Chapter 7
Data Storage

Edward M. Smith, Thomas Jay Crawford, and Katherine P. Andriole

Contents

E. M. Smith (Deceased)

T. J. Crawford (✉)
University of North Carolina School of Medicine, Department of Radiology, Chapel Hill, NC, USA
e-mail: jay_crawford@med.unc.edu

K. P. Andriole
Brigham and Women's Hospital, Harvard Medical School, MGH & BWH Center for Clinical Data Science, Boston, MA, USA
e-mail: kandriole@bwh.harvard.edu

B. F. Branstetter IV (ed.), *Practical Imaging Informatics*,
https://doi.org/10.1007/978-1-0716-1756-4_7

7.1 Philosophy of Storing Electronic Protected Health Information (ePHI)

- The storage and retrieval of ePHI is a multifaceted function ranging from archival of data within the healthcare facility and related off-site practices, to the referring physician's office, to the patient and to regional, national, and international repositories.
- Requires the purchasing of and adherence to interoperability standards for all hardware, software, and applications that touch the institution's digital environment.
- Image objects consume 95% or more of the storage requirements versus 5% or less for non-pixel data.
- This includes not only radiologic data but visible light images, digitized pathology slides, and some genomic data.
- Estimating storage requirements is at best an educated guess due to new innovations generating increasingly larger amounts of information and the expanding digital imaging environment in health care.
- Storage management components should be purchased separately from the clinical and administrative applications to reduce cost and ensure vendor independence (**vendor-neutral archives**) while requiring interoperability between all components in the digital environment.

7.1.1 Enterprise Storage Versus Departmental Silos of Storage

- Isolated independent silos of departmental storage are difficult to manage.
- Divorcing storage components from administrative and clinical applications:
  - Allows administrative and clinical departments to acquire applications that optimize workflow and productivity.
  - Creates independence from application vendors by eliminating proprietary storage formats, thus minimizing future cost of data migration and costs due to obsolescence of technology.
- Reduces cost of hardware, software, licensing, and maintenance fees as well as personnel cost.
- Provides an integrated storage and storage management system.
- Provides high availability and redundancy of information at a reduced cost.
- Satisfies state and federal retention periods for medical information and HIPAA requirements in the most cost-effective manner.
- Allows for consistent Information Lifecycle Management Policies to be enforced.
- Allows for a more complete view of the patient's imaging studies, across departments or across affiliations.

FURTHER READING: Economics of Enterprise Storage

Cecil RA. Solve the enterprise archive puzzle. Imaging Economics. 2007 Jan; 38–43.

Langer SG. Global PACS archiving. Enterprise Imaging and Therapeutic Radiology Management. 2008 Sep; 61–7.

Smith EM. Integrated implementation revamps information storage. Diagnostic Imaging. 2005 Jan; 39–43.

Smith EM. Storage management: one solution doesn't fit all. Imaging Technology News. 2004 Nov/Dec; 80–3.

7.1.2 Storage Management Responsibilities

- **Information Technology (IT) Department** is primarily responsible for storage management including funding, providing necessary infrastructure, hardware, and software management and meeting administrative and clinical departmental requirements, including the following:
  - Meeting storage volume requirements.
  - Periodically upgrading hardware and software due to technological obsolescence and migrating information to new infrastructure.
  - Providing 24×7 accessibility to all information.
  - Meeting required clinical and administrative **response times** for retrieval of information.
  - Adhering to HIPAA and other mandated federal, state, and local security and retention requirements for health-related information.
  - Consistently providing required disaster recovery (DR) and business continuance (BC) processes including timely backup and replication of required information (see **Chap. 28**).
- **Administrative and Clinical Departments** are responsible for:
  - Working with IT to ensure that their applications adhere to the purchasing and interoperability standards mutually agreed upon.
  - Being realistic in their requirements for response times for information queries.
  - Estimating clinical storage requirements for a 12- to 24-month period.
- **Chief Financial Officer** understands that:
  - Storage management is an on-going cost.
  - Life cycle of many of the components is 3–5 years due to technical obsolescence.

7.1.3 Economics of Storage Management

- The cost of storage management is an operational cost that must be budgeted annually.
- Cost of storage management includes
  - Hardware, infrastructure, and software
  - Initial cost
  - Maintenance and licensing fees
  - Replacement cost due to technological obsolescence
  - Storage media cost
  - Personnel
  - Utilities
  - Physical space in a datacenter
  - Costs related to data backup and replication, business continuity, and disaster recovery
  - Data migrations due to technology obsolescence or overgrown capacity
- To minimize the cost of storage management:
  - Reduce the number of isolated silos of storage.
  - Implement a scalable storage management design, reduce the use of proprietary components, and adhere to archival and integration standards.
  - Purchase media based on projected volume and growth needs for periods of 12–24 months since **storage costs decrease** with time.
  - Purchase subsystems and associated media based on performance (e.g., data retrieval time) requirements of applications.

7.2 Datacenters

The IT department is responsible for funding, managing, and supporting the datacenter.

- Providing power and cooling for a datacenter with high-density computing resources is a significant cost.
- Healthcare facilities computing resources should reside in a secure datacenter providing a level of availability of medical information that meets or exceeds the clinical requirements of the institution.
- Clinical departments must specify the characteristics of the computing resources and availability requirements for medical information meeting the clinical requirements.

FURTHER READING: Datacenters

Updated tier classifications define infrastructure performance, www.uptimeinstitute.org.

A quick primer on data center tier rating, www.itconsultant.boblandstrom.com.

7.2.1 Tier Rating of Datacenters

- The Uptime Institute provides industry recognized and accepted standards for high-density computing and mission critical facilities in a vendor neutral manner.
- The primary attribute of the Tier rating of a datacenter is the availability of the computing resources it contains.
- Tier rating may be too expensive and rigorous for many healthcare facilities, but it does provide a guideline and may assist in selecting a datacenter for outsourcing some aspect of a center's computing resources or for choosing disaster recovery services:
  - Tier 1 datacenter has no redundant components such as on-site generator, uninterruptible power supply (UPS), fire suppression system, heating, ventilation and air conditioning (HVAC), etc. in which availability is affected regardless of whether the outage was planned or unplanned.
  - Tier 2 datacenter has redundant components, but only a single-distribution system. This provides availability for planned outages, but not for unplanned outages.
  - Tier 3 datacenter has redundant components and distribution systems plus dual public power supplies that provide the capability to operate self-sufficiently. All computer resources have two independent power sources. These features provide system availability during routine maintenance as well as for unplanned outages.
  - Tier 4 datacenter has all the features of Tier 3 plus the topology is configured such that any single component or distribution failure has no negative impact on availability (Table 7.1).

Table 7.1 Tier characteristics of data centers

Characteristics	Tier 1	Tier 2	Tier 3	Tier 4
Year first deployed	1965	1970	1985	1995
Single points of failure	Many + HE[b]	Many + HE[b]	Some + HE[b]	Fire, EPO[a] + Some HE[b]
Planned maintenance shutdowns	2 per year at 12 hours each	3 per 2 year at 12 hours each	None	None
Typical site failures	6 over 5 years	1 every year	1 every 2.5 years	1 every 5 years
Annual downtime due to site unavailability	28.8 hours	22 hours	1.6 hours	0.8 hours
Availability based on site caused downtime	99.671%	99.741%	99.982%	99.995%

[a]*EPO* emergency power off
[b]*HE* human error

7.3 Types of Medical Data

- From a storage perspective, imaging data can be divided into:
  - Data that may change after initial storage and are thus managed as a variable content file (VCF).
  - Data that will not change once stored and are thus managed as a fixed content file (FCF).

7.3.1 Variable Content Files (VCF)

- VCF or transactional data consist primarily of databases that comprise approximately 5% of the total stored image-related data.
  - Examples: radiology information system (RIS), hospital information system databases, and the demographic database of the PACS.
- The frequency of read/write commands to the database dictates the storage technology used to manage and replicate or backup the database.

7.3.2 Fixed Content Files (FCF)

- FCF imaging data consist primarily of DICOM objects such as images, structured reports, and curves that comprise approximately 95% of the total stored image-related data.
- Typical storage scenarios for image data include:
  - Storage on the acquisition modality for one or more days.
  - Forwarding of an image copy from the modality to "Tier 1 storage" (on-line storage) and possibly to one or more workstations for local storage of emergent studies.
- Tier 1 storage is managed by the PACS and is typically configured to store between 3 and 15 months or more of study volume depending on the clinical setting (e.g., outpatient, inpatient).
  - The PACS verifies the information in the DICOM header of the study against the information in the Radiology Information System (RIS) order for that study.
  - The PACS forwards a copy of the study to "Tier 2 storage" (long-term storage) where it is retained for the legal life of the study; a duplicate copy is either:

    Forwarded to another storage system at a remote location or

    Copied to tape or other media and manually carried off-site for "Tier 3 storage" (disaster recovery) and retained for the legal life of the study.

- Initially, four copies of a study exist until deleted from the modality
  - Then, three copies until eliminated from Tier 1 storage.
  - Finally, one copy will remain on Tier 2 and one on Tier 3 storage.
  - These Storage Tiers are different from the Datacenter Tiers discussed above.

7.4 Storage Requirements for ePHI

7.4.1 Health Information Portability and Accountability Act (HIPAA)

- The security regulations of HIPAA effective April 21, 2005 cover the confidentiality, integrity, and availability of ePHI.
- PHI has the potential to enable identification of an individual and includes: details of past, present, or future physical or mental health
  - Provision for health care or
  - Past, present, or future payment for health care

DEFINITION: Retention Period

Time, mandated by federal, state, or local statute, that medical information must be retained in its original and legal form.

- HIPAA covers ePHI stored on any type of storage media or through any means of information delivery including:
  - Portable computers and related devices
  - Electronic transmission via the Internet
  - Via e-mails or other related methods

Confidentiality is the assurance that ePHI is available to and viewable by only authorized persons or organizations.

- Requires that ePHI that is stored on media or transmitted electronically be encrypted to avoid access by unauthorized individuals (e.g., individuals in a facilities datacenter) or organizations.
- Both when in transit and when stored on physical media.

Availability is the assurance that systems responsible for delivering, storing, and processing ePHI are accessible in a timely manner by those who need them under both routine and emergency situations.

- HIPAA requires that two copies of ePHI must exist, so if one copy is accidentally destroyed during its legal life (retention period), a second copy will be available in a secure and accessible location.

Integrity is the assurance that ePHI is not changed unless an alteration is known, required, documented (via audit trial), validated, and authoritatively approved.

- When ePHI has been authoritatively approved, it should be stored electronically in a format that inhibits unauthorized alterations, e.g., Write Once, Read Many (WORM).

  See **Chap. 30** for more information about HIPAA.

7.4.2 Storage Requirements for Clinical Studies

Typical storage requirements for radiologic studies are listed in Tables 7.2 and 7.3.

- Storage requirements vary widely depending on
  - Image size
  - Number of images
  - Slice thickness
  - Protocols
  - Sequences
  - Modality vendor

CHECKLIST: DICOM Attributes Needed to Determine Image Size

DICOM Tag	Attribute	Description
0028,0010	Rows	Number of Rows in the image (X dimension)
0028,0011	Columns	Number of Columns in the image (Y dimension)
0028,0100	Bits Allocated	Number of bits allocated for each pixel sample (Divide by 8 (for 8-bit bytes) to get bytes per pixel)
0028,0008	Number of Frames	Number of frames in a multi-frame image

Table 7.2 Representative uncompressed storage requirements for various modalities

Modality	Image size			Per study basis			
				No. of images		Uncompressed MB	
	X	Y	Bytes	Avg.	Range	Avg.	Range
CR	2,000	2,500	2	3	2–5	30	20–50
DR	3,000	3,000	2	3	2–5	54	36–90
CT	512	512	2	60	40–300	32	21–157
Multi-slice CT	512	512	2	500	250–4,000	262	131–2.1 GB
MR	256	256	2	200	80–1,000	26	11–131
Ultrasound	640	480	1[a]	30[b]	20–60	9.2	6.1–18.4
Nuc Med	256	256	2	10	4–30	1.3	0.3-3.8
Digital Fluoro	1024	1,024	1	20	10–50	20	10–50
Rad Angio	1024	1,024	1	15	10–30	15	10–30
Breast Tomosynthesis			2	50	10–100 per breast	1800	450 MB–3GB

[a]3 Bytes for color

[b]For Multi-frame Series' multiply by number of frames per series

Table 7.3 Typical uncompressed storage requirement per 100,000 studies for a general radiology practice

(Excludes CTA, 3 Tesla MR, MRA, PET/CT and mammography studies)			
Modality	% of Studies	Avg. MB/Study	GB/Year
Angiography	3	15	45
CR and DR	64	42	2,688
CT	20	52	1,040
MR	5	39	195
Nuclear Medicine	3	1.3	3.9
Ultrasound	5	18	90
Total TB per 100,000 studies			**4.1 TB**

Table 7.4 Average mammography storage requirement by type of detector

Resolution (microns)	Study type	Uncompressed storage requirement (MB)	2.5-1 Lossless compressed storage requirement (MB)
50	Screening	197	79
50	Diagnostic	296	118
70	Screening	96	38
70	Diagnostic	144	58
100	Screening	52	21
100	Diagnostic	78	31

Table 7.5 Representative storage requirement for breast MR

Views	Number of sites	Uncompressed (MB)		Mean – lossless compressed 2.5-1 (MB)
		Range	Mean	
Axial w/o contrast	5	63–132	98	39
Axial with contrast-typically, 5 sequences	5	223–336	294	118
Sagittal	5	42	42	17
Typical lossless storage requirements				174

- Protocols requiring contrast increase storage requirements substantially.
- Tertiary care and/or specialty settings may have a greater percentage of cross-sectional modalities in their total study volume.

Breast Imaging

- CR: 50-micron resolution—storage requirement **4 times greater** than 100-micron resolution.
- Average mammography storage requirements based on 70% large cassette/paddle and 30% small cassette/paddle and 4 images for screening and 6 images for diagnostic studies (Table 7.4).
- Protocols and exam sizes vary by vendor (Table 7.5).

- Don't forget:
  - To account for outside imaging studies imported into PACS.
  - To consider growth
    Acquisition of or merger with new facilities or imaging centers
    New modality technology which may increase number of images and/or file sizes acquired (e.g., breast tomosynthesis)
    Increased volumes due to reduced scan times and efficiency improvements

7.5 Retention and Destruction Requirements for ePHI

7.5.1 Retention Requirements for ePHI

- Depending on the local, state, or federal statutes, ePHI must be retained in its original form (from which diagnosis was made) for a period of from 5 to 7 years or longer.
- Healthcare facilities should establish and document retention periods for all types of ePHI based on the requirements imposed by various statutes.
- Images stored for purposes of complying with regulatory backup (DR) requirements must be of the same quality as images used for diagnostic purposes.
- ePHI for minors may have to be retained until the individual reaches the age of 21 or beyond, depending on when the ePHI was acquired.
- Regulations exist (typically state statute of limitations) that require ePHI be retained for a period of 2 years after the death of a person.

KEY CONCEPT: Retention

The retention period for ePHI varies and

- Is dependent on type of ePHI
- Is dependent on the age of the individual
- Is specified by federal statute
- Varies by state and type of provider
- Complies with state's statute of limitations
- Mammography has special regulations
  - Federal Register 900.12©(4)(i),(ii)

7.5.2 Film Screen and Digital Mammograms

- Retained for not less than 5 years.
- Not less than ten years if no additional mammograms of patients are performed at the facility
- Longer if mandated by federal, state, or local statutes.
- Digitized film screen mammograms cannot be used for legal retention purposes (not original) but can be used for comparison with current digital mammograms.

- Recommended (but not required) that computer-assisted diagnosis (CAD) reports be retained as part of the mammography study.
- Lossy compressed mammograms cannot be used for legal retention purposes.

7.5.3 Destruction Requirements for ePHI

- ePHI must be destroyed so there is no possibility of reconstructing identifying information.
- Resulting from the complex rules governing retention of ePHI, automation of its destruction is not currently possible.
- The most cost-effective solution to manage outdated ePHI may be permanent retention.

KEY CONCEPT: Digital Image Destruction

When a study is deleted from a PACS storage system, the study is deleted from the demographic database so it can no longer be located or retrieved. The actual study itself *should not be deleted* from the digital media.

- Destruction of ePHI must be documented including
  - Date of destruction
  - Method of destruction
  - Description of disposed records
  - Inclusive dates covered
  - Statement that the records were destroyed in the normal course of business
  - Signature of individual supervising and witnessing the destruction
- Magnetic and optical media
  - CD, DVD, and other magnetic or optical media should be shredded.
  - Degaussing is the preferred method to destroy data on magnetic disk or tape.

7.6 Storage Technology

7.6.1 Types of Storage Media

- Parameters used to select storage media
  - Functionality—does it meet the clinical requirements for the application?
  - Response time (to read/retrieve and write information) and storage capacity.
  - Longevity—how long will the media and the components used to read and write to the media be available and be supported (technology obsolescence?),

FURTHER READING: Storage Media

Kerns R. Information archiving: economics and compliance. Boulder; 2012.

will there be backward compatibility, or will information have to be migrated? Information migration is expensive and can be time consuming. For most media, technical obsolescence is between 3 and 6 years depending on when it is purchased in the technology life cycle.
  - Total cost of ownership (TCO) of the entire storage management system for the storage media used is the total cost of purchasing and supporting the storage management system and includes hardware, software, maintenance and licensing fees, personnel, utilities, space, and information migration costs.
  - Cost—media is typically less than 5–10% of the TCO of the storage management system.
  - Durability—information stored on media must be accessible for a long period of time, factors that affect durability include wear and tear from read/write components, temperature, and humidity changes, etc.
  - Compliance features—HIPAA requirements regarding immutability and integrity of ePHI (e.g., storing in WORM format).
  - Remove ability—media that can be removed from the read/write source; evaluate the pros and cons for the specific application.
  - Storage media cabinets should have redundant power supplies and fans to minimize system failures.
- Optical Media
  - Used primarily to provide a transportable copy of individual patient study and as a low-cost, long-term storage, and disaster recovery media.
  - Inferior performance characteristics to spinning magnetic disk with respect to read/write speed capabilities and storage capacity.
  - Optical media maintains integrity of stored information for the long term if properly handled but limited useful life of 6 years due to technology obsolescence.
  - Data are typically written in Write Once Read Many (WORM) format which is advantageous for purposes of long-term storage and HIPAA requirements.
  - **Compact disk (CD)** used primarily as transportable media for an individual patient study stored in DICOM Part 10 format as WORM with a DICOM viewer.

    CDs are available as CD-R, WORM version, and also CD-RW, a re-writeable version that **must not** be used for storage of medical information or ePHI.

    Storage capacity is up to 700 MB.

    FURTHER READING: Optical Media

    www.en.wikipedia.org/wiki/computer_data_storage#optical.
  - **Digital versatile disk (DVD)**

    Available as DVD-R and DVD-RW; only the DVD-R version should be used to store medical information or ePHI.

    Single-sided, single-layer DVD will store up to 4.7 GB and the single sided, double-layer will store 8.5 GB.

    These disks are typically used for inexpensive long-term storage or DR within a robotic storage system.

- Spinning magnetic disk
  - Also known as Hard Disk Drive (HDD)
  - Most storage today uses spinning media
- Performance
  - To decrease time to access data, increase rotational speed (measured in rpm).
  - To increase throughput and storage capacity, increase media storage density (measured in Terabytes). This number is likely to rise as technology improves.
- Redundant Array of Independent Disks (RAID)
  - RAID writes parity data across the array of disks, which are organized so that the failure of one disk in the array will not result in loss of any data.
  - A failed disk can be replaced by a new one, and the data on it reconstructed from the remaining data and the parity data.
  - As a result of this redundancy, less data can be stored in the array.
  - Selection of the appropriate RAID level depends upon the application, degree of protection against data loss, storage capacity, and performance (number of write/read per second).
  - RAID is not a substitute for backing up data on another media located remotely from the RAID.
  - There are several different types of RAID configuration (Table 7.6).

FURTHER READING: Hard Drives and RAID

www.en.wikipedia.org/wiki/hard_disk_drive.
www.en.wkpedia.org/wiki/redundant_array_of_independent_disks.

CHECKLIST: Commonly Used Types of RAID

RAID 0—(Striping) Data are striped across several disks to improve performance and obtain 100% storage capacity. No redundancy is provided; if one disk fails, data on that disk are lost.
RAID 1—(Mirroring) Two groups of similar disks store exactly the same data. No data will be lost as long as one group of disks survives.
RAID 5—(Striped disk with parity) Combines three or more disks in a way that protects data against loss of any one disk. Storage capacity is reduced by one disk's worth. A "hot spare" disk should be added to the array so data can be quickly and automatically restored. If an additional disk fails while data are restored, data in the entire array can still be lost.
RAID 6—(Striped disk with dual parity) Can recover if two disks in the array are lost. A "hot spare" disk should be added to the array so data can be quickly and automatically restored.
Raid 1+0 or RAID10—uses both striping and mirroring, consists of a striped set of mirrored subsets, provides fault tolerance and improved performance.

- Content-addressable storage (CAS) is a method for storing information based on its content (object-based) rather than its storage location and is used for fixed content files of ePHI to meet regulatory requirements.
- Solid state media, as used in USB flash drives, has begun to replace HDD on most computers, but is not yet inexpensive enough for mass storage.

Table 7.6 Characteristics of RAID levels in common use

Feature	RAID 0	RAID 1	RAID 5	RAID 6	RAID 1+0
Minimum number of drives	2	2	3 + hot spare	4 + hot spare	4
Data protection	None	Single drive failure	Single drive failure	Two drive failures	One drive failure in each sub-array
Read performance	High	High	High	High	High
Write performance	High	Medium	Low	Low	Medium
% Storage utilization	100	50	67–94	50–88	50
Typical applications	Not applicable for storage of ePHI	Operating systems, transactional databases	Tier 1, 2, and 3 storage of ePHI	Tier 1, 2, and 3 storage of ePHI – high availability required	Fast transactional databases, application servers

7.6.2 Storage Management Infrastructure and Hardware

- Data storage is more than just the right choice of hardware. An entire system for storage management must be built around that hardware. Considerations include
  - Capacity—amount and type of data (file level or block level) to be stored or shared
  - Performance or availability—input/output and throughput requirements
  - Scalability—long-term data growth
  - Availability and reliability—how mission-critical is the application
  - Data protection—backup and recovery requirements
  - IT staff resources and capability
  - Initial and annual budget availability
- Direct attached storage (DAS)
  - Simplest and least expensive storage technology where computer (server) is directly attached to storage device such as RAID or tape system.
  - Workstations must access server to connect with storage system.
  - Since the server must handle processing for the applications as well as servicing the storage device, availability of stored data is impacted and thus system performance, for example, queries for clinical studies are compromised.
  - Other disadvantages of DAS are scalability and the inability to automate backup and minimize planned system downtime.
- Network attached storage (NAS)
  - Developed to address the inherent weaknesses of a server-based infrastructure such as DAS.
  - NAS is a file-based storage system with management software that is 100% dedicated to serving files over a network.

  - NAS eliminates the need for the server supporting storage and responding to read/write responsibilities.
  - Uses industry standard IP network technology and protocols (Transmission Control Protocol/Internet Protocol [TCP/IP], Common Internet File System [CIFS], and [NFS]).
  - NAS can provide:
    Simple and cost-effective ways to achieve fast data access for multiple clients at the file level.
    Performance and productivity gains over DAS.
    Data protection features such as replication and mirroring for business continuance.
    Ability to consolidate DAS resources for better utilization.
    Scalability with storage capacity in the multi-terabyte range with efficient use of datacenter space.

- Storage area network (SAN)
  - A SAN is a dedicated, high-performance storage network that transfers data between servers and storage devices, separate from the local area network.
  - SAN moves data at the block level rather than at the file level as does DAS and NAS, thus it is ideal when large quantities of data must be moved.
  - Since the SAN operates on a block level and workstations operate at the file level, the PACS or other application must provide a block level to file level conversion.
  - A SAN environment is more costly, complex to manage, has many components, and requires sophisticated management software than other storage management environments; however, it provides superior:
    - Performance
    - 24/7 data availability
    - Reliability, with a high degree of fault tolerance
    - Scalability
    - Data protection
    - Storage virtualization
  - In a SAN infrastructure, storage devices such as DAS, NAS, RAID arrays, or tape libraries are connected to servers using fiber channel.
  - Since a SAN provides a high-speed connection between a storage device and a server,
    - The server can respond to the computations required by the applications without supporting the storage system.
    - Copies of the server's operating system and applications can reside on the storage systems supported by the SAN and can be rapidly restored should they fail.

FURTHER READING:
Network Storage

Alabi D. NAS, DAS or SAN? Choosing the right storage technology for your organization. www.storagesearch.com/xtore-art1.html.

- Fiber channel (FC)
  - FC is a gigabit-speed network technology used primarily for storage networking.
  - FC uses special cabling to move large volumes of data without the distance and bandwidth limitations of SCSI.
  - FC has transmission rates of 1, 2, 4, 8, 10, and 20 Gbps.
  - FC is highly reliable and enables simultaneous communication between workstations, mainframes, servers, data storage systems, and other peripherals.
- Hybrid SAN/NASs
  - Adds file interface to SAN
  - Supports NAS standards
  - Leverages a common storage infrastructure
- Internet Small Computer System Interface (iSCSI) Storage system
  - iSCSI uses IP networks rather than fiber channel to transmit data.
  - Unlike FC that requires special-purpose cabling and is more expensive, iSCSI can run existing, less expensive network infrastructure.
  - iSCSI enables data storage and retrieval from remote and independent storage systems because IP networks such as LANs, WANs, and the Internet are widely available.
  - iSCSI has transmission rates of 1 Gbps and 10 Gbps but require more overhead than FC.

7.6.3 Storage Management Software

Storage management software is primarily used for data protection, including:

- File backups
  - Daily backups of entire system, with continuous data redundancy
  - Used for VCF such as RIS or demographic database
  - Make daily manual backup to tape and stored in a secure off-site location
  - Various automate backup methods available; however, secure off-site copy must have
    - Replication
    - Point-in-time (PIT) copies
    - Continuous data protection
- Archiving
  - Used for FCF such as patient studies.
  - Can be automated using hierarchical storage management (HSM) application, grid storage, or clinical information lifecycle management application (CILM).
  - Can be performed manually by copying studies acquired each day to tape or other removable media and stored in a secure remote off-site location.
- Grid storage
  - Software application used to manage Tier 2 and Tier 3 storage of FCF at multiple geographically separated locations.

  - Each location may have one or more nodes of the grid software operating on off-the-shelf hardware that manages a storage resource.
  - Nodes are interconnected via network links and can process DICOM store and query requests from any node.

7.6.4 Cloud Storage

Cloud technology is transforming the architecture of imaging application delivery and storage. In many cases, facilities have determined that the purchase of on-site physical storage is not a cost-effective investment. The high upfront capital costs and physical space requirements combined with rapidly increasing media capacities make the cloud solutions an attractive alternative. Furthermore, newer storage methods (e.g., magnetic tape to magneto-optical Jukeboxes to spinning disk RAIDs) have the potential to render established archives obsolete or lacking support for the clinical workflow.

- Cloud storage is service based.
  - The service is designed to serve the specific needs of a set of consumers.
  - The performance of cloud storage is based on service level agreements and response time, not on technology limitations.
- Is scalable and elastic.
  - The service can scale capacity up or down as the consumer demands.
  - This scalability is a feature of the underlying storage infrastructure.
  - Elasticity is a trait of shared pools of resources.
- Is shared.
  - The underlying infrastructure, software, and/or platforms are shared among the storage users. IT resources are used with maximum efficiency.
- Is metered by use.
- Applies the standards of Internet technologies.
- Types of Clouds
  - SaaS (Software as a Service): Cloud providers install and operate application software in the cloud and cloud users access the software from cloud clients.
  - PaaS (Platform as a Service): Cloud providers deliver a computing platform including operating system, programming language execution environment, database, and web server.
  - IaaS (Infrastructure as a Service): Cloud providers offer computers—as physical or virtual machines, raw storage, firewalls, load balancers, and networks.
- Introduces unique challenges in PHI and privacy compliance.
- Application Service Provider (ASP)
  - Many vendors now utilize cloud technologists to offer an ASP Contract Version.
  - These ASP contracts can be structured as a SaaS or an IaaS model.
  - Contract imposes a charge per study and/or a fixed monthly recurring charge.
  - May be penalties for exceeding contracted study volumes.

7.6.5 Vendor Neutral Archive (VNA)

Most PACS vendors would prefer to control the data storage of images. But storage technology is commoditized, so that third-party storage solutions can be appended to the PACS.

- Advantages include:
  - Provides a comprehensive imaging record for patients across an enterprise
  - Patient Centric—view of all image types from different departments
  - Scalable image and data storage with Life Cycle Management
  - Ability to query and retrieve images and related information
  - Use open standards
  - Transparent to vendor changes and upgrades
  - Does not require data migrations, conversions, or changes to data formats or interfaces
- Caveats include:
  - Must maintain patient privacy and security
  - Any viewing, acquisition, and workflow management components use compatible interfaces

7.7 Compression of Medical Images

7.7.1 Basic Concepts

- Reasons to compress medical images
  - Decrease transmission times for medical images
  - Decrease storage requirements for medical images
  - Decrease bandwidth requirements for the transmission of medical images
  - Reduce cost for storage management and infrastructure
- Caveats
  - The primary justification for compressing FCF DICOM objects is reducing the time to transmit an image from one location to another that directly affects productivity and reduces storage requirements and cost.
  - Studies interpreted based on a lossless compressed image **must** be stored in lossless compressed format.

KEY CONCEPT:
Compressed Storage

The image format from which the diagnosis is made *must* be the format in which the study is stored.

- Studies interpreted based on a lossy compressed image MUST be stored using the same lossy compression algorithm.
- All medical images must be stored in either a lossless or lossy DICOM compliant compression format.
- Proprietary compression formats.
  Will negatively impact interoperability of images and data migration
  Can lock you into your current PACS vendor

7.7.2 Lossless Compression

- Run length encoding (RLE)
  - Uses the redundancy within the image to decrease the image size
  - Replaces sequences of the same data values within a file by a count number and a single value
  - Reduces image size between 1.8 and 2.8 depending on modality and body part

DEFINITION: Lossless Compression

Digital data compression in which all the original data information is preserved and can be completely reconstituted.

7.7.3 Lossy Compression

- Primarily used for web distribution of images to the enterprise for review purposes rather than primary interpretation.
- Lossy compression ratios must be stated on each image of a study.
- Study can be restored without clinically significant data loss.
- Lossy compression ratios depend on body part and modality.
- Typical compression ratios used when interpreting lossy compressed images.
  - DR and CR: 20 to 1
  - CT: 10 to 1
  - MR: 5 to 1

DEFINITION: Lossy Compression

Methods of digital compression in which the original information cannot be completely reconstituted.

PEARLS

- Eliminate isolated silos of storage and implement an enterprise storage solution such as a vendor-neutral archive.
- There are various storage media types and storage architectures to choose from. Be aware of the major acronyms in this field.
- Estimating storage requirements is at best an educated guess. Plan for the future, not the present.
- Understand the privacy requirements that apply to stored medical data.
- Retention and destruction of medical data require well-reasoned policies and strict adherence. Do not delete data to make space—buy more space to make space.

Self-Assessment Questions

1. Which of the following is an advantage of departmental storage over enterprise storage?

   (a) Vendor independence and best-of-breed purchasing
   (b) Reduced cost of hardware and licensing fees
   (c) Ease of providing redundancy
   (d) Ease of initial setup

2. Which of the following is true about the economics of storage management?

   (a) Once storage is purchased, it is no longer an ongoing annual budget item.
   (b) Media itself accounts for approximately 50% of the total cost of storage.
   (c) Storage costs include personnel, utilities, datacenter space, and data backup.
   (d) Technological obsolescence will require upgrades approximately every 10 years.

3. Regarding datacenters, which of the following is true?

   (a) Tier 4 is the highest Tier rating for datacenters.
   (b) The IT Department is responsible for specifying the characteristics of the resources and the availability requirements.
   (c) Power and cooling represent a small cost fraction of datacenter costs.
   (d) 24/7 information availability is desirable, but not necessary.

4. The retention period for PHI depends on all of the following EXCEPT:

   (a) Age of patient
   (b) Gender of patient
   (c) Federal statues
   (d) State statutes

5. Which of the following must be documented when destroying PHI?

   (a) Method of destruction
   (b) Inclusive dates covered
   (c) Statement that the records were destroyed in the normal course of business
   (d) All of the above

6. Which of the following is an example of optical media?

   (a) DVD
   (b) Hard drive
   (c) Tape drive
   (d) Flash drive

7. Which of the following is **not** a form of network storage?

   (a) RAID
   (b) SAN
   (c) NAS
   (d) VNA
   (e) Cloud

8. A Tier 4 datacenter with a guaranteed availability of 99.995% allows for how much downtime per year?

   (a) None
   (b) 5.25 minutes/year
   (c) 26.28 minutes/year
   (d) 52.56 minutes/year

9. How frequently would you schedule an event to remove imaging data that was due for deletion?