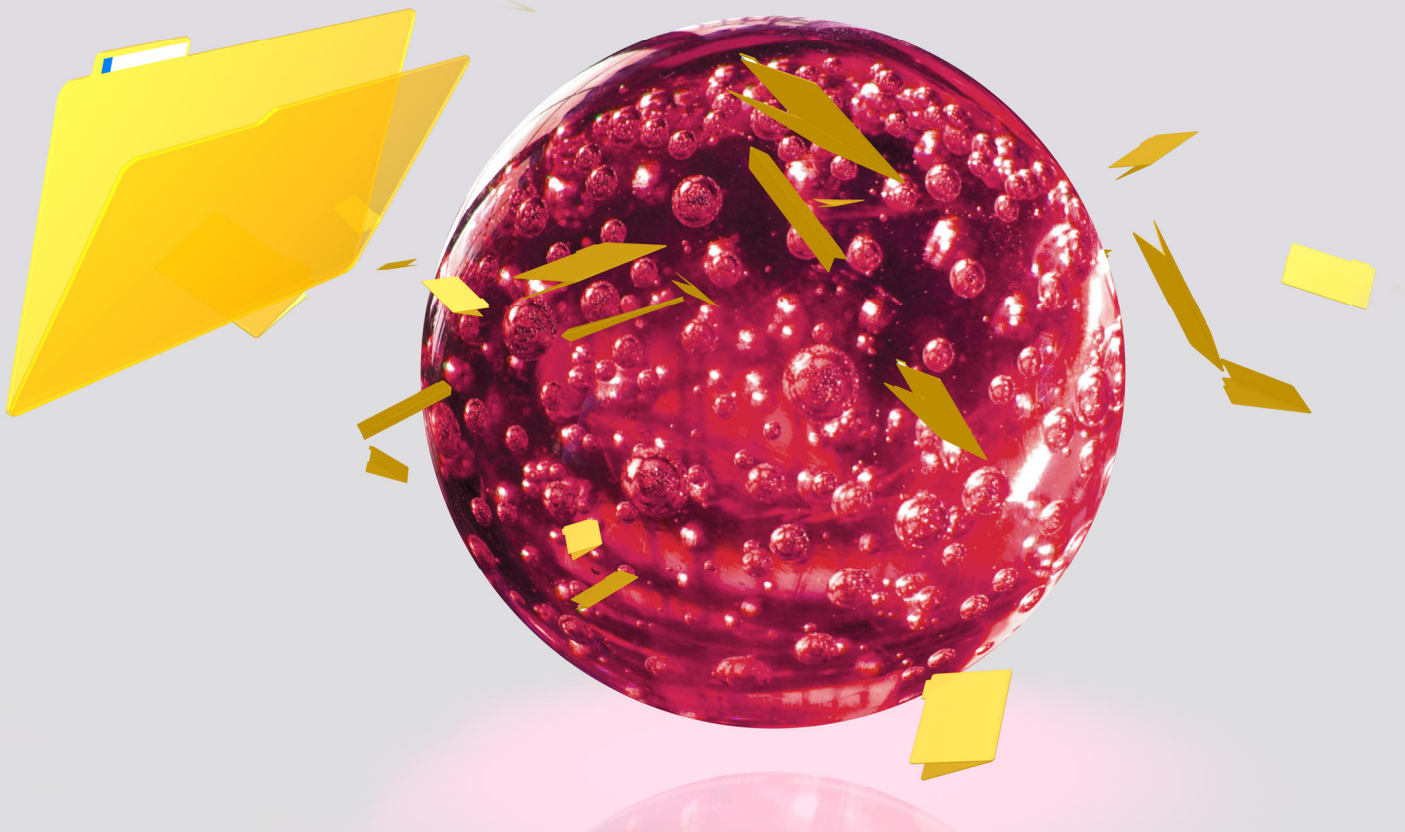


## **PAM for FILES** **file server archiving at its best**

### **Whitepaper**

Information Lifecycle Management  
with HP-RISS



#### **CONTENT**

Nowadays the importance of storing various data for online access is growing very quickly. One of the concerns is that a high percentage of this data is very hard to manage by means of security and backup. This is only a technical issue. The other concern is the offloading of the overloaded system, while assuring transparent access to the data.

Author: Jozef Végh

All Rights Reserved, including all rights concerning reproduction, copying or any other use or transmission of this document and its contents or parts of it. No part of this publication may, no matter in what form, be reproduced without written permission by H&S Software AG, passed on to third parties, edited by electronic retrieval systems, copied, distributed or used for public presentations. H&S Software AG reserves the right to change and update the content at any time. All data shown on screenshots is solely for demonstration purposes of the software. H&S is not responsible for this content.

PAM-STORAGE® is a trade mark of H&S AG. Microsoft®, Microsoft Windows® and the names of other Microsoft Products are registered trademarks of Microsoft Corporation. All Rights Reserved. Other product names are being used for identification purposes of products and can be registered trademarks of the according manufacturers. Doc.No. exp-wp-0704-001

H&S Heilig und Schubert Software AG  
A-1150 Vienna, Staettermayergasse 30

Tel. ++43.1.21555  
Fax. ++43.1.21555-200  
email: [issteam@hs-soft.com](mailto:issteam@hs-soft.com) | [support@hs-soft.com](mailto:support@hs-soft.com)

URL: [www.hs-soft.com](http://www.hs-soft.com)  
URL: [www.hs-soft.com/www.hs-soft.com/FileServerArchiving](http://www.hs-soft.com/www.hs-soft.com/FileServerArchiving)



---

## Introduction

---

Nowadays the importance of storing various data for online access is growing very quickly. One of the concerns is that a high percentage of this data is very hard to manage by means of security and backup. This is only a technical issue. The other concern is the offloading of the overloaded system, while assuring transparent access to the data. Today companies face an issue of priority: the main question is how to be compliant with legal rules established by government policies as well as internal policies. These rules mostly consider the obligation of keeping company-related data for a defined time period, to establish a transparent access to this data, to provide authenticated access to it, and to assure that the data cannot be changed or tampered with. It is also important to enable version tracking of this data, in order to track any changes made by authorized access. Providing such a solution means that not only is an appropriate SW for data lifecycle management needed, but an appropriate HW based storage solution is required too.

---

## **PAM for FILES. PAM for EXCHANGE and HP RISS**

---

Company information is kept mostly in the form of e-mails, files and various data stored in ERP, ECM, CRM, DMS or other information systems. Business to business (B2B) communication today is made in the form of email exchange. The amount of data communicated in this information exchange is growing exponentially each year. The growth in the exchange of electronic information inherently means a growth in the importance of the information transferred, as it contains business related data. For storing this important information, an archiving solution is needed. **PAM for EXCHANGE** is that solution, a scalable software solution for archiving emails from MS Exchange. It enables transparent access for users on Outlook and OWA, as well as providing a third point of access via its own Web interface. It supports various methods of archiving, such as:

- Archiving by a scheduled task
- Archiving by a direct archive process
- Manual archiving by the user
- Journal mailbox archiving
- Automatic versioning

**PAM for EXCHANGE** splits the archived emails into two streams:

- Metadata stream
- File stream

The metadata stream contains metadata from the e-mail for indexing and further searching purposes. This metadata is usually stored in a database. The file stream contains the attachments and the e-mail body itself. These files are stored on physical HW storage and are managed by the HSM module, which provides write, retrieval, single instance and retention time setting capabilities. HSM is in communication with its logical SW Stores. The lowest level of the HSM architecture are the SW stores, which are directly communicating with the underlying HW storage system through Windows API, storage API or even third party drivers.

Another form of storing data is to store it in the form of files from file servers or other IS. The growth of such files is also exponential. For the purpose of archiving data in the form of files either directly from OS or other IS, **PAM for FILES** was developed. It supports MS Sharepoint Server archiving as well. It also enables transparent access for users directly from the OS level. The process is transparent to the end-user, who sees no difference when opening an archived file in comparison to opening a non-archived one. It offers the same flexible archiving capabilities as **PAM for EXCHANGE**, such as:

- Archiving by a scheduled task
- Archiving by a direct archive process
- Manual archiving by the user
- Automatic versioning

Similarly to the **PAM for EXCHANGE** solution, **PAM for FILES** splits the archived files into two streams:

- Metadata stream
- File stream

The metadata stream contains metadata from the file for indexing and further searching purposes, such as file type, ACL, access date, creation date, modification date etc. This metadata is typically stored in a database. The file stream contains the file to be archived and its metafile for recovery purposes. These files are stored on physical HW storage and are managed by the HSM module, as mentioned with **PAM for EXCHANGE**.

The files extracted from various systems are stored finally on HW storage. The underlying HW storage is responsible for securely saving these files and assuring their consistency and availability. Various HW storage solutions can be applied, depending on which data access security, revision security, availability, fault tolerance and auto recoverability is needed.

For smaller installations, where no higher security and availability is needed, RAID arrays can be used as the HW store. For applications where high security, data consistency and availability is needed, HP RISS together with @PAM solutions represents an ideal HW-SW solution for archiving.

HP RISS represents a HW platform, which assures high availability, security, scalability, fault tolerance and disaster recovery for enterprise crucial data. It is a new storage system that stores objects as single instances and, according to the business and compliance policies of customers, prevents data from being manipulated.

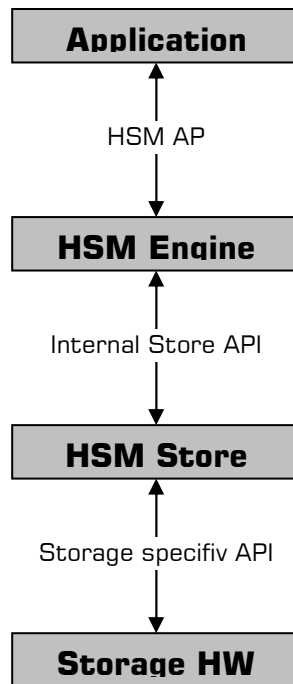
For integrating SW solutions HP RISS offers several application programming interfaces (APIs). Today the ILM Object Storage API and Query Web Service API are supported.

The ILM Object Storage API allows applications to integrate into RISS, by means of storing and retrieving objects and its metadata in RISS.

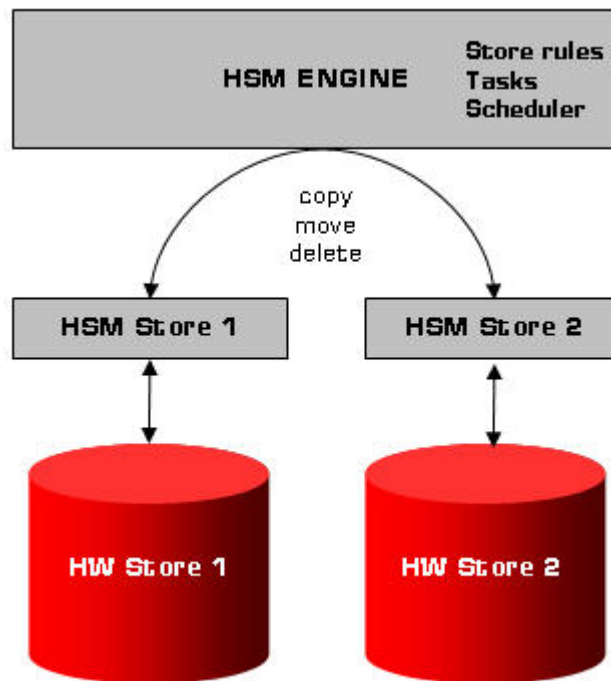
The Query Web Service API provides applications with an easy interface to search and discover information associated with data stored in RISS by other applications. An application can have the ability to search for file contents stored in RISS, because HP RISS has an indexing facility.

## Solution

As mentioned above both **PAM for EXCHANGE** and **PAM for FILES** are using the HSM (Hierarchical Storage Manager) engine as a primary interface for storing their data on various HW platforms. HSM is also responsible for lifecycle management of the stored files. It has the capabilities to move and copy files according to defined rules between any stores managed by the HSM.



HSM APIs and architecture

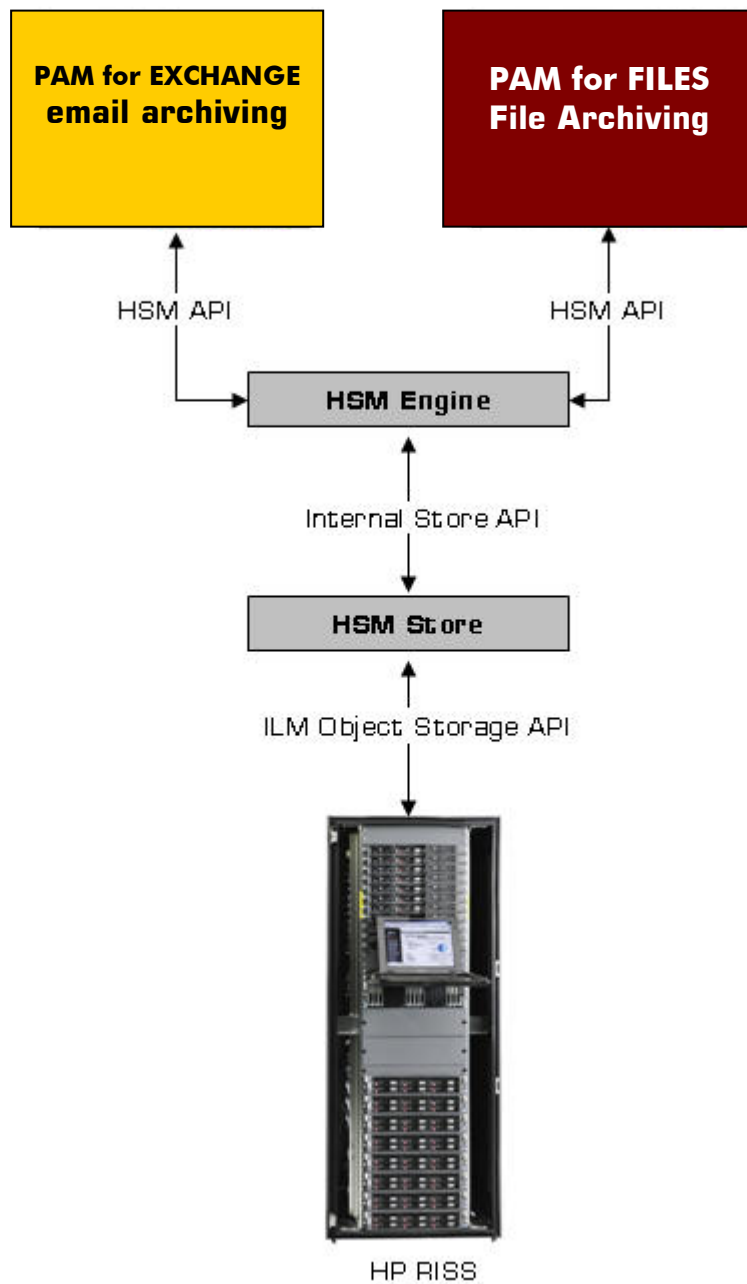


HSM APIs and architecture

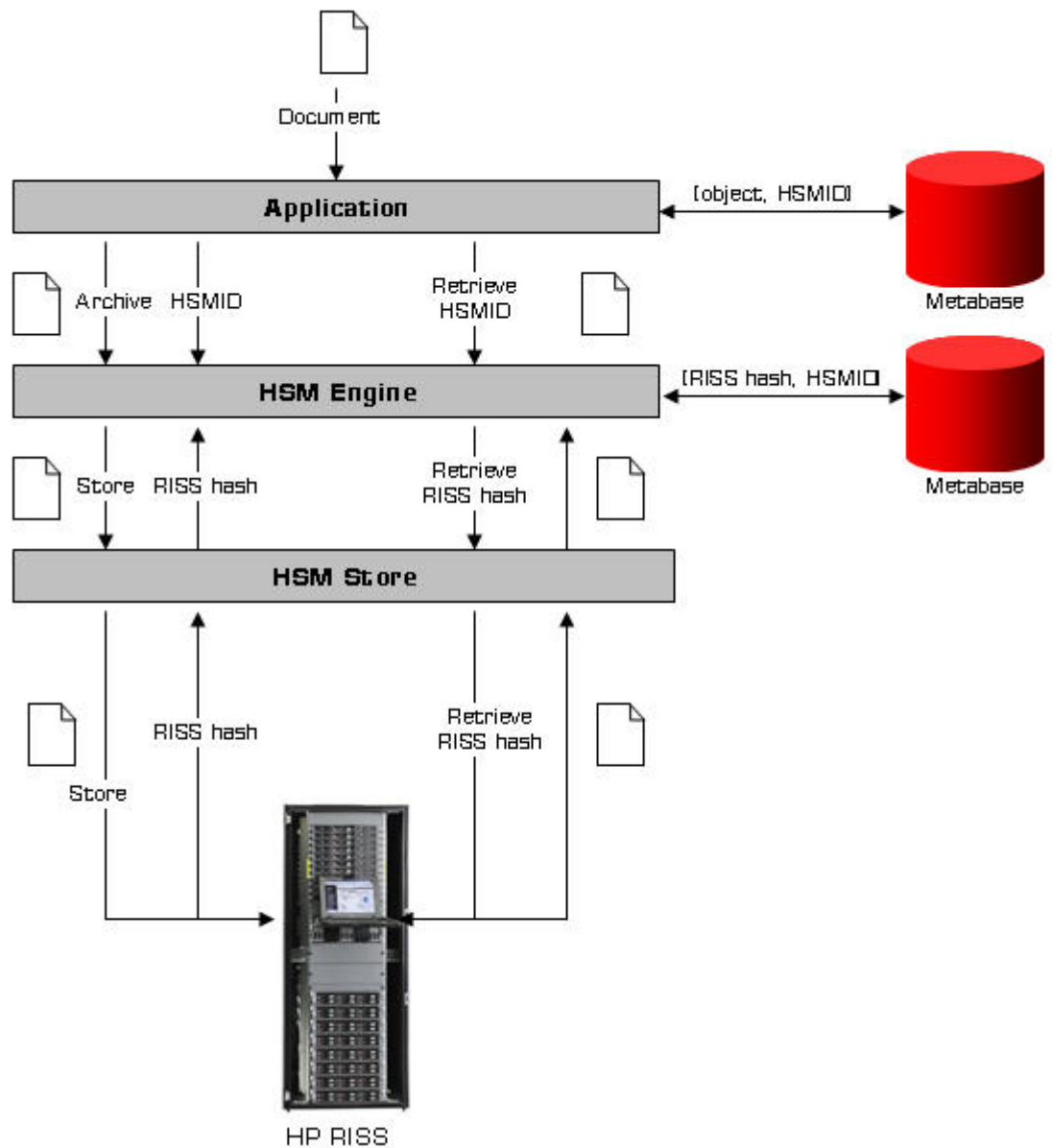
HP RISS is integrated into **PAM for EXCHANGE** and **PAM for FILES** on the store layer to provide a secure, highly available and scalable storage system. The communication between the HSM store and HP RISS is based on the ILM Object Storage API. The HSM store level is responsible for storing and retrieving files from RISS.

In the process of archiving, the HSM Engine passes the files to be archived to the HSM Store, which stores them into HP RISS through the Object Storage API. After the files have been stored in HP RISS it returns a unique hash to each of the files, which is then stored in the HSM Engine’s metadata for accessing the archived files in the future. The application level (**PAM for EXCHANGE, PAM for FILES**) receives a unique identifier called HSMID, which is in this case an indirect pointer to HP RISS stored files. The HSMID is stored on the application’s metabase layer in order to access the files.

While retrieving a file from the store, the application level has to provide the HSMID to the HSM engine. The HSM engine identifies the store, where the requested file resides and maps the HSMID into the appropriate HP RISS unique hash. The unique hash is passed to the HSM Store, which communicates directly with HP RISS through the Object Storage API. HP RISS then returns the requested file back to higher levels.



HSM – HP RISS Infrastructure



Storing and retrieving of documents with HP RISS

## Conclusion

Enterprise data archiving has become a crucial task in enterprises to assure high availability, security and compliancy of enterprise documents and data. **PAM for EXCHANGE** is addressing the issue of archiving emails and providing transparent access to them through existing email applications. **PAM for FILES** is addressing the issue of archiving files from the file system and supporting seamless integration into the OS. These software products have to ensure data consistency, security and availability. HP RISS is chosen as a HW storage system, which can provide the above mentioned capabilities together with fault tolerance, disaster recovery and discovery services.

**Appendix 1**

Possible solution infrastructure with HP RISS

