White Paper

# Medical Image Archiving

Tony Turner, RT., NetApp
November 2020 | WP-7333


## Abstract

Medical images, such as radiology and cardiology images, must be retained as part of a patient's medical record. Since the early 2000s, medical images that were once printed on film and stored in physical film libraries were produced in digital format. Also, Picture Archive and Communication Systems (PACS) were in widespread use for interpretating and retaining digital imaging studies. The retention policies for medical imaging vary from state to state, and many states have not updated their retention policies since the inception of digital imaging and PACS. This lag in updating, coupled with a misperception that storage is inexpensive, has resulted in most healthcare organizations retaining imaging data indefinitely. Today, digital medical imaging makes up approximately 70% of all clinical data stored worldwide, and with the recent influx of digital pathology imaging, that amount is sure to grow. Managing this data places an increasingly huge burden on health IT organizations. Understanding the clinical and medicolegal requirements for imaging retention and how to use new data storage technologies can make the difference between effectively managing your imaging data and letting your imaging data manage you.

## TABLE OF CONTENTS



 

## Medical Imaging Workflow Overview

Digital medical imaging has been in use since the early 2000s. Back then, the average study size across all of the modalities used to create medical images in radiology departments was around 40MB. Technology has advanced the capabilities of imaging modalities, and today the average study size is around 100MB. A study usually consists of several image files ranging anywhere from 64KB to 300MB, depending on the type of medical procedure and body part being imaged. Also, medical images created outside of traditional care areas in the form of digital pictures or videos are commonly stored. Examples of these images are digital photographs captured in Dermatology, documents scanned by a document imaging system, and videos captured during surgical procedures.

State-mandated retention requirements require healthcare organizations to make imaging studies available for an average retention time of 7 years, with exceptions for mammography and pediatric studies. Retention times vary from state to state, and most healthcare organizations don't delete anything. Medical images make up 70% of all data stored by healthcare organizations, which puts a huge burden on hospital IT departments to efficiently manage the data sprawl. To make matters worse, the likelihood of these images ever being retrieved again after 60 days is less than 10%, yet they still need to be available in just seconds by the hospital's PACS.

All medical archiving solutions have a small database, application virtual machines, short-term cache, and an archive. The database is typically Oracle, SQL, or Sybase and is around 2 to 3TB in size. When an imaging study is stored, it writes patient and study metadata into the database and adds a pointer in the form of a NAS UNC folder path. Fast disks (SSD or SAS) are used for the application, database, and image short-term cache. Depending on the vendor, the database and application portion is usually connected via Fibre Channel, although a handful of imaging vendors use NFS. The images are stored elsewhere via NAS-based protocols like CIFS/SMB, or NFS in the storage array, typically on large-capacity drives or via S3 in object-based storage solutions.

Most imaging systems have a short-term cache, although some vendors don't have this requirement. Typically, the cache holds anywhere from 30 days to 12 months of the healthcare organization's imaging volumes, and the PACS application manages the image movement between the archive and cache. Management is usually done on performance drives, because these are the images most recently acquired and therefore most frequently accessed by radiologists and other clinicians. When a study is stored in the PACS, a temporary copy is kept in the cache and a permanent copy is sent to the archive. The cache runs a delete process and an algorithm that keeps the most recently accessed images in the cache.

When the cache hits the upper watermark level, usually around 90% to 95%, the delete process deletes the temporary images down to the low watermark, usually around 80% to 85%. This constant process makes sure that the cache never fills up.

 

The next part of the system is the long-term archive or permanent archive. This archive grows constantly, and careful attention is required for capacity planning. The archive typically uses 7.2K capacity drives via CIFS/SMB or NFS. All imaging systems need two full copies of their images. The first copy is located in the archive and the other is in an offsite location. The offsite imaging system can be used purely for disaster recovery or for clinical business continuity, depending on how it was architected.

Depending on the imaging vendor's capabilities, the replication can be handled by the imaging application at the application level. Or it can be done at the block level by the storage vendor. The secondary storage array should be geographically dispersed, and at a minimum it should be located offsite in a separate data center.

Most imaging vendors can support Simple Storage Service (S3) natively in their archive product, so there are several storage methodologies available to provide an offsite copy. Object-based storage solutions, like NetApp® StorageGRID® are good options for archiving in larger environments. NetApp FabricPool is also an option, using StorageGRID on the premises for imaging vendors that cannot support S3 natively. Concerns about latency and the imaging application timing out have caused most imaging vendors to shy away from using FabricPool in a cloud service provider for the S3 bucket. Hybrid-cloud solutions are also a viable solution for customers who want to leverage a public cloud service provider to move their image storage out of their data centers and to use their operating budget to cover the costs.

## Archiving Requirements and Considerations

### Data Retention and Information Lifecycle Management

Data retention in medical imaging is one of the main challenges that IT organizations face. Every state has its own interpretation and rules for how long medical images need to be retained, and these rules can be specific to each type of modality. Some modalities, like mammography,  can never be deleted, and most states have special provisions for pediatrics.

When most imaging professionals hear the phrase *information lifecycle management* (ILM), they think of policy-based deletion. Actually, policy-based deletion is just one option in ILM.  Information lifecycle management—or in this case, *image* lifecycle management—is a data retention efficiency tool that is used to manage the imaging study over its lifespan.  It is the process of moving older imaging studies to a cheaper tier of storage to keep the cost of managing the study as low as possible. *Policy-based deletion* is the process of deleting older images based on policy. Most imaging archives have this capability built into their software. The policy is typically a combination of age of study, modality type, and one or more DICOM metadata tags. Safety nets are built in to protect mammography and pediatrics studies.

Policy-based deletion has been available for more than 10 years, but most healthcare organizations don't delete studies, for compliance and legal reasons. These days, customers are starting to view data as having value for analytics or artificial intelligence (AI) training, so another justification to keep all of their imaging data forever is that some day it might be useful to their organization.

Another question that often comes up when discussing ILM is, what is the cheapest tier of storage? Historically, several different types of media have been used, such as LTO tape, DVDs, and Blu-Ray jukeboxes. Now, healthcare organizations have many different options, such as 7.2K spinning hard disk drives, object storage, and cloud-based storage. However, with all of these options, organizations need to consider how realistic the final tier is in terms of the images being used for patient care. The images must be easily accessible in just seconds.

Even more importantly, medical images need to be protected and secure, and they must remain unchanged. Most storage environments now offer data encryption, at either the hardware or software level. Also, some storage vendors offer write once, read many (WORM) compliance that safeguards images from being altered. NetApp accomplishes with SnapLock® compliance software. Another aspect of reducing the amount of data an organization archives for medical imaging is to use data compression techniques. Most data manipulation and compression is managed by the medical image application, and most imaging vendors use JPEG2k lossless compression. JPEG2k compresses the images to 2.5:1, the most a medical image can be compressed without losing medical fidelity. It is discernable by the naked eye and still allows the image to be used for diagnostic medical interpretation by a radiologist.

Some modalities that require the highest resolution of detail, such as mammography, do not typically get compressed. Once an image has been interpreted, some imaging applications have the ability to compress it further for archiving. These images will be used only when comparing new studies to previous studies. Because medical images are precompressed by the imaging application, once they reach the storage level, further compression techniques, like deduplication, are typically not deployed. Deduplication is not effective on precompressed images and consumes more compute processing than the storage savings it typically achieves. Savings can be 5% to 10% with deduplication. In large multipetabyte archives, 5% to 10% savings could be 100s of terabytes, so in that situation it may make sense to use deduplication. Typically, it is turned off in smaller deployments.

NetApp offers flexibility with several options for deploying ILM techniques that are effective in archiving and managing large amounts of medical imaging data. Customers have flexibility when planning an archive for medical imaging. They can use on-premises storage, private or public cloud, and/or hybrid cloud.

The first consideration in building a medical imaging archive is to understand the image workflow clinically and the size of the environment. It’s easy to overcomplicate an image archiving environment, especially in a smaller environment. Sometimes the best solution is to keep it simple.

The next consideration is to fully understand the capabilities of the image archiving software and what it supports. Can the software manage the data movement between multiple tiers of storage? Does it have robust ILM capabilities to intelligently manage the data movement? How does it compress the images? Does it support S3 connectors into cloud-connected storage environments? All of these questions need to be taken into consideration.

Depending on the capabilities of the image archiving software and the customer's needs, NetApp can recommend products to make the archive environment most efficient, keep the data secure, and improve patient outcomes at the least cost for the organization. If an imaging vendor can't provide robust ILM rules or support S3 connectors to cloud-connected storage, NetApp FabricPool offers this functionality. Hybrid-cloud solutions are also available for customers who want to use a public cloud service provider to move their image storage out of their data centers and use their operating budget to cover the costs.

### NetApp FabricPool

If an imaging vendor can't provide robust ILM rules or support S3 connectors to cloud-connected storage, NetApp FabricPool offers this functionality. FabricPool is also an option for imaging vendors that can't support S3 natively. FabricPool can look at hot data blocks versus cold data blocks and based on policy move the cold blocks into an S3 bucket. This bucket can be on the premises, like StorageGRID, or in a public cloud service provider.  Concerns about latency and the imaging application timing out need to be addressed when choosing the S3 tier. NetApp generally recommends using StorageGRID with FabricPool on premises for the S3 object-based storage for the primary copy.

### Vendor-Managed Tiering

If a medical imaging vendor has the ability to manage multiple imaging storage tiers, it's important to understand what types of storage tiers they support. How and where a study is stored and the number of copies stored are determined by ILM rules built in to the medical imaging software. It's also important to understand how the software manages ILM rules and how easy it is to create ILM rules.  Most ILM rules are a combination of time (when the study was last used), modality types, application entry titles, and any DICOM tags present in the metadata of the study. Flexibility in building ILM rules is important because medical imaging objects come from many different care areas in a healthcare organization.

### Object-Based Storage

Most imaging vendors support native S3 in their archiving products. Object-based storage is ideal for large imaging environments because of its ability to handle large amounts of unstructured data without having to spend time crawling through directories as is necessary with file storage. The metadata that is stored along with image files in object storage can also be used for query searches and data analytics.  NetApp StorageGRID is scalable into billions of objects. It provides triple erasure coding for the highest data protection and uses a robust set of policy-based ILM policies at scale.

### Cloud Technologies

Healthcare providers typically do not view cloud as an option for medical imaging, but those views are changing. Cloud is slowly being adopted in medical imaging. Historically, it was used only for offsite secondary archiving and was mostly managed by the medical imaging vendors in a pay-per-study model. Today, cloud has become mainstream and cloud-based technologies now provide the performance, security, and reliability required for medical imaging.

Cloud has multiple advantages, but the sheer physics of moving large imaging files across a small WAN has always been a concern. Healthcare providers struggle to understand the performance capabilities of the cloud as they relate to the time value of images. There is a perception that old images are looked at much more than they actually are. This is part of the fear that it might take 20 seconds to retrieve an 8-year-old chest CT scan. Understanding clinical workflows and cache sizing can mitigate some of these concerns, and the case for cloud and hybrid cloud can be bolstered by researching the actual recall rates of the organization.

The reality is that image archive workload characteristics are a natural fit for cloud because most images are seldom recalled after 60 days, and managing the data sprawl of image growth can be an organizational nightmare. Also, S3 protocols are much more suitable than CIFS/SMB for moving data across WANs. Several different models are used in medical imaging. Hybrid cloud is gaining in popularity because it is a better fit for most clinical workflows and gives the most flexibility for IT departments. A growing number of cloud PACS and PACS as a service providers are emerging, but they mainly focus on smaller standalone hospitals and imaging centers.

Cloud technologies offer several advantages, and the flexibility of cloud is very attractive for hospital providers. Providers that are looking for flexibility in paying for storage for medical imaging may be required to use operating budgets rather than capital budgets to grow their storage footprint. Some smaller IT departments, especially those in areas where it's difficult to recruit top IT talent, struggle to maintain their technology environments. Not having to maintain an offsite archive for their medical images can be very valuable for those IT departments. Also, some organizations struggle to keep up with their image growth and capacity planning, and they like the idea of a boundless archive in the cloud. Hybrid cloud offers the ability to choose some or all of these benefits.

Hybrid cloud for medical imaging can be implemented in a few different ways. One way is to put the entire production imaging environment on the premises and the secondary imaging environment in the cloud. Another way is to put part of the production imaging environment on the premises and the tier part into a private or public cloud provider. The secondary imaging environment can mirror this environment, or it can be entirely in the cloud.

Because of these different scenarios, careful planning is required and clinical considerations must be taken into account. The provider needs to have a good understanding of the clinical workflows in the imaging department, in order not to disrupt how the radiologists access images for clinical interpretation. This consideration has a direct impact on production and has clinical ramifications throughout the facility.

Another planning consideration is to look at the costs of using the cloud for imaging. Because images are pulled back from the archive into the PACS every day for comparison, the higher egress rates may not have been factored in, and the monthly costs may be significantly higher than the amount that was budgeted.

### Data Replication

Depending on the customer's RPO and RTO, NetApp offers several products for secure data replication. The NetApp SnapMirror® suite of replication software products offers flexibility and efficiency when replicating data to support remote backup and recovery, disaster recovery, and data distribution. SnapMirror offers integrated remote backup and recovery and disaster recovery with incremental asynchronous data replication; preserves storage efficiency savings during and after data transfer; and provides asynchronous replication. It's possible to replicate data for the entire array or only specific volumes, and the target can be another NetApp array on the premises or in the cloud.

For hybrid cloud environments, SnapMirror Cloud can replicate data from the customer's on-premises NetApp system into their public service cloud provider, such as AWS, Azure, Google Cloud, or IBM Cloud, leveraging S3 cloud resources. SnapMirror has WAN optimization tools to configure performance and backup windows to meet the customer's needs. For customers who need to maintain business continuity for critical enterprise imaging applications and workloads, the NetApp MetroCluster feature provides synchronous replication between two NetApp storage peers. This robust infrastructure combines array-based clustering with synchronous mirroring to deliver continuous availability and zero data loss for SAN and NAS workloads, with up to 700km distance between nodes.

### Compression

Another aspect of data management is to use data compression techniques to reduce the amount of data an organization archives for medical imaging. Most data manipulation and compression is managed by the medical image application, and most imaging vendors use JPEG2k lossless compression. JPEG2k compresses the images to 2.5:1, the most a medical image can be compressed without losing medical fidelity. This compression is not noticeable to the naked eye and still allows the image to be used for diagnostic interpretation by a radiologist.

Some modalities that require the highest resolution of detail, such as mammography, do not typically get compressed. Once images have been interpreted, some imaging applications have the ability to compress them for archiving. These images will be used only when comparing new studies to previous studies.

Because medical images are precompressed by the imaging application, once they reach the storage level, further compression techniques, like deduplication, are not typically deployed. Deduplication is not effective on precompressed images and consumes more compute processing than the storage savings it typically achieves. Savings might be 5% to 10% with deduplication. In large multipetabyte archives, 5% to 10% savings might be 100s of terabytes, so it might make sense to use deduplication. Typically, it is turned off in smaller deployments.

 

### Data Protection and Data Security

One of the most important aspects of archiving medical images is the need to protect and secure the images and ensure that they remain unchanged. NetApp's portfolio of security capabilities helps healthcare organizations move to a Zero Trust model: Deploying integrated data security across the medical imaging archive prevents unauthorized data access and combats ransomware attacks. With built-in NetApp Volume Encryption features, you can easily and efficiently protect your at-rest data by encrypting any volume without using special encrypting disks. Internet Protocol Security (IPsec) is a security protocol suite that provides in-flight encryption for backup and replication to protect your data in transit.

Other features such as multifactor authentication, role-based access control, and onboard and external key management increase the security of your data. To meet healthcare's stringent compliance and data retention policies, NetApp SnapLock software enables WORM-protected data to prevent medical images from being altered in any way. NetApp offers superior integration with enterprise backup vendors and leading applications. In addition, NetApp offers state-of-the-art "right-to-erasure" capability for General Data Protection Regulation (GDPR) to address changing data protection regulations. Finally, NetApp FPolicy enables partner applications to monitor and set file access permissions to determine how the storage system handles requests to respond to file access and file operation requests.

## Other Considerations

### Managing the Environment

Managing storage for medical image archiving requires detailed planning, budgeting, and communication with other stakeholders in the healthcare organization. Medical images are created by almost every care area in the organization. For planning purposes, it's important to understand the volumes and image sizes for each care area and their annual growth rates.

Communication channels with stakeholders in each care area are also key. New digital modalities and advanced technologies image sizes have a direct impact on the amount of storage required to store these images. It's also important to consider which areas in the organization are experiencing large growth, as well as corporate initiatives like mergers and acquisitions.

Once the amount of storage that is required to maintain growth of the organization is determined, budgeting and timing of storage expansions should be considered. Storage projections and budgeting should assume a 2-year budget cycle. Considerations should include the sum of the run rate of all individual locations that are contributing data for 2 years, any migration upcoming or in process, and any additional volume that might be added from new modalities, new departments, or new sites.

Another vital step is to obtain quotes and set expectations with the storage vendor to ensure that the storage is procured and implemented when it's needed. This process can avoid last-minute panic or worse, outages due to storage arrays being out of capacity.

 

As image volumes and the amount of data flowing through hospital networks continue to grow, discussions should also be held to make sure that the networks can keep up with the demand of the additional growth.

NetApp data management infrastructure software can centrally monitor the health of your environment by viewing metrics on capacity usage, performance, availability, and data protection. It can also help automate your storage processes and integrate them into your data center orchestration platform. NetApp Active IQ® is a digital advisor used by all NetApp solutions that employs artificial intelligence to simplify and automate the proactive care and optimization of your NetApp environment. Active IQ manages continuous risk assessments, predictive alerts, and automated actions to help prevent problems before they occur, resulting in improved system health, higher system availability, and enhanced security.

### Data Migrations

Data migrations in medical imaging have unique requirements that need to be considered. Unlike typical archives, healthcare organizations and storage vendors can't simply migrate files from one storage system to another, as in other NAS environments. Whenever a medical image is ingested by the imaging solution using Digital Imaging and Communications in Medicine (DICOM), the DICOM study includes two key pieces of information. The first is the DICOM metadata that includes specific patient information from the hospital information system (HIS). The second part is the actual image files.

During the ingestion of the DICOM study, the DICOM metadata gets indexed into the imaging database. At the same time, the actual image files get stored on the NAS array. The file location on the storage is also written in the database in the form of a UNC file path (often referred to as a database pointer). This convention allows the imaging application to locate the image file whenever the application recalls the image for clinical purposes. Therefore, if the file is ever moved to another location or deleted and the file path or database pointer has not been updated to reflect the new location, the imaging application will fail on file retrieval. This means that whenever medical images are migrated from one storage array to another, all of the file paths in the database need to be updated. Imaging vendors typically don't allow access to their databases, so only certain resources are capable of providing medical imaging migrations.

Over the years, imaging vendors have created profit centers in their professional services organizations, and they count on revenue every year to perform data migrations. Also, several third-party companies that work closely with imaging vendors have created standalone businesses to perform medical imaging migrations.

Data migrations often get overlooked during planning for implementing or refreshing medical imaging storage environments. The requirements associated with moving medical images mean that imaging vendors put a premium cost on their professional services, which can make it cost prohibitive for the customer to change their storage provider. Storage vendors often require their customers to do forklift upgrades due to lack of compatibility between new storage products, forcing their customers to migrate their data as their storage arrays come to the end of their lifecycle.

When customers move to NetApp, they can avoid costly DICOM-based medical image migrations by using integrated data management tools to move data between old and new NetApp arrays. And they can migrate nondisruptively to avoid downtime. Nondisruptive migration saves the customer money and avoids planned change control windows and system down time. NetApp Professional Services has the experience and tools to move the image file data from non NetApp storage to NetApp, but the customer needs to engage the imaging vendor to update the database pointers in their database.

### The Value of Data

The mindset around the value of data has shifted. Historically, deleting data was a common method of managing the amount of data an organization was archiving. Now, with advances in artificial intelligence and data analytics, all data is considered valuable beyond its initial use. With all of an organization's data aggregated in a single imaging archive, technology like AI algorithms can be applied to gain insights to help create more efficient clinical care paths to improve patient outcomes. Viewed as augmented intelligence, AI can also improve productivity. Radiologists can use AI to assist them in identifying suspicious pathologies when interpreting medical imaging studies, or to immediately flag emergent and time-sensitive cases such as stroke. AI can also be used to comb through the countless normal medical imaging exams and allow radiologists to focus on more difficult or abnormal cases.

These are just a few examples of how AI and medical imaging data can add value to healthcare organizations. Analytics from mining medical imaging metadata is another way in which organizations can gain powerful insights into their imaging environments.

## Conclusion

Archiving medical images can be a daunting task, especially for large healthcare organizations. Technology advances provide tools that allow organizations to manage this data while ensuring that it fits the clinical requirements for medical imaging. Now that legacy data is recognized as useful for organizations, medical imaging data can drive better patient outcomes and help to identify trending within the organization. NetApp has the knowledge and products to help healthcare organizations manage and protect their imaging data so that they can focus on the next innovation in healthcare IT.

## Where to Find Additional Information

To learn more about the information that is described in this document, review the following documents and/or websites:

- NetApp Medical Imaging
  https://www.netapp.com/pdf.html?item=/media/17141-sb-3906-1017-pacs-solution-briefpdf.pdf
- Security Features in ONTAP 9
  https://www.netapp.com/pdf.html?item=/media/8128-ds-3846.pdf
- NetApp Data ONTAP 9
  https://www.netapp.com/pdf.html?item=/media/7413-ds-3231.pdf
- AI in Healthcare

https://www.netapp.com/pdf.html?item=/media/7393-na-369pdf.pdf
https://blog.netapp.com/ai-solutions-healthcare-imaging

- NetApp Product Documentation
  https://docs.netapp.com

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.