

# NDMA : Storage, retrieval, and clinical use of digital mammography

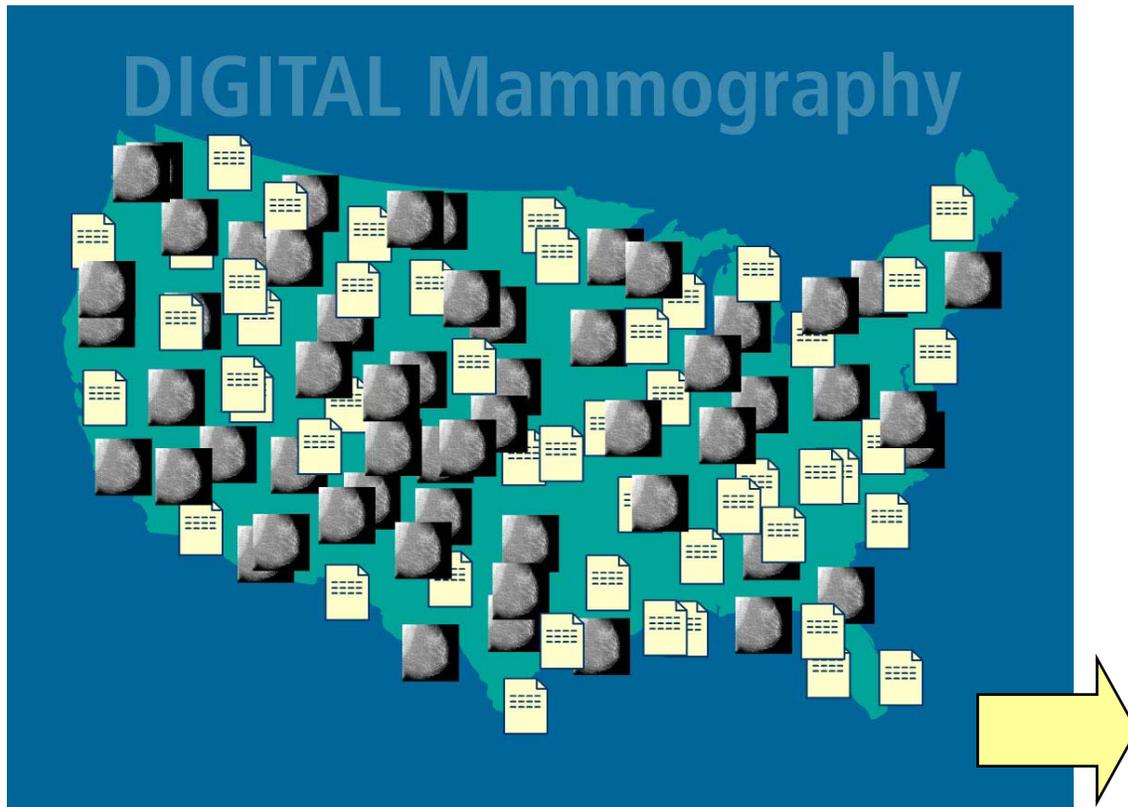
National Cancer Institute  
Bethesda, Md, Sept. 26, 2002

Robert Hollebeek  
Professor, University of Pennsylvania

Mitchell Schnall  
Hospital of the University of Pennsylvania

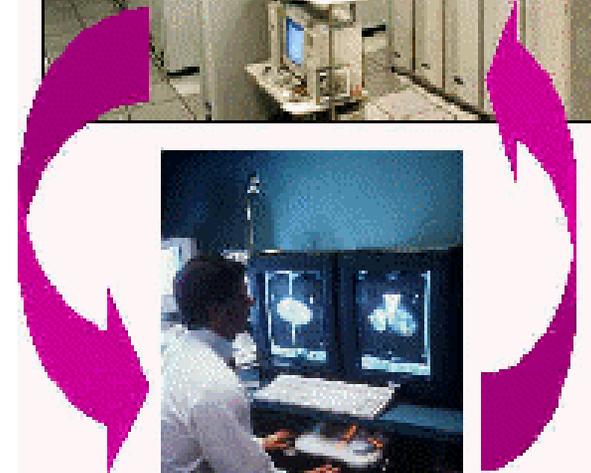
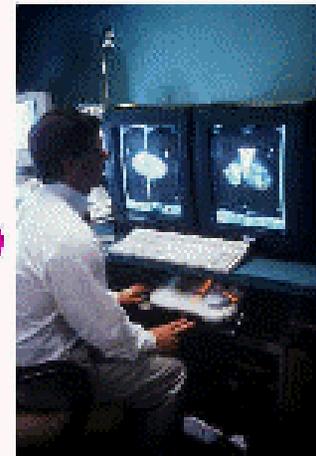


# The Challenge: Collect, Manage, Store, Retrieve, Index, and Mine



Highly Distributed and Massive Source

Use High Performance Networks, Hierarchical Storage and Indexing



# Major Challenges

- **Huge Quantity of Data**
- DICOM database and Indexing
- Hospital Web Interface
- Hospital DICOM Server
- Local Caching and forwarding
- Internet 2 VPN Connectivity
- front-end portal systems (wall plug)
- security for hospital connections
- **HIPPA compliance**
- High speed bulk transfer
- high capacity storage
- **load balancing and scalability**
- XML query specification
- Applications

# NDMA : Wide Area Data Infrastructure for storage/retrieval/use of medical records (focused on radiology)

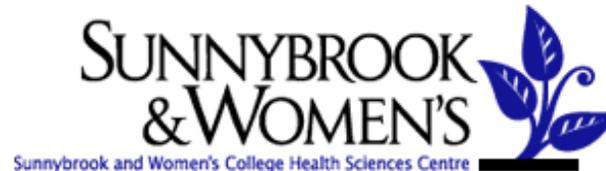
Funded by

- Next Generation Internet / Internet 2
- National Institutes of Health and the  
National Library of Medicine



# ***NDMA Team Members***

- University of Pennsylvania
  - Hospital at UPenn
  - National Scalable Cluster Project
- University of Chicago
- University of North Carolina
- University of Toronto (Sunnybrook and Women's College Health Science Centre)
- BWXT-Y12 (Advanced Computing Technologies), Oak Ridge, TN
- Web Site:



**<http://nscp.upenn.edu/NDMA>**

# Digital Radiology GRID

- externally managed attachment portals
- static networks with VPNs and hardware encryption
- system wide query and retrieve but with virtual filerooms separating hospitals
- turnkey operation, roles, smartcards

Improved functionality and lower cost through Wallplug-like Grid services.



# Focus: Digital Radiology

- **Hospital Digital Radiology Data**
  - Very large data sources - great clinical value to digital storage and manipulation and significant cost savings
  - 7 Terabytes per hospital per year
  - dominated by digital images
- **Why we chose Mammography as a starting point**
  - clinical need for film recall and computer analysis
  - large volume ( 4,000 GB/year ) (57% of total)
  - storage and records standards exist
  - great clinical value to this application

# System is DICOM compliant: i.e. it can accommodate

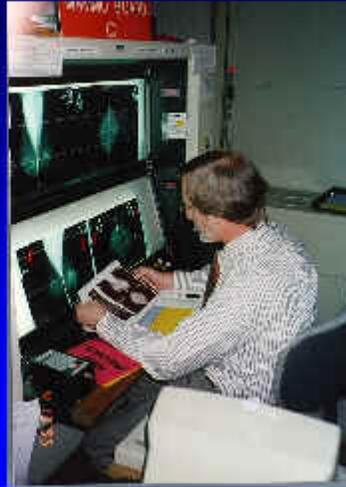
- Mammograms
- BiRADs reports
- X-rays
- MRI
- CAT scans
- endoscopies
- ...

# NDMA: Three Initial Applications

- Archive Storage and retrieval for clinical use
- Training and Teaching tools and collections for Radiology Departments
- Computer Assisted Diagnosis as a network accessible service

## Computer-Aided Diagnosis

- Application of algorithms to reduce variability
- Front-end pre-processing using Focus-of-Attention Region algorithms
- Potential centralized service over the network
- Auto-training of CAD tools

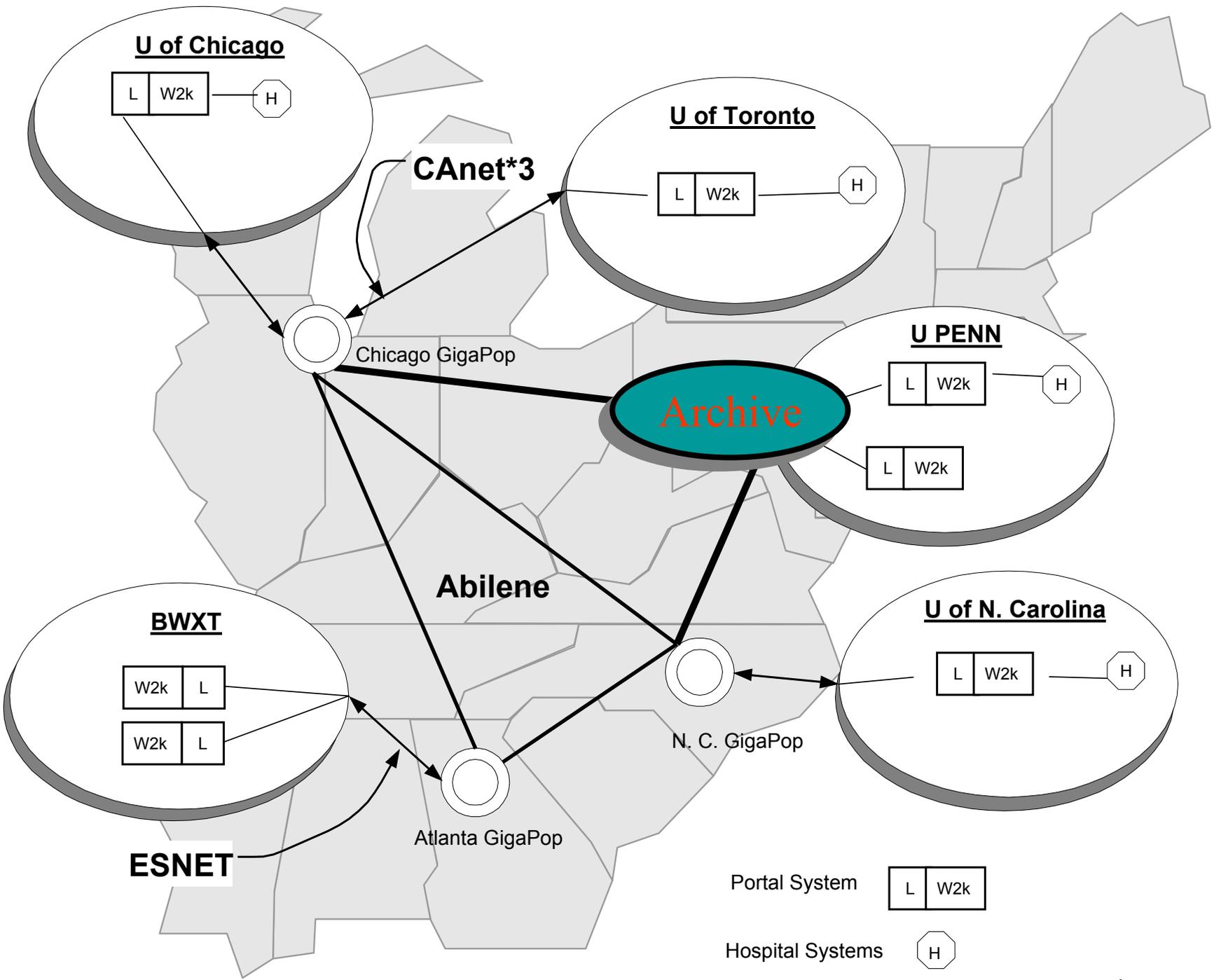


## Training Algorithms and computers

### Training, Teaching, Evaluation

## Training Radiologists and Students



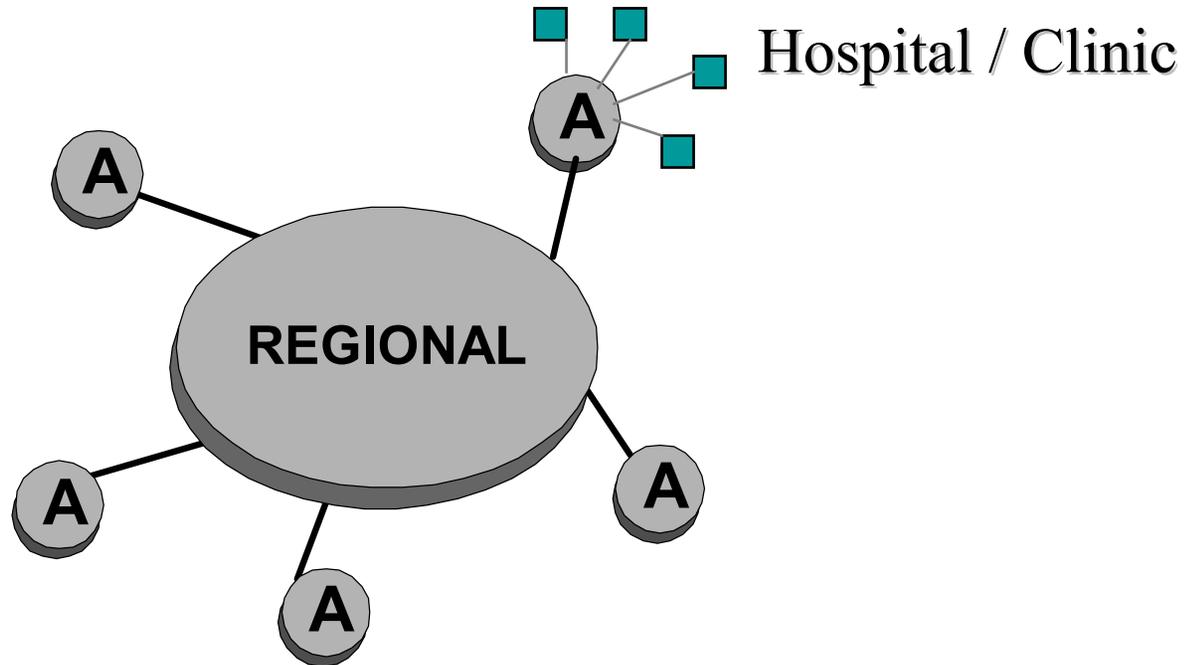


# Storage Hierarchy: small, medium and very large

7 R @ 4,000 TB/yr

20 A @ 100 TB/yr

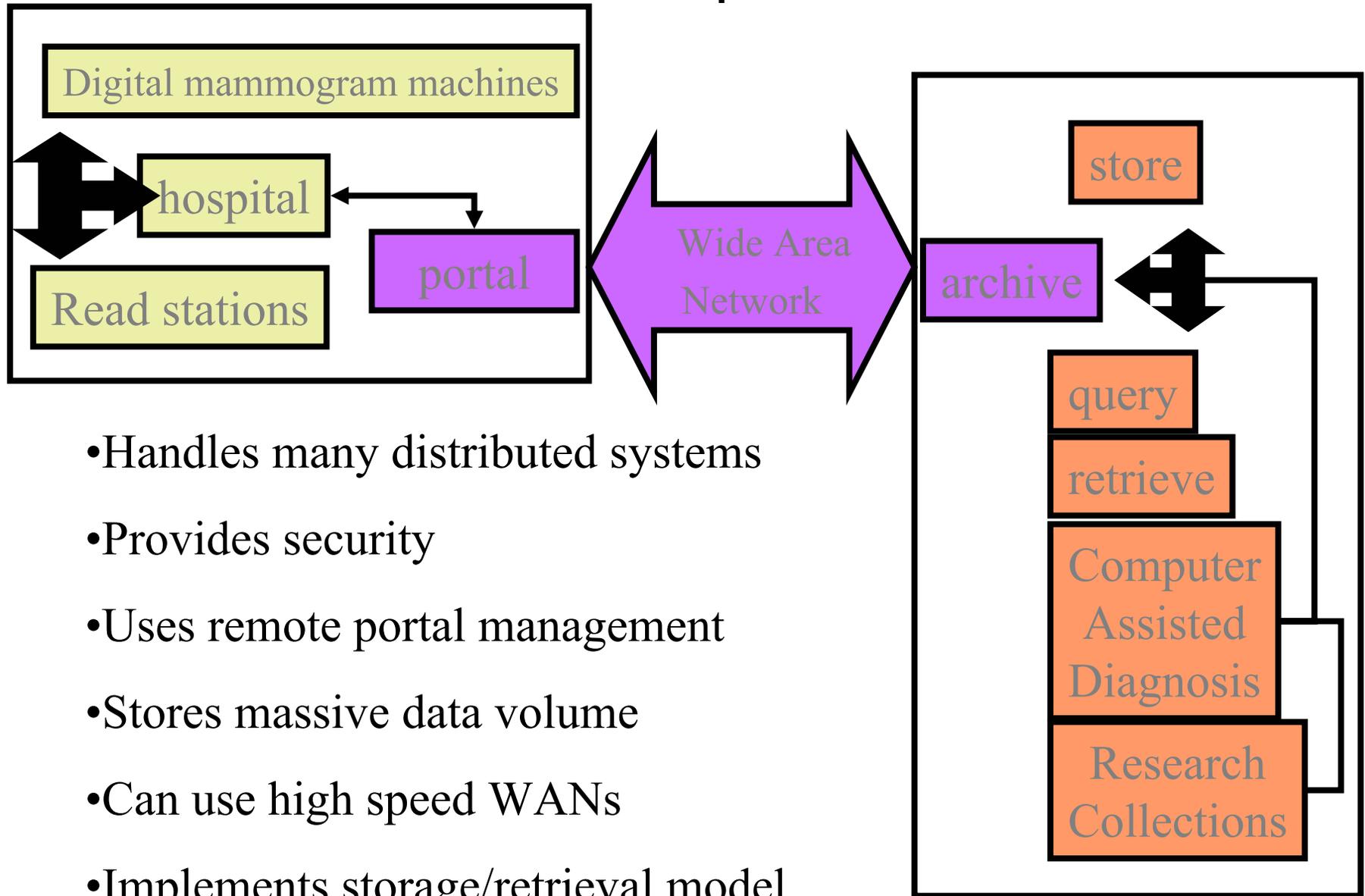
15 H @ 7 TB/yr



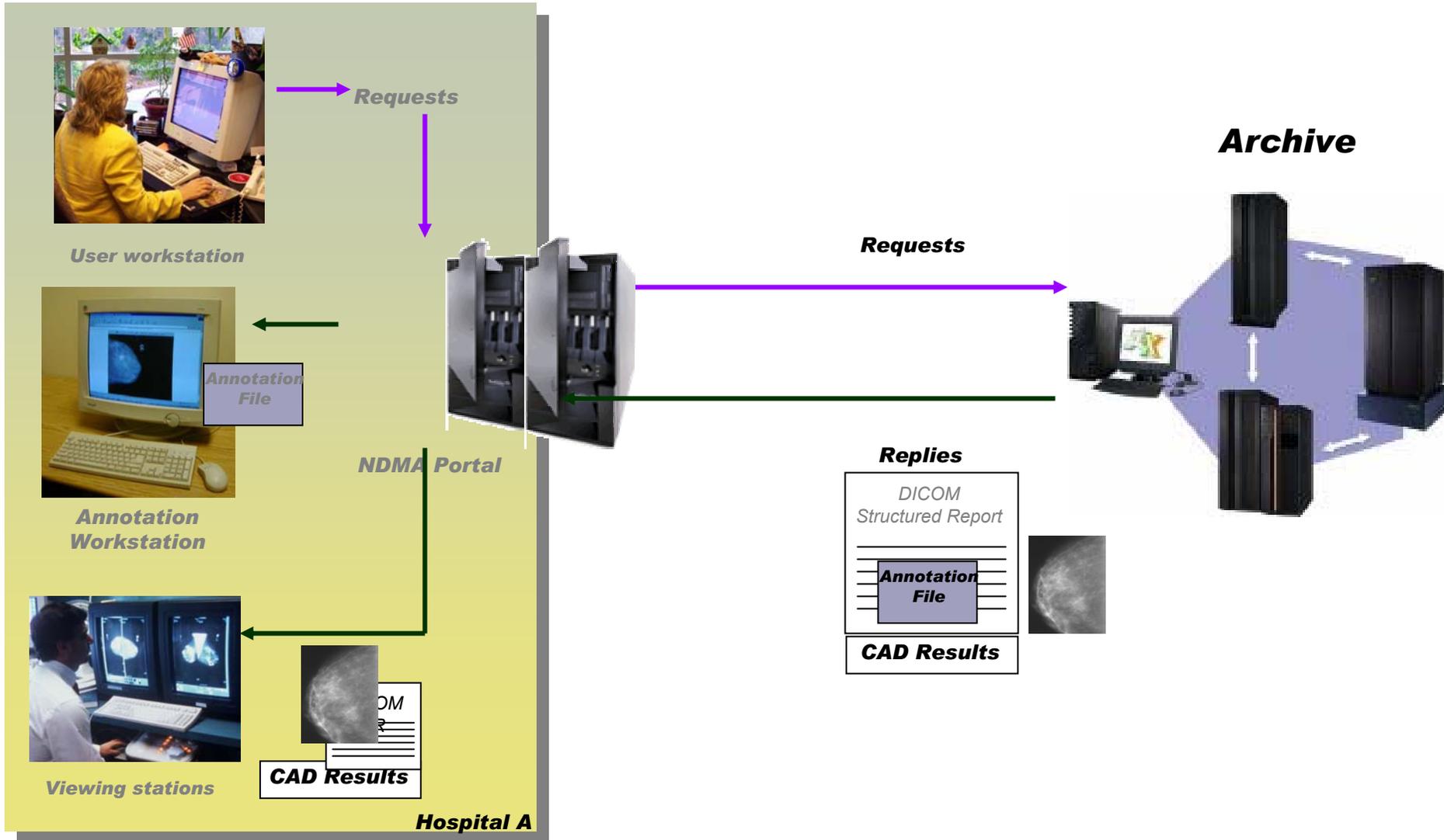
Goal: Distribute Storage Load and Balance Network  
and Query Loads

# Connect the Hospital to the Archive via the Grid

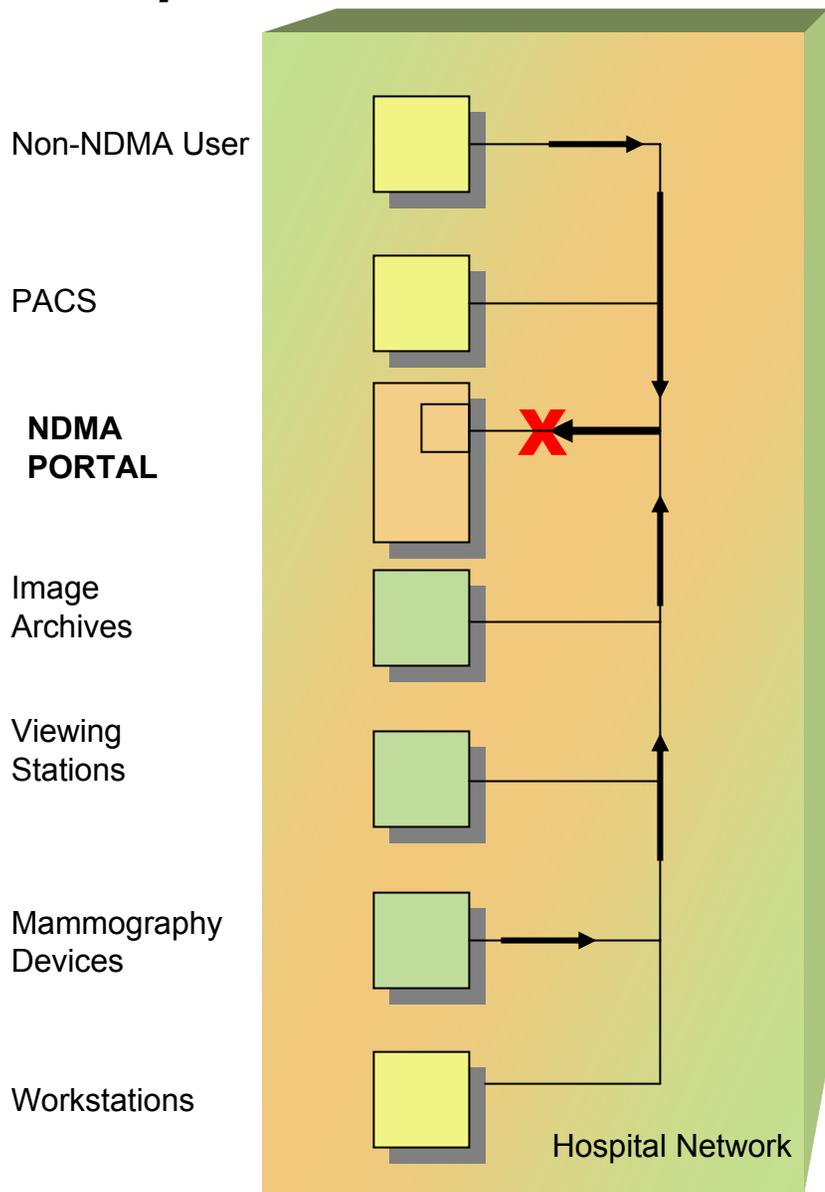
:



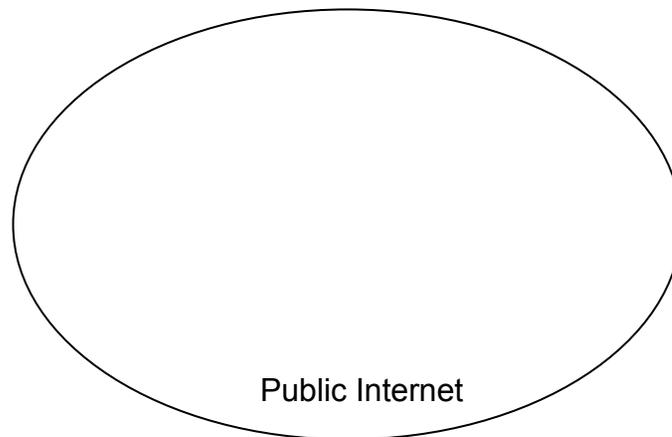
# National Digital Mammography Archive



## ***The portal controls access from the hospital***



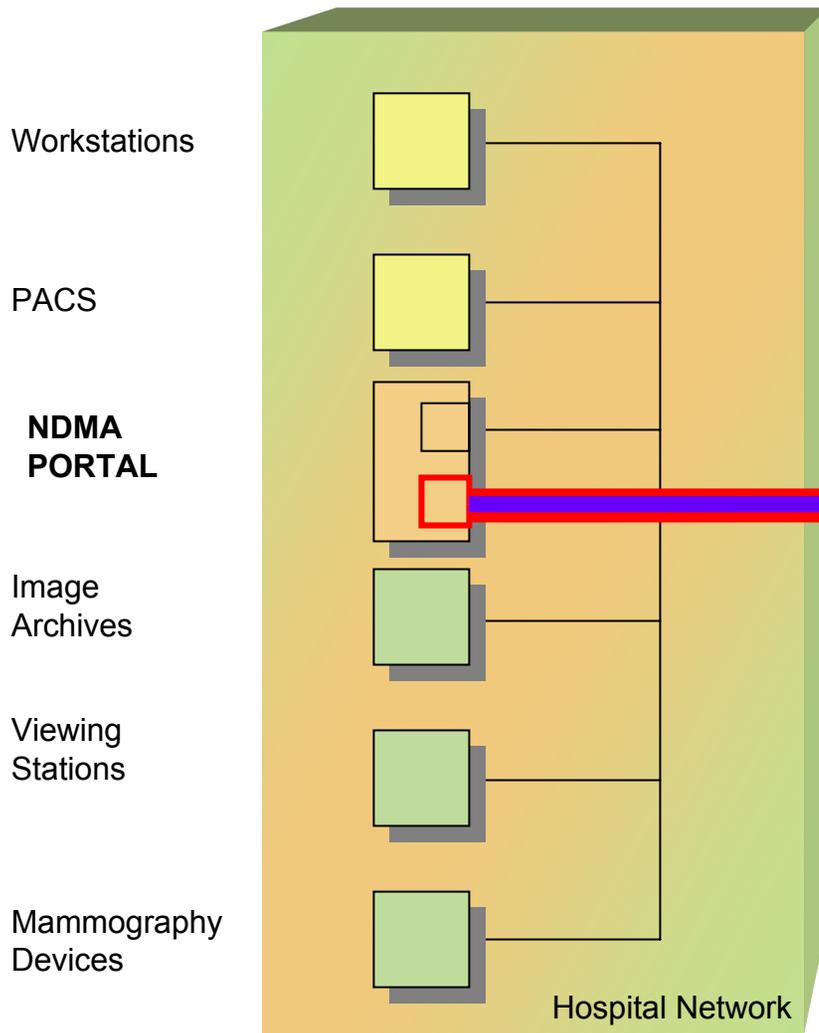
***The NDMA Portal sits on the Hospital Network and will communicate with approved devices but block others***



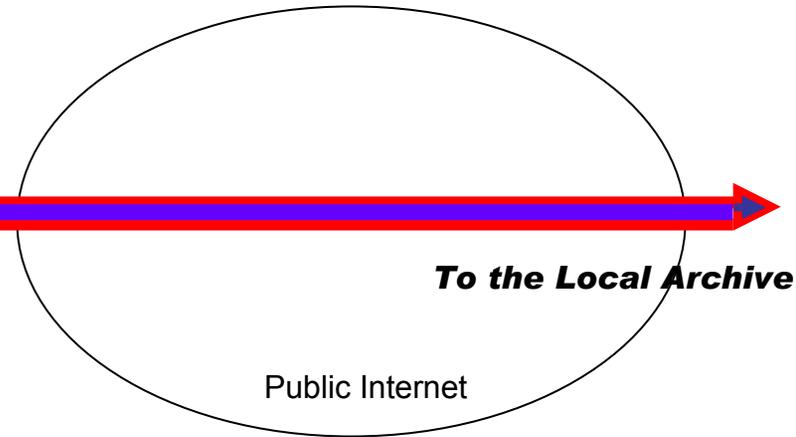
***The portal is itself a firewall type device.***

***The portal is very selective about what communications it allows***

***Only verified https and DICOM are allowed between the Portal and other Hospital devices***



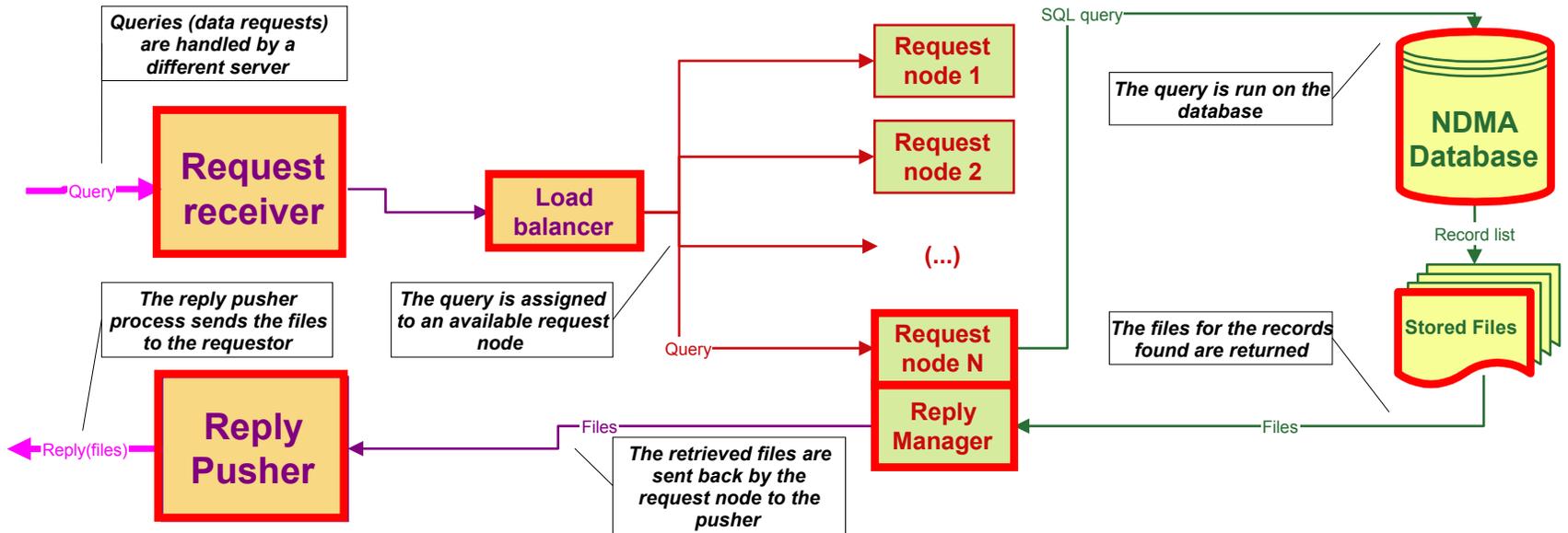
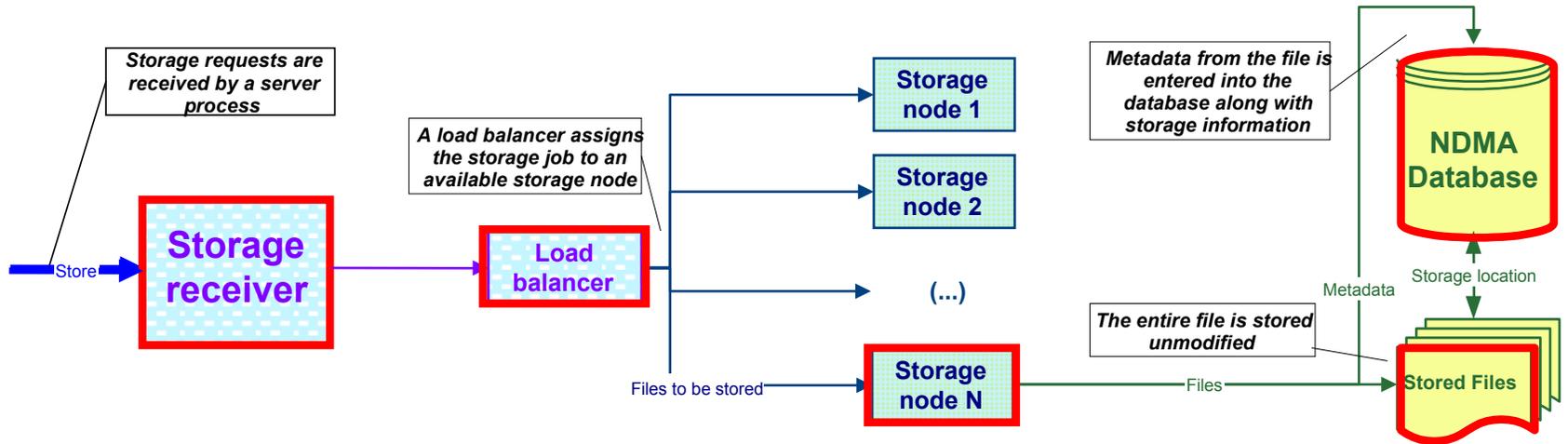
***The portal communicates with the Archive via a dedicated network card across a “tunnel”.***



***All communications both to and from the Archive are over a “Virtual Private Network” (VPN) providing hardware encryption.***

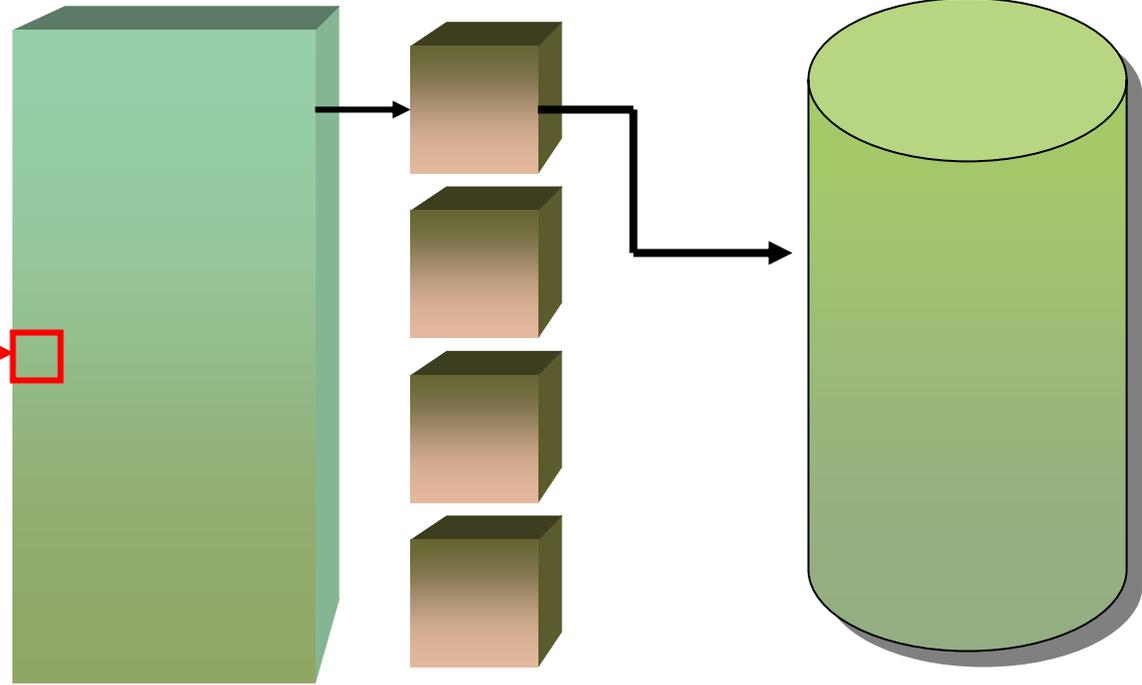
# Scalable NDMA Archive System

## Basic Function



**The requests (both to store and to request records) are received at the Archive's Portal**

**The Archive further processes the requests**



**Archive Portal**

**Archive Processors**

**Database**

**Secure**

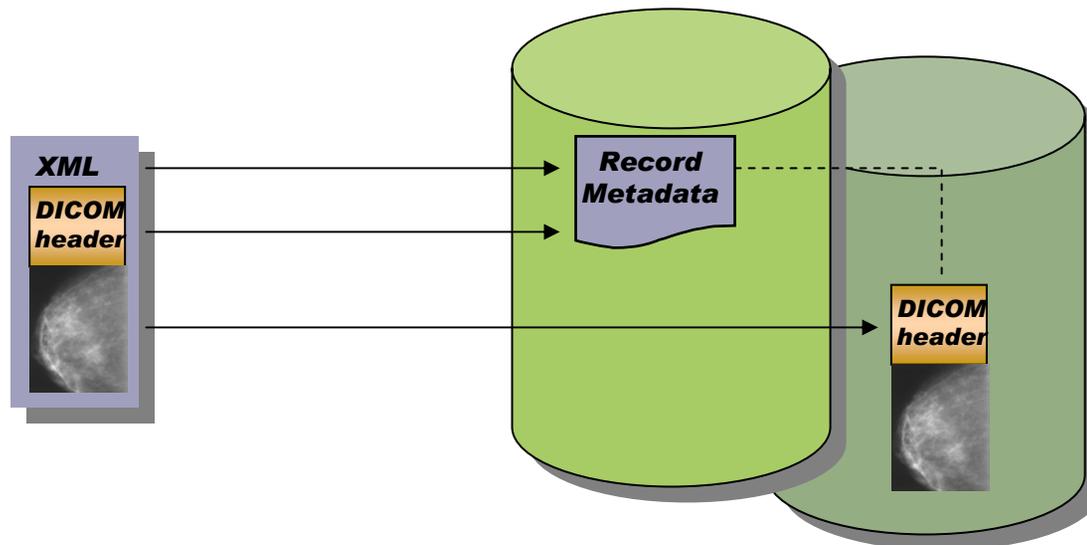


# STORAGE

**Information is parsed from the DICOM Header and the XML for the metadata and stored in the Archive Database**

**Original DICOM objects are stored with links to the database metadata**

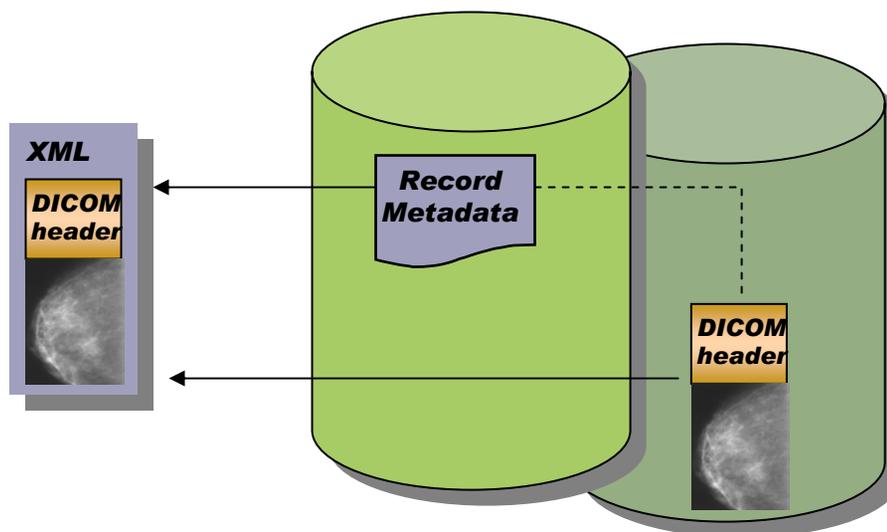
**(DICOM Objects can be images, structured reports, annotations, or CAD results)**



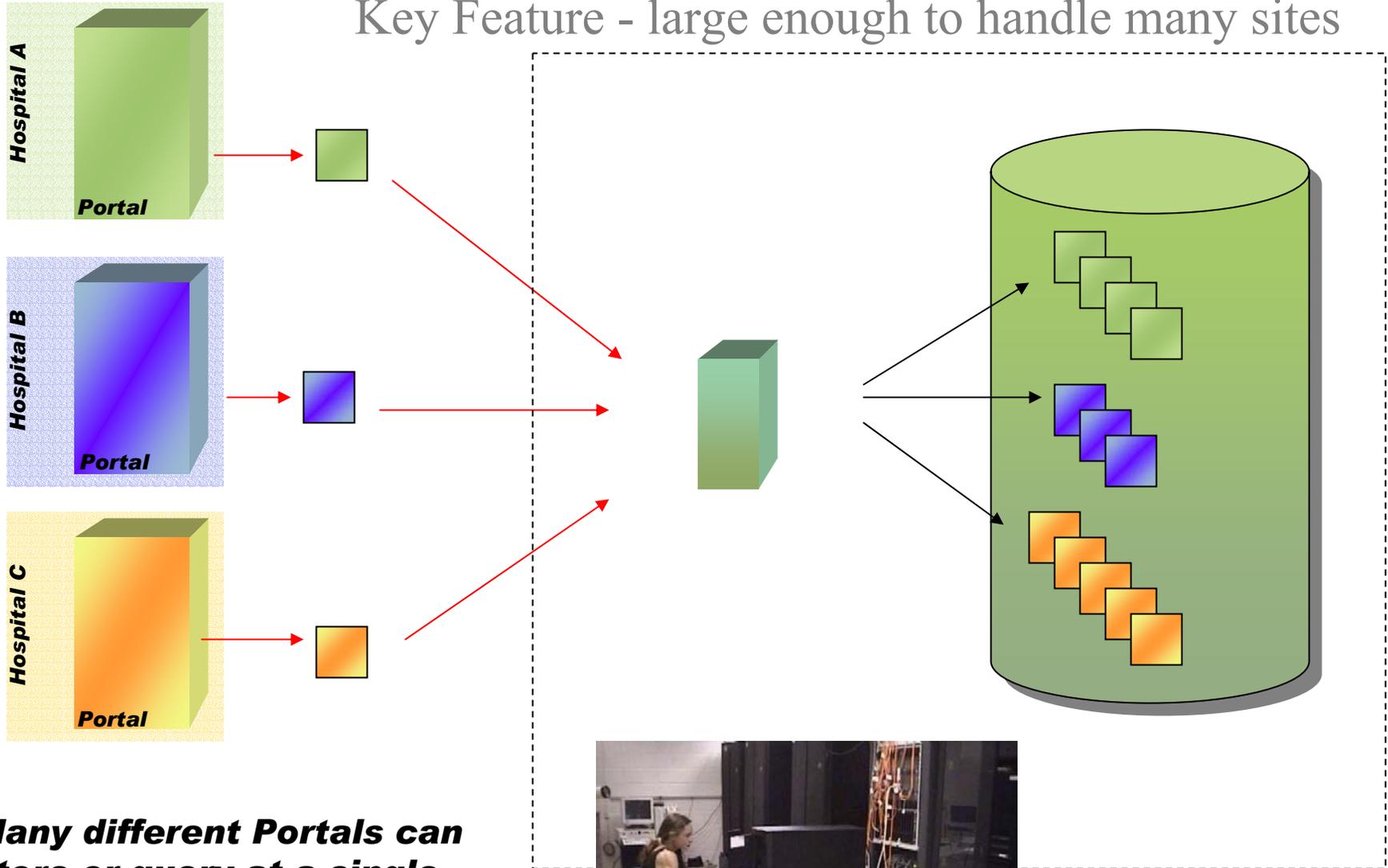
# RETRIEVAL

**Images are found by searching the metadata for specified criteria**

**Matching DICOM objects are packaged with identifying metadata in the XML and returned to the Portal**

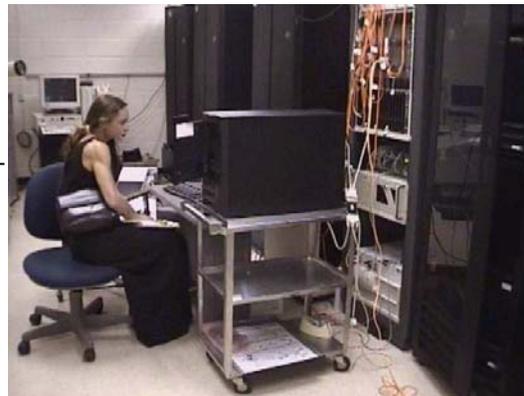


Key Feature - large enough to handle many sites



**Many different Portals can store or query at a single Archive**

**Archive**



# IBM-Penn regional pilot hardware



32 IBM x342's



16x Myrinet

32x 100BT switched

4x Gigabit Ethernet

Archive will provide data for

- computer algorithm training for CAD
- radiologist training and teaching collections
- Research and clinical use

GRID:

Deliver storage, clinical, and CAD services to many hospitals simultaneously

Provide easy and quick access to records (with the proper authority and HIPPA tracking)

Provide secure networked systems for hospital to hospital information transfer