

Wagar, Horowitz & Siegal's

LABORATORY ADMINISTRATION FOR PATHOLOGISTS

Second Edition

Elizabeth A. Wagar, MD
Michael B. Cohen, MD
Donald S. Karcher, MD
Nicole D. Riddle, MD
Gene P. Siegal, MD, PhD
Editors

Copyright © 2019 College of American Pathologists (CAP).
All rights reserved.

Library of Congress Control Number: 2019930389
ISBN: 978-194109646-8

College of American Pathologists
325 Waukegan Road
Northfield, Illinois 60093
800-323-4040
cap.org



COLLEGE of AMERICAN
PATHOLOGISTS

Management of Pathology Information Systems

James H. Harrison Jr, MD, PhD

Contents

- Definition of Pathology Informatics
- Early Development and Growth of Laboratory Information Systems
- Maturation and Adoption
- Recent Developments and Challenges
- Computer Basics
- Data Representation
- Image Compression
- Data Standards
- General Characteristics of Laboratory Information Systems
- Database Applications
- Computer Networking
- System Interfaces and HL7
- System and Data Security
- Disaster Recovery
- Systems and Personnel
- Application Management
- Interface Management
- Other Systems
- Record Retention
- Procedure Manual
- System Evaluation, Selection, and Installation
- Performance Monitoring and Quality Assurance
- Laboratory Analytics and Computational Pathology
- Software Tools and Local Software Development
- Diagramming Tools for Requirements Analysis and Workflow Redesign
- Appendix 6-1. Site Visit Questions for LIS Acquisition

Definition of Pathology Informatics

Pathology informatics is a subdomain of biomedical informatics, the science that underlies the academic investigation, optimization, and practical application of computing and communications technology in health care, health education, and biomedical research. Biomedical informatics includes, as broad areas: (1) *bioinformatics*, the use of computer technology in basic research, especially for the study of gene, protein, and cellular structure and relationships; (2) *medical or clinical informatics*, the application of information technology to health care delivery and clinical research; (3) *public health*

informatics, which focuses on population health; and (4) *health knowledge informatics*, which is concerned with managing clinical and research literature and health information for patients. Pathology informatics is primarily a subset of clinical informatics that focuses on optimizing the acquisition, management, communication, and use of information related to anatomic and clinical pathology analyses and laboratory operations.

Early Development and Growth of Laboratory Information Systems

The use of information systems in pathology predated the formal recognition of the field of biomedical informatics and contributed to its development. The initial impetus for clinical laboratory computerization occurred in the 1960s as a result of the transition to primarily third-party (insurance) payment for medical care and the adoption of automated analyzers for chemistry and hematology testing. These developments created a demand for increased testing and the capacity for laboratories to meet that demand, except that manual specimen tracking, data management, quality control, and reporting were limiting factors. Automation of these functions using early computers was a natural step, and the capital to create those systems was available because laboratories were a revenue source at the time. The initial work in the 1960s led to a consensus by the end of the decade that laboratory computer systems could be cost justified and should be implemented as in-laboratory minicomputers with local terminals (a laboratory information system [LIS]) rather than as a component of hospital information systems implemented on large mainframe computers.¹ The first vendor systems began to appear at this time.

Interest in LIS development was stimulated in the 1970s by national workshops and increasing hardware and software capability. Systems became more flexible, allowing users to manage test names, normal ranges, and report formats without reprogramming. The concept of “turn-key” vendor systems was developed, ie, LISs that were almost complete and required only relatively simple configuration and “a turn of the key” to use. The systems remained primitive by modern standards, though, and installations were not always successful. A review in 1975 defined an “optimal” laboratory system as one that

had electronic connections to at least some chemistry and hematology instruments and at least three of the following five features²:

- Test requests could be imported or entered through a terminal.
- Collection lists with labels could be printed.
- Test results could be entered without manual re-entry of a patient identifier.
- Ward reports could be printed.
- Cumulative summaries could be printed.

Only about half of installed systems met this benchmark, and many attempted system installations failed completely, with substantial loss of effort and funds.

Technical progress in the latter part of the 1970s substantially improved the uncertain success of early systems. LIS programming moved from assembly language (LCI) and FORTRAN (Medlab, CHC) to MUMPS (Meditech, Sunquest) and COBOL (Cerner), which freed systems from ties to particular computer models, handled variable-length text efficiently, and facilitated faster software development. Systems were modularized around key laboratory functions, new modules were developed to handle nonautomated laboratory sections such as microbiology and blood bank, and system interfaces to HIS and billing systems became more common. Recognizably modern vendor systems appeared during this time. By the end of the 1970s, the basic design of LISs had been established, with a modular architecture supporting multiple laboratory sections and working from a common database, and greater than 90% of installations were successful. Separate systems supporting anatomic pathology (APLIS) began to appear in the late 1970s and early 1980s, with features such as specimen accessioning, workload recording, capture and display of patient demographics and history, in-terminal text editing and report formatting, searching and printing of Systematized Nomenclature of Medicine (SNOMED)-coded reports, and billing.³⁻⁵

The challenges and progress of LISs provided background for a paper published in *Science* in 1980 by Lincoln and Korpman titled “Computers, Health Care, and Medical Information Science.”⁶ The authors used the clinical laboratory as an illustration of the larger challenges in the application of computer technology to health care delivery and argued that meeting these challenges required a new approach that blended medical knowledge with information science and engineering. They termed this approach *medical information science*, foreshadowing the field that ultimately became prominent as *medical informatics* in the 1980s and 1990s.

Maturation and Adoption

As automation of laboratory procedures advanced in the 1980s, LIS adoption became widespread and was essentially a requirement for clinical laboratories of any size. Commercial LISs added features and refinements that took advantage of increasing computer hardware capabilities, though at the cost of increased complexity and occasional reliability problems. Standardization of electronic communications improved the ability to interface laboratory instruments and other systems such as the hospital information system (HIS) to the LIS. The American Society for Testing and Materials developed standards for instrument interfaces, and the Health Level Seven International (HL7) standards organization was created. In 1987, HL7 released the second version of its messaging standard, which became widely used for instrument and system interfaces in the United States. A shift from terminals to networked microcomputers (single-user personal computers) running terminal emulators occurred by the end of the decade. Microcomputers increased the usefulness of workstations and saved space by effectively combining multiple terminals into one device, but they also increased end-user support requirements and decreased overall reliability and security.

The important role of pathologists in managing pathology information systems and using pathology information to optimize patient care was highlighted by Korpman in 1987,⁷ and these concepts formed the basis for the field of pathology informatics. Friedman^{8,9} and Buffone and Beck¹⁰ subsequently provided strong quality, strategic, and financial arguments for establishing pathology informatics as a distinct component of pathology services and a subspecialty of pathology. This early work, which defined a role for pathologists in ensuring the accurate and understandable communication of electronic data to clinicians, is reflected in the current College of American Pathologists (CAP) laboratory accreditation checklist requirements that pathologists validate both laboratory computer system operation and pathology data display in downstream information systems such as electronic health records (EHRs).

By the 1990s, clinical laboratory systems were relatively mature, and development centered on workflow and system integration. Improvements included automated specimen handling, calculations and rules storage and execution, and communications and data standards to support LIS incorporation into developing medical enterprise information architectures. The SNOMED medical terminology system, which grew out of the pathology community, was substantially expanded through several releases, culminating in

SNOMED-CT in 2002.¹¹ Development of the Logical Observation Identifiers, Names and Codes (LOINC) code set for identifying laboratory tests began in the mid-1990s.¹² As microcomputer hardware and operating system capabilities advanced, LISs offered client software that provided graphical user interfaces (GUI) as a replacement for scrolling textual displays on terminal emulators. Although GUIs provided easier training and aided in the use of unfamiliar functions, they were not clearly more efficient—and were often demonstrably less efficient—for laboratory workflow than textual displays with expert users. For this reason, terminal-style displays remained in use until relatively recently. Anatomic pathology systems (APLIS) largely adopted GUI client software, with some systems embedding commercial word processors for text editing and printing, and report design focused on printing on paper. Continuing orientation to paper reports is a limitation in APLIS as medicine transitions to fully electronic systems; however, report formatting recommendations as late as 2008 dedicated only limited space to formatting for electronic use and noted that most ambulatory reporting in anatomic pathology remained on paper.¹³

Advances in hardware speed and storage capacity in the 1990s enabled experimentation with telepathology and digital imaging of microscope slides. Telepathology may be classified in order of increasing computing and communication requirements as static, in which selected images from a slide are transmitted to a remote pathologist for diagnosis or consultation; dynamic, in which remote real-time microscopic images are viewed; or virtual, in which the entire tissue content of the slide is digitized, often in multiple focal planes. During the 1990s, radiology underwent a transition from film-based imaging to digital imaging, but pathology did not. Though the technical challenges in digitizing routine pathology practice are substantial, the primary difference between these two applications is related to financial impact and turnaround time. Digital radiology eliminated the use, processing, storage, and transport of film, with substantial financial savings, and decreased the time to image availability. In contrast, digital pathology did not eliminate slide processing and, because it added digitization time to the existing workflow, tended to lengthen the time to image availability. With potential benefits limited to the possibility of more rapid consultation, and with some regulatory barriers related to interstate licensing, there were inadequate financial or quality incentives to spur investment in the engineering and process change required to develop routine digital pathology for widespread use. In settings where access to necessary pathology expertise is limited, however, telepathology can provide substantial value for diagnosis and consultation.^{14,15}

Technology Acronyms

APLIS: anatomic pathology laboratory information system; ASCII: American Standard for Computer Information Interchange; ASP: application service provider; ASTM: formerly American Society for Testing and Materials; DBMS: database management system; DICOM: Digital Imaging and Communications in Medicine; DSL: digital subscriber line; EHR: electronic health record; GUI: graphical user interface; HIPAA: Health Insurance Portability and Accountability Act; HIS: hospital information system; HL7: Health Level Seven International; HL7 CCD: HL7 Continuity of Care Document; HL7 CDA: HL7 Clinical Document Architecture; HL7 RIM: HL7 Reference Information Model; IT: information technology; JPEG: Joint Photographic Experts Group; LIS: laboratory information system; LOINC: Logical Observation Identifiers Names and Codes; MUMPS: Massachusetts General Hospital Utility Multi-Programming System; NIC: network interface card; NIST: National Institute of Standards and Technology; PHI: protected health information; PNG: Portable Network Graphics; RAM: random access memory; RFID: radio frequency identification; SNOMED: Systematized Nomenclature of Medicine; SQL: Structured Query Language; SSL: Secure Sockets Layer; TCP/IP: transmission control protocol/internet protocol; TLS: transport layer security; UML: Unified Modeling Language; VPN: virtual private network; XML: Extensible Markup Language

Recent Developments and Challenges

The progression to greater automation in the laboratory continued in the 21st century. The advancement of automated sample handling systems with associated instrument clusters created a need for specialized instrument management software that was not met by the standard LIS. Instrument management software, often referred to as “middleware” because it mediates between the LIS and instruments, is a class of LIS software that specializes in the management and quality control of high-volume analyzers. These systems’ modern user interfaces and databases, quality control capabilities, and ability to execute stored operational rules have proven valuable outside the limited setting of automated sample handling; thus, they are now encountered as part of automated testing sections in many laboratories. LISs and related systems are also being extended to improve testing-related workflow in nonlaboratory settings. Point-of-care (POC) systems support the download of data from mobile devices, quality control management, and the upload of patient results to the LIS for final reporting. Patient recognition

and label printing has also become automated; “positive patient identification” tokens such as barcoded arm-bands and radio frequency identification (RFID) tags have improved sample acquisition workflow and simplified in-lab sample processing.

Unfortunately, data in pathology systems continues to be represented primarily as locally defined codes and free text, and report formatting and content in anatomic pathology is idiosyncratic. The use of standard terminologies to represent patient characteristics and diagnoses or test identities is limited. While idiosyncratic systems may function adequately for local patient care—particularly when reports are printed on paper—they do not optimally support emerging health system priorities such as communicating data across care delivery systems, providing data that can be summarized automatically, or contributing data to large-scale comparative effectiveness and postmarketing surveillance studies of therapeutics. Current APLISs are particularly weak in providing reports that communicate well when displayed electronically in downstream systems such as EHRs. Recent progress in standardizing the content and presentation of anatomic pathology reports—for example, the CAP cancer protocols¹⁶—is a positive sign that suggests increasing support for report standardization. As the use of paper reports declines, the need for accurate data sharing and effective display within these downstream systems will increase. Appropriate data representation and formatting standards exist to meet these needs; their incorporation will likely be one of the next steps forward for pathology systems.

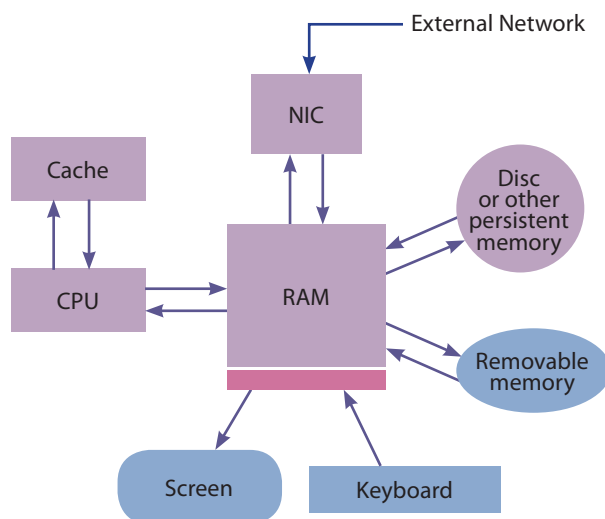
More profound changes in pathology informatics are on the not-too-distant horizon. The migration of microarray technologies, whole genome sequencing, and other high throughput methods from research to the clinical laboratory is generating an enormous amount of data for which computer storage, analysis, and

diagnostic interpretation are essential. Management and processing of the raw data from these methods to yield predictions of disease risk and therapeutic response requires a fundamentally different architectural and processing strategy than exists in current LISs and APLISs. Software and data standards development in this area is very active and will ultimately yield a new class of data processing, storage, and reporting applications. Pathologists will manage these applications and integrate their output into the LIS and ultimately the EHR.

Computer Basics

Digital computers are available in a wide variety of sizes and configurations, from large mainframe systems and warehouse-sized computing clusters to devices that can fit in the palm of one’s hand or be embedded in other small devices. Across this wide physical range, the basic principles and features of computers are remarkably consistent (Figure 6-1). Data contained in a computer’s working memory (random access memory [RAM]) are transformed in a central processing unit according to instructions contained in application programs, which are also stored in RAM (for example, word processors, spreadsheets, and LISs). Application programs are managed by a controlling program called the *operating system*. Windows, Mac OSX, Linux, and Unix are familiar examples of operating systems. Data and programs that need to persist when the computer is shut down are written as files from RAM to persistent storage, such as magnetic or optical disks, flash memory cards, or tape, and the files are read from those media back into RAM when they are needed. Data files can be exchanged between computers using removable persistent storage media or by transmission across communication networks. Physical devices are generally referred to as *hardware*, while programs are referred to as *software*. Data stored semipermanently in computers in flash memory, usually

Figure 6-1. Simple diagram of the major components of a digital computer. Binary data stored in random access memory (RAM) are transformed in the central processing unit (CPU). A smaller set of high-speed RAM called cache memory is used to store commonly accessed instructions and data temporarily. If the computer is connected to a screen and keyboard, sections of RAM (which may be separate from primary RAM) hold data that are written to the screen and receive data from the keyboard. Data for long-term storage are transferred to persistent memory, such as magnetic or optical discs, flash memory, or tape, which can be removable and portable. Data for transmission across networks to other computers are processed by a network interface card (NIC) and passed to the network using communication protocols such as Ethernet.



used for basic device configuration settings or system startup, are sometimes called *firmware*.

Data Representation

Computers store and process sequences of binary digits, or bits, that have values of either 0 or 1. For speed and simplicity, bits are usually used in groups of 8, called bytes, that together can represent integers from 0 to 255. Larger integers and other kinds of data, such as floating-point (decimal) numbers, text characters, and images, are represented by using defined patterns to string bytes together in long sequences. Those sequences can then be processed in RAM or saved as files. Large data sets may require many bytes, and thus the byte sizes of files and computer memory capacity are generally referenced in thousands (kilobytes, KB), millions (megabytes, MB), billions (gigabytes, GB), or trillions (terabytes, TB).

Medical systems generally store patient demographics, diagnosis codes, test codes, test results, textual descriptions and interpretations, etc, as characters—including any numbers these data types contain. Previously, most systems represented characters with the American Standard for Computer Information Interchange (ASCII), which uses one byte per character and assigns numerical values to 128 characters, including the lower- and uppercase Roman alphabets; Arabic numerals from 0 to 9; and a limited number of punctuation and “whitespace” characters including the space, tab, and carriage return. ASCII is very limiting, particularly in non-English settings, and has been replaced in most systems by a newer standard called *Unicode*, which can use two bytes per character. Unicode supports a much larger character set, including all alphabets and number representations in use; a large set of business, mathematical, and scientific symbols; and a large number of Japanese and Chinese characters. Current microcomputers support Unicode, and medical information systems are transitioning to Unicode support.

Computers can represent images as vector graphics or bitmaps, also known as raster images. Vector graphics are drawn as combinations of shapes from mathematical instructions contained in a data file and are commonly used for graphic art. Bitmaps are two-dimensional arrays of dots called *pixels* that vary in brightness and recreate an image when displayed in aggregate (Figure 6-2). Images captured by digital cameras or scanners, including most images used in medical and pathology systems, are bitmaps. In a full-color image, each pixel is defined by three bytes whose values represent the brightness of red, green, and blue; the remaining colors are created as mixtures of these primary colors. Modern systems may include a fourth byte that can be used for additional purposes, such as indicating whether the pixel should represent transparency if it is layered over another pixel

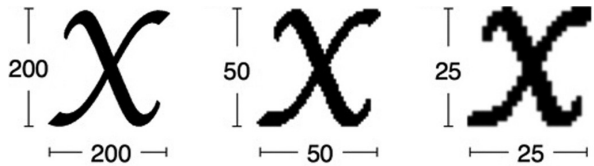


Figure 6-2. Bitmap image resolution. The images of the letter X are formed from black-and-white dots (pixels) at resolutions of 200 x 200 (left), 50 x 50 (center), and 25 x 25 (right). As resolution decreases, the image loses its ability to represent fine detail. Full-color images of these resolutions would require a minimum of 120 KB, 7.5 KB, and 1.9 KB, respectively, assuming 3 bytes per pixel.

by allowing a variable amount of the underlying pixel’s color to be averaged with the color of the overlying pixel. The resolution of a bitmap image is the number of pixels it contains and is usually expressed as the vertical and horizontal pixel dimensions (Figure 6-2). The minimum size of an image’s data set is the product of the pixel dimensions and the number of bytes used per pixel; for example, a full-color, medium-sized image (1024 × 768 pixels) would require a minimum of about 2.4 megabytes. Higher-resolution images appear sharper at a given image size (Figure 6-2) but require more storage space and take longer to transmit over a network. Images that represent pathology slides at multiple levels of magnification appropriate for diagnosis usually have a very high resolution and are thus very large.

Image Compression

Image compression techniques, which mathematically describe groups of pixels in an image rather than representing each pixel separately, reduce the size and transfer time of images. Image compression may be lossy or non-lossy. Lossy compression methods eliminate less important data and reduce the size of the remaining data. If the method is applied correctly, the regenerated image, while not identical to the original image, does not show noticeable visual differences. Non-lossy methods compress all the data, and regenerated images are identical to the original images. The most common lossy method is called *JPEG* because it was developed by the Joint Photographic Experts Group. This method is very effective for compressing continuous tone images such as photographs or scans—ie, images that contain color gradients with few abrupt edges. The JPEG method allows variable compression with greater artifacts at higher compression levels; thus, compression can be adjusted to suit a particular use. Because lossy methods introduce artifacts each time they are used, images that are repeatedly compressed will degrade. It is best to use a lossy compression method for the “final form” compression of an image that will not be further edited and re-saved.

The most common non-lossy compression method is PNG (Portable Network Graphics), which is used optimally for bitmap graphic art with large areas of identical color and few gradients. PNG allows one compression level; the method compresses photographic images, albeit much less effectively than JPEG. However, because it is non-lossy, PNG allows images to be repeatedly compressed with no loss of quality. JPEG and PNG files are typically denoted by *.jpg* or *.png* file-name extensions, respectively. JPEG is most useful for pathology images as long as they are not overly compressed and repeated compression is avoided.

Data Standards

ASCII, Unicode, JPEG, and PNG are data standards that have been defined and approved by standards organizations, and their specifications are generally available. A number of organizations are active worldwide in producing data standards, including professional societies, government agencies, and consortia comprising vendors, users, and other interested parties. Standards development typically includes a requirements definition, an initial trial implementation, a period of comment and refinement, and a vote leading to approval for general use and publication. Standards may be copyrighted and licensed for fees that support their ongoing maintenance, or they may be freely available. Any vendor or programmer can review the specifications in a data standard (subject to applicable licensing fees), design a computer program to work with the standard, and share data with other programs that adhere to the same standard. Pathology information systems use several standards for medical data representation that support the basic interchange of information between systems for regulatory reporting and electronic billing. The most important of these standards are the International Statistical Classification of Diseases and Related Health Problems (commonly known as ICD) and the Current Procedural Terminology (CPT). Both are controlled vocabularies that associate well-defined medical concepts with numeric codes. ICD, currently ICD-10, is a hierarchical list of diagnoses and inpatient hospital procedures published by the World Health Organization. In the United States, ICD has been separated into a diagnosis terminology list, ICD-10-CM, which contains about 72,000 codes, and a procedure terminology list, ICD-10-PCS, which contains about 30,000 codes.¹⁷ CPT represents professional medical services and outpatient procedures, such as laboratory testing and pathology interpretation, and is maintained and licensed for use by the American Medical Association. In billing claims, procedures must be identified by CPT codes and justified by ICD codes.

Though ICD and CPT have been useful for communications related to billing and reporting, they are not

complete medical vocabularies capable of describing laboratory tests, results, and interpretations in adequate detail for clinical summaries. Future health-data-sharing goals including the transport of clinical summaries between care providers, and population health surveillance will require more comprehensive medical data standards such as LOINC and SNOMED-CT. LOINC is maintained by the Regenstrief Institute and is available for use at no cost.¹⁸ It provides standard names and codes for clinical observations, including laboratory tests as well as components of the physical exam, radiology exams, EKG studies, etc. A laboratory test is designated in LOINC by a “fully qualified name” that includes the analyte (component), the property being measured, the timing of the measurement, the type of specimen (system), the general measurement scale, and an optional method descriptor (Table 6-1). Each name defines a type of test and is designated a numerical code with a hyphenated last digit. SNOMED-CT (Table 6-2), a large clinical vocabulary with excellent coverage of pathology topics, was originally developed under CAP management and is currently maintained and developed by the SNOMED International organization. It has been licensed for national use in a number of countries including the US.¹⁹ The vocabulary contains over 300,000 concepts represented by numeric codes and is organized into 18 topic-oriented hierarchies (clinical finding, body structure, organism, specimen, social context, etc). It also includes rules for combining its codes into higher level concepts. SNOMED-CT defines relationships between concepts such as “is-a,” “finding-site” (body location), and “associated-morphology” and includes over 1 million specified relationships within and across its 18 hierarchies. Notably, the terminology includes the notions of uncertainty and severity, which are absent from ICD.

General Characteristics of Laboratory Information Systems

Modern LISs are deeply embedded in laboratory operations, managing data and supporting workflow required for preanalytic tasks, testing, reporting results, and laboratory administration.²⁰ LISs receive patient demographics and test orders, and print work lists with container requirements and specimen labels for phlebotomy teams or clinicians. Specimens are tracked, usually by scanning a barcoded label at each step, from creation through transport, accessioning in the laboratory, preparation, aliquotting, analysis, and storage. LISs provide testing instructions to automated analyzers and worksheets for manual methods, and receive results from both sources. Quality control data are tracked and compared with acceptable values on a run-by-run basis and accumulated into reports for periodic review. Initial

Table 6-1. LOINC Example for Cardiac Troponin I

Code	Component	Property	Time Aspect	System	Scale	Method
10839-9	TROPONIN I.CARDIAC	MCNC	PT	SER/PLAS	QN	

MCNC: mass concentration; PT: point in time; SER/PLAS: serum or plasma; QN: quantitative.

Fully-qualified name: TROPONIN I.CARDIAC:MCNC:PT:SER/PLAS:QN; Short name: Troponin I SerPl-mCnc (LOINC short names are for convenience but may change over time).

Table 6-2. SNOMED-CT Data Elements for the Concept of Bronchial Pneumonia (Simplified)

Concept ID	Name	Synonyms*	Relationships**	
67814005	Bronchopneumonia	Bronchial pneumonia	is-a	pneumonia
		Lobular pneumonia		
		Segmental pneumonia	associated-morphology	inflammation
		Bilateral bronchopneumonia	finding-site	lung
		...	finding-site	bronchus

* Each synonym also has a numerical ID.

** Representative examples shown; a SNOMED record would contain additional relationships. Relationships may connect two SNOMED concepts within a hierarchy (is-a) or across hierarchies (associated-morphology, finding-site). A SNOMED record connects concepts by associating the primary concept code with a second concept code through a relationship code.

results of tests may be processed using rules that calculate derived results, initiate reflex testing, or provide decision support; interpretive or explanatory comments from stored libraries of text may be manually or automatically added to selected results. Reports containing results and associated text in electronic or paper form are routed to a clinical information system, the patient's location, and/or the ordering physician. LISs may create requisition forms or electronic orders to accompany send-out tests to reference laboratories and receive and report the results of such tests. They may also receive orders from and transmit results to other laboratories and physician offices to support outreach activities. A variety of predefined scheduled reports support quality assurance, workload analysis, and resource use, and ad hoc reports are often created to address special questions from laboratory management, an associated health care system, or medical researchers. LISs provide data necessary for billing pathology services, and billing systems may be closely associated with an LIS under pathology management, or the LIS may provide procedure information and diagnostic codes to an institutional billing system through an electronic interface.

LISs receive, create, store, and report many different data elements as they carry out the tasks listed above. While some of these data elements are represented using standard vocabularies such as ICD and CPT, many more are defined locally and thus are not meaningful outside their local environment. For example, most laboratories use locally invented codes to designate laboratory tests. This practice developed because in the past there was no standard way to represent tests, and because short,

mnemonic, locally memorable codes were useful as commands for terminal displays. The disadvantage of local codes is that they are meaningless if they are sent to another system unless explicitly transformed to the appropriate data representations used by that system—a process that requires effort and time. Standard vocabularies such as LOINC and SNOMED-CT, which could potentially solve this problem, are not yet widely used for data representation within LISs.

Database Applications

LISs store standard and locally defined data elements in databases, which are organized data files that are managed by application programs called *database management systems (DBMSs)*. DBMSs provide access control, audit trails, transaction support for data modification, and searching/reporting tools. User access is typically managed via password-protected accounts. Good password practices are crucial for system security (Table 6-3). The network environment may provide additional security such as a virtual private network (see “System and Data Security”). Descriptions of users' actions, including viewing or updating data, are captured to log files called *audit trails* that identify users by log-in account. Transactional databases allow multiple data items to be updated together so that all data are confirmed to be correctly updated; otherwise, the state of the database is rolled back to the point prior to the transaction so that the transaction succeeds or fails as a whole. Concurrent transactions are managed in such a way that they do not conflict with one another. Transaction support

Table 6-3. Good Password Practices

Use at least 8 characters (more is better).
Do not use single words from the dictionary or names (even with numbers appended).
Do not use common number sequences.
Do not use current or past personal information.
Use upper and lower case letters, punctuation, numbers, and spaces.
Passwords derived from phrases can be good, especially if numbers, punctuation, and spaces are added.
Passwords should be changed periodically to limit access if a password is broken.
Never share your password with others.
Do not enter your password in front of an observer.
If you write down your password, keep it in a safe place; do not keep passwords where they can be seen or easily found.

is important in receiving an order, which adds a number of new data elements to the database that should be accepted only as a complete set. Medical databases have several additional transaction requirements: once accepted, transactions should not be deleted because the information could have triggered actions that affect patients. Corrections should be accomplished via an annotated replacement record while maintaining the previous version. DBMSs also provide a method for defining database searches, formatting and scheduling reports, and saving searches and report formats for future use. LISs typically provide predefined modifiable reports that are useful for laboratory management and allow the creation of additional scheduled and ad hoc reports.

Several types of DBMSs are used in transactional LISs. *MUMPS databases* first appeared in commercial LISs in the 1970s and are still used in a number of systems. MUMPS databases—efficient, fast, and able to support high transaction rates in large laboratories or multiple sites with a consolidated LIS—are essentially hierarchical data structures that update and report very rapidly when the update or search “fits” the design of the hierarchy (Figure 6-3). Ad hoc searches that do not fit the hierarchy may be very slow, lasting for hours, and may require substantial system resources. Several LISs and instrument managers use an alternative relational database design that organizes data in multiple tables that are associated through shared data elements (Figure 6-3). *Relational databases* also use structured query language (SQL), a powerful, relatively

standard programming language for specifying searches and reports. Because they are flexible and applicable in many different settings, relational databases are the most widely used type of database outside of medicine. However, relational databases are slower and more demanding of computer hardware than MUMPS databases when the latter are used optimally for their hierarchical design, and thus have some disadvantages for very large applications and high transaction rates. On the other hand, their flexibility allows relational databases to provide much better average performance than MUMPS databases across a range of arbitrary searches. Some MUMPS databases provide adapters that allow searches to be specified in SQL to take advantage of SQL’s convenience and familiarity, but the performance of those searches remains subject to the hierarchical structure of the database.

A *data dictionary* is the set of individual data elements a database contains. The internal organization of data elements in different databases with the same data dictionary can differ, yielding advantages or disadvantages for particular applications. The structure of a database is called its *schema*, which is often depicted graphically, similar to Figure 6-3. A broader and more general term for the definition of a set of data elements and their relationships is a *data model*. Databases have data models that describe their data elements and organization; on a smaller scale, individual reports, insurance claims, and any other grouping of data elements designed for a particular purpose also have data models. Database searches usually yield lists of data elements in meaningful groups; for example, a work list of specimens with requested tests or a list of test results with patient names and medical record numbers. The entries in these lists are often called *records*, and their constituent data elements are called *fields*. A list of records is often displayed as a table, with one record per line and one field per column, but records also might be displayed as one record per page or screen, with the content of fields arranged logically in the display area. The fields of a record may contain single data elements with well-defined representation and meaning that are either local or derived from a standard, known as *structured data* (eg, dimensions of a tissue specimen captured in separate fields using a defined syntax), or they may contain uncontrolled narrative text with embedded data elements that have variable position and wording, known as *unstructured data* (eg, dimensions of a specimen captured as part of a textual gross description). Unstructured data are difficult to search quickly and reliably because their presence and content may vary unpredictably across large numbers of records. Methods have been developed to add structure to narrative text, including the use of defined templates or markup (eg, Extensible Markup Language [XML]),

and to process these textual patterns in databases; however, use of this technology in pathology systems has been limited to date.

Although LISs and anatomic pathology systems share many similarities, the systems differ in several important ways. Because LISs support a flow of clinical data that includes orders, specimens, discrete results, and locally standardized textual comments, most data in LISs are structured. In contrast, anatomic pathology systems support extended textual descriptions and interpretations with fields representing general report sections, such as clinical history, gross description, final diagnosis, and comments—a design that yields a substantial amount of unstructured data. Anatomic pathology systems usually allow standard coding of diagnoses (using SNOMED-CT, for example), but there is little incentive to perform detailed coding since it is not required for billing or patient care, and because clinical data sharing has consisted primarily of transporting complete reports. In lieu of using defined data elements to search results, anatomic pathology systems generally offer “free text” (character string) searches of unstructured fields whose reliability is dependent on the local site’s ability to standardize text descriptions by convention. The use of uncontrolled text has serious drawbacks, but it does have the benefit of flexibility. In recent years, this flexibility has enabled anatomic pathology systems to support

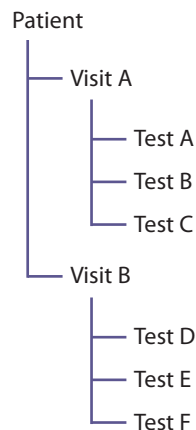
relatively complex reports from, for example, molecular diagnostics and flow cytometry, when LISs did not offer structured data models adequate for these purposes.

Some anatomic pathology systems also associate selected images with cases. Since images are often large and do not need to be searched internally, it is most efficient to store them as files outside the textual case database, with links in the database that lead to the image files. In some cases, images are maintained on a separate image server, a computer optimized for the storage and transmission of large data files. Data standards for representing pathology images with associated information such as measurements and interpretations have been developed recently by DICOM, the organization that also develops standards for digital radiology.²¹

Databases such as LISs that are used in daily patient care are typically transactional systems with schemas designed to support efficient data lookup and updating for individual patients. Databases are also useful for finding populations of patients with similar characteristics, for example, for retrospective research, population health surveillance, or quality improvement studies. Because transactional schemas are not efficient for these types of tasks, data is often exported from transactional systems to large databases called *data warehouses* that have schemas optimized for population searches over an extended time span.²² Some commercial clinical systems include

Figure 6-3. Database structure. Hierarchical databases (A) connect data elements in tree structures. Looking up test results on a particular patient by traversing the tree is very fast. Finding all instances of Test E across a population of patients in an ad hoc report, in contrast, would be very slow because each patient and visit would need to be evaluated to see if Test E was present. Relational databases (B) maintain data in multiple tables that share item identifiers. Each patient in the Patients table is assigned a unique patient ID (pID), and the table also contains other information about patients. Each visit has a unique visit identifier (vID) in a separate Visits table, and also the pID from the Patients table and other information about the visit. Each test has a unique test identifier (tID) in a separate Tests table, along with the pID from the Patients table and the vID from the Visits table, plus other information about the test and result. Test results for a particular patient can be found by searching the tests table using the patient’s pID. This is reasonably efficient but not as fast as traversing the hierarchical tree in (A). Instances of Test E can be found by searching the Tests table, and any necessary information about visits or patients associated with Test E can be found by searching the other tables with the ID values associated with Test E in the Tests table. This is much faster than evaluating all patients and visits, and illustrates the flexibility of the relational design.

A. Hierarchical Design



B. Relational Design

Patients			
pID	Other Information ...		

Visits			
vID	pID	Other ...	

Tests			
tID	vID	pID	Other ...

an integrated database for this purpose that supplements the transactional system. In other cases, organizations may construct a data warehouse as a separate database that receives periodic data transfers from other systems. In addition to better population query performance, data warehouses often integrate data from multiple systems and thus support correlative analyses that would be otherwise difficult to accomplish.²³ Data warehouses may be searched directly to find patients with particular attributes, or statistical techniques collectively known as *data mining* may be used to discover previously unknown associations in data or to create data-driven predictive models.²⁴ These techniques appear to have substantial potential for application to pathology data,²⁵⁻²⁷ but so far they have seen limited use in the field.

Computer Networking

LISs were developed initially as stand-alone systems, with most data entered manually through multiple, directly connected terminals and results communicated through printed reports. This situation has changed dramatically, and most laboratory systems now communicate directly with instruments and other computer systems through computer networks. Networked computers are connected to each other by media such as coaxial cables, twisted pair copper wires, fiberoptic cable, or radio frequency signals. Network interface cards (Figure 6-1) inside the computers send and receive binary data to and from the media in the appropriate form (eg, impulses of electricity, light, or radio waves). For convenience, local networks that use physical media are usually configured in star patterns, with a network hub or switch at the center and computers at the ends of the branches (Figure 6-4). The hub or switch is located in a protected spot such as an electrical closet and may service 48 or more branches leading to workstations, instruments, the LIS itself, other hubs, and/or external connections. Basic hubs allow all network communications to travel to all branches. Switches are more expensive and allow communications to travel only to the branches containing the devices to which they are addressed. Switches have speed and security advantages because they reduce overall network traffic and prevent devices from reading each others' network communications surreptitiously (for example, eavesdropping). Computers and application software that provide shared resources or data to a network are called *servers*; computers and application software that users work with directly to access server-provided resources across a network are called *clients*. Thin clients are a class of client software that is designed primarily to display server-generated views of data rather than receive and process data themselves.

Successful communication between two computers requires that both computers use the same strategy for

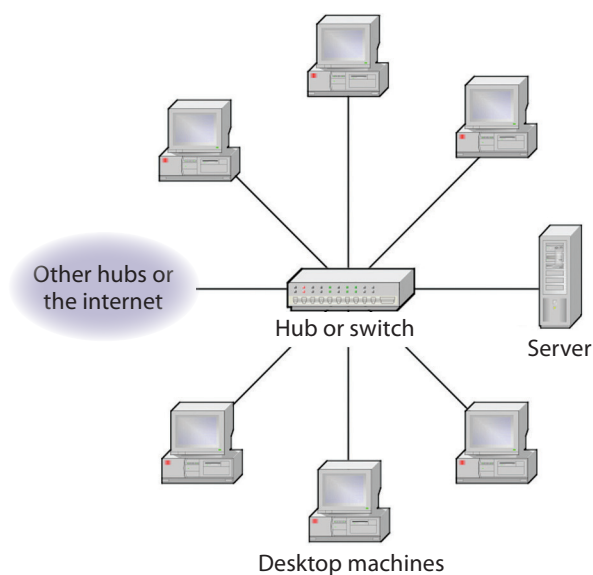


Figure 6-4. Basic “star” network. The connected devices are arranged like the points of a star around a central hub or switch. Workstations, servers, printers, scanners, and other devices may be connected to the network, and a branch of the network may connect to other local networks or to an internet connection point. Hubs support a common network in which all communications go to all branches, and devices ignore communications addressed to others. Switches actively sort communications so that data only enter a branch if they are addressed to a device on that branch. Wi-Fi base stations function similarly to standard hubs but use radio frequencies instead of wires. All devices within range can “hear” all communications but ignore communications addressed to other devices.

encoding, addressing, and exchanging data, and these shared network protocols are established by standards organizations. Ethernet and Wi-Fi are two of the most common data transport protocols used in local networks. Most current devices implement these protocols directly, but some older instruments and printers may use other protocols such as the RS-232 serial communications protocol for direct connection to computers. RS-232 devices can be connected to local networks using a hardware adapter, which is essentially a small computer that gives the device a network address and translates the protocols. Additional protocols are used to transmit data greater distances between local networks, and data transmitted across the internet between widely separated computers may cross many different media and be translated through many protocols. Computers or local networks in homes and small businesses may be connected to the internet via digital subscriber lines (DSL) from telecom vendors, cable television networks, or satellite connections in remote locations. These low-cost, relatively high-speed connections for end users are collectively termed *broadband*. Larger organizations may

lease higher-capacity lines from telecomm vendors at higher cost.

Network performance is generally measured as data transmission rate in bits per second and is referred to as *bandwidth*. “Standard” Ethernet over twisted pair copper wire runs at a theoretical maximum of about 10 megabytes per second. Updated versions of Ethernet available in most networks and current desktop and laptop computers run at 10 to 100 times that speed (eg, gigabit Ethernet). Protocols running over the national network backbones employ even higher bandwidth. Radio frequency networks are somewhat slower, with Wi-Fi (the 802.11 family of standards) running at about 10 to 150 megabytes per second depending on the specific protocol in use. In reality, the theoretical maximum achievable with one computer transmitting at the maximum rate is rarely approached in local networks. In the usual situation, multiple computers on a network communicating individually occasionally interfere with each other by transmitting simultaneously. When these network collisions occur, both devices wait a random length of time and retry their transmissions. Such collisions reduce the efficiency of communication, and they become more common as overall network usage increases. When this happens, the network appears to slow down from an individual device’s perspective. Network usage and collision frequency can be monitored, and excessive numbers indicate a need to reduce network usage through policy, segment the network to separate traffic, or upgrade the network to a higher overall speed.

A suite of standard communication protocols called *transmission control protocol/internet protocol (TCP/IP)* was originally developed for use on the internet but is now almost universally used across a variety of networks. TCP/IP works in combination with data transport protocols and defines how data is addressed to remote computers (IP addresses), how data is transmitted, and how transmission errors are resolved. In TCP/IP communications, data is divided into packets (short bursts of network activity) that have destination and return addresses plus a data payload, and the packets are numbered in sequence. They are transmitted independently to the destination, and the data is reassembled into the correct sequence by the receiving device.

System Interfaces and HL7

TCP/IP and related protocols allow binary data to be transmitted reliably across networks, but data standards are necessary to allow the binary data to carry meaningful information between application programs such as an LIS and an instrument control program or an EHR. A connection between application programs is called a *system interface*. Interfaces may be unidirectional (one system always transmits and the other always receives)

or bidirectional (systems can exchange data in both directions). Bidirectional interfaces are typical in modern systems; unidirectional interfaces are used with laboratory instruments that require manual test setup but can transmit results and with some hospital administrative systems that only broadcast admit/discharge/transfer information. In the United States and increasingly in the rest of the world, standards for both instrument and system interfaces used in health care are developed and approved by the HL7 International standards organization. The American Society for Testing and Materials (ASTM) also provides standards for instrument interfaces.

Current system interfaces use version 2 of the HL7 standard, which defines a set of text message formats for transferring information related to particular health care tasks. Important message types for pathology include admit/discharge/transfer (patient demographics and location), order entry (new orders to the lab), and results (pathology data for reporting). Each message type has a data model that defines its overall structure, the type of data elements (fields) it contains, and the placement of the fields in the message (Figure 6-5). HL7 v 2 defines the permissible content of fields but also permits some locally defined content. This content flexibility allows HL7 interfaces to be used without full data standardization, but it also means that the interface must transform nonstandard data elements between representations used in the connected systems. This is usually done with hand-constructed mapping tables that associate equivalent data elements and must be edited and tested whenever data elements in either system change. The interface must also catch and log transformation errors that occur when mappings are omitted or incorrect. In the absence of an HL7-compatible standard for text display formatting, HL7 messages carry textual reports as a series of separate text lines without formatting instructions.

These characteristics mean that current HL7 interfaces are unique for a particular installation, expensive to create and maintain, and severely limited in transmitting textual reports.²⁸ A new version of HL7, HL7 version 3, addresses these problems by implementing a comprehensive data model for medical care, the Reference Information Model (HL7 RIM), as a basis for defining both the structure of a message and the allowable content of its fields. Messages can be expressed in several standard forms including XML markup,²⁹ and a standard XML-based Clinical Document Architecture (HL7 CDA) is available that supports textual reports including clinical summaries (HL7 CCD) containing both formatted narrative text and structured data.³⁰⁻³² A messaging architecture based on the HL7 RIM called Fast Healthcare Interoperability Resources (FHIR, pronounced “fire”; <http://www.hl7.org/fhir/>) is in development and prototype use that may resolve many of these


```

MSH|^~\&|IDXLAB|UVAHSC|DADD||DateTime||ORU^R01|DateTime|P|2.2|||
PID|||xxxxxxx||lname^fname^mi||Date|M|^|3116488572||
PV1||OP|STFM^|||483230^lname^fname mi.(4201)|^|OP|...
ORC|RE|320532-0^0|||4201^lname^fname mi.(|||^|
OBR||320532-0^0||CPBAS^BASIC METABOLIC PANEL^CPBAS||DateTime|^|...
OBX|1|NM|NA^SODIUM|1|137|mmol/L|136-145|||F||DateTime|UVA^|2886^AUTO^VERIFY|
OBX|2|NM|K^POTASSIUM|1|4.0|mmol/L|3.5-4.5|||F||DateTime|UVA^|2886^AUTO^VERIFY|
OBX|3|NM|CL^CHLORIDE|1|101|mmol/L|98-107|||F||DateTime|UVA^|2886^AUTO^VERIFY|
OBX|4|NM|CO2^CO2|1|24|mmol/L|23-31|||F||DateTime|UVA^|2886^AUTO^VERIFY|
OBX|5|NM|BUN^UREA NITROGEN|1|23|mg/dL|8.4-25.7|||F||DateTime|UVA^|2886^AUTO^VERIFY|
OBX|6|NM|CREAT^CREATININE|1|1.0|mg/dL|0.7-1.3|||F||DateTime|UVA^|2886^AUTO^VERIFY|
OBX|7|NM|GLUC^GLUCOSE|1|135|mg/dL|74-99|H|||F||DateTime|UVA^|2886^AUTO^VERIFY|
OBX|8|NM|CALCM^CALCIUM|1|9.8|mg/dL|8.4-10.2|||F||DateTime|UVA^|2886^AUTO^VERIFY|
OBX|9|ST|GFR^CAL^CALC GFR (mL/min/1.73m2)|1|>60|||F||DateTime|UVA^|2886^AUTO^VERIFY|

```

Figure 6-5. Representative HL7 version 2 message. HL7 messages are textual and adhere to a structure defined in the HL7 v 2 standard. This is an Observation Result message (ORU-R01) for a basic metabolic panel. Names and dates have been replaced to protect identities, and two long lines were truncated (indicated with "..."). Each line of an HL7 message is called a segment, and each segment starts with a three-letter identifier and then contains fields delimited by vertical bars. Fields may be subdivided with additional delimiter characters to produce components (carats are used as delimiters here). Not all fields are used, and unused fields are left in place but are empty (indicated by contiguous bars). The segment structure of a message is defined by the HL7 standard; an ORU-R01 message contains a Message Header (MSH), Patient ID segment (PID), Patient Visit segment (PV1), Common Order segment (ORC), Observation Request segment (OBR), and a variable number of Observation Result (OBX) segments. It may also include Note (NTE) segments, which are not shown here. The type of information that a field should carry is also defined in the standard, for example, field 3 of an OBX segment should carry the service identifier (test name), but there is no specification of how that name should be expressed. Thus HL7 v 2 is a partial standard defining message structure, but some of the data carried by these messages is not standardized.

limitations. Future systems based on FHIR are intended to transmit and receive data using fully standardized syntax (data representation) and semantics (data meaning). Such systems would be interoperable and easily interfaced using a generic HL7 data connector, with minimal configuration and data mapping.

LISs that serve large health care delivery enterprises typically implement many (mostly HL7) interfaces.²⁸ Within the laboratory, interfaces connect the LIS with automated instruments and instrument manager middleware. From an enterprise perspective, LISs are one of a number of ancillary systems that support specialized services and are interfaced to hospital information systems, clinical information systems (eg, EHRs), and billing systems. Enterprises often implement an automated communications manager, sometimes called an *interface engine*, to provide a consistent connection point for communications and to organize and support a large number of HL7 interfaces. In aggregate, the core systems, interface engine, ancillary systems (as well as their local environments), and external connections make up the physical architecture of a health care enterprise information system (Figure 6-6).

LISs may also use HL7 interfaces to exchange orders and results with external systems via the internet. Such systems may include reference laboratory or other LISs, EHRs in physician offices or remote clinics and community hospitals, and/or specialized laboratory outreach support systems. Remote health care delivery sites may

also offer local laboratory services with workstations and instrument interfaces that connect to the central LIS. Some vendors, called *application service providers* (ASPs), provide off-site LIS services and support by subscription, usually to smaller sites. Participating sites connect instruments, workstations, and local administrative and clinical systems to the remote LIS via HL7 interfaces provided by the vendor. This arrangement reduces the local requirement for managing hardware, the LIS application, and interfaces, and may be financially attractive depending on a site's size. However, the ASP vendor must be able to provide an ongoing adequate level of service, meet regulatory and accreditation requirements, and provide a viable plan for exporting data in an interoperable form in the event that the level of service is inadequate or if a site wishes to change vendors or bring the LIS in house.

System and Data Security

System security is important because of the importance of LISs in supporting laboratory workflow and the importance of pathology information in diagnosis and patient care. Security has two aspects: (1) pathology systems must deliver data reliably and correctly or patients may be harmed, which means that LISs should not fail or allow data to be deleted or modified surreptitiously by malicious activity or inadvertent errors; and (2) pathology data contain patient identifiers and thus qualify as protected health information under the

Health Insurance Portability and Accountability Act (HIPAA), which means that identified pathology data is confidential. HIPAA indicates that confidential data should be displayed or transmitted only when necessary for medical care/billing, quality control/improvement, or formally-approved research; and should be viewed only by the individuals involved in those tasks. As a practical matter, most inappropriate data access involves local personnel and local devices. These problems can be limited by good information system practices such as user training, good password security (Table 6-3), timely removal of retired accounts and passwords, automated logout and screen blanking of inactive workstations, regular review of audit trails listing data accesses and changes by personnel, and the use of network switches rather than hubs to limit eavesdropping.

Connection to the internet raises the possibility of remote attacks that divert computers from LIS tasks, steal data, and/or disrupt network communications. Recent tests indicate that susceptible computers on unprotected networks connected to the internet are compromised in about an hour on average.³³ Attacks may attempt to take

over control of computers or capture sensitive data such as passwords or credit card numbers by targeting operating system or application defects, or by tricking users into installing malicious software (*malware*). Compromised computers may be diverted to unwanted or malicious processing, communication, or data storage tasks, which reduces their performance and may render them unreliable. Alternatively, malicious external computers and compromised local computers may intentionally flood a local network with data packets, preventing legitimate communication (*denial-of-service attack*). Any of these events can be very disruptive to laboratory operations and requires substantial effort for recovery, and they should be prevented by good computer and network security practices. Basic good practices include keeping software updated to fix identified security problems (small corrective updates are sometimes called *patches*; note that LIS vendors may need to verify operating system and other software patches and updates before they are applied), doing routine work in user accounts with limited access privileges rather than in administrator accounts, using malware detection programs to block and

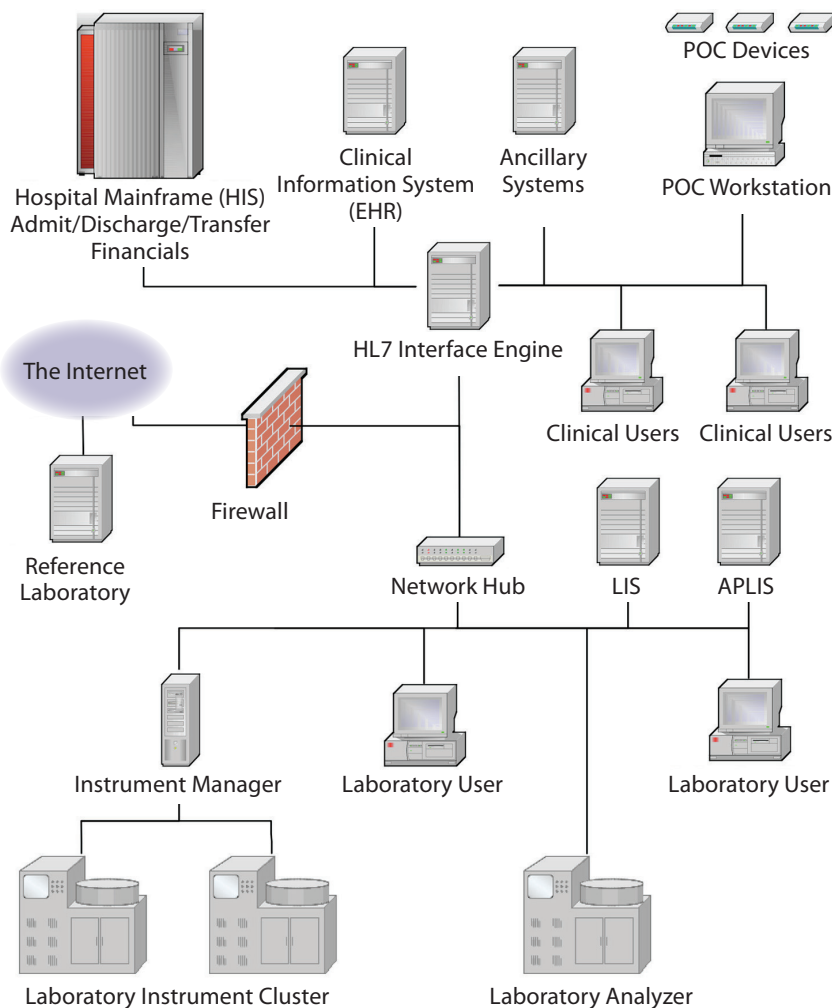


Figure 6-6. Health care enterprise system architecture. This diagram illustrates the participation of laboratory systems in the larger health care information ecosystem. The laboratory information system (LIS) and anatomic pathology system (APLIS) on the right side of the figure are accessible to laboratory users and instruments over the network. This laboratory implements an instrument manager ("middleware") that manages a cluster of instruments and connects to the LIS. HL7 interfaces between the major systems (hospital information system [HIS], electronic health record [EHR], and ancillary systems such as radiology, point-of-care [POC] devices, and the laboratory systems) are managed by an HL7 interface engine. Non-HL7 network traffic such as email and web communications also travel on this network. One of many network hubs in a system this size is illustrated. External communications with a reference laboratory, other outside systems, and the web are available through a firewall that protects the internal network.

remove known problem software, and limiting the use in the laboratory of removable data storage media (eg, flash memory drives) and portable computers that have been connected to external networks.

Computers and networks can be protected from malicious communications using firewalls that filter and route communications packets (Figure 6-6). The simplest firewalls are application programs that block all types of communication to the computer they are running on except for those that are explicitly allowed. More complex firewalls are used to protect local networks, and they may be separate devices or computers that are located between sections of a network or between a local network and the internet. These firewalls may examine the target addresses, return addresses, sequence, size, and even content of communication packets, and block those without authorization or those with evidence of malicious content or usage patterns. These firewalls can keep track of communication sessions and, for example, allow external packets to pass only when they are responses to internally initiated communications. Firewalls may implement a proxy server through which all communications to and from the outside must pass. These proxies prevent an outsider from gaining knowledge about specific devices inside the network and thus make it difficult to target devices and vulnerabilities for attack. Firewall configuration and network security should be managed by an expert and may be handled by the general information technology support unit at enterprise sites or a security consultant at smaller sites.

Routine internet communications, such as email, file transfer, and data submission using web forms, are insecure. The packets that carry those communications are stored and forwarded between many devices as they pass from origin to destination. Logs from those devices may be copied, or eavesdropping software may be used to copy the contents of packets to text files. Those files can then be searched automatically to find information of interest. Packets could even be altered or replaced prior to reaching their destination. For many types of communication, a low level of security is acceptable. It is not acceptable, however, for the transmission of protected health information, online commerce, or many other types of business communications. Such data requires encryption, a process by which the bit sequence carrying the data is scrambled using a special second sequence called a key and, depending on the form of encryption, can be unscrambled only with the same or a related key. The protocol currently used to encrypt internet communications is called *Transport Layer Security (TLS)*, which is an extension of the *Secure Sockets Layer (SSL)* protocol developed by Netscape in the mid-1990s.

A full review of cryptography and TLS is beyond the scope of this discussion, but the following is a brief sketch

of the logical process. Pairs of cryptographic keys can be created such that if one is used to encrypt data, the other can be used for decryption (asymmetric encryption). These keys are incorporated into a digital certificate that is provided to a recipient by a certificate authority that also verifies and registers the recipient's identity according to a standard protocol such as NIST SP 800-63.³⁴ The certificate is stored on the certificate owner's computer, and one key is made public, while the other is kept private (public key encryption). When another computer wishes to communicate securely, it requests the public key, verifies it with the certificate authority, and uses it to encrypt the data to be sent (Figure 6-7). After that, only the matching private key can decrypt the data, and only the computer holding the certificate has the private key. Thus the certificate provides an encryption method that both protects the data and authenticates the identity of the certificate holder. In online commerce, the vendor's web server holds the certificate, because the goal is to authenticate the vendor so that the client's payment is directed correctly. (The client's successful credit card validation is regarded as adequate authentication for the vendor's purposes [Figure 6-7, A].) For access to protected health information or other enterprise network resources, the goal is different; the greatest need is to authenticate the client, though both sides should ideally be authenticated. Thus to identify and encrypt communications from trusted outsiders, the latter should download and install digital certificates on the client computers they will use to access the sensitive data or resources (Figure 6-7, B). Some health care systems implement their own certificate authorities to provide these certificates, or they may be purchased from commercial certificate authorities. In addition, many sites require two-factor authentication for external access to protected health information. Two-factor authentication requires something the user knows (a password) and an identifiable object the user has (a hardware token that can be plugged into a computer, a smart card with a number that changes according to a unique pattern, or an assigned cell phone with an application that responds to an access confirmation request).³⁴

Encrypted communications from trusted clients may be sent directly to a server, or they may be passed through a firewall into the local network of the enterprise. The latter essentially extends the local network to the authenticated client through an encrypted "pipeline" called a *virtual private network (VPN)*. Alternatively, a server at one location with a local network may establish a VPN connection to a second server at a remote local network, creating a virtual network that connects the two local networks securely and allows free communication between their computers. Encrypted communications are thus very useful for connecting enterprise

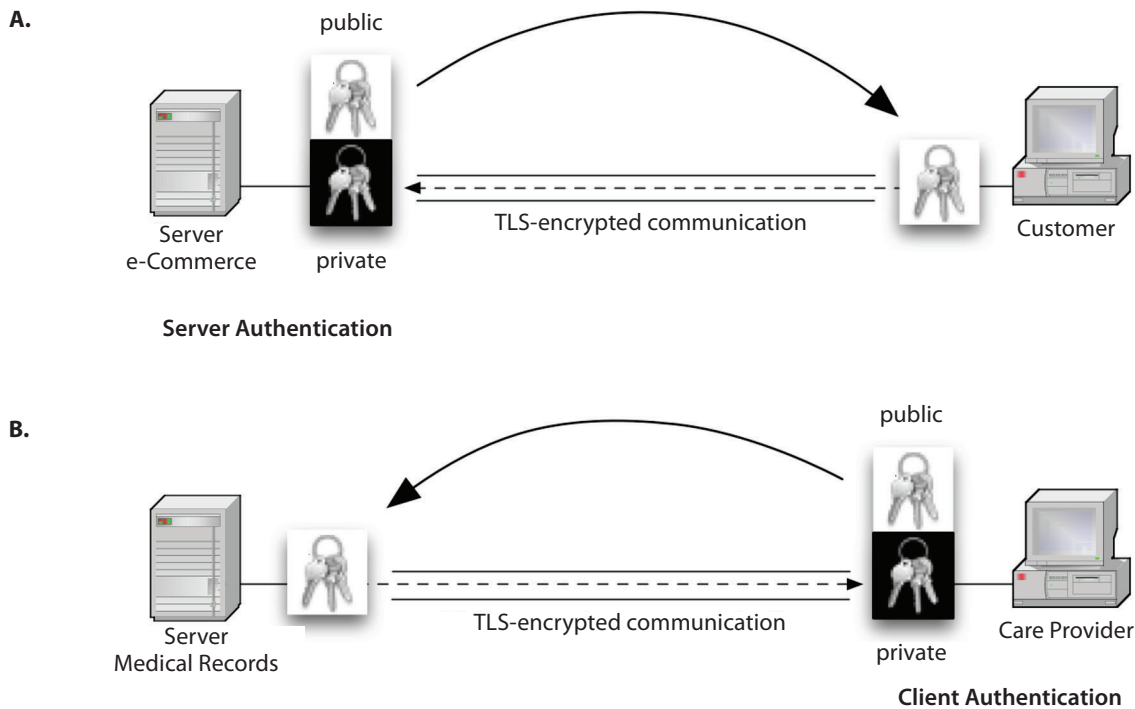


Figure 6-7. Public key encryption. In typical e-commerce applications (A), it is important to authenticate the vendor server because the customer sends the server sensitive information such as credit card payment instructions. The server has a digital certificate containing public and private keys. The public key is sent to the customer, who may check it for authenticity with a certificate authority. The public key is then used to encrypt the data, and the data are sent to the vendor. Only the private key that matches the public key and is owned by the vendor can decrypt the data. In health care applications (B), it is usually most important to authenticate the client, such as an external health care provider, because they will be sent sensitive medical records information. The client must have a digital certificate installed on their computer, which will provide its public key to the medical data server. The server can validate that key with a certificate authority or its own records and then use it to encrypt sensitive data. The data are then sent to the client and can be decrypted only with the private key that matches the public key, which only the client owns. Virtual private network (VPN) connections are established similarly, creating a persistent encrypted connection to a remote network. TLS, transport layer security.

networks with workers from home, physician offices, affiliated health care systems, vendors, or application service provider (ASP) sites that provide subscription services. Public key encryption is also used in digital signatures, in which documents are signed by encryption with an individual's private key. If a document can be read with an individual's public key, it is proven to be signed by the individual (authentication) and not modified after signature (nonrepudiation, ie, the individual cannot deny signature or claim modification). The actual protocols for encrypted communications and digital signatures are a bit more complex than described here, but the general idea is accurate.

Disaster Recovery

Problems may occur with hardware, software, and security even in systems that are optimally designed and managed, and there can be physical plant failures or natural disasters. Thus, it is critical to have a well-designed plan for data backup, interim operation without the LIS (downtime operation), and rapid rebuilding of systems

(disaster recovery). Backups are often made to tape cartridges because tapes are relatively inexpensive and reliable. A typical backup plan will rotate seven tapes (one for each day of the week), with tapes that are not in use rotated among several storage locations, including at least one off-site storage facility. Some systems may copy data directly to off-site storage. Because replacing hardware and restoring a system from a tape backup takes significant time, depending on the nature of the problem, some sites implement two complete systems with data replicated on both systems. Optimally, the systems are in different locations. One system is used as the primary production LIS, and the other is a "hot backup" to which operations can switch rapidly. Making this switch still requires setup time because interfaces and clients must be reconfigured to point to the backup system. The most critical systems that require very high availability can implement automatic failover, which manages a rapid, automated shift of operations to the second system if a switch is necessary. The laboratory should have a planned downtime operations mode, usually a paper

process, that can be used for ordering, accessioning, analysis, and reporting if the LIS is down for any significant period. The disaster recovery plan should include procedures for evaluating the nature and scope of the problem, establishing an appropriate management process, determining when downtime operations should be activated, returning the system to functional status, deactivating downtime procedures, and entering all work done during the downtime into the recovered system.

Systems and Personnel

An LIS installation may include the main clinical laboratory system, the anatomic pathology system, and related software such as instrument managers, fax and print servers, etc. Because blood bank systems manage therapeutic products, they must be approved by the Food and Drug Administration (FDA) and are thus sometimes separate from LISs. Not all LIS vendors provide an FDA-approved blood bank module; sites using those systems may implement a stand-alone, FDA-approved blood bank system that communicates with the LIS and other local information systems through system interfaces.

Laboratory systems are usually managed by a technical group that should be led by a pathologist and includes a supervisor with one to eight staff depending on the size of the site. There should be at least two technical staff members even for small sites to provide continuity during staff turnover and support after-hours call schedules. LIS staff may be recruited from the medical technology or IT fields. Because the design and implementation of LISs are so deeply intertwined with laboratory operations, it is useful for at least some of the staff to have medical technology backgrounds. These individuals may be identified as strong LIS users among laboratory staff and further trained internally. The working relationship between the laboratory system group and the hospital information group is critical, and one way to cement good relations is for the pathologist to volunteer to be a medical consultant to the hospital's chief information officer. Pathologists are uniquely qualified to do so because they are physicians experienced with instrumentation and automation, have a broad knowledge of medical practice, and are familiar with quality management.

Application Management

At larger sites, LIS staff specialize in application management tasks, for example, managing the LIS software. The LIS computers are often located in a protected environment such as a machine room or data center with other enterprise information systems. System administration, the management of the hardware and operating system, and network management tasks may be carried out by enterprise IT teams or consultants who collaborate with the LIS staff. At smaller sites, LIS staff

may be responsible for LIS system administration tasks, the laboratory network, laboratory workstation support, management of associated printing devices, and other technical tasks. Thus the requirements of different sites may vary widely, and it is important to match the skill set of the LIS staff with these requirements. Some sites that implement laboratory systems that are part of larger EHR systems recently have moved LIS management to the general-enterprise IT support staff. The early consensus seems to be that this arrangement may reduce the quality of LIS support and that it is beneficial to have individuals with laboratory backgrounds manage the LIS application under laboratory medical leadership.

The primary job of the LIS staff and medical leadership is to provide an information management application that is adequate in capability and reliability for patient care and meets the local laboratory's workflow and data communication needs. Evaluating and installing LISs and related systems is part of this job, though LIS installation is such a large project that these systems often remain in place for 10 to 20 years. After installation, the LIS staff is responsible for maintaining the LIS and its connections to other systems, training laboratory staff in its use, resolving problems that users encounter, and selectively enhancing the LIS and associated systems as needed. A formal change control process must be followed for hardware and software modifications and should include validation and approval prior to releasing new systems or modifications into routine use. Validation includes formal testing using a selected testing library of data that demonstrates all functions of the system, with documented review and approval of results. Systems that support transfusion have particularly extensive validation requirements.

Table 6-4 provides an overview of a typical LIS staff portfolio of responsibilities. LIS staff must monitor error logs/messages and system performance on an ongoing basis and document problems and their resolution. System maintenance includes maintaining software (patches and upgrades) and contained data, which include a large amount of information that the LIS requires to operate in the local environment. These data are held in a section of the LIS database often called *maintenance tables* (Table 6-5), and managing these maintenance tables is an ongoing activity as the laboratory and health care organization evolve. Feedback is important to determine whether the LIS is meeting the needs of its users, and most LIS management groups follow user feedback provided directly through the LIS and in regularly scheduled meetings between the LIS leadership, laboratory section supervisors, and laboratory leadership. Laboratories that have an LIS off site, either through affiliation with a larger laboratory or by use of an ASP vendor, will have fewer technical responsibilities but will still need

Table 6-4. Key Responsibilities of the Laboratory Information System (LIS) Management Team

Evaluation, recommendation, installation, configuration, testing, and validation of software and hardware that supports laboratory operations, including the LIS, anatomic pathology system, and related systems
Hardware and software installation or updates to improve performance, add new capabilities, or fix problems, with documented formal testing, review, and approval prior to implementation
Creation and maintenance of a comprehensive LIS procedure manual, with documented biennial review (see Table 6-6 for additional detail)
Documented regular system checks and proactive maintenance to confirm that the LIS and associated systems are stable, are operating within established parameters, and have adequate data storage capacity
User account management, including addition of new users, deletion of inactive users, and resolution of password and other account access problems
Problem and issue resolution for LIS users, including 24-hour night and weekend support
Maintenance of data stored in the LIS, including definitions of laboratory tests, worksheets, and reports; system configuration data; information about the local laboratory and reference laboratories; personnel data; information about the medical practice environment; etc (see Table 6-5 for additional detail)
System interface management, including initial implementation, updating, and data mappings maintenance with documented testing, validation, and acceptance
Documented description, follow-up, and resolution of computer system and interface error logs and error messages, performance problems, and unscheduled downtime
Ensuring or carrying out an appropriate backup plan, and restoring and verifying the system from backup files if necessary
Managing software licenses, hardware acquisitions, and service contracts, including need forecasting and budgeting
Management and updating of test reference information for online or print distribution to laboratory users
Training of laboratory technologists in use of the LIS and any software or hardware modifications, with documentation of training
Data validation and quality assurance (discussed in the “Tools, Metrics, and Outcomes” section)
Demonstration of procedures and documentation for laboratory accreditation inspections

to perform data maintenance, application configuration, error resolution, and documentation related to local use.

Interface Management

System interfaces that are used in health care are built using a partial standard (HL7), and much of the transmitted data use local representations that may need to be translated between systems based on mapping tables. When the interfaces are between commercial systems, the vendors of each system are contracted to develop their sides of the interface at an additional licensing and maintenance cost. If the site uses an HL7 interface engine (Figure 6-6), the vendors will develop the portion of the interface that runs between their systems and the interface engine. For LIS interfaces, the LIS staff will take delivery of the LIS side of the interface and collaborate on its installation, develop a test data library, and carry out the interface testing and validation, confirming

correct data transfer to and from the interfaced system. The LIS staff may then maintain the data mapping tables to ensure that they continue to correctly represent all the data elements that need to be transmitted across the interface. As discussed previously, future developments in HL7 standards may support “plug-in” data communication modules that are applicable at any site and can be configured easily by local staff, with limited or no need for mapping tables.

When data are transmitted to other systems—eg, EHR systems—for patient care, laboratory accreditation requirements specify that the laboratory must confirm and periodically re-check that the data are being transmitted and displayed correctly and completely in a form adequate for clinical decision making. In practice, this requirement usually means that the LIS staff must periodically validate both the data transmission and data display for each system that is interfaced. This validation

Table 6-5. Example Data Tables that are Contained in the LIS and Maintained by LIS Support Staff

Table	Description of Contents
Tests and batteries	List of tests and test groupings with ID codes, CPT codes, orderable status and names (may want to add test catalog)
Test limit values	Reference ranges, delta checks, technical limits, auto-validation limits
Cumulative headers	Result summary formats (becoming less used with EHR displays)
Worksheets	Specimen lists for analyzers or manual workstations
Text comments	Standard comments that can be added to results
Default text	Comments that automatically add to results
Laboratory departments	Main laboratory sections
Laboratory locations	Sites performing testing that are part of the local laboratory
Terminals	Data entry and display devices connected to the LIS
Diagnoses	ICD-10-CM codes
Tube types	Specimen container descriptions
Reference laboratories	Outside laboratories performing testing not done locally
Tech accounts	Laboratory technician demographics, ID codes, and passwords
Locations	Sites where samples will be obtained and/or results will be reported
Physicians	Physicians who may order tests on patients and receive results
Workstations	"Benches" or work areas of the laboratory
Test logs	Listings of specimens, tests, and results
Quality control	Definitions of controls and target values
Calculations and rules	Instructions for automatically calculated values, reflex testing, etc.
Reports	Definition of the names, content, and formatting of reports
Event types	Inpatient, outpatient, and other designations
Alternate facilities	Other health care providers
Species	Codes for veterinary specimens
Sex	Codes for gender
Performing laboratories	Other laboratories from which data may be entered or received

LIS, laboratory information system; EHR, electronic health record

is most commonly carried out by identifying a set of test data that includes representative data elements (textual and numerical results, categorical results, result flags, comments) and then documenting medical director review and approval of the data display (obtained as screen prints or report printouts) of the "downstream" systems. Though a receiving system could transmit data to another system, the display validation is required only for the first downstream system that would reasonably present the data to a clinician for patient care decisions.

Each separate system must be validated individually; however, if several sites use a common display system, the system need not be validated for each site.

Other Systems

LIS staff may also have responsibilities outside the laboratory. Laboratories that offer outreach programs may provide software that allows outside laboratory users to order tests and receive results electronically. Such software may support interfaces between the LIS and

external systems such as a physician office's EHR, and/or it may support a user interface for manual order entry and results review. Outreach systems may be installed locally and managed by LIS staff, or the LIS staff may manage a contract with an ASP vendor to provide these services using a VPN connection to the LIS. LIS staff may also manage other systems that operate in conjunction with the LIS, including positive patient ID systems that use armband barcodes to identify patients for specimen procurement, dictation or voice recognition systems that support pathology report creation, and point-of-care testing systems that require data capture and processing from base stations for laboratory comparisons and quality control. Depending on the site, the LIS staff may provide application and device support for these systems in addition to managing a system interface to the LIS.

Record Retention

Patient records and laboratory documentation must be accessible for periods of time that are defined in the CLIA regulations; these regulations apply to both electronic and paper records.^{35,36} Test orders and quality control data should be kept for at least 2 years. Routine laboratory results should be kept for 2 years from the time of reporting. Immunohematology and blood bank results should be kept for 5 years after reporting; records related to products with lifetimes longer than 5 years should be kept for 6 months past their outdate. Anatomic pathology reports should be available for 10 years after the reporting date. Medium-sized laboratories with a modern LIS and hardware can maintain data over these spans, and often longer, in their primary database, allowing immediate access. Very large laboratories that may need to purge data from their main databases before these spans are reached can fulfill this requirement by moving older data to an archival system.

Procedure Manual

The LIS staff is responsible for maintaining a complete, up-to-date procedure manual that covers the routine activities discussed in the previous section as well as downtime and disaster recovery procedures. The manual must be reviewed and approved according to standard clinical laboratory practice and must include the dates of initiation, yearly review, revision, and discontinuation of each procedure. Procedures must be available for 2 years after they are discontinued. The major sections and general topics that should be included in an LIS procedure manual are shown in Table 6-6.

System Evaluation, Selection, and Installation

The evaluation, selection, and installation of an LIS³⁷ is a very challenging task, and the LIS team will generally be deeply involved in or lead any major IT project involving the laboratory. After the EHR, the LIS is usually the second-largest clinical system in a health care enterprise. The LIS is probably also the second-most challenging system to implement, with large installations often requiring 1 to 2 years to complete. Other laboratory information technology projects, such as installing instrument managers, outreach systems, and specialty laboratory systems, are of smaller scope but have analogous management requirements. Because laboratory procedure manuals do not normally include standard procedures for managing software evaluation, selection, and installation, these projects are approached on an individual basis by laboratory leadership. Nevertheless, there are well-described common patterns in successful projects. Some organizations maintain a project management office that employs people who have expertise in large projects and can be a valuable resource in designing a management strategy, helping keep the project on track, or collaborating on project leadership.

The administration, pathologists, and laboratory personnel should articulate the organizational goals and general scope of the project to provide a framework for discussion and identification of leadership. Large projects should have both enterprise and laboratory goals. Using the established goals, the project leadership should identify stakeholder groups who will be affected by the system and gain support from the leadership of these groups and enterprise leadership. As planning moves forward, continue to identify and include stakeholders proactively.

The leadership group should form a project management structure that includes a steering committee with stakeholder representatives and work groups for the major areas that will interact with the system. They should also identify sectional leadership and "champions" (project advocates) from each of these areas and from enterprise leadership. The sectional leadership should define the functional and support requirements for the new system based on current workflow and project goals. Diagramming techniques for use cases and workflow sequences described in the next section of this chapter may be helpful in this task. The primary leadership should identify potential vendors using generally available reference material, peers at other sites, and industry contacts. If open source (community-developed) software will be considered, an internal group or a commercial support vendor may act as the software vendor for the purposes of the evaluation.

Table 6-6. Typical Laboratory Information System (LIS) Procedure Manual Contents

Section	Topics*
General	Description of systems, operating environment, and operating requirements
	Overall change control policies for hardware and software
	System installation and updating requirements for hardware and software
	Personnel training (LIS staff and laboratory users)
	Monitoring and evaluating system and network performance
LIS	System description and vendor contacts
	Description of system functions and their use for laboratory tasks
	System backup and data archiving
	System startup and shutdown
	Downtime procedures (laboratory operation during downtime)
	Software error resolution
	Printer maintenance and error resolution
Interfaces	Installation, testing, acceptance, and periodic re-testing
	For each interface: purpose, data elements, operating conditions, start and stop procedures, maintenance procedures, vendor contacts
Reports	For each report: purpose, design, schedule, run instructions, distribution
Quality assurance	Periodic system monitoring and documentation tasks
	Database verification for accuracy
	Report verification for accuracy and formatting
	Calculation/rule creation and review
Security	User account and password management
	Client computer configuration and software installation
	Procedure for software updates related to security
Associated systems	Topic lists similar to LIS but at a smaller scale, for example, an instrument manager
Disaster recovery	Integrated plan including decision-making framework, initiation, limiting damage, downtime, restoration of hardware, restoration of software and data, startup, verification, entry of downtime data, and resuming operation

* This list is topic oriented, and related procedures have been combined to save space. The organization of the actual manual will vary among laboratories.

The primary leadership group should create a *request for information (RFI)* to be sent to the most promising vendors based on the initial review. An RFI is a document that requests an initial assessment from a vendor of the suitability of their product for a customer's environment and goals. It is used for informational purposes in identifying vendors to review in detail. An RFI will generally contain the following:

- A description of the customer's organization and the current characteristics of the environment into which the new system would go
- The goals of the project and the desired outcome
- An overview of known functional requirements for the software, constraints under which it will operate, and general support requirements
- Criteria for evaluation of software

- Questions about the philosophy, history, size, financial status, and overall stability of the vendor
- Questions about the number and characteristics of other installed sites, and contact information for sites willing to discuss their experiences with the software

An organizational project management or procurement office can be helpful in organizing and writing the RFI. If vendors believe their product is a good match for the environment and requirements, they will respond to an RFI with answers to the questions, a description of how their product would meet the organizational goals, and general descriptions of the installation process and the future working environment with their product.

On-site vendor demonstrations with a limited number of vendors (usually six vendors or fewer) can be scheduled based on the review and ranking of the returned RFIs. These demonstrations give a general feel for how the product supports workflows, but because they are controlled by the vendors, their utility in revealing product quality is limited.

Working group members and other staff who will be system users should visit installed sites with similar complexity to their own to review the system under operational conditions and discuss the sites' experiences with the product and vendor support. Because the quality of software engineering is difficult to measure in feature checklists, descriptions, and vendor-controlled demos, visitors should pay close attention to how well their key functional requirements are met in the installed software. Visited sites can include vendor-recommended references, and some contact with other sites can be beneficial. Appendix 6-1 lists a number of questions that should be answered during a site visit.

A *request for proposal (RFP)* is submitted to the top vendor or vendors after the demonstration evaluation and visit results. RFPs are generally prepared in collaboration with the enterprise procurement office and may include vendor-supplied templates. RFPs may contain similar information at a greater level of detail than the RFI, with the addition of requests for cost, installation, and support proposals from the vendor. The vendor will respond with a commitment to the functional requirements stated and a plan for hardware and software installation and support. The cost proposal may include licensing, installation, travel, training, and maintenance costs plus additional costs such as migration of data from older systems, any required custom development, and interfaces. A vendor selection is made on the basis of the RFPs, followed by contract negotiations that are usually supported by the organizational procurement or business office.

When the contract is complete, the vendor should deliver a detailed implementation plan with an overall

timeline, milestones for customer and vendor deliverables, and a projected "go-live" date. Components of the implementation plan generally include the following:

- Installation, configuration, and testing of new computing and network hardware.
- Software installation and initial testing.
- If old data will be brought into the new system, they must be exported, transformed, checked for consistency, and imported into the new database. Data representations commonly differ between systems, so data from a previous system must be converted to the new representations, and that conversion process must be validated automatically and manually. In LIS conversions, old data are often stored in an accessible backup system rather than being imported into the new lab system.
- Analysis and modification of the local workflow to take advantage of the new system. In addition, systems usually provide some workflow and screen display flexibility, and these local options—for example, the sequence of screens associated with a task, the fields displayed on those screens, default values, and the terms used on tabs and menus—must be defined and configured.
- Maintenance table data collection and loading, for example, personnel, locations, reference laboratories, and other local data.
- Testing, initially of the new system's modules, and then of the full system, using a testing plan provided by the vendor and approved locally.
- Installation and testing of new system interfaces.
- Training for application managers, advanced users, and routine users.
- A go-live plan that provides a period of extra support and contingencies for problems.

Testing is generally carried out with test scripts supplied by the vendor, and the local site may supply data libraries for use in testing. Testing may be done in two stages: (1) unit testing, which evaluates a newly installed software module independently for internal problems; and (2) integration testing, which tests the module in the setting of the whole system to catch problems in data transfer to and from the module. Testing, validation, and approval of the system and its interfaces must conform to established laboratory procedures.

Vendors generally provide three levels of training. Administrator or system manager training is designed for the LIS staff who will manage the application. This is usually offered relatively early in the install process so that the staff can contribute to data loading and configuration. Superuser training is often offered about 6 weeks before go-live. Superusers are individuals chosen from key user roles who will become very knowledgeable about the system and will later serve as resources

or trainers for routine users. Standard user training is offered about 2 weeks prior to go-live, and superusers may participate as instructors or training aides. If user training is carried out too early, users will not retain the training information at go-live.

The new system is activated for production use at go-live, and old systems may be deactivated. Go-live is challenging and optimally is done at a time with lower workload. Extra staff and support personnel should be available to help work out problems, and contingency plans should be in place in case go-live cannot be completed successfully. After go-live, there is generally an extended period of reduced productivity until users gain familiarity with the system. Issues may appear with workflow or system configuration that require adjustment. As problems are resolved and the system becomes more familiar, the LIS staff will transition to routine maintenance and enhancement.

Performance Monitoring and Quality Assurance

LIS staff monitor the performance of the LIS primarily through routine quality measurements and regression testing. Routine measurements of LIS and network performance include:

- Network throughput, collision rate, and downtime
- Scheduled and unscheduled system downtime, total incidents, total time, and time per incident
- Interface transfer rates
- Stored data volumes and available disc space
- Total accounts used and available licensed accounts
- Resolution of error messages and error log entries for systems and interfaces

Benchmarks and response thresholds for these values are established by individual laboratories based on the local environment and are included in the LIS quality assurance procedure (Table 6-6). Some sites capture issues and report the problems to a separate issue-tracking system, which allows the enumeration and classification of issues for follow-up, resolution, and monitoring. The LIS may also be configured to identify and report cases that are useful for general laboratory quality assurance follow-up, such as values that are absurd or out of reportable ranges, and apparent duplicate tests.

Regression testing is the re-testing of previously tested functions to ensure that interim changes or upgrades have not introduced errors into data processing or transmission. Regression testing generally requires a test data library that contains a variety of data elements for which the correct system output has been established. Alternatively, it may be convenient when performing regression testing to identify current data with the necessary elements and compare output with expected behavior.

Output is generally reviewed in printed reports or screen shots, signed on approval, and filed.

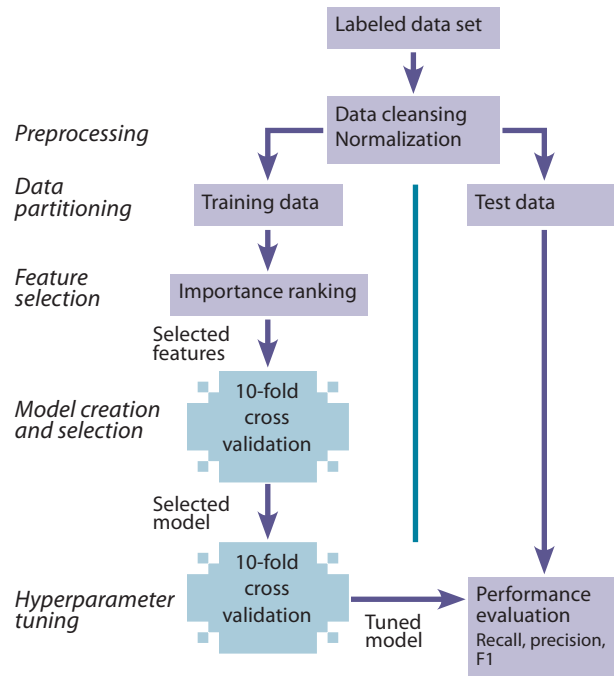
Regression testing is carried out biennially for calculations and executable rules, for data transmission across interfaces, for reports, and for the display of data in interfaced systems. The performance of calculations, interfaces, and reports are evaluated by comparison to expected performance using a test data set. Data display tests usually require the review of printed screen shots from the interfaced system and must include two examples of surgical pathology reports, cytopathology reports, clinical laboratory textual reports, quantitative results, categorical results, microbiology reports, and blood bank reports. The reports should include examples of all data elements and flags as well as corrected results. The results of testing should be available for 2 years past the life of the system tested.

Laboratory Analytics and Computational Pathology

LISs typically provide basic tools and reports for monitoring laboratory operations, but their built-in analytics capabilities are limited. An increasing number of laboratories are experimenting with new techniques for analyzing large and complex data sets using special-purpose systems and software.^{38,39} This expansion of analytics is driven partly by the computational demands of high-throughput genomics including next-generation sequencing (NGS) and partly by the potential for machine learning and related techniques to analyze, classify, and make useful predictions from detailed clinical, operational, and imaging data. The application of these techniques in the pathology domain is known as *computational pathology*.⁴⁰

NGS is computationally demanding, generates large amounts of data, and requires staff with specialized knowledge.⁴¹ Its raw sequence and data quality files may occupy hundreds of gigabytes (GB) per specimen, depending on the details of the analysis, and require a high-performance computer or computing cluster for processing. Data analysis is carried out via a processing pipeline that uses a chain of software tools to align millions of overlapping DNA fragments to each other and to a reference sequence. Variants from the reference are cataloged in a variant call file (VCF), which may occupy several GB. The VCF is interpreted to yield a limited number of clinically relevant variants that are documented in the LIS and reported to physicians. Some NGS instrument vendors provide data processing as a service, which reduces the requirement for local hardware and technical expertise but necessitates secure transfer of high volumes of data between the laboratory and the vendor. Currently, best practices for retention of raw and intermediate data files are controversial. Maintenance of

Figure 6-8. Supervised machine learning workflow. The starting data set contains instances, such as cases, that are correctly classified (eg, by diagnosis, outcome, or cost category). The goal is to create a machine learning model that can accurately classify new instances based on learned patterns in their data elements. The data is preprocessed to correct erroneous and missing values, and to normalize the data if necessary. The data is then partitioned into a training set and test set, with similar incidences of the class labels. Feature selection on the training data ranks the likely importance of the individual data elements for the classification task, using one or more methods based on correlation with the class labels or impact on the performance of basic machine learning algorithms. The most important features are chosen to train machine learning models. Several models based on different algorithms may be trained, with the best-performing chosen for optimization. Model training may use stratified cross-validation, in which the data is divided into similar-sized subsets with equal incidences of class labels. In 10-fold cross-validation, the model is trained repeatedly against 90% of the data, with a different 10% of the data held out for performance evaluation each time. The overall model performance is the mean of the 10 performance evaluations. The best performing model is chosen for optimization by hyperparameter tuning. Hyperparameters are human-defined, algorithm-dependent values that may represent starting or finishing conditions for learning, or overall constraints on the learning process, and they may be optimized using a second round of 10-fold cross-validation. The final optimized model is carried forward to performance evaluation against the held-out test data. Performance metrics for machine learning models often include recall (equivalent to sensitivity), precision (equivalent to positive predictive value), and the F1 score (a weighted average of precision and recall).



very large files in an accessible form for extended periods is expensive, and as the cost of sequencing declines it may become more cost effective to reanalyze samples as needed rather than maintain the original data.

Laboratories are also beginning to use advanced analytics with clinical and operational data.^{42,43} These techniques are derived from “big data” analyses in other domains, ie, analysis of data sets that may be large, complex, heterogeneous, and/or dynamic. These characteristics, which are common in the real world, make data challenging to analyze by traditional statistical approaches commonly used in controlled experimental settings. In the laboratory and in medicine overall, even data sets that are not extraordinarily large may be quite complex and heterogeneous, with each subject (for example, patient) characterized by many data elements with many possible relationships and frequent gaps in the data. Some of the most promising advanced analytic techniques for use in health care are based on *machine learning*.⁴⁴ Machine learning has been used with clinical laboratory data to improve the simultaneous interpretation of multiple laboratory tests,³⁹ predict the values of particular tests based on other test values,⁴⁵ identify and classify biomarkers,⁴⁶ and forecast demand for laboratory testing services.⁴⁷ Machine learning is most useful in settings where instances within data sets have many different features or dimensions, for example, patients

with many different laboratory test results, demographics, diagnoses, and other characteristics. The machine learning task is to identify automatically the important features and use them to recognize groups of instances with similar patterns of features (unsupervised learning) or to classify/predict a new instance’s membership in known groups or “classes” (supervised learning). Recognizing similar groups or clusters in the data may yield new knowledge about the characteristics or behavior of a population or workflow whereas classification and prediction may be useful in resource planning, disease progression monitoring, decision support, or outcomes prediction.

Supervised learning algorithms are “trained” on a data set in which the classes are known and their labels are disclosed to the algorithm, and validated against a separate data set in which the labels are withheld (the training and test data sets, respectively, Figure 6-8). During training, the algorithm adjusts its internal parameters progressively to weight the values of the various input dimensions so that the output label for an instance is correct. The trained algorithm with its weighted parameters is referred to as a *model*. After training, the model is validated by assessing its accuracy in classifying or predicting instances in the test data set. Supervised learning models are subject to overfitting, in which the model may excessively weight irrelevant features that happen

to be correlated with a label in a particular training data set. The risk of overfitting can be reduced by limiting the features to those likely or demonstrated to be important (feature selection or dimensionality reduction) and by cross-validation (Figure 6-8) in which training/validation is carried out repeatedly with the data divided differently into training and test sets each time, and performance is optimized across all trials. There are multiple unsupervised and supervised machine learning algorithms that may perform differently based on the characteristics of a data set, and therefore machine learning projects often test the performance of several algorithms before settling on an optimal approach. The accuracy of supervised learning models is subject to the quality of the training and test data sets, and also the degree to which those data match the data that the model will use when it is deployed. Some systems using pre-trained machine learning models are sold for implementation in settings different from those in which the model was developed, with guidance that the model should be validated on local data. When this is done, it is important to (1) use a validation data set large enough to cover most data patterns expected and (2) make sure that the characteristics of the individual local data elements (dimensions) match those of the training and test data sets. If data elements behave differently from the training data, the performance of a machine learning model may be difficult to predict.

Effective application of machine learning and related techniques requires specialized knowledge. The recent broad interest in these techniques across multiple domains has stimulated growth of training programs in *data science*, an interdisciplinary field focusing on techniques for extracting knowledge from data. Data science differs from traditional statistics by having an increased emphasis on the characteristics of heterogeneous data, digital computational techniques, and machine learning. Many larger health care providers have established analytics units housing staff with advanced training in data science. These staff can be valuable collaborators with the laboratory in applying advanced analytics to laboratory problems. Larger laboratories with ambitions in analytics may consider adding this skill set to their LIS team.

Software Tools and Local Software Development

LIS groups may effectively extend the LIS by using additional software and programming tools to process reports or data extracts to offer greater analytic, data presentation, or automation capabilities. Spreadsheets are useful for the processing, formatting, and simple statistical analysis of tabular data. Spreadsheet plug-ins may offer additional capabilities useful in the laboratory,

such as ROC analysis.⁴⁸ The most widely used software tools for “big data” analysis and machine learning are R⁴⁹ and Python.⁵⁰ Both are open source (available at little or no cost) and provide convenient working environments, extensive reference material, and code libraries that support routine and advanced statistics as well as general-purpose and genomic data processing, machine learning, and data visualization. Commercial software packages designed for specialized clinical laboratory data analyses are also available.⁵¹⁻⁵² Some laboratories have written libraries of their own software to enhance their LIS or anatomic pathology systems; programming in MUMPS, Microsoft’s Visual Basic, or .Net⁵³; or programming in cross-platform scripting languages such as Perl, Ruby, or Python (which is a popular general programming language in addition to its use for data analysis).⁵⁴ Microsoft’s development tools and Python are approachable and provide large libraries of prewritten code that enable self-taught part-time programmers to create limited but very functional programs for local use. Commercial tools or locally developed software that become part of the routine laboratory operation have maintenance and change control requirements similar to those of the LIS, but if used judiciously, they can provide analysis capabilities and workflow improvements that justify the maintenance effort.

Diagramming Tools for Requirements Analysis and Workflow Redesign

Requirements and workflow analyses are important parts of planning the installation of a new system and improving existing systems. To effectively support a workflow with an information system, pathologists must understand the key tasks and actors in the workflow, their association with each other, and the data elements that are captured or communicated as part of that workflow. Documenting this information accurately is very challenging because it requires collaboration between individuals who normally play different roles in different work domains and often have different vocabularies and assumptions. These communication problems were the impetus for the development of the *Unified Modeling Language (UML)*,⁵⁵ a standardized set of diagramming techniques designed for collaborative documentation and cross-domain communication of work processes and information system design concepts. Workflows and information models captured in UML diagrams clarify complex real-world environments and relate directly to information system requirements and design elements.

The most useful diagrams for clinical laboratory settings are use case models, activity diagrams, and class diagrams. *Use case models* (Figure 6-9) are very simple and are designed to clearly depict key tasks and actors (human and technological). They are easy to create

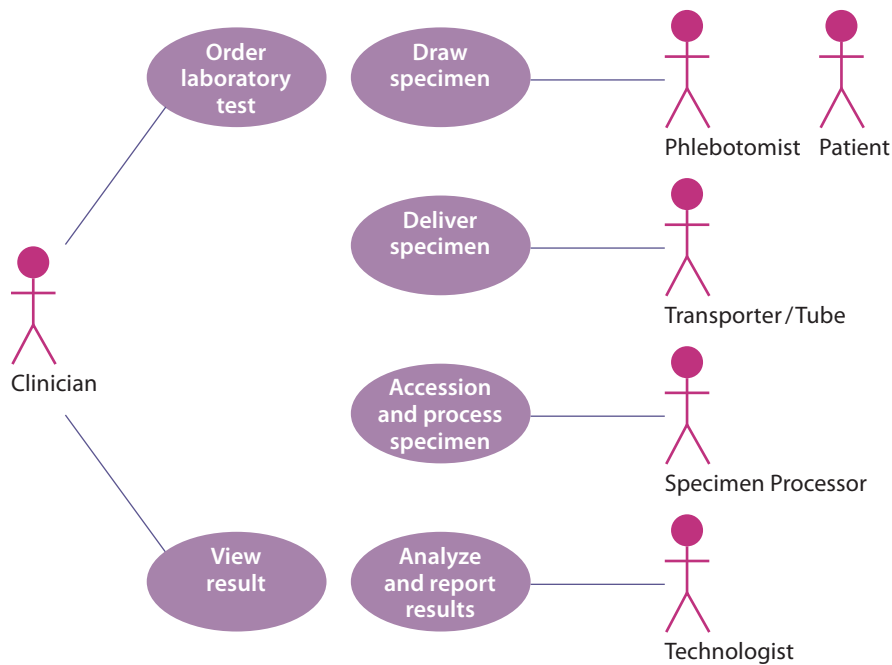


Figure 6-9. Use case diagram of preanalytic, analytic, and post-analytic laboratory workflow. Key tasks are captured in bubbles, and actors (personnel and technical systems) are shown as stick figures linked to the tasks they perform. Use case tasks may be broad, as shown here, or specific, and analysis typically progresses from broad to specific. Task bubbles are often arranged in sequence, and brief textual descriptions may accompany diagrams.

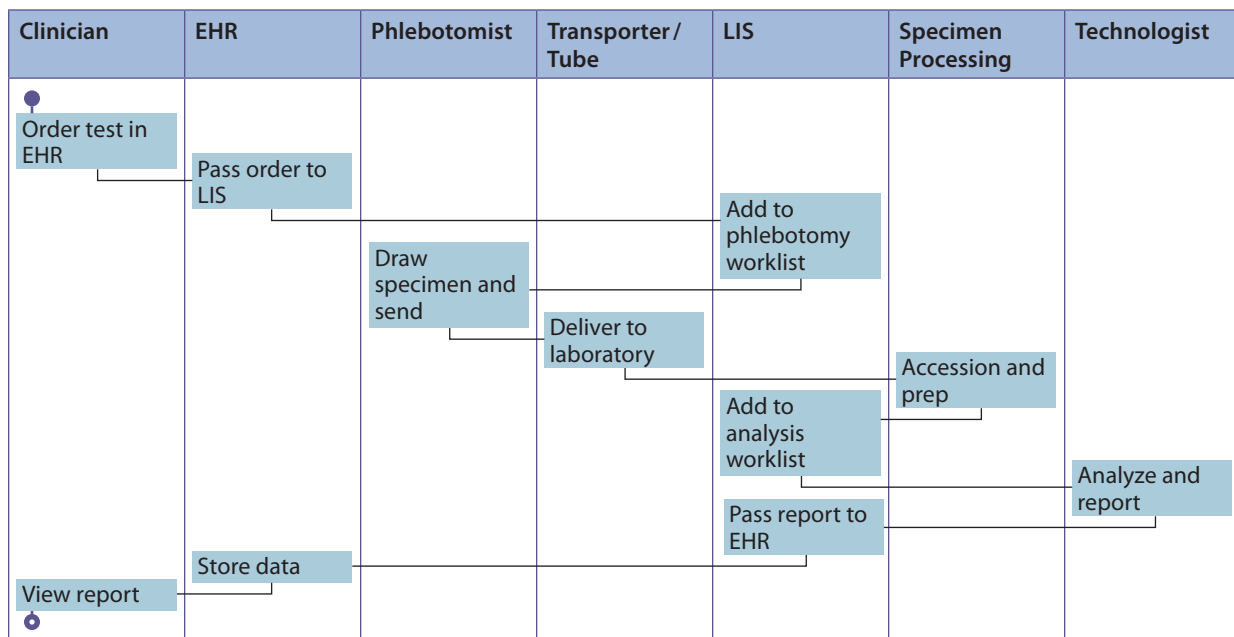


Figure 6-10. Activity diagram of laboratory workflow. The actors and tasks from the use case diagram in Figure 6-9 are mapped in sequence to an extended flow chart, with one lane for each actor. Some additional actors that were not shown in the use cases (electronic health record [EHR] and laboratory information system [LIS]) are included here to complete the information flow. Activity diagrams show the flow of tasks and information both in time and across participants and systems.

during a discussion, and they may be supplemented with a brief paragraph providing details for each task. All requirements analysis and workflow redesign should be rooted in use cases. *Activity diagrams* (Figure 6-10), also known as *swimlane diagrams*, are flow charts in which each actor has a lane, and the sequence of events moves down the chart and across the lanes to indicate the flow of actions and information between the actors.

Class diagrams (Figure 6-11) display data models that map all the data elements used in a workflow and their relationships with each other. Unlike activity diagrams, class diagrams are static; they do not specify a particular sequence of use of the data elements they depict. The combination of these three diagram types supplemented with some textual description can capture most of the information needed for the discussion and analysis of

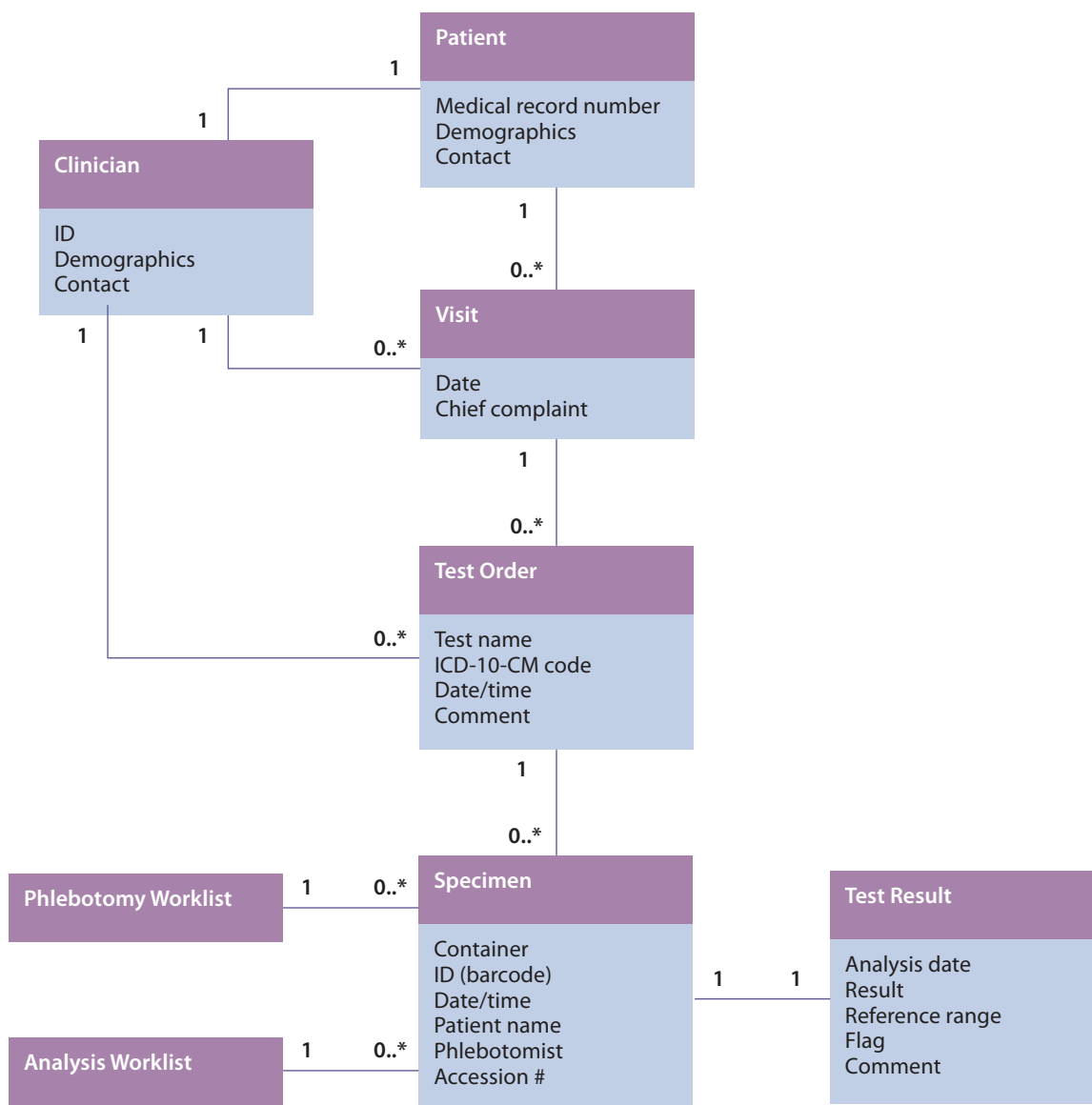


Figure 6-11. Simplified class diagram for laboratory workflow. This diagram shows data elements and their relationships from the simple workflow shown in Figures 6-9 and 6-10. The boxes represent classes (entities for which data will be collected). Items within the boxes are simple data elements, sometimes called *attributes*. Classes can also contain other classes as complex data elements, and these associations are shown as lines. Multiplicity relationships are shown as numbers on the associations; for example, a patient can have any number of visits (the permissible range is shown as 0..*, where the asterisk indicates “any number”), but a visit can have only one patient. The clinician class is used in several settings; it represents the patient’s primary physician, and it also represents the physician of record for a visit and the ordering physician for a test. These may be different physicians, but the clinician class provides a data model for all of them. Class diagrams express data models only and do not depict data flow or action sequences.

information flow in laboratory work processes. Most technical diagramming software includes templates for UML diagrams, and dedicated UML diagramming software is available commercially and in open source.

Discussion

LISs were an early and cutting-edge application of computer technology in medicine, and work with LISs

contributed to the development of the broader field of medical informatics. LIS technology has been continually developing for 40 years, and well-managed LISs are now integral to the operation of modern laboratories. Pathology informatics is a subset of medical informatics that seeks to optimize the use of LISs and other pathology computing resources for pathology services, physician decision making, and patient care. Several recent books and monographs have covered aspects of

Case Example

Dr Henry Little is associate laboratory director and medical director of the LIS at a 400-bed hospital that has been actively pursuing relationships with regional clinics and small hospitals. A clinic about 30 miles away is forming a relationship with the hospital and is interested in sending testing to Dr Little's laboratory but would like to be able to order tests and review results within its EHR.

What are the options for supporting the clinic?

Comments: There are several possibilities. The simplest, which could be appropriate for a physician's office or small clinic, would be to establish a broadband internet connection such as a DSL line to the remote site, configure one or more of their office computers to establish a VPN connection with the LIS over the internet, provide appropriate clinic staff with LIS accounts, and allow them to order tests and review results directly in the LIS user interface. The LIS should be configured to recognize users from that location and allow them to view results only on patients from that location. This approach would not satisfy the desire to use the clinic's EHR, and adding accounts to the LIS could require additional licensing fees that would need to be balanced against the anticipated test revenue.

Another approach would be to establish an HL7 interface running over a secure connection (similar to a VPN) between the LIS and the clinic's EHR. This would require creating and validating an HL7 interface with the participation of both systems' vendors, and Dr Little's LIS group would probably spend significant time working with the remote site. However, this could also provide access to systems other than the LIS,

which might be a benefit depending on the closeness of the business relationship, and some of the cost thus might be picked up by other parts of the organization.

A third option would be to contract with a company that acts as an HL7 interface aggregator, ie, one that connects to systems at different sites and passes HL7 messages between them. Such a company would work directly with the clinic site and the LIS group to help set up interfaces to their locations, and the overall interface development time and cost might be decreased if the company has previously developed interfaces to the systems used by the clinic and the laboratory.

Finally, some reference laboratories offer a service similar to the HL7 aggregator companies at favorable pricing, if esoteric testing from the clinic will be sent to the reference laboratory. In this case, the reference laboratory establishes the interface with the clinic's system. If the hospital laboratory has an existing interface with the reference laboratory, it can be used to receive information on tests being sent from the clinic to the hospital laboratory and return results to the clinic through the reference laboratory system. This can be a good option if both the clinic and the hospital laboratory are willing to use the reference laboratory for esoteric testing.

Dr Little should remember that laboratories are not allowed to offer material inducements or "kickbacks" to gain testing business and thus should be careful about, for example, low- or no-cost placement of computers and other communications equipment in the clinic. He should also consider that reporting directly to the clinic system means that his LIS group will need to do regression testing and validation of the interface and the data display in the clinic every 2 years.

pathology informatics in detail.⁵⁶⁻⁵⁸ Concepts from the pathology informatics community—for example, the necessary role pathologists have in ensuring that pathology information is presented to the clinician accurately and in a useful form—have become part of routine laboratory operation and accreditation requirements.

Most current pathology system designs are based on stand-alone systems that printed reports on paper for delivery to clinicians and for monitoring of laboratory operations. This previous environment allowed each system to define its own data representation, data models, and data presentation strategy. Without an incentive to standardize, vendors and local sites developed independent approaches to solve laboratory computer problems. As LISs and other systems became integrated into enterprise architectures (Figure 6-6), these once-independent systems required complex, expensive, and error-prone interfaces for translating their various differences. Accurately translating complex formatted reports between

systems remains an unsolved problem. The third decade of the century will see a national push to make recently deployed EHR and other health systems truly interoperable; goals will include transferring clinical summaries containing actionable information between health care systems, enabling decision support across all data in clinical workflow systems such as EHRs, and aggregating large data sets across systems for quality assurance and process optimization, population health research and surveillance, postmarketing studies for therapeutics and devices, and comparative effectiveness research.

The next major evolutionary step for LIS is achieving syntactic and semantic interoperability between pathology information systems and health care enterprise systems based on shared data representations and data models. Such interoperability will substantially reduce the cost and effort of creating and maintaining system interfaces and will increase the reliability of these interfaces. It will also allow pathology data of all types

Table 6-7. Data Standards Working Groups Pertinent to Laboratory Information Systems and Anatomic Pathology

Group and Contact	Product/Activity
CAP Informatics Committee http://www.cap.org/	Monitoring and promoting standards development and general informatics in anatomic and clinical pathology
DICOM Working Group 26 (Pathology) http://medical.nema.org/DICOM/minutes/WG-26/	Imaging and image data standards for whole slide digitization
HL7 Clinical Genomics Working Group http://wiki.hl7.org/	Communication standard for genomics testing results, family history, and disease risk calculations, using genomic and other terminology standards
HL7 Orders and Observations Working Group http://wiki.hl7.org/	Communication standards for anatomic and clinical pathology orders and results, using terminology standards
IHE Pathology and Laboratory Medicine (PaLM) Domain* http://wiki.ihe.net/	Profiles** for anatomic pathology reporting, and clinical laboratory ordering and reporting based on terminology and communication standards
SNOMED International http://www.snomed.org/	SNOMED-CT medical and pathology terminology system
Logical Observation Identifiers Names and Codes (LOINC) http://loinc.org/	Standard terms for naming laboratory tests and other observations

Definitions: CAP, College of American Pathologists; DICOM, Digital Imaging and Communications in Medicine; HL7, Health Level Seven International; IHE, Integrating the Healthcare Enterprise International.

* IHE Technical Frameworks for laboratory and anatomic pathology are available at http://www.ihe.net/Technical_Framework/.

** Profiles are sets of existing standards that can be used to accomplish a particular information systems goal in the health care environment.

to be accurately communicated and clearly presented in systems that support clinical workflow, and will enable those systems to more effectively incorporate clinical decision support. Data standards are available, but they are not yet complete or organized into generally accepted solutions to data-sharing problems. A number of working groups are actively developing and extending pathology-related standards in a collaborative effort that creates pathways for standards implementation by vendors (Table 6-7).⁵⁹ For example, groups within HL7 are developing data models and messaging syntax that incorporate existing terminologies such as LOINC and SNOMED. Representations of pathology reports that take advantage of the HL7 Clinical Document Architecture are being developed to carry both formatted text and processable data elements in sharable forms. The Integrating Healthcare Enterprise (IHE) PaLM is gathering these standards into profiles (groups of standards used in particular ways) that accomplish real-world communication tasks.⁶⁰ Multiple system vendors participating in IHE will be able to implement these profiles simultaneously, with confidence that a critical mass of systems using the standards is forthcoming.

These changes will ultimately simplify the operation of LISs and anatomic pathology systems, and allow pathology services to contribute more actively to patient

care through more flexible data display, increased decision support options, and systems that are able to respond in useful ways to test ordering and result patterns. Further development of pathology systems will include more sophisticated data analyses and data mining capabilities that monitor routine data to provide useful and actionable information, perhaps through incorporation of widely used open-source analytics tools. These capabilities will enhance pathology services' ability to tailor their offerings to clinical needs and individual patient characteristics, and will improve analysis and decision-making in the management of pathology services. There will be demands and challenges associated with necessary changes in systems and work processes over the next decade, but there is also great potential for progress if these challenges are met.

References

1. Johnson JL. *Clinical Laboratory Computer Systems. A Comprehensive Evaluation*. Northbrook, IL: Lloyd Johnson Associates; 1971.
2. Johnson JL. *Achieving the Optimal Information System for the Clinical Laboratory*. Northbrook, IL: Lloyd Johnson Associates; 1975.
3. Aller RD, Robboy SJ, Poitras JW, et al. Computer-assisted pathology encoding and reporting system (CAPER). *Am J Clin Pathol*. 1977;68(6):715-720.

Case Example

Dr Little's hospital recently finished installing a new EHR system. About 6 weeks after go-live, the laboratory received a call from a clinician questioning several calcium values over 20 mg/dL. When the laboratory staff checked the results in the LIS, the values were normal, or in some cases no calcium assay had been done, but the clinician insisted that the high results were present in the EHR display. The laboratory staff could not review the display because the role-based security of the EHR had been configured so that only staff caring for a patient could see data on that patient. Laboratory staff members were not regarded as patient care staff and thus did not have access to the EHR data display. All interface validation had been completed with no problems before the EHR go-live. Further investigation revealed that a calculation rule had been implemented in the EHR to correct total phenytoin levels for albumin, based on clinician requests and without the knowledge of the laboratory, and the result of that calculation was being inserted erroneously into the calcium result display. The calculation was turned off.

Comments: This case raises several important issues. The most important is that a rule was created that produced a result, and the performance of the rule was not adequately validated. It is tempting to speculate that one reason for this outcome was that the rule was created by hospital computing personnel who were not

familiar with patient data calculations and their validation. Though there are well established procedures in pathology for validating calculations with appropriate test data sets, the advent of EHR with rules engines that are managed outside of pathology raise the possibility of calculations and rules using pathology data that are not managed according to pathology standards. A case could be made that pathologists should participate in the implementation and review of any rule or calculation in a clinical system that uses pathology data and produces an actionable result. In this case, the hospital committee in charge of EHR rules previously did not include a pathologist; this incident yielded an invitation to Dr Little to join the committee.

The inability of laboratory staff to review the data display is a second issue. Practically speaking, when problems are perceived in the display of pathology data, the pathology service will receive the call for assistance even though the display system may not be under pathology management. To resolve issues rapidly in support of quality patient care, the pathology service should be able to see the same display that the clinician sees and compare its content to their service records. This is also true for any other data-producing clinical service. Security of patient data is critically important, but it should be handled through methods, including appropriate training and access monitoring, that do not create barriers to problem resolution.

4. Foulis PR, Norbut AM, Mendelow H, Kessler GF. Pathology accessioning and retrieval system with encoding by computer (PARSEC): a microcomputer-based system for anatomic pathology featuring automated SNOP coding and multiple administrative functions. *Am J Clin Pathol*. 1980;73(6):748-753.
5. Pearson JM. A comprehensive software subsystem for anatomic pathology and cytology. *Proc Ann Symp Comput Appl Med Care*. 1982;2:274-281.
6. Lincoln TL, Korpman RA. Computers, health care, and medical information science. *Science*. 1980;210(4467):257-263.
7. Korpman RA. Using the computer to optimize human performance in health care delivery: the pathologist as medical information specialist. *Arch Pathol Lab Med*. 1987;111(7):637-645.
8. Friedman BA. The laboratory information system as a tool for implementing a strategic plan. *Am J Clin Pathol*. 1989;92(4 Suppl 1):S38-43.
9. Friedman BA. Informatics as a separate section within a department of pathology. *Am J Clin Pathol*. 1990;94(4 Suppl 1):S2-6.
10. Buffone GJ, Beck JR. Informatics: a subspecialty in pathology. *Am J Clin Pathol*. 1993;100(1):75-81.
11. Cornet R, de Keizer N. Forty years of SNOMED: a literature review. *BMC Med Inform Decis Mak*. 2008;8(Suppl 1):S2.
12. McDonald CJ, Huff SM, Suico JG, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clin Chem*. 2003;49(4):624-633.
13. Valenstein PN. Formatting pathology reports: applying four design principles to improve communication and patient safety. *Arch Pathol Lab Med*. 2008;132(1):84-94.
14. Dunn BE, Choi H, Recla DL, Kerr SE, Wagenman BL. Robotic surgical telepathology between the Iron Mountain and Milwaukee Department of Veterans Affairs Medical Centers: a 12-year experience. *Hum Pathol*. 2009;40(8):1092-1099.
15. Horbinski C, Wiley CA. Comparison of telepathology systems in neuropathological intraoperative consultations. *Neuropathology*. 2009;29(6):655-663.
16. College of American Pathologists. CAP Cancer Protocol Templates. www.cap.org/cancerprotocols. Accessed June 20, 2018.
17. CDC National Center for Health Statistics. International Classification of Diseases, (ICD-10-CM/PCS) Transition – Background. https://www.cdc.gov/nchs/icd/icd10cm_pcs_background.htm. Published 2015. Accessed December 20, 2017.
18. Regenstrief Institute. LOINC: Logical Observation Identifiers Names and Codes. loinc.org. Accessed June 20, 2018.
19. SNOMED International. What is SNOMED CT?. Available at: <https://www.snomed.org/snomed-ct/what-is-snomed-ct>. Accessed December 20, 2017.
20. Henricks WH. Laboratory information systems. *Clin Lab Med*. 2016;36(1):1-11.
21. Le Bozec C, Henin D, Fabiani B, et al. Refining DICOM for pathology: progress from the IHE and DICOM pathology working groups. *Stud Health Technol Inform*. 2007;129(Pt 1):434-438.
22. Lyman JA, Scully K, Harrison JH Jr. The development of health care data warehouses to support data mining. *Clin Lab Med*. 2008;28(1):55-71.

Challenge Questions

Which of the following statements regarding the HL7 version 2 standard is false?

- A. It is a standard that supports data exchange between medical devices including information systems and laboratory analyzers.
- B. The standard defines the structure of text messages.
- C. The standard specifies standard terminologies for use in all message fields.
- D. HL7 version 2 interfaces usually require mapping tables to support data conversion between systems.
- E. HL7 version 2 messages carry the character content of textual reports but not the page formatting.

Answer: C

Which of the following are requirements for laboratory accreditation?

- A. Validation of data display in the first downstream system used for clinical decision making.
- B. Maintenance of routine test results and quality control for 2 years.
- C. Regression testing of interface data transmission every 2 years.
- D. Verification of calculation results and executable rules every 2 years.
- E. All of the above.

Answer: E

Which of the following is not part of the usual responsibilities of the LIS staff?

- A. Regression testing.
- B. Validation of data display in clinical systems interfaced to the LIS.
- C. Monitoring interface transfer rates.
- D. Carrying out method linearity and reportable range checks.
- E. Resolution of error messages.

Answer: D

23. Krasowski MD, Schriever A, Mathur G, Blau JL, Stauffer SL, Ford BA. Use of a data warehouse at an academic medical center for clinical pathology quality improvement, education, and research. *J Pathol Inform.* 2015;6(1):45.
24. Brown DE. Introduction to data mining for medical informatics. *Clin Lab Med.* 2008;28(1):9-35.
25. Aller RD. The clinical laboratory data warehouse: an overlooked diamond mine. *Am J Clin Pathol.* 2003;120:817-819.
26. Brossette SE, Hymel PA Jr. Data mining and infection control. *Clin Lab Med.* 2008;28(1):119-126.
27. McDonald JM, Brossette S, Moser SA. Pathology information systems: data mining leads to knowledge discovery. *Arch Pathol Lab Med.* 1998;122(5):409-411.
28. Wilkerson ML, Henricks WH, Castellani WJ, Whitsitt MS, Sinard JH. Management of laboratory data and information exchange in the electronic health record. *Arch Pathol Lab Med.* 2015;139(3):319-327.
29. W3C. XML Technology. <http://www.w3.org/standards/xml/>. Accessed June 20, 2018.
30. Dolin RH, Alschuler L, Boyer S, et al. HL7 Clinical Document Architecture, Release 2. *J Am Med Inform Assoc.* 2006;13(1):30-39.
31. Ferranti JM, Musser RC, Kawamoto K, Hammond WE. The clinical document architecture and the continuity of care record: a critical analysis. *J Am Med Inform Assoc.* 2006;13(3):245-252.
32. HL7. Product Brief - CCD - Continuity of Care Document. http://wiki.hl7.org/index.php?title=Product_CCD. Accessed June 20, 2018.
33. Harthun K. What's your system's survival time? TechTarget. IT Knowledge Exchange. <http://itknowledgeexchange.techtarget.com/security-corner/whats-your-systems-survival-time/>. Accessed August 18, 2010.
34. Burr WE, Dodson DF, Polk WT. Electronic authentication guideline. NIST Special Publication 800-63 [online]. 2006. http://www.csrc.nist.gov/publications/nistpubs/800-63/SP800-63V1_0_2.pdf. Accessed September 11, 2010.
35. CLIA. Records and Reports. 21 CFR Part 606, Subpart I (606.160). December 2005.
36. CLIA. Standard: Retention requirements. 42 CFR Part 493, Subpart J (493.1105). October 2007.
37. Kaplan KJ, Rao LKF. Selection and implementation of new information systems. *Clin Lab Med.* 2016;36(1):113-131.
38. Analyze-it Software L. Analyze-it Method Evaluation Edition. <http://www.analyze-it.com/products/method%5Fevaluation/>. Accessed June 20, 2018.
39. Richardson A, Signor BM, Lidbury BA, Badrick T. Clinical chemistry in higher dimensions: machine-learning and enhanced prediction from routine clinical chemistry data. *Clin Biochem.* 2016;49(16-17):1213-1220.
40. Louis DN, Gerber GK, Baron JM, et al. Computational pathology: an emerging definition. *Arch Pathol Lab Med.* 2014;138(9):1133-1138.

41. Yohe S, Thyagarajan B. Review of clinical next-generation sequencing. *Arch Pathol Lab Med*. 2017;141(11):1544-1557.
42. Shirts B, Jackson B, Baird G, et al. Clinical laboratory analytics: challenges and promise for an emerging discipline. *J Pathol Inform*. 2015;6(1):9.
43. Tolan NV, Parnas ML, Baudhuin LM, et al. "Big data" in laboratory medicine. *Clin Chem*. 2015;61(12):1433-1440.
44. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science*. 2015;349(6245):255-260.
45. Luo Y, Szolovits P, Dighe AS, Baron JM. Using machine learning to predict laboratory test results. *Am J Clin Pathol*. 2016;145(6):778-788.
46. Swan AL, Mobasher A, Allaway D, Liddell S, Bacardit J. Application of machine learning to proteomics data: classification and biomarker identification in postgenomics biology. *OMICS*. 2013;17(12):595-610.
47. Mohammad EA, Naugler C. Open-source software for demand forecasting of clinical laboratory test volumes using time-series analysis. *J Pathol Inform*. 2017;8:7:1-10.
48. ROC.KIT. Formatio Reticularis. <http://www.formatio-reticularis.de/roc/>. Published 2006. Accessed June 20, 2018.
49. The R Foundation. The R Project for Statistical Computing. <https://www.r-project.org>. Accessed January 3, 2018.
50. The Python Software Foundation. Python. <https://www.python.org>. Accessed January 3, 2018.
51. Visiun Inc. Visiun - The leader in laboratory analytics. www.visiun.com. Accessed August 23, 2018.
52. Data Innovations, Inc. EP Evaluator. <https://www.datainnovations.com/ep-evaluator>. Accessed June 20, 2018.
53. Microsoft Corp. Visual Basic developer center. <http://msdn.microsoft.com/en-us/vbasic/default.aspx>. Accessed June 20, 2018.
54. Berman JJ. *Methods in Medical Informatics. Fundamentals of Healthcare Programming in Perl, Python, and Ruby*. New York: CRC Press/Taylor and Francis Group; 2010.
55. Unified Modeling Language. UML Resource Page. uml.org. Accessed June 20, 2018.
56. Cowan DE. *Informatics for the Clinical Laboratory. A Practical Guide*. New York, NY: Springer; 2003.
57. Pantanowitz L, Henricks WH, Beckwith BA. Medical laboratory informatics. *Clin Lab Med*. 2007;27(4):823-843.
58. Sinard JH. *Practical Pathology Informatics. Demystifying Informatics for the Practicing Anatomic Pathologist*. New York, NY: Springer Science+Business Media; 2006.
59. Daniel C, García Rojo M, Bourquard K, et al. Standards to support information systems integration in anatomic pathology. *Arch Pathol Lab Med*. 2009;133(11):1841-1849.
60. Integrating the Healthcare Enterprise. www.ihe.net. Accessed May 2, 2018.

Acknowledgements

The author thanks James C. Boyd, MD; Lynn Hamilton; Alexis B. Carter, MD; Carol Beth Casto, MT; Paul G. Catrou, MD; and Walter H. Henricks, MD for helpful reference materials and comments.

Appendix 6-1.**Site Visit Questions for LIS Acquisition**

A site visit is crucial to making a good decision on an LIS and is part of standard due diligence prior to selecting a system for installation. An LIS is a complex, engineered system and, like other engineered systems such as buildings and bridges, the quality of its engineering is better indicated by performance over time than feature lists and superficial appearance. Ideally, a site that is similar in complexity and other characteristics to the planned install site should be visited. Vendors generally recommend sites for visiting that are having good experiences with the product. It may be beneficial to visit or at least call an additional site “off the list” if one is available. A site visit team should include representatives from all major laboratory areas and administration, and team members should visit each of these areas and watch the system in operation. It is important to speak with the people in the laboratory who are actually using the system on a daily basis, as well as the pathologist, administrators, or hospital CIO who were involved in the acquisition.

Questions to consider:

- Was the vendor forthright in presenting the system’s capabilities and its ability to meet the goals established in the site’s RFI/RFP? Does the system operate as expected?
- Were there any unexpected outcomes or “gotchas” related to installation or use?
- Did the vendor’s system installation and training plan and support services meet the needs of the site?
- How much time did the installation process require?
- How well was go-live handled, and how long after go-live did it take to get comfortable with the system?
- What is the extent of connectivity to other systems (instruments and instrument managers, EMR/HIS/billing, remote locations, reference laboratories), and have there been any problems in developing and maintaining interfaces or interface performance?
- Is the overall responsiveness of the system adequate (user interface responsiveness, query time, printing time)?
- Is the system reasonably efficient to learn and use on a daily basis? Do technologists, supervisors, pathologists, and lab administrators like it (review each subsystem)?
- If appropriate, how well does the system support laboratory automation including specimen transport tracks and storage repositories?
- If appropriate, how well does the system handle multiple laboratories, hospitals, and clinics that may have differing patient identification, ordering, sample handling, and reporting requirements?
- What are the downtime requirements, and how much unscheduled downtime has occurred? How long does recovery from downtime take?
- What are the strategies for high-availability operation, backup, and disaster recovery, and are they compatible with the needs of the visiting team?
- Does the vendor’s ongoing technical support meet the needs of the lab? Is the response to contacts timely? How quickly are issues resolved? Are there any unresolved issues or bugs?
- Does the vendor provide good quality technical and user training materials for ongoing use, such as printed material, online learning modules, or webinars?
- Is there a user group for the software, and is it active and useful?
- What are the strongest and weakest points of the system?
- Are operating costs as projected?
- Does anything at the site suggest that the system will not be able to support the requirements listed in the RFI/RFP of the visiting team?
- Would the site buy or lease the software again? Why or why not?