documents of all types (scanned paper documents, faxes, electronic documents of various formats)
- Print lists (generated by SAP system reports)
- SAP data archiving objects.
- Other SAP ArchiveLink and SAP KPro documents created by the SAP system or users in different SAP modules and applications.

Open Text also provides a comprehensive product portfolio for all document archiving and document management needs in an SAP environment:



Figure 11:
Overview: Open Text ECM
Suite for SAP® Solutions

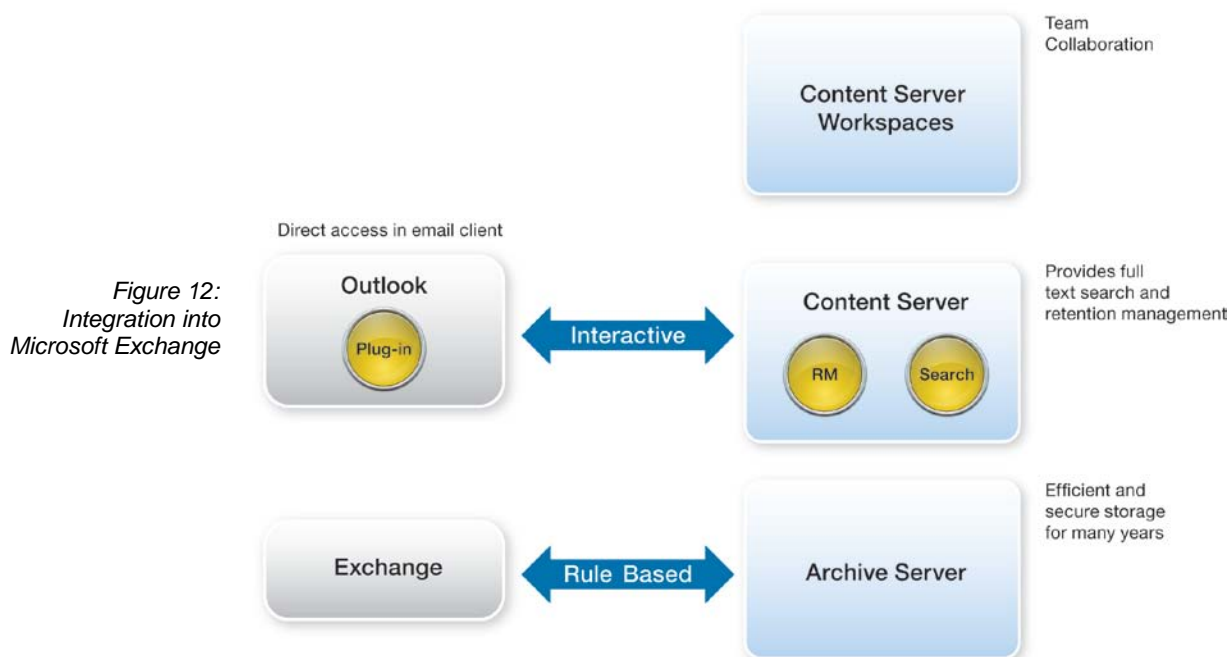


Integration into groupware systems

Integration into Microsoft Exchange

Email traffic becomes more and more complex in the daily business, which has led to a high volume of documents that must be stored in the email systems. Some of these documents even need to be stored for several years because of legal requirements, so deleting documents is not always a solution to save disk space. Furthermore, deleting emails can be time consuming and tedious.





An archiving solution helps save disk space on email systems, and integration into Microsoft Exchange speeds up operations. Integrating Archive Server into Microsoft Exchange considerably reduces the amount of data stored on the MS Exchange servers, enabling them to perform better. Significantly less hard disk capacity is required, resulting in additional savings. Backup and data storage activities are reduced, as well as the amount of administration time, by archiving emails, attachments, PST files, and public folders.

Your archiving environment can be customized: you may archive your emails automatically or interactively. Interactive archiving and the display of archived objects are based on MS Exchange custom forms. No extra software is necessary on MS Outlook® clients, whether for automatic or manual archiving.

To save even more disk space, we provide Single Instance Archiving. An attachment that several users want to archive is archived once only and is referenced with the individual emails.

Integration into Lotus Notes

With the add-on product, Open Text Email Archiving for Lotus Notes, it's easy to archive Lotus Notes emails simply by selecting



them and using the respective menu option or the Archiving Toolbar options.

It is also possible to automatically archive emails using predefined criteria such as time stamps, size limits, or Lotus Notes formulas. A single Lotus Notes email can either be archived completely or only the attachments. For multiple emails, Open Text Email Archiving for Lotus Notes offers the possibility to archive the emails separately or together as a structure of documents.

Via a Lotus Notes client or Domino Web Access, the emails can later be retrieved for viewing, restored to their original condition, or copied back to the Lotus Notes database. It is also possible to delete archived documents from the archive and to retrieve archived documents into a local replica.

Archiving from customer solutions

Open Text provides a general Archive Server API that enables customers to develop their own archiving solutions. The Server API is available for all supported server platforms in C and Java.

High-volume management

The value of ECM is strategic. It's an important factor in the organization's overall financial performance and a competitive advantage. An ECM system must be planned very carefully in order to meet performance requirements—especially, when the daily created volume of documents or the number of users is very large.

Storage need is growing exponentially. Archive Server is specifically designed to handle an ever increasing amount of documents, content lifecycle, and large numbers of users.

Archive Server provides long-term storage for high volumes of data. Since storage technology has a lifecycle, Archive Server takes responsibility for the reliable, seamless, and transparent migration of content from outdated storage to recent storage technology. Archive Server helps you to adopt your content and storage strategy to changing requirements and new technology in a cost-effective way.

Open Text customers already manage high volumes of data, as these figures show:



- Current document stock up to 15 TB, expected in the next three years up to 150 TB
- Current document volume up to 200 million, expected within the next three years up to two billion
- Current daily document filing up to 1 million

Access speed optimization

Archive Server grants fast (within seconds) and efficient access to very large document volumes—as much as billions of business documents handled by one Archive Server.

The scalability of Archive Server—horizontal as well as vertical—guarantees fast access to documents for large numbers of users. Load balancing horizontally scales the system. Vertical scalability is implemented by a configurable number of threads and connections. In addition, various caching mechanisms, designed to suit different business scenarios and system configurations, provide speedy access to documents—such as network and media caching, SSL session caching, and attribute caching. However, speed of access depends on the underlying storage medium. It also depends on the access technology used. The archive system provides a metadata layer that is used by leading applications to efficiently retrieve documents. The metadata layer is also used to organize the storage location by logical information (document lifecycle class) rather than only technical information (such as file size or last accessed date, like HSM systems).

Fast access is among the tasks of Archive Server, which require high performance. Archive Server fulfills performance requirements for filing (store), backup, replication, migration, deletion, and administration.

Optimized data handling

Container files

The size of business documents may vary from a few kilobytes up to several gigabytes, and both sizes challenge storage systems. Very small documents may waste much space due to big block size of storage media and decrease filing performance. Very large documents may exceed physical partition limits.



Furthermore, high-end storage systems and modern file systems cannot handle an unlimited number of files. Limitations are the number of files within one directory up to a total number of entries within the index of a storage system. Archive Server addresses these limitations with a special container file technology. Depending on the document type, the business scenario and the storage media, Archive Server supports several types of container files:

- Archive Server uses ISO9660-Images as container files. A container file may contain several thousands of documents but occupies just one index entry. This technique dramatically relieves the index of a storage system and increases write performance. ISO-images are best suited for mass data that will not change after it has been archived.
- MTA documents: The document pipeline may deliver thousands of records within one so-called MTA document. The retrieval of one record is transparent for the clients. This has been done to save both disk space and index database space and to dramatically increase write performance. MTA documents not only save index space on the storage system but also in Archive Server's database. They are best suited for large document lists that will not alter.
- To overcome partition limitations Archive Server stores big documents, up to 100 GB (tested), in several chunks.

Single instance archiving

Especially in groupware scenarios, identical documents may be at risk of being stored several times, if emails with attachments are sent to hundreds of recipients and all of them want to archive this email. Archive Server enables single instance archiving (SIA), keeping the same document only once in the connected storage devices. Dependent on the amount of expected redundancy of email attachments, SIA may reduce required storage space significantly.

Compression

In order to save storage space, you can activate data compression for each individual logical archive or content type. All important formats including email and office formats are compressed by default. Compression rates depend on file format and content and correspond roughly to gzip level 6.



Document lifecycle management services

The main focus of a content store used to be trusted storage—i.e., never lose a document. Deletion of documents was only an exception. Today, SOX and other regulatory acts bring new requirements to content storage and document management.

Content storage exponentially increases. Document lifecycle management, including retention management, becomes essential for an ECM system. When the retention period of a document expires, it occupies not only wasted space in a company's content store, but also its value may invert. Before a document expires, it is important not to lose it. However, after a document expires, it is important to delete it since it may give negative evidence.

Retentions handling

As physical storage may not allow immediate physical deletion or even physical destruction of documents.

Archive Server provides policies (depending on capabilities of physical storage) to logically delete a document immediately on request and to do the physical deletion or destruction asynchronous within a committed time frame.

Archive Server implements retention handling, not retention management. Retention handling enables a leading application to implement retention management. A retention period of a document defines a time frame, in which it has to be impossible to delete or modify this document. For compliance reasons, it's not enough to set a flag that enables the software to reject any deletion request against the document. The content of the document needs to be physically protected instead (or protected by a system supporting the WORM capability). So for a storage manager fulfilling regulatory compliance, this means that it is not sufficient to store the components with a specified retention period on a simple hard disk. Storage systems with a WORM capability have to be used.

After the retention period has expired, the document can be deleted. This has two aspects. The first is if an administrator wants to delete documents to get rid of old volumes. In this scenario, it is sufficient to delete sooner or later. The other aspect is that the content of a document could compromise someone. In this case, the document must be deleted immediately after the retention period has been expired.



Retention handling in Archive Server is designed as a top-down concept: a leading application sets the retention period in Archive Server. Archive Server, in turn, sets the retention period on the storage system. After the retention expired, the leading application has to trigger the purge of the content. Then, Archive Server triggers the purge of the files on the storage system.

A leading application may specify a retention period (and a retention behavior) during the creation and migration of a document. If nothing is specified, a default period and behavior is used, configured per logical archive within the administration client.

Retention management

Retention management is performed by the leading application that accesses Archive Server's Retention Handling functionality. For instance, Records Management requires classification, retention management, audit trails, and deletion of documents. Though most of these requirements have to be met by a records management application, Archive Server handles retention periods and keeps track of all changes on document content.

Furthermore, Archive Server provides logical archives and monitoring functionality for retention management. For example, all invoices from the current year are grouped together into a logical archive so they can be deleted after the retention period has expired.

Volume migration

Volume Migration is a very important function needed for a long-term ECM strategy and to assure compliance. Compliance requires not only the storage of documents in a safe place but also the need to purge them once the retention period has been expired. Therefore, Volume Migration is important to retention handling if documents are stored on WORM media.

For this purpose, Archive Server administration compiles a list with all volumes containing mostly expired documents. Numerous volumes with mainly expired documents can be reduced to a handful via automatic migration. When the migration is completed, the expired volumes can be removed or purged, thus saving jukebox slots or storage space, depending on the media.



Volume Migration also provides the flexibility to adjust the storage strategy or to move from outdated storage media/devices to recent technology with more capacity (e.g., WORM to NetApp SnapLock).

Strong focus on compliance and security

Various laws and regulations require document and data retention to prove services rendered, orders placed, and so on. Moreover, many documents and forms are crucial to the company's success, so it's vital to protect and secure these documents against unauthorized access and alteration throughout creation, transmission, long-term archiving, and retrieval. The following sections contain information on how Archive Server handles security issues.

Authorization and authentication

Secure user authorization

It is essential to protect business documents against unauthorized access. But that's not always easy or efficient when managing billions of documents over decades. Access control to documents via users, groups, and access control lists (ACL) can create high administrative efforts as users leave the company, move, and others join.

Business documents are always accessed by business applications (and are mostly worthless without their business context), but Archive Server follows a different concept. The business application itself (SAP, Content Server, Transactional Content Management)—and not single users—authenticates at Archive Server (signed URL resp., secKey, certificates). Archive Server expects that the business application has authorized the user of the corresponding request and grants access to documents.

User authorization in Content Server applications

Open Text Content Server applications (for instance, Open Text Transactional Content Processing) use their own refined security concept. It is possible to configure different access privileges for specific user groups and individual users. For security reasons, only the working memory contains the configuration files for the archive client. No local files are displayed in the process. Access to Archive



Server is required before an end user can display or modify the configuration files.

Authentication with secKey

A very effective mechanism in identifying unknown and unauthorized requests is using access with signed URLs. In this case, Archive Server accepts only those requests that were signed by a trusted source (e.g., a special application server). The signature from this trusted source guarantees that the request was initiated by an authorized user.

When a client sends a request to the application server, the trusted source checks the access rights, and if they exist, signs the URL and sends it to the client. The client can then access Archive Server with this URL. The signed URL contains an expiry time (e.g., two hours), after which it is no longer valid.

Within Archive Server, the URL signature is called secKey, which is part of the Server API and used by all leading applications, such as Exchange Archive and Production Document Management. Archive Server can be configured so that unsigned requests are rejected; i.e. only requests from the explicitly authorized SAP application server are accepted. Thus, even if an attacker obtains a document ID, unauthorized access to the document will be denied.

Secure data transport

SSL “Secure Sockets Layer” Communication

By enforcing SSL, authorized and encrypted access to all or individual logical archives can be ensured.

Client-server transport secured with checksums

Checksums are used to recognize and reveal unwanted modifications to the documents on their way to and through the archive. When clients archive or display documents, checksums are used to identify whether transmission was complete and error free. The checksums are not signed, as the methods used to reveal modifications are directed towards technical failures and not malicious attacks.



Open Text Imaging Enterprise Scan generates checksums for all scanned documents and passes them on to the Document Service. The Document Service verifies the checksums and reports errors. On the way from the Document Service to the storage the documents are provided with checksums as well, in order to recognize errors when writing to the media.

Server-client transport secured with timestamps

How does the client know that a document is authentic and has been sent by the Archive Server? Clients can check the document's timestamp in order to prove data integrity and authenticity of the document.

Digital signatures

We distinguish two types of digital signatures: personal signatures to handle authentication and timestamp signatures to ensure data integrity. Although personal signatures are stored with Archive Server, the handling is controlled by the leading application. Timestamp signatures provided by Archive Server are described below.

Secure, long-term archiving and data integrity

Generally, Archive Server archives documents on unalterable media. These can only be written once, providing excellent security against accidental as well as intentional deletion or alteration. However, to ensure document integrity, timestamp signatures are required.

Timestamps

In order to avoid any unnoticed data loss, even the transmission of a document is secured on its way with the help of checksums. From there, the integrity is secured with the help of timestamps. Timestamps ensure that document components cannot be modified unnoticed after they have been archived. Timestamps guarantee the authenticity of archived business documents. When tax auditors examine a document several years later, the company can prove that it was saved at a certain time and hasn't been changed since.

A timestamp is a signed datagram containing the document's hash value, the current time and date, and additional information. The



Archive Server supports interfaces to external, certified timestamp service providers like timeproof and Authentidate.

To put a timestamp on every document, Archive Server needs a service to request timestamps for a document. This can be a special hardware device or a Timestamp Service.

A timestamp is valid for about eight years. After a certain time, it loses its security because it's based on a hash algorithm, which may be identified by hackers. Thus, after a certain period of time, signature renewal must be performed.

ArchiSig: Signature renewal for long-term digital signature

In contrast to paper-based documents, the value of digitally signed documents as legal evidence decreases over the course of time. This is particularly due to the following reasons: the employed cryptographic algorithms and the keys lose their security qualification over time. It also cannot be guaranteed that the directories and documents needed for the verification of certificates are available for 30 years or more. In addition, the use of digital signing procedures is often insecure, and information for the subsequent evaluation of the actual security is missing. Concepts to solve these problems have only been developed to a certain extent.

The solution to meet these shortcomings is the ArchiSig concept. Archive Server supports the ArchiSig concept. An ArchiSig-generated timestamp is valid for an unlimited period of time.

An example scenario for ArchiSig can be found in the public services area. Masses of historical and new documents have to be handled and stored. Some communities switch to electronic processing of these documents. That also means a huge capturing effort for historical documents. The digital signature during the capturing process keeps the legal integrity of these documents. ArchiSig keeps the integrity of the digital signature. The paper-to-electronic transformation is a secure process, and the electronic documents are having the same legal force as their corresponding paper documents. Electronic documents can now be integrated in such processes as SAP. By this, SAP users have the full information overview for every transaction.



Auditing: Long-term traceability

All actions of the Archive Server are monitored in audit trails. Audits on document actions can be enforced for compliant retention classes. Typical actions to be audited are create, copy, migrate, timestamp, and delete. Administrative changes will always be audited. To access audit information, Archive Server provides a tool to extract audit information from the database, as well as http-based calls for leading applications to display audit information documents.

Encryption of the stored data

By encrypting the document data on the storage medium, critical data such as salary tables can additionally be stored in an encrypted manner. Thus the documents cannot be read without an archive system. A symmetric key (system key) is used for document encryption. The system key is encrypted on Archive Server with Archive Server's public key and can then be read only with the help of Archive Server's private key. SSL is used to exchange the system key between Archive Server and the backup server.

Always available through various backup scenarios

Backup

Power outages, physical damage, outdated media, hardware faults, or usage errors can unexpectedly shut down IT operations at any time. Archive Server provides a variety of options to optimize the availability of the business documents.

Archive Server backup concept provides maximum reliability. This includes backing up all the hard disk partitions that contain archived documents before they are stored in the optical archive, as well as the operating system and the application software. The system can also generate backups of all the entries in the archive database and duplicate the optical media, largely as automated functions. Furthermore, Archive Server can create copies of volumes as backups.

To avoid losing data in the event of a hard-disk failure and resume using Archive Server immediately, we recommend using Redundant Array of Independent Disks (RAID) technology as an additional data backup mechanism.



In addition to document content, administrative information is synchronized between original and backup systems.

High availability

To eliminate long downtimes, Archive Server offers high availability via “hot standby server.”

The hot standby server is a cluster solution, in which a fully-equipped secondary Archive Server monitors the current production system. If a server fails, the secondary server automatically assumes all activities, with full transparency for end users. Archive Server clusters run through a fast LAN and respond to end users in the same way as a single, high-availability Archive Server.

If the production system fails, users can continue to work normally on the secondary archive system. In contrast to the remote standby server scenario, both read (retrieval) and write (archiving) access to documents is possible in this configuration.

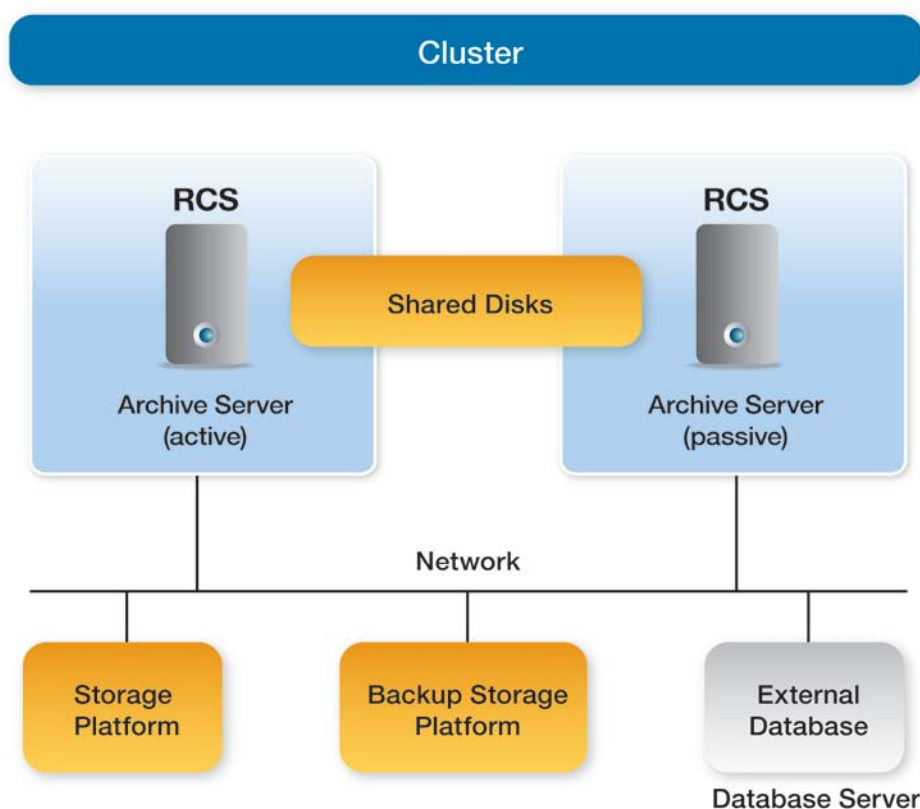


Figure 13:
High availability scenario



Remote standby

With a remote standby server, all the documents in an archive are duplicated on a second Archive Server—the backup server—via a WAN connection for geographic separation. The remote standby server's configuration is identical to that of the original Archive Server. The archives and hard disk buffers of the original server are replicated asynchronously.

The remote archive system generates backups of the original optical media. If the production Archive Server fails, the backup server continues to provide read-access to all the documents. Physically separating the two servers also provides optimal protection against fire and other catastrophic loss.

Disaster recovery

The Archive Server stores the available meta data together with content on the storage media (e.g. DocId, aid, timestamp). This allows Archive Server to completely restore access to archived documents in case the Archive Server hardware has a major breakdown or has been destroyed. Technically, the entire database can be restored from the information that is stored on the media. Consistency checks are supplied to check database versus volumes and volumes versus database. Also, support for a fast delta import after a server crash is provided.

Storage and resources

Logical archives

A logical archive is an area on Archive Server in which documents can be stored. Archive Server may contain many logical archives. Each logical archive may be configured to represent a different archiving strategy appropriate to the types of documents archived exclusively there. A logical archive may contain one or more storage pools. Each logical archive is assigned its own exclusive set of partitions, which make up the actual storage capacity of that archive.

Documents are related to a business process that is handled by a leading application. For example:

- All invoices from the current year are grouped together so that they can be easily deleted after the retention period has expired.



- HR documents have to be kept separate from financial documents, and special treatment such as encryption may apply.

Logical archives make it possible to store documents in a structured way. You can organize archived documents in different logical archives according to the following criteria:

- The leading application and the module to which it belongs
- The contents of the document
- The retention period
- The archiving and cache strategy
- Storage media types
- Customer relations (for ASPs)
- Text versus productive context
- Protection of documents (authentication certificates per archive)

Storage resource management

Hardware abstraction

Key tasks of Archive Server include hiding specific hardware characteristics to leading applications, providing transparent access, and optimizing storage resources.

The Archive Server looks like a “Janus”—on the one side, it can handle complex hardware; on the other side, it provides hardware abstraction by offering a unified storage. If a hardware vendor’s storage API changes or if new versions come up, it’s not necessary to change all the leading applications using the hardware—only the storage manager’s interface needs to be changed.

Storage reorganization

Content lifecycle may be different depending on the document type, thus imposing different requirements on the storage sub-system. For example, many working copies will be created until a conceptual document (such as a product specification or contractual work) is finalized. Often, it is not necessary to store working copies in a long-term archive; sometimes they even may be deleted once the content has been finalized. The finalized version,



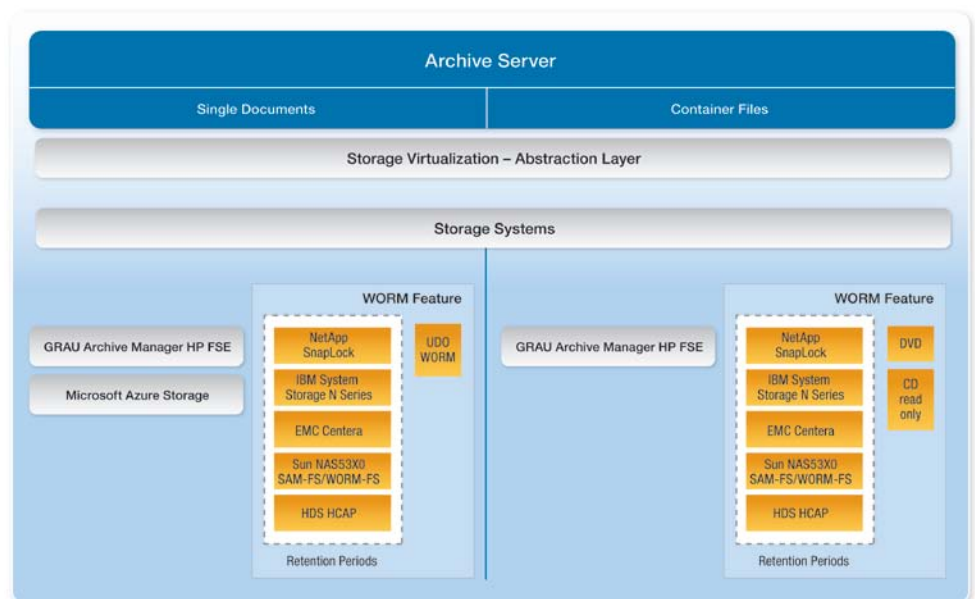
however, needs to be stored on a safe, non-alterable, long-term storage medium.

Another example is incoming invoices. They must be immediately filed on a non-alterable medium. Only during invoice processing, the documents are cached on high-speed storage in order to guarantee very fast access. If retention periods change for existing archived documents in regulated scenarios, storage needs to be reorganized. Other causes for storage reorganizations are changes in storage strategy (e.g., move on to hard disk system from opticals), organizational changes, or legacy decommissioning.

Supported storage media

Archive Server supports a wide range of different storage media and devices. Supported storage media are hard disk write-once or optical media or media. Archive Server connects to Hard Disk Write-Once media devices from different vendors. Furthermore, also Hierarchical Storage Management (HSM) systems and cloud storage are supported. (For the most recent information, see the Storage Platform Release Information of the Archive Server.)

Figure 14:
Supported Storage Media



Caching and Cache Server

Local cache scenarios on the Archive Server

On the Archive Server, cache areas can be assigned to logical archives. These caches can be filled upon purging Disk Buffer and by read requests following the First In, First Out (FIFO) rule. Old documents are removed from cache if cache area is full. Disk Buffers are also used as read cache as long as document copies are in Disk Buffer.

Cache Server

Archive Server supports caching via the Cache Server. It gives users fast access to archived documents. This is especially important in distributed network environments (such as WAN) because it greatly reduces the network load. It stores locally all the recently read documents and displays them on the client on request. When displaying documents, the Cache Server ensures that the document in the cache is the most current archived original. If several Cache Servers are used, even the logical archives and subnets of the network can be individually configured.

The Cache Server normally operates in a write-through mode, where all documents that are created locally are stored on the Cache Server and at the same time directly written through to the Archive Server. The Cache Server can be switched into a write-back mode. In this mode all the documents are cached in the local store of the Archive Cache Server only. An administrative job will later transfer these documents to the central Archive Server. This mode is intended for architectures with low network bandwidth.



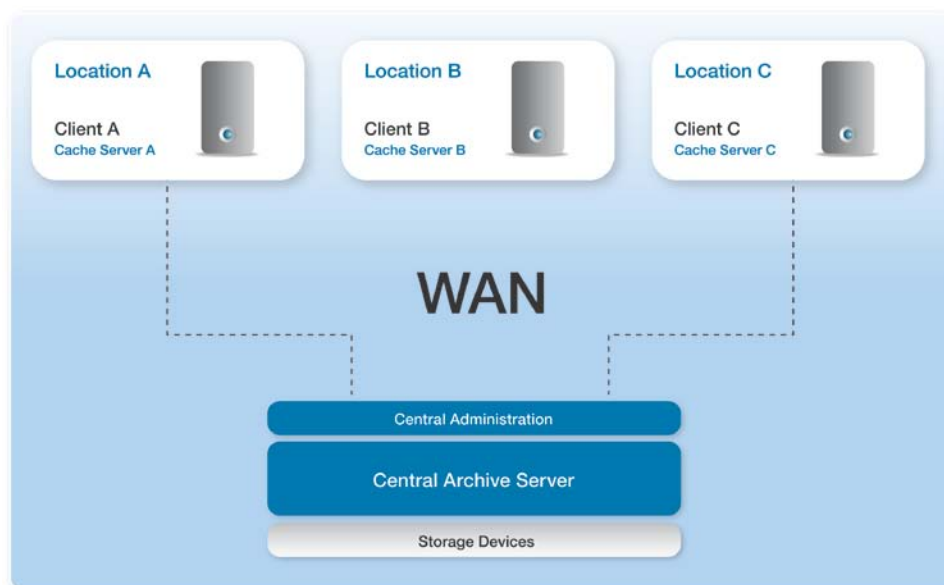


Figure 16:
Cache Server Scenario

The cache of the Cache Server is filled upon reading and writing documents (e.g., when scanning with Enterprise Scan or importing documents via the Document Pipeline). Also all applications using the Archive Server API will make use of the Cache Server scenarios.

Built-in accounting functionality

It may be more profitable for smaller firms to lease an archive from a provider rather than buy it. In this case, the provider must be able to make a list of all data concerning the costs of using the archive so that a precise invoice can be created. Thus, Archive Server has a built-in accounting system.

Administration and monitoring

Administration server

The Administration server of Archive Server is used to manage and configure the following system components:

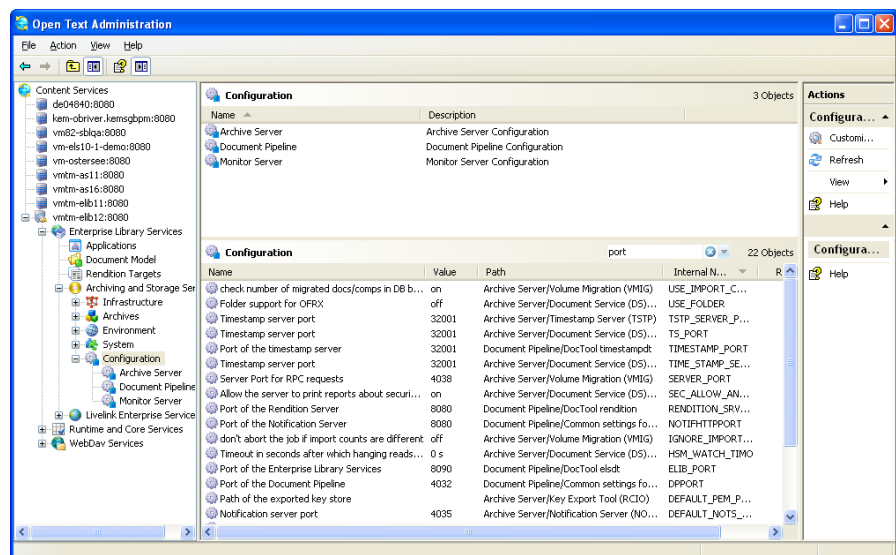
- The logical archive, which can be used to group the documents by department, physical location, document type, etc. A retention period can be specified for each logical archive.



- Archive Server-Cluster, in which several Archive Servers (possibly in different locations) are combined to function as one system for high-availability scenarios.
- The optical media and media pools (e.g., automatic WORM finalization)
- Archive Server users
- The timestamp certificate
- The definition and scheduling of the archive jobs.

The entire archiving system can be managed either locally or remotely using the Administration Client of the Enterprise Library.

Figure 17:
Enterprise Library
Administration



Server monitoring

Monitoring ongoing processes helps maintain optimal system performance. For this reason, Archive Server includes various monitoring systems that help control the overall system—from the resources for the storage hardware to the individual archiving components' processes.

The Monitor Server helps administrators locate and correct potential problems by using remote procedure calls, SQL queries, and operating system calls to collect and monitor data from the individual components. It continuously saves data about the components' status and the available storage space.



The Monitor Server has a Web-based monitor client that lets the administrator monitor processes. The individual components' processes appear in an intuitive graphical user interface. System resource status and availability appear as symbols.

Moreover, log files offer another powerful method for diagnosing Archive Server. All the archive components generate log files, which record the activities of the different processes. The log levels' default setting records a minimum of information. If the administrator suspects a problem with a certain component, however, he/she can increase the log level for that component.

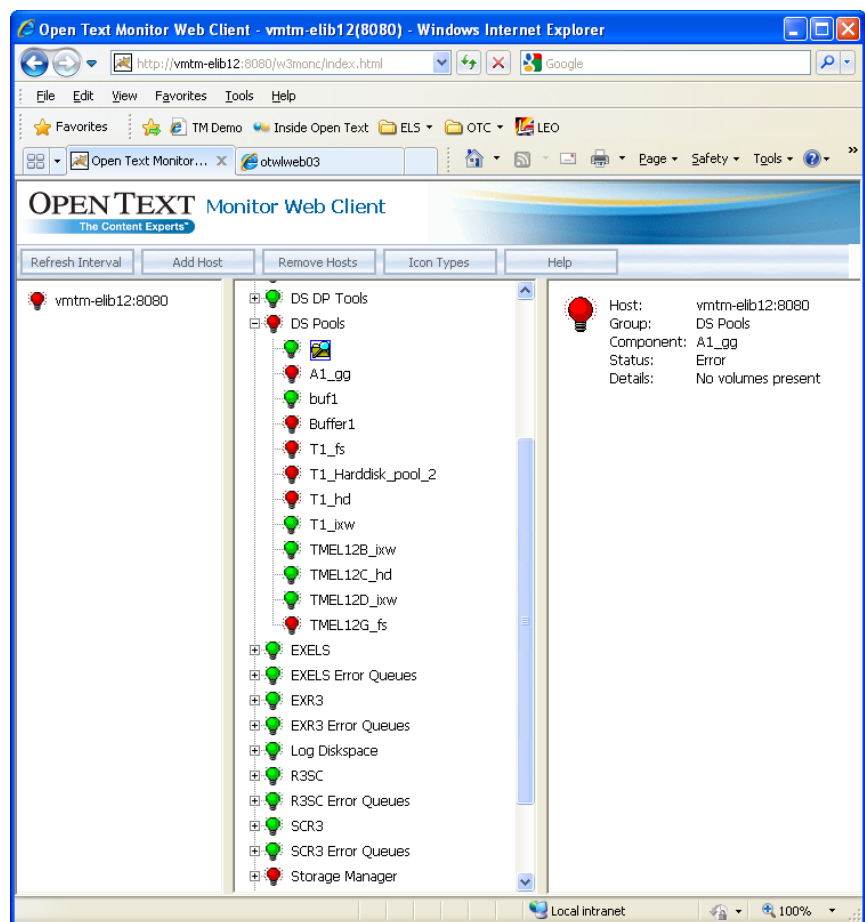


Figure 18:
 Archive Server
 Web Monitor

Active monitoring with the Notification Server

The Notification Server sends notifications, via mail or message, when certain server events (errors, access violations, etc.) occur. You can define these notifications in the Archive Administration.



Figure 19:
Events and Notifications

Events and Notifications					11 Obj			
Event Filter Name					Component	Severity	Class	Message Codes
Any Fatal Error	ANY	Fatal	Any					
Any Message from Admin Server	Administration Server	Any	Any					
Any Message from Document Service	Document Service	Any	Any					
Any Message from Monitor Server	Monitor Server	Any	Any					
Any Message from Storage Manager	Storage Manager	Any	Any					
Any Non-Fatal Error	ANY	Error	Any					
Cache Server Request Error	Administration Server	Any	Server	8				
ISO volume has been written	Administration Server	Information	Administration	2				
IXW volume has been initialized	Administration Server	Information	Administration	3 - 4				
Jukebox error: Jukebox detached	Storage Manager	Error	Server	205045 - 205047				
More blank media required in jukebox	Administration Server	Information	Administration	5				

Cache Server Request Error				1 Obj
Assigned Notifications				
Notification Name	Type	Scheduled Time	Scheduled Days	
Cache Server Alert	Alert	Full-time	Sun Mon Tue Wed Thu Fri Sat	

Logging

For long-term monitoring, you can have performance data written to log files. Logging for each component of Archive Server can be individually switched on or off within the Server Administration.



Conclusion

Sophisticated infrastructure and methodology are vital when it comes to archiving as they provide key capabilities such as seamless integration with your existing business applications and secure long-term archiving of your business information.

Archiving enables you to address business requirements by means of compliance and governance—with internal, industry, and legal regulations and standards—reducing risk. Last but not least, and certainly most dominant, it drives down the cost of business operations on both levels processes (operational efficiencies) and assets (TCO of real estate, IT legacy systems, paper, etc).

It's a catalyst for your business' sustainability—that is why archiving matters.



About Open Text

Open Text is a leader in Enterprise Content Management (ECM). With two decades of experience helping organizations overcome the challenges associated with managing and gaining the true value of their business content, Open Text stands unmatched in the market.

Together with our customers and partners, we are truly The Content Experts,™ supporting 46,000 organizations and millions of users in 114 countries around the globe. We know how organizations work. We have a keen understanding of how content flows throughout an enterprise, and of the business challenges that organizations face today.

It is this knowledge that gives us our unique ability to develop the richest array of tailored content management applications and solutions in the industry. Our unique and collaborative approach helps us provide guidance so that our customers can effectively address business challenges and leverage content to drive growth, mitigate risk, increase brand equity, automate processes, manage compliance, and generate competitive advantage. Organizations can trust the management of their vital business content to Open Text, The Content Experts.

www.opentext.com