

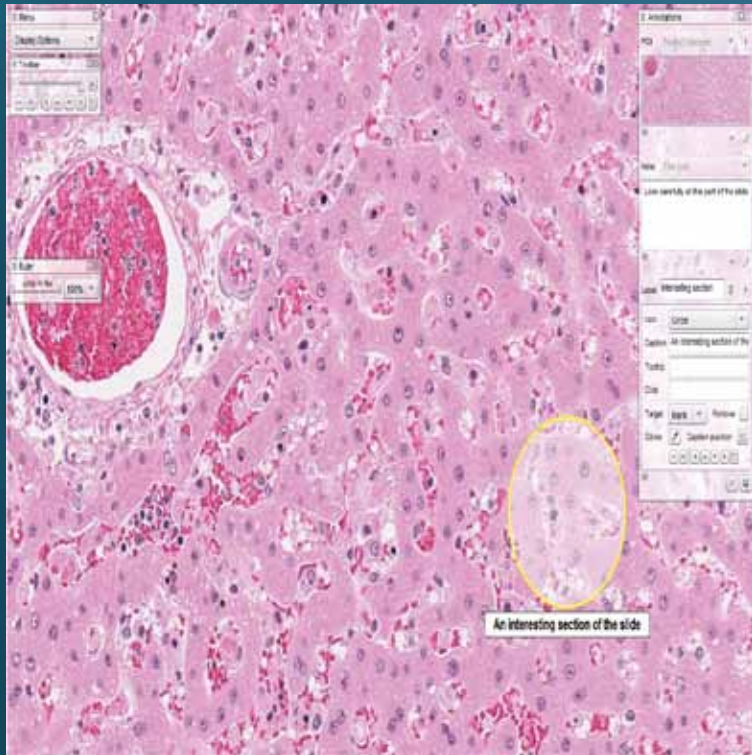
# A Flexible, Open, Decentralized System for Digital Pathology Networks



Robert Schuler, David E. Smith, Gowri Kumaraguruparan,  
Ann Chervenak, Anne D. Lewis, Dallas M. Hyde, Carl  
Kesselman

HealthGrid 2012

# Introduction



- What is Pathology?
  - Study and diagnosis of diseases
  - Microscopy is a critical component
- What is Digital Pathology?
  - Management of pathology data using computers
  - Emerging technology transforming sciences
- What is a Whole Slide Image?
  - High-resolution capture of microscope slide (20x to 100x)
- What is a Digital Pathology Network?
  - Sharing pathology data outside local environment
  - Platform to improve workflows, education, and collaboration

# Motivation

- Collaboration
  - Share pathology data across sites
  - Allow users to search across the entire participating network from a single request
- Imaging
  - Instant access to whole slide images over the web
  - Rich annotation capabilities
  - Minimal client-side software
- Security
  - Local storage and data entry
  - Control over which data to publish

# Challenges - Collaboration

- Application data schema and ontology differences
  - e.g. immunology research of animals vs. clinical oncology of humans
  - Differences even between labs within single organization
  - Requirements change as pathologists gain experience with the system
- Proprietary and closed architectures
  - Limited data sharing
  - Difficult to integrate applications together
- **Distributed system with flexible data schema and ontologies to easily share data**

# Challenges - Image Size

- Digital slide images are MUCH larger than typical digital images
- Low network bandwidth
- Images must be instantly viewable without full download



**1/4<sup>th</sup> megapixel  
300 KB each**



**14 megapixels  
4 MB each**



**80,000 megapixels  
40 GB each**

# Challenges - Image Format

- No standard data or whole slide image format
  - DICOM supplement 145 for Whole Slide Imaging published but not yet implemented
- Multiple closed image formats
  - Pixel storage and compression differences
  - Annotation storage, capabilities, coordinates
  - Metadata locations and implementation
- Vendor-specific imaging clients
  - Limited platform, browser support
  - Feature-rich
- Pixel data, annotations, and metadata must be converted to a common format



# Challenges - Security

- Sensitive data sets
  - Regulatory requirements
    - e.g. HIPAA
  - Protect access to data
    - Even to other parts of the organization
- Network firewalls
- Retain control over local environment
- Data must be protected, maintained locally, and selectively published

# Related Systems

**No reusable infrastructure for digital pathology networks**

**Standalone virtual slide servers**  
OMERO

**Online digital pathology resources**  
Pathbase mutant mouse database  
Zebrafish Atlas

**No whole slide image and annotation sharing**

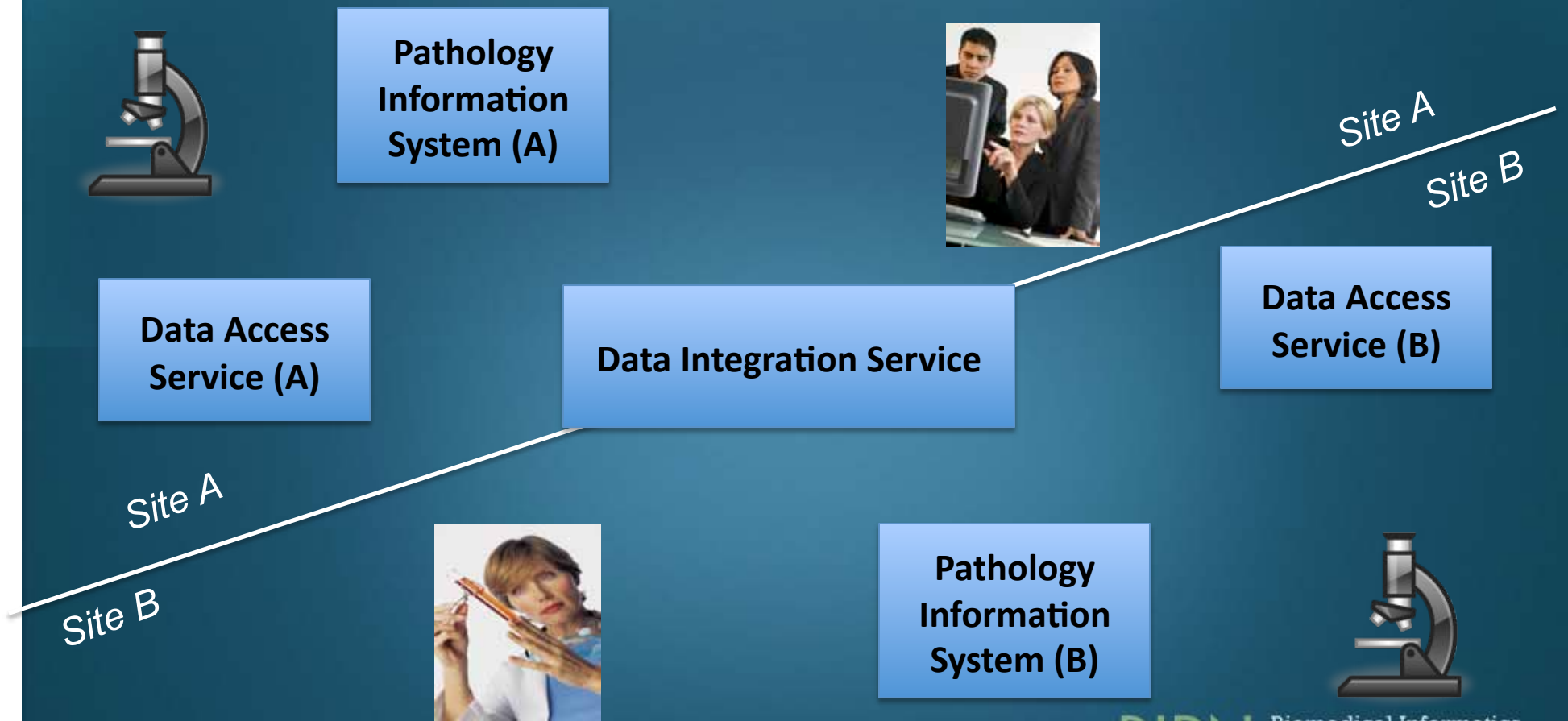
**Specimen Inventory Management**  
caTissue

**Neuroimaging Systems**  
Human Imaging Database (HID)  
eXtensible Neuroimaging Archive Toolkit (XNAT)

**Limited platform and image support**

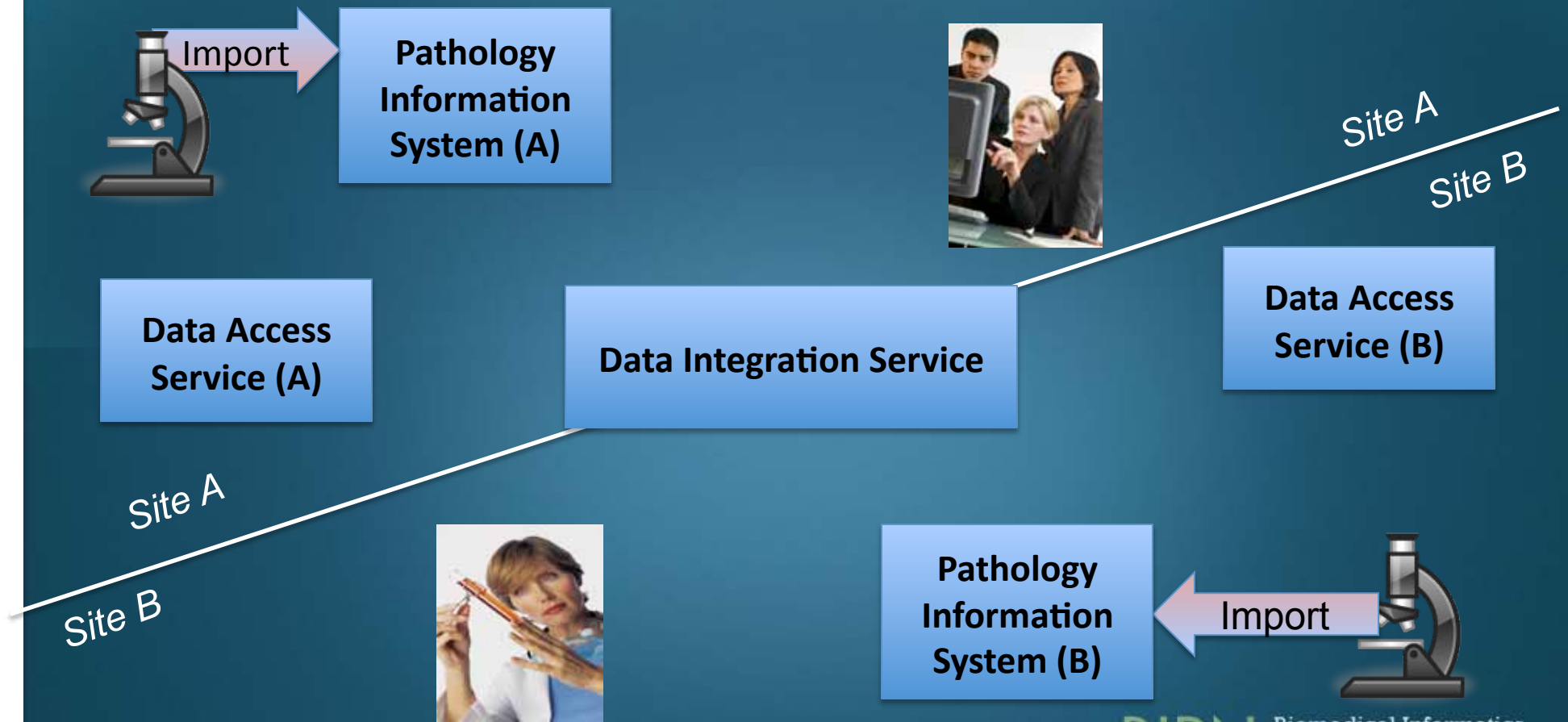
**Scanner Vendor Systems**  
Olympus OlyVIA  
Aperio ImageScope

# Architecture Overview



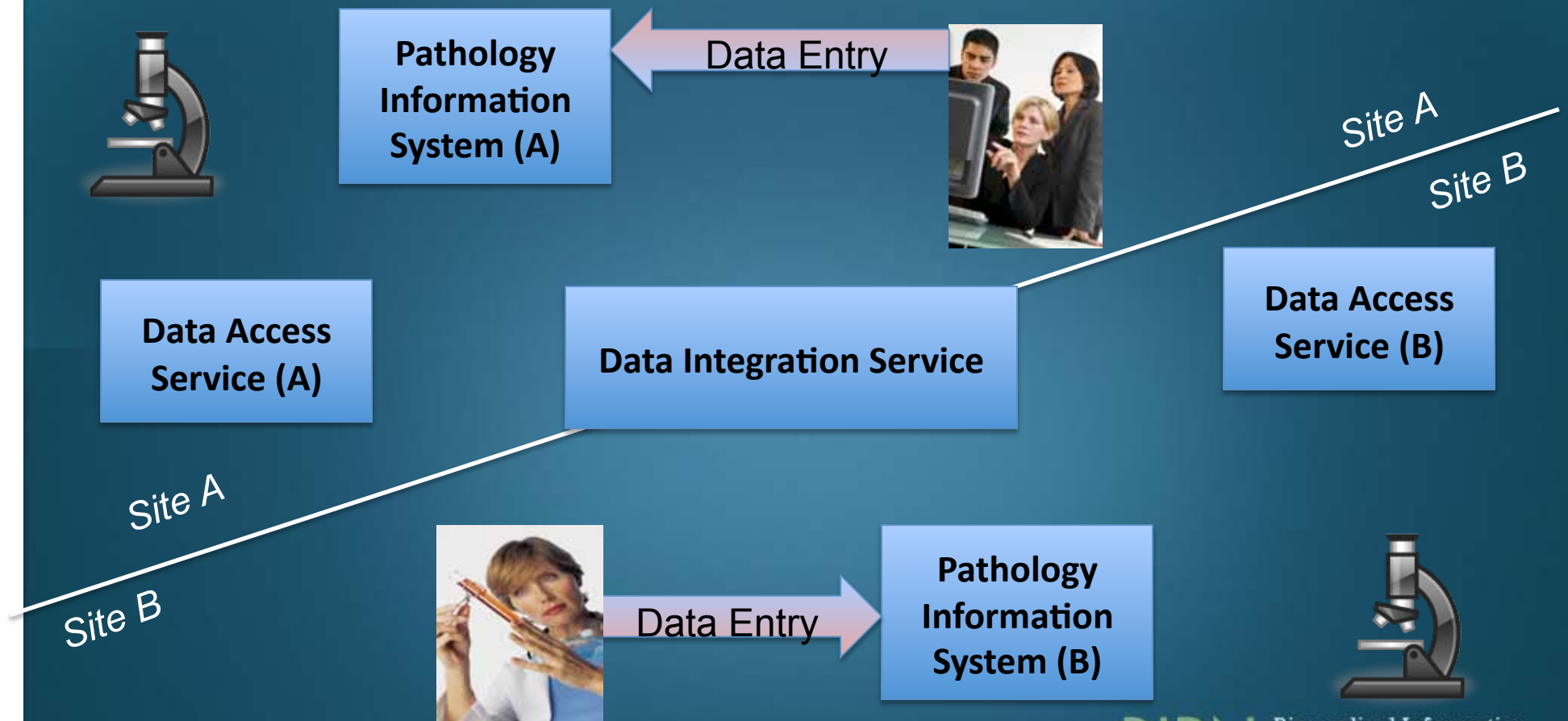
# Architecture Overview

*Scan images at local centers*



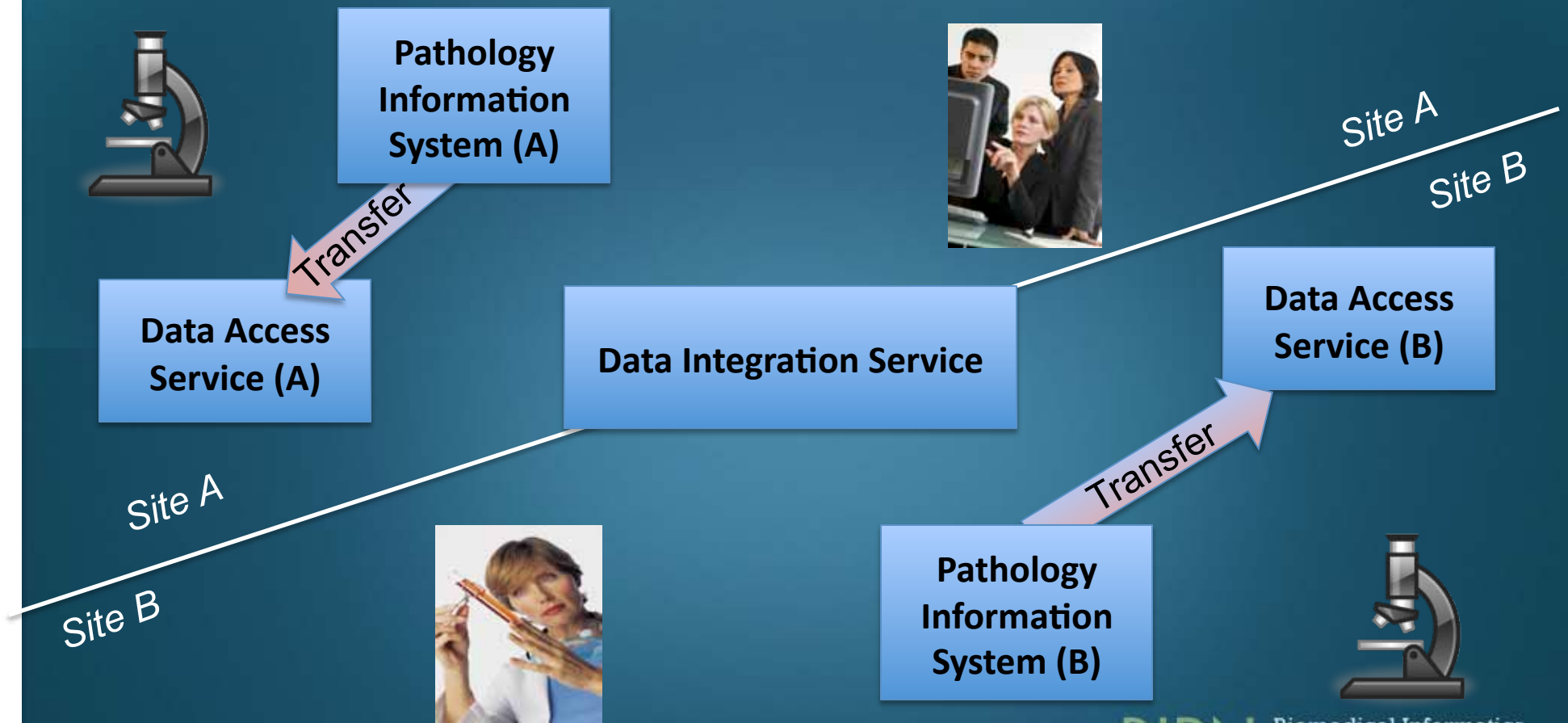
# Architecture Overview

*Curate data at local centers*



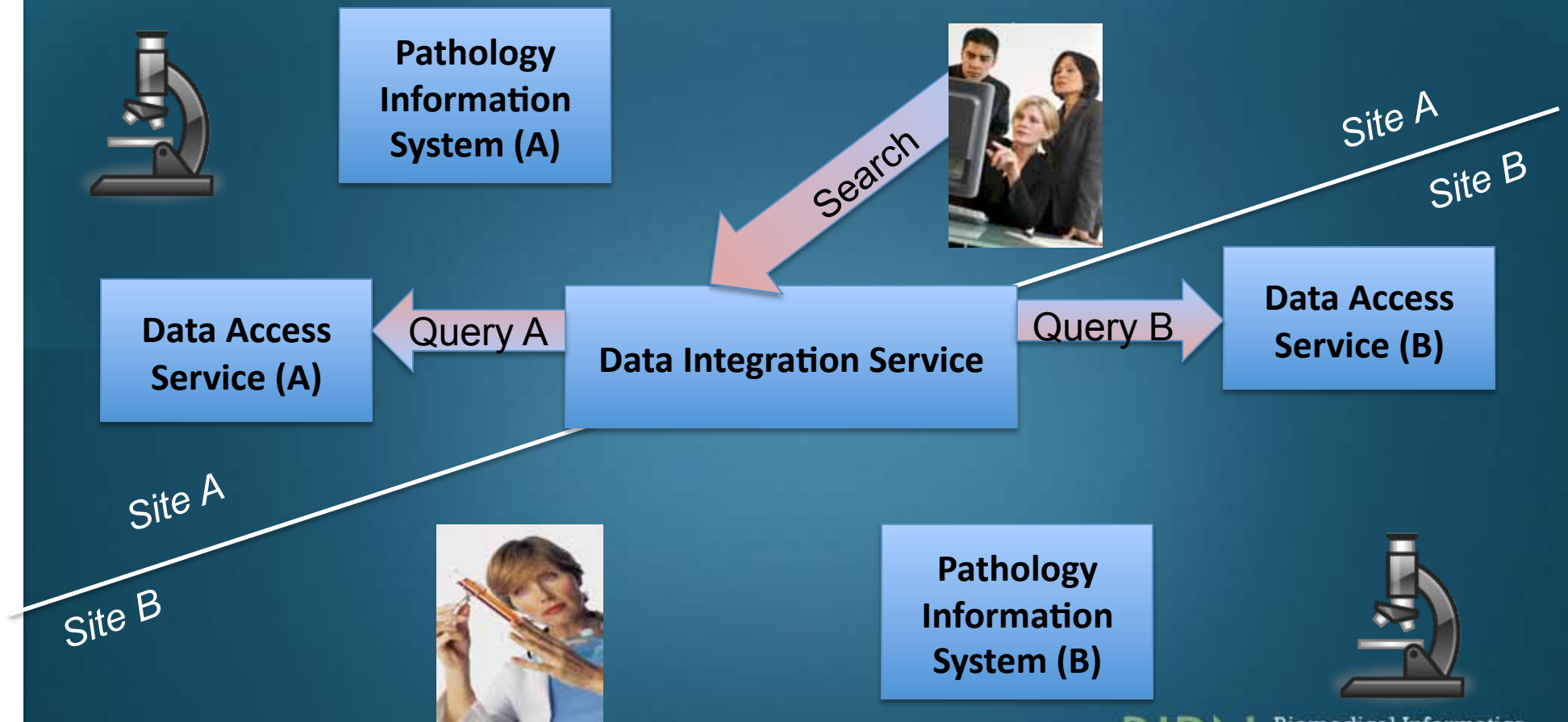
# Architecture Overview

*Push published data to access services*

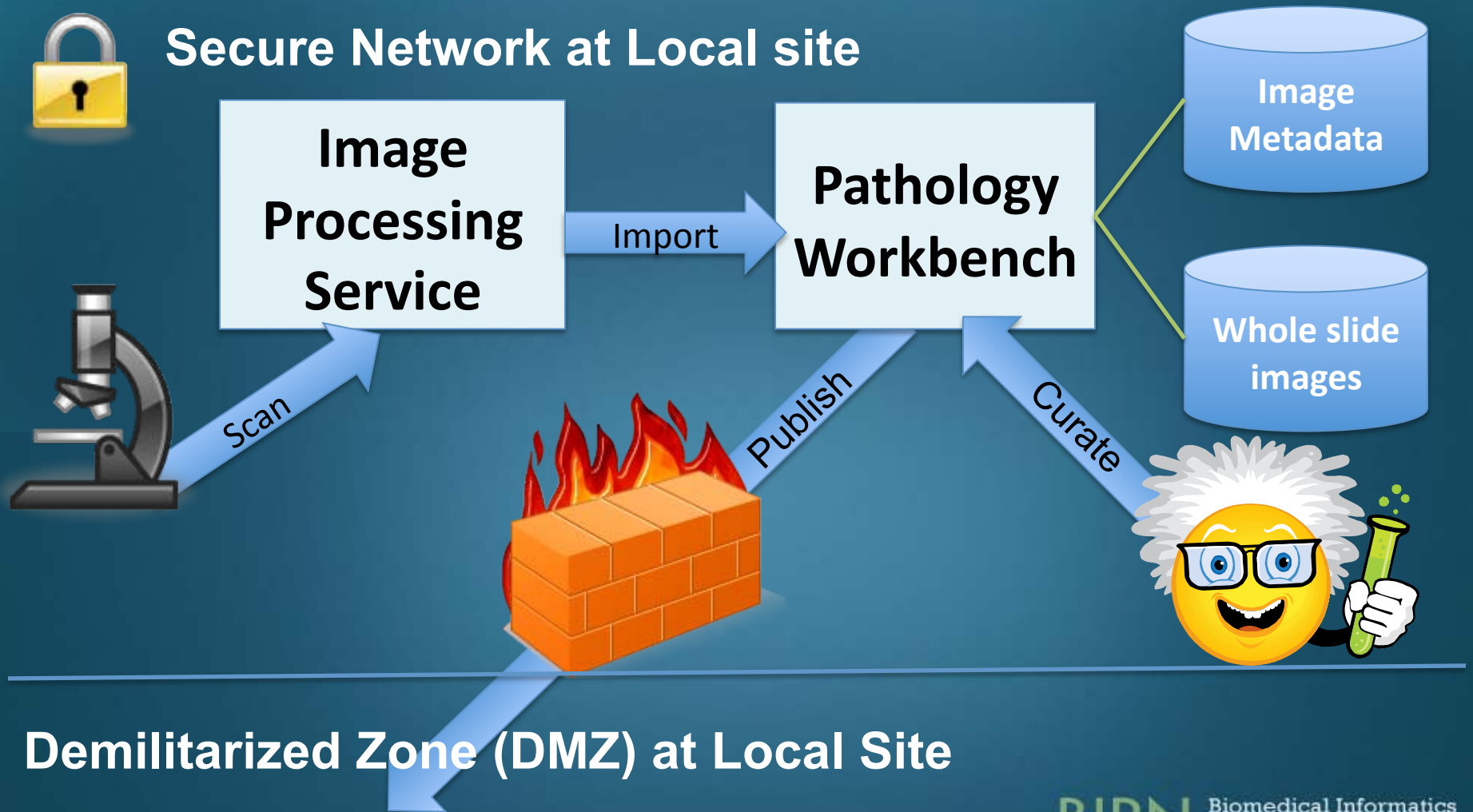


# Architecture Overview

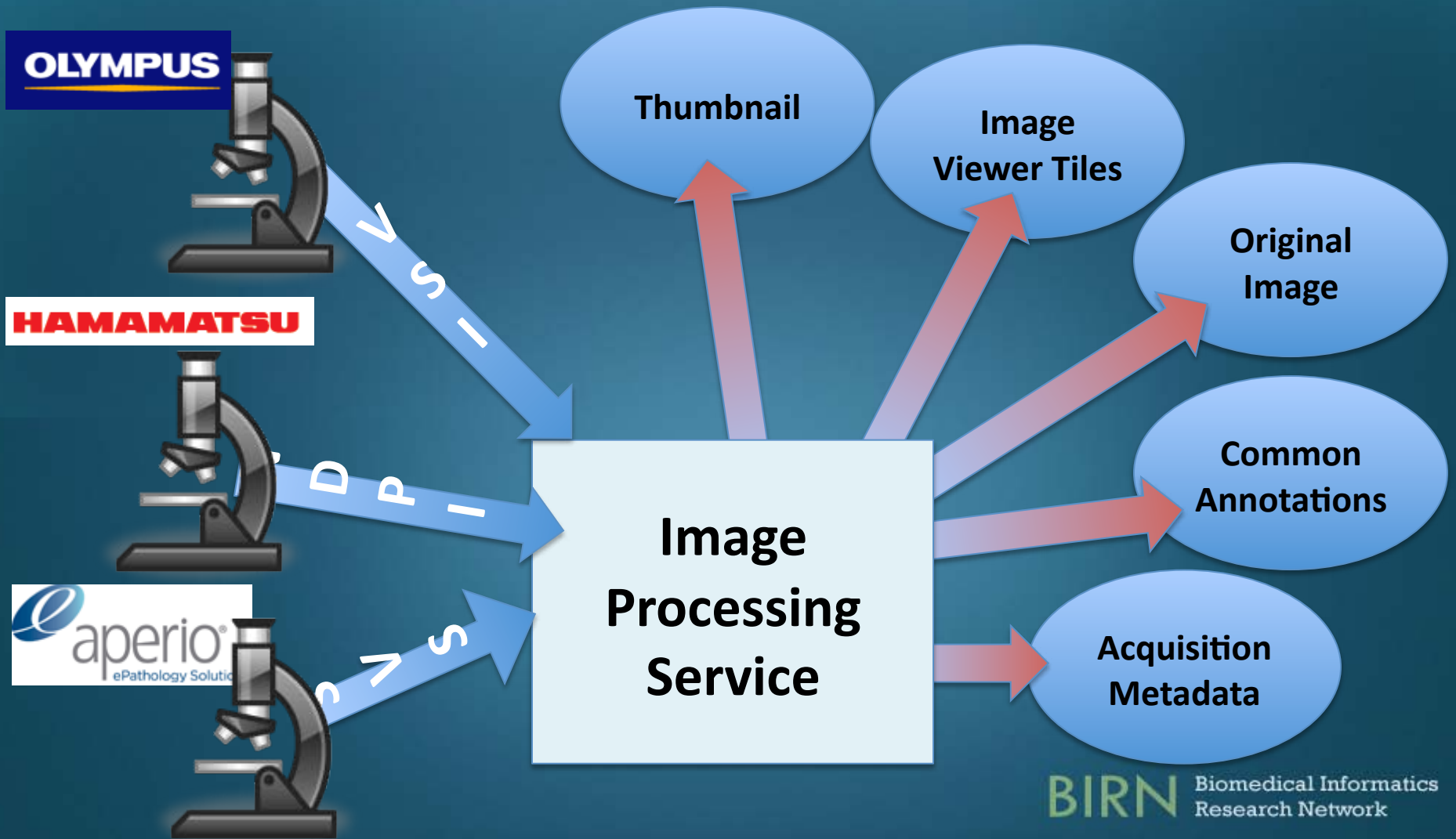
*Search for data across all sites*



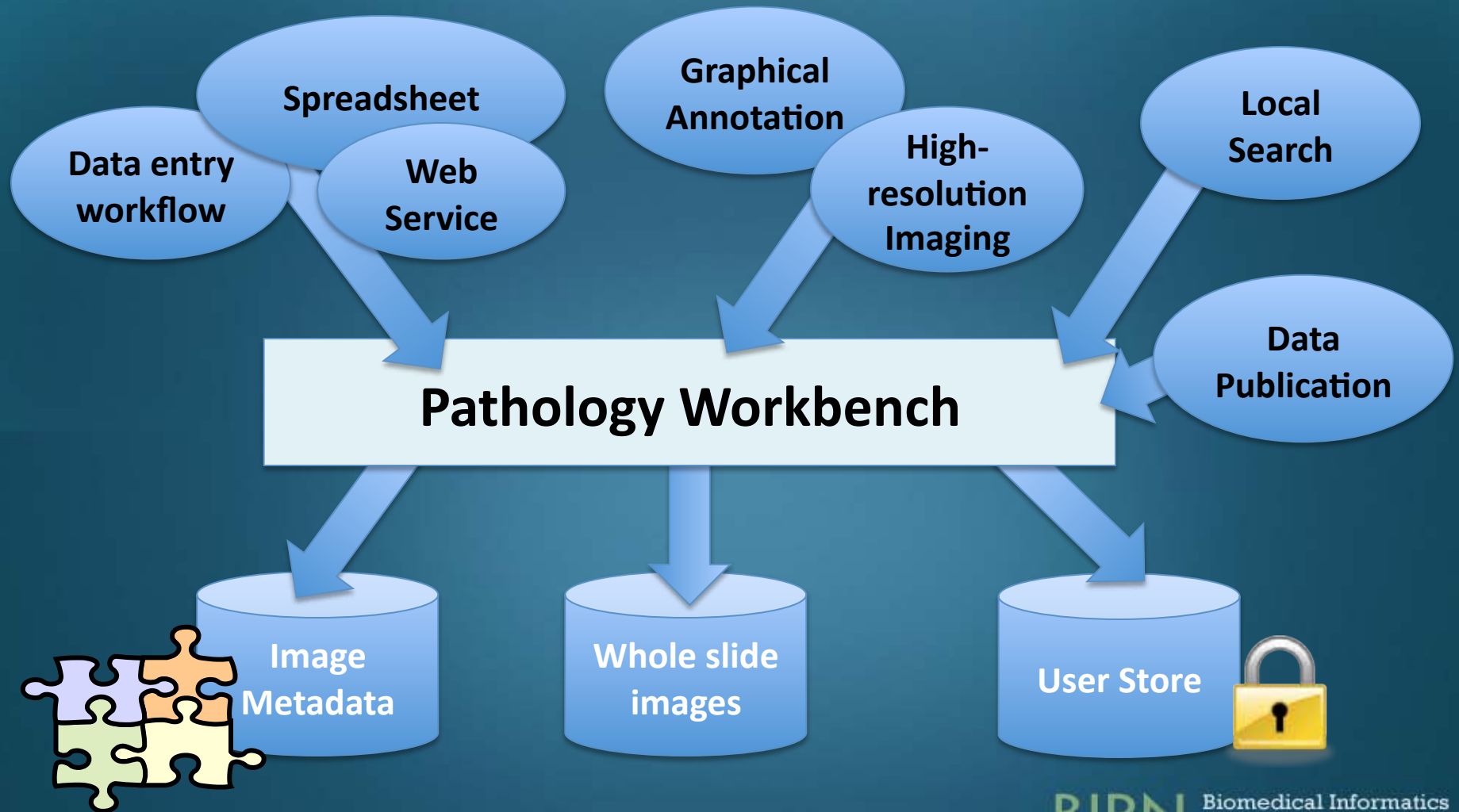
# Pathology Information System



# Pathology Information System - Image Processing Service



# Pathology Information System - Pathology Workbench

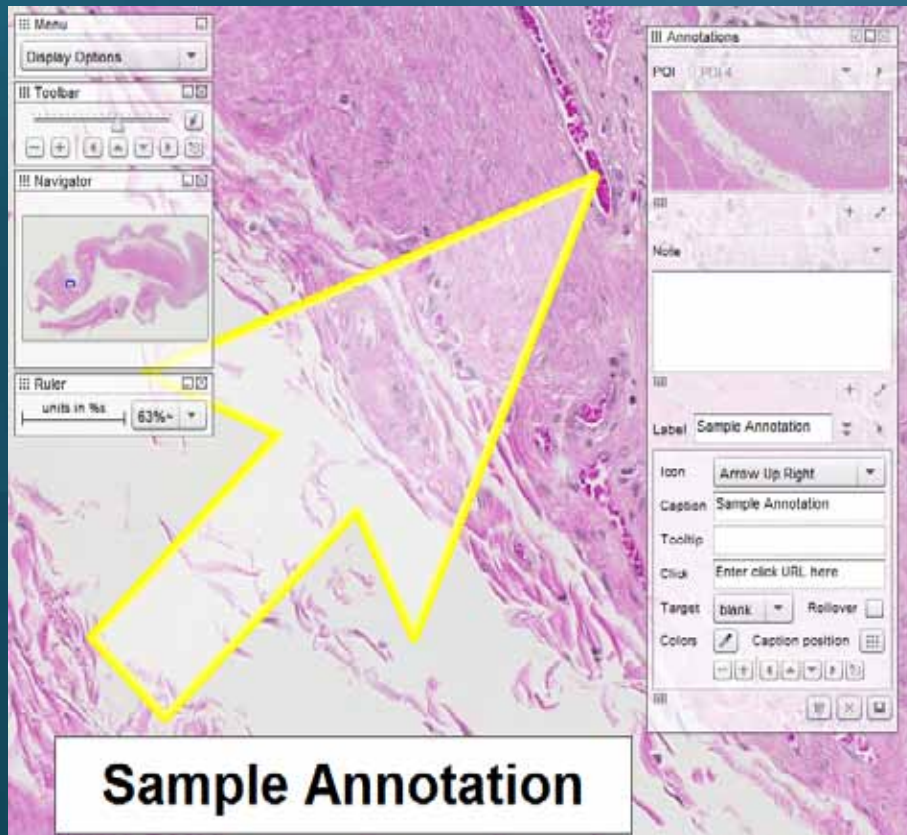


# Pathology Information System - Pathology Workbench

The screenshot displays the 'Case A-1111-C1 - Case History: Case/Clinical Overview' form. It includes fields for Case ID (A-1111-C1), Age (1 year), Age Category (Juvenile), Case Date (2011-11-02), Age (in days), and Body Weight (12 kg). The 'Procedure(s)' section lists 'CSF COLLECTION (P-1C1B)', 'ABORTION, INDUCED, NOS (P-1750)', and 'ACID FAST (P-YY13)'. The 'Disease' section shows 'ECHINOCOCCOSIS, NOS (P-0910)' with an etiology of 'CUNICINE (E-7672)'. The 'Disease Processes' list includes 'Immune-mediated', 'Infectious', 'Inflammatory', and 'Inherited'. The interface also features a 'Case History Comments' field, a 'Published?' checkbox, and 'Save' and 'Save and Add Disease' buttons.

- Secure
  - Local access and data administration
    - One per center
  - Authenticate with username and password
    - Supports common credential store (LDAP)
  - Application and image-level permissions
- Flexible, complex data schema and ontologies
  - Based on SNOMED
- Data entry methods
  - Workflow-based web pages
  - Spreadsheet import
  - Web-service (REST)



# Pathology Information System - Pathology Workbench



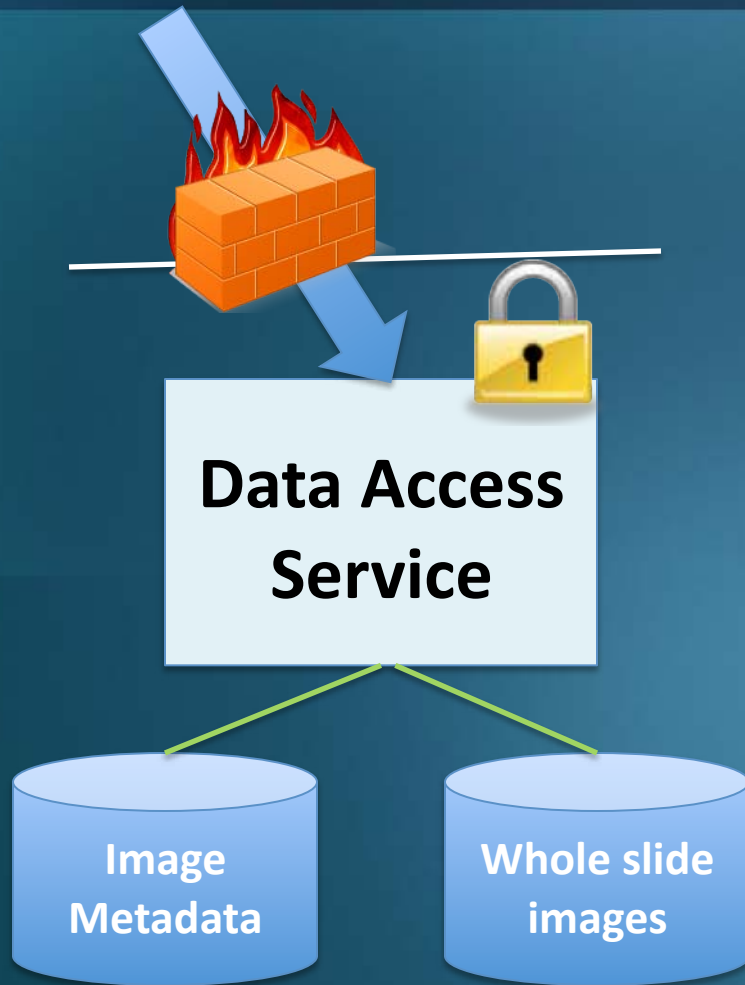
- Zoomify Enterprise Viewer
- Image viewing
  - Browser-based, loads instantly
  - Progressive-rendering
    - View whole image at low resolution
    - Zoom to high resolution area
- Annotations
  - Mark, label points of interest
  - Share with other users

# Pathology Information System - Pathology Workbench

- Search local data
  - Specific metadata
    - e.g. find all images that belong to a lung specimen
  - Full-text
    - e.g. find all images that contain the phrase “Ringworm” in all metadata
- Publication
  - Choose the data that gets pushed to collaborators

Search Results						
	Image Description	Accession ID	Subject ID	Case ID	Specimen ID	Diseases
			A-1111	A-1111-C1	A-1111-C1-S1	RINGWORM, NOS (D-05750)
	An image from the first subject's case		A-1111	A-1111-C1	None	ECHINOCOCCOSIS, NOS (D-06830)

# Data Access Service

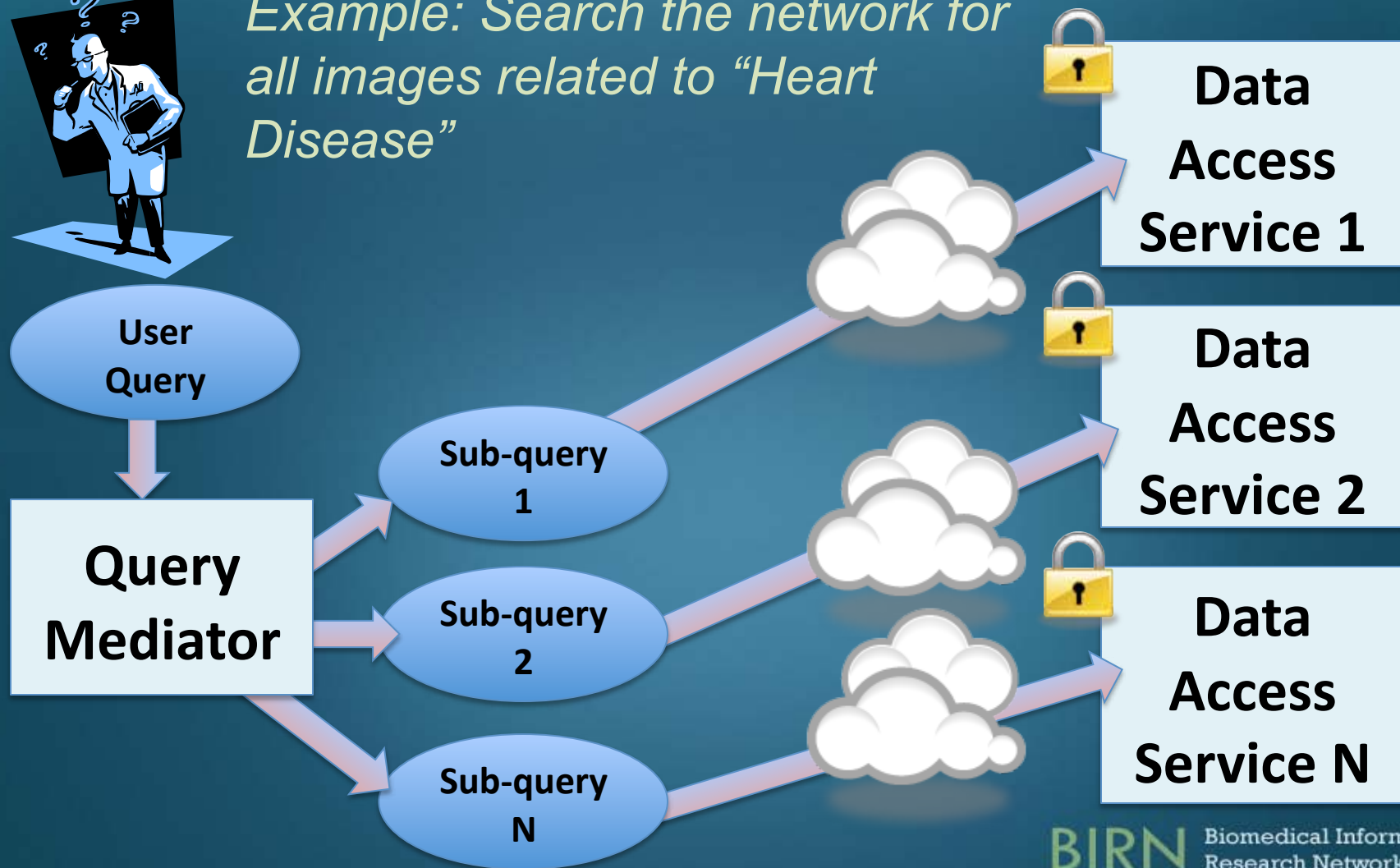


- Operated outside the firewall (DMZ)
- Receives data securely from a single Pathology Information System
  - e.g. using Globus File Transfers (GridFTP) with host or service certificates
- Manages its own copy of metadata and image files

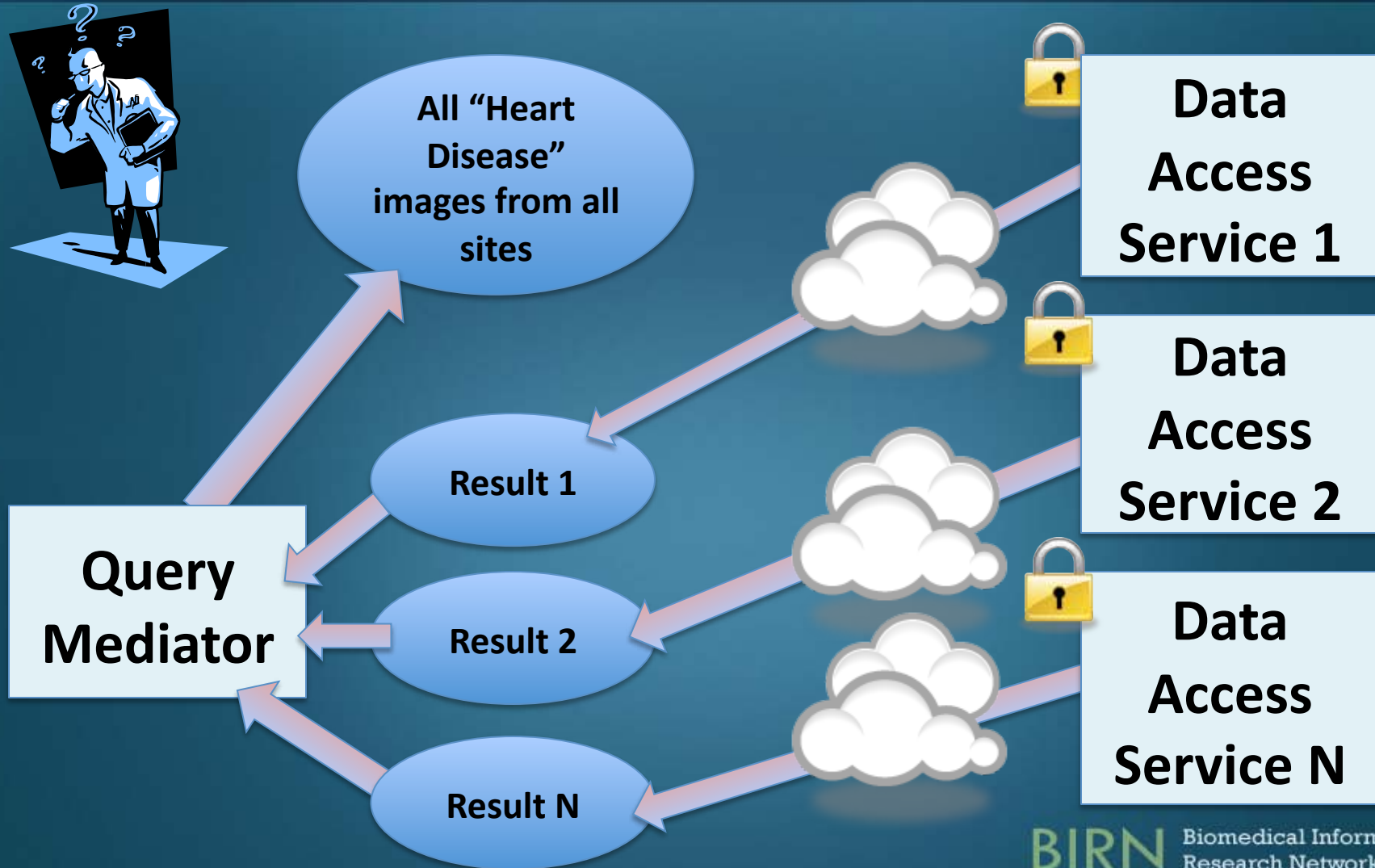
# Data Integration Service



*Example: Search the network for all images related to "Heart Disease"*



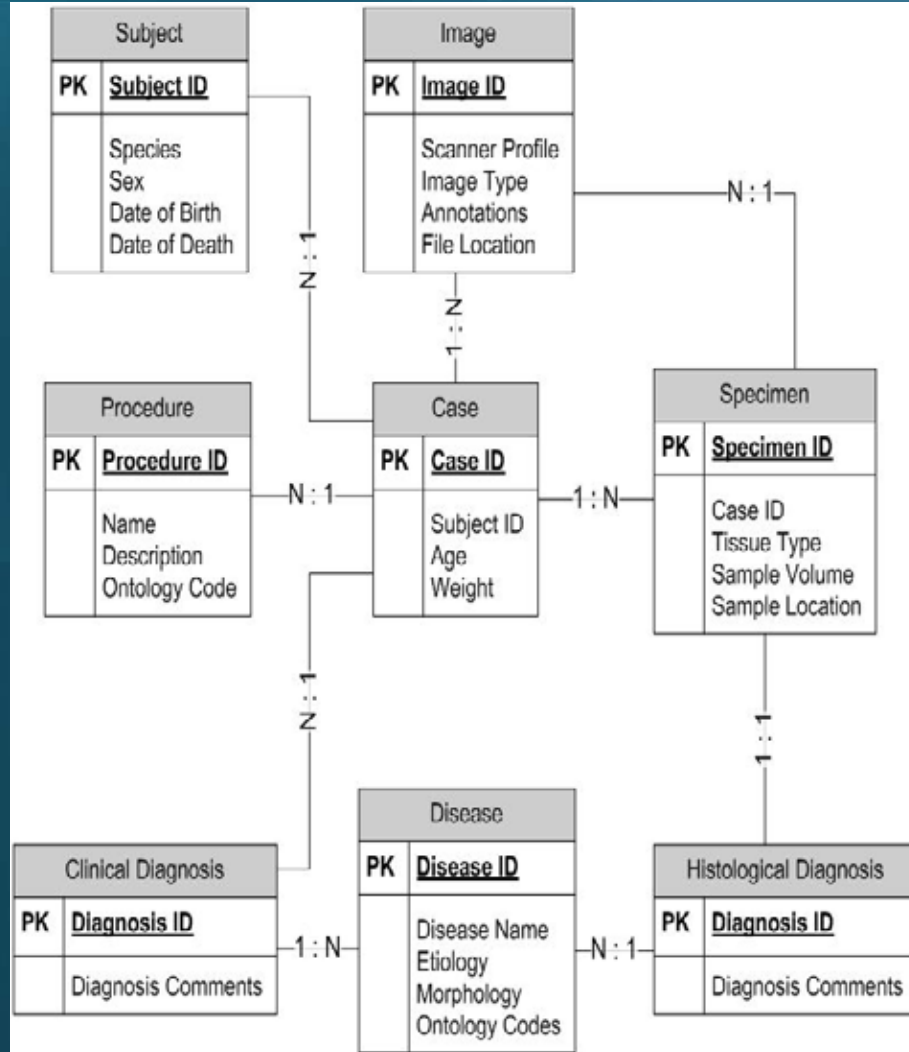
# Data Integration Service



# Testbed - Initial Deployment

- Two National Primate Research Center (NPRC) sites
- 27,000+ images processed
  - Olympus VSI, Hamamatsu NDPI, Aperio SVS, JPEG, TIFF
  - Rate of conversion ~1 GB/hour per compressed image data
    - Acceptable for processing archives
    - Not ideal for instant access to 20+ GB slide from scanner, but okay for working with pre-defined sets
    - Processing images simultaneously for better capacity
  - Collection is over 1.5 TB in size
- Pathologists curating data and assigning to images
- Currently publishing data to share with other NPRC sites

# Testbed - Data Schema



- Developed over 18 months
- Identified concepts and relationships
- Rapid database implementation based on schema
- Continual feedback and improvements
- Combination of standard and custom ontologies
  - e.g. SNOMED

# Future Work

- Expand testbed beyond initial deployment
  - More National Primate Research Centers to collaborate in their network
  - New collaborative networks
    - Human pathology domain
    - Clinical use cases
- Improve data curation experience
- Further data schema refinement
- Federate between different schemas
- Image processing automation improvements
  - Scalability
  - Performance
  - Management



# Conclusion

- Digital Pathology Network system architecture and implementation met user requirements
  - Flexible, open, decentralized
  - High-resolution whole slide imaging with annotations
  - Complex pathology schema and data workflows
  - Local administrative control
  - Share data with collaborators
- Successful initial deployment
  - Two large-scale research centers
  - Active use by pathologists and related research and support staff
- Learn more about BIRN Pathology

<http://www.birncommunity.org>