Archival and Retrieval in Digital Pathology Systems

Authors: Elizabeth Chlipala (Premier Laboratories), Jesus Elin (Yuma Regional Medical Center), Ole Eichhorn (Aperio Technologies), Malini Krishnamurti (Omnyx), Richard E. Long (Charles River Laboratories), Bikash Sabata (bioImagene), Matthew Smith (Aurora Interactive, Ltd.)

Abstract

This white paper frames the issues related to Archival and Retrieval of images and associated data as Anatomic Pathology laboratories adopt a digital imaging workflow in a research or clinical setting. This overview also includes a brief discussion of some of the solutions being developed and offered in digital pathology systems (DPS's).

DPS's start with creation of a whole slide image and provide an imaging workflow for pathologists by associating the images with a patient and case and furnishing the tools to review the cases digitally. Whole slide images (WSI) are very large and will drive the need for extensive storage and information life cycle management. Customers seek options for fast access to high quality and highly available data, when they use the DPS in a business critical application. Most importantly, since DPS's are deployed in regulated environments, data reliability, privacy, and security need to be built into the solution and its management process.

Solutions that are being developed and deployed in digital pathology systems are also presented.

Contents

Introduction

Digital Pathology describes the creation, viewing, management, sharing, analysis, and interpretation of digital images of glass slides and includes workflow considerations unique to a digital imaging environment.

Several vendors now offer scanning products that support digitization of entire glass slides – often called Whole Slide Images (WSI's), digital slides, or virtual slides. Scan times continue to improve and DPS's increasingly provide pathologists and AP laboratories with new tools to improve the workflow and quality of patient care. These trends make increased use of DPS's in routine clinical settings inevitable.

Full adoption of the digital workflow cannot be achieved unless the DPS can store high quality images and associated information in a secure, reliable, and scalable manner and provides fast access to them while utilizing the hospital infrastructure efficiently and supporting the lab's ability to comply with regulatory requirements for information management.

This whitepaper will describe the issues pertaining to archival and retrieval of digital information that arise with adoption of digital pathology and solutions that are being developed to address them.

Overview

Before discussing the impact of a digital pathology workflow on archival and retrieval of information, let us review how the digital workflow is enabled.

Whole Slide Images and Digital Cases

Whole slide scanners are used to scan a microscope slide at high resolution to create an image of the slide. These images are called whole slide images, digital, or virtual slides. The scanner, typically, uses a proprietary method to capture several resolutions of the whole slide, compress, and store them in an image file designed to rapidly reconstruct the digital slides at different resolutions. The capture resolution in the scanner determines the maximum effective viewing magnification of the digital slide on a computer monitor. Most scanners support capture resolutions of 0.5 microns/pixel (effective viewing magnification: 20X) or 0.275 microns per pixel (effective viewing magnification: 40X). The image file associated with a 20X scan of a 15mmx20mm tissue specimen is as large as 3.6GB and a 40X scanned image can be as large as 14.5GB. The images are compressed to more manageable sizes (25:1 compression or greater) such that there is an optimization between image quality, image file size, network bandwidth usage, and server and client resource utilization. For example, the 20X scan could be stored in a JPEG2000-compression file of size 144MB. The 40X image described above could be stored in a JPEG2000-compressed file of size 576MB. **Digital Pathology images are about 10X that of Radiology images and will require more storage management through their useful life cycle.**

In order to facilitate the digital workflow, the DPS also stores, retrieves, displays, and manages metadata associated with digital slides. At the scanner, metadata such as the macro image, scan parameters, scan plan, and tissue map are generated and associated with the image. When the DPS is integrated with an AP-LIS, then patient, case, and histology (part/block/stain) data can also be retrieved from the LIS. A digital case can be created in the DPS by associating digital slides with the case and histology information. In addition, if bar-coding of slides s deployed in the lab, then the bar-code can be scanned and decoded by the whole slide scanner and used by the DPS to automatically associate the scanned image with the glass slide, previously accessioned in the LIS in the case context. The bar-code is a unique ID for the digital image, to ensure a valid electronic chain of custody. Finally, when the pathologist views the images and adds annotations and results of quantitative analyses to them, then these are also stored in association with the digital images (in a case context, in some DPS's). **The DPS needs to ensure that this data remains associated with the images and can be retrieved easily for use in the clinical workflow.**

Image Viewing

The digital pathology workflow enables the pathologist to view digital cases from any location regardless of where the images and associated information are stored. Pathologists want to be able to load and navigate the images in 0-2 seconds. Although the digital image can be as large as 200MB, when a pathologist is viewing the slide at a workstation, the entire image does not need to be downloaded. Instead, the image data for the region of interest and magnification level at which it is being viewed is streamed to the workstation and displayed. The image streams comprise about 5-20% of the whole slide. **DPS's are designed to balance the speed of the stream, efficiency of the network bandwidth use, and image quality.**

The optimal viewing experience for a pathologist is a 24" monitor with 1920x1200 resolution and 0.27 pixel pitch. On such a display, the viewing magnification of the image is similar to that under the microscope. Computer software used for the workstation are specified by DPS vendors to best present the performance of their application.

Use Cases

Digital pathology systems have been used in research applications and for education and conferences at academic medical centers. Other current and future use cases include:

- Research Applications
- Tumor Boards
- Inter-operative consults / Frozen section workflows
- Consults with sub-specialists with remote image access
- Reference Labs (QA/QC)
- Single Lab Organization – small to medium volumes (100-500 Slides/day)
- Multiple Lab Enterprise – large volumes (500-3000 slides/day); intra and inter departmental consults, histology and pathology services to outreach centers

As the scale of use increases, several issues related to information archival and retrieval present themselves for examination.

Regulatory Considerations

When used in healthcare environments such as hospital labs, digital pathology is subject to various regulatory considerations.

HIPAA

The Health Insurance Portability and Accountability Act (HIPAA) imposes a number of security and privacy requirements on digital pathology systems. Central to this act is the concept of Protected Health Information (PHI). The metadata associated with digital slides usually constitutes PHI under the meaning of this act, and must be protected accordingly.

Systems which manage digital slides and associated metadata must be secure and must authenticate all access to information by verifying user access credentials. User access to the system must be logged and be auditable. Display of information such as slides and metadata must be restricted to authenticated users.

When data are transmitted across networks, secure protocols must be used for authentication and for transmission of PHI. In general, digital slides themselves do not constitute PHI and their data need not be

encrypted, but any metadata including slide labels, hospital / patient / case / specimen information, etc. is PHI and must be encrypted.

Additionally, the procedures used to manage the components of the system must be designed to protect the information in the system. Frequently digital pathology management systems are designed such that all PHI is kept encrypted, to reduce the risk of unauthorized physical access to the computer systems.

FDA

The U.S. Food and Drug Administration (FDA) regulates vendors of instruments and systems for hospital labs. For DPS's to be deployed in a FDA regulated lab, safe guards will need to be built for to comply with FDA requirements for:

- Proper storage of digital slides, metadata, and information such as annotations and analysis results in order to ensure data integrity for the duration of image retention (2-5 years, depending on the use case)
- Deploying data management processes which include responsibility for and access rights to the archive
- Data security and protection of all data including any electronic records archived in the DPS. This entails existence of time stamped audits of activity that create, modify, or delete data from the archive.

CLIA

The Clinical Laboratory Improvement Amendments (CLIA) imposes a series of procedures on labs, including the need to periodically validate all instruments and systems used for testing. Labs will need to have trained individuals who can work with the DPS applications. In addition, they would need to ensure that CAP or JCAHO checklists include proper documentation for deployment, maintenance, and security of DPS's and are used to validate digital pathology systems before they are used in hospital labs for diagnostic / testing applications. The DPS may be integrated with the LIS to retrieve patient, case, and specimen data to constitute a digital case. In this case, a plan needs to be implemented to regularly ensure that the DPS information remains synchronized with the LIS.

Information archival and retrieval

Storage and interoperability

Online storage for digital slides is usually recommended to be connected to the scanner through a dedicated network. In a reasonably big lab, a scanner could be used all day to generate up to 500-1000 slides. When a 200MB image is scanned every minute, then the storage subsystem of the DPS should be able to store at least as fast. Therefore, the sustained network usage will be 30 Mbps. The hospital will need to dedicate at least 50 Mbps capacity for this ingestion. To be conservative, the hospital would need to plan for sustained ingestion of 40X scans which would require up to 175Mbps. A dedicated GB full duplex connection between the scanner and the DPS online archive would ensure high ingestion performance.

The performance requirements from the DPS for the storage platform are very nominal. Dense capacity is critical so that large amounts of data can be stored in a smaller and lower-cost footprint. Storage vendors offer 1TB or 2TB SATA disk drives in their storage systems. In order to ensure that the data is stored in a redundant manner, best practices suggest that the disk storage is configured in RAID 6 with 1 hot spare disk. This ensures that up to three disk failures can occur without affecting data integrity; however it does

reduce the available capacity on the disk for image storage. If the image size is 200MB, then one TB of disk storage holds about 3500 digital slides when configured in RAID 6. Random read I/O performance of the storage platform should be able to support image streaming although this is not observed to be a limitation in high performance storage solutions.

Other key functionalities, offered by storage vendors, that are useful when included in DPS's are:

- Tiered storage management software that enable automated movement of images from higher cost disk to lower cost disk or magnetic tape systems.
- Replication of data between storage systems (in one or more data centers) to provide highly available data access.
- Disaster recovery solutions that can be designed to recover data to a customer specified recovery point (how much data loss is acceptable) and recovery time (how soon should the data be available again). Note: data replication can also be used as a disaster recovery solution.

DPS's should be designed to work with direct attached storage (DAS), storage area networks (SAN), and network attached storage (NAS) hardware.

DPS's need to be built to be interoperable with storage platforms, which provide high-performing, scalable, cost-effective storage. The application should be able to store and retrieve images and other data without needing to understand how the data is managed by the storage system.

Enterprise view

In an Enterprise, there could be distributed histology labs where the different stains are being prepared. It would be preferred if the digital slides are stored near the scanner and histology lab. Ideally, the DPS should be able to stream images from any of the digital archives to the Pathologist without having to move the images/data from its source location.

Within the hospital local area network (LAN), DPS vendors recommend a 100 Mbps network from the DPS archive to the workstation to provide the desired level of viewing performance.

Remote viewing of cases provides high-value for digital pathology since it enables something which is not possible with physical glass slides. However, if the pathologist is viewing images on the Wide Area Network (WAN), then network bandwidth will have a big impact on the viewing performance. The minimum required WAN bandwidth is 2Mbps.

Remote access to cases raises the issue of hospital security because hospitals' policies often prohibit external access to servers inside a hospital or lab network. This issue can be addressed in a couple of different ways:

- Installing a web server in the DMZ of the network. The firewall in the DMZ can be configured to allow the remote user to only access the server in the DMZ whereas the web server is configured in a restricted way to access the DPS within the hospital private network. The web server can proxy information retrieval and store transactions in a much more secure manner.
- Cloud replication of data. The operation is outbound and does not compromise network security. Once copied to the external server, slides can be viewed from there. The external server typically provides authorization checking and authentication mechanisms, as well as a secure datacenter environment, to preserve HIPAA compliance. This option will require a dedicated, large bandwidth network from the hospital to the cloud to be able to scale in volume. In addition, management of two separate copies of the data carries more overhead for the DPS solution to support an integrated digital workflow.

System and Data Reliability

When a scanner sends a scanned image and associated data to the DPS, data needs to be transmitted and stored reliably. The image data needs to have a failsafe way to be associated with the relevant case so that if there is a loss of sub-components of the system then there is the ability to rebuild the association. It is also critical that there is no data loss or corruption when the storage hardware has component failures. DPS's are building redundancy both into their software and the hardware configurations to eliminate single points of failure.

For a DPS to scale reliably to support a large volume workflow, there needs to be an abstraction between how the images are accessed and the storage is managed, which can be achieved by associating the data stored in the repository with a key in the database in the case context.

Image Life Cycle Management

As digital pathology gains adoption, the trade-off between image access and the cost of storage becomes an issue. Even though the glass slide may remain the, legal record in most jurisdictions, the ubiquity of the digital image will place new requirements for its retention. This will require costly storage infrastructure unless Image Life Cycle Management (ILM) policies are developed by pathology departments and applied intelligently in the DPS. Decision criteria that influence the ILM policy for the institution are: reasons for retention, duration of retention, speed of image access during the different phases of the lifecycle, and associated cost of storage.

- Clinical review – images must be stored for up to 2 months. Image access during this phase will need to be very rapid and, therefore, need fast, costlier storage media.
- Prior cases – different pathology benches may retain prior cases for varying amounts of time. For example, in high volume dermatopathology, it may be less likely for pathologists to view prior cases than in transplant cases, where prior cases may be needed for 2-3 years. Pathologists may be satisfied if prior images are available in 4-5 minutes. In many cases, regions of interest inserted into the report may suffice for future reference.
- Education and research purposes – specific cases may be flagged to be retained indefinitely and to be accessed rapidly.
- Legal or regulatory reasons – Analogous to ILM policies in Radiology, the average case may be kept for 7-10 years; although regulatory requirements in different countries will vary. These cases may only be retrieved rarely, if ever, and could be stored in very slow storage media.

As pathology departments start to develop their image retention policies, DPS's can deploy rules-based auto-archival and leverage tiered storage management solutions that are provided with most high-performance storage platforms.

IT Infrastructure and Management Costs

IT managers are increasingly more influential in determining how healthcare IT (HCIT) solutions are deployed. They may require HCIT solutions to integrate with their existing infrastructure or work on a specified hardware platform. DPA applications will be required to integrate with the IT authentication systems such as the Lightweight Directory Access Protocol (LDAP) Servers or Active Directory services. An increasing need within the HCIT is the single sign-on capability that will be enabled using the integration with the enterprise wide directory services. Institutions have enterprise management tools for tiered storage management and disaster recovery and require new HCIT solutions to be able to fit into this framework in order to keep management of the new solution affordable. Many have a central data repository for all their imaging and clinical data, which can be presented to referring physicians in a patient

context; and will require their DPS vendor to be able to support the platform for long-term storage of pathology images. One key trend is that the cost of storage hardware has been declining by 25-40% annually and coupled with intelligent storage management, the cost of the image explosion can be alleviated.

Future Directions Affecting Information Archival

Digital Pathology comprises rapidly advancing technologies. These trends will affect the archival solutions being deployed in DPS's.

Scanning speeds and image resolutions

Whole slide scanners are likely to be able to digitize a slide in less than 30 seconds and digital image resolutions are likely to increase driving the need for larger slides. However, concurrently, the densities of storage media are likely to grow exponentially. Image archives will grow efficiently to meet the increased throughput of the lab.

Viewing Advancements and Clinical Decision Support

New tools that allow the pathologist to quickly and efficiently review and analyze digital slides will evolve. This could reduce the number of slides that need to be reviewed by the pathologist; but will also shift the balance of server and client resource requirements in the performance – cost equation.

Standards for Archival

Working Group 26, a subcommittee of the DICOM standards body, has ratified a proposal for storing digital slides within a PACS archive in DICOM Supplement 145. In the near future, PACS vendors and DPS vendors will support this standard. This will increase the development of vendor neutral archives for the Enterprise.

Multi-Site Collaboration Applications

As Digital Pathology enables simpler and more rapid access to sub-specialty opinions, consultation networks will continue to proliferate. This will in turn underscore the acceleration of open archives that enable third party application to access to images at the source.