ABSTRACT

The University of Kentucky (UK) School of Library and Information Sciences proposes a collaboration with the UK Markey Cancer Center, UK College of Health Sciences, and Massachusetts General Hospital to create a standardized metadata framework for pathologic images so that a set of well-described, integrated biomedical imaging information can be efficiently stored, managed, retrieved, and shared. The proposed three-year project, to be conducted from July 1, 2008 through June 30, 2011, is designed to address critically important needs of the biomedical imaging community for metadata tools supporting comprehensive biomedical image libraries. For any types of information-bearing entities, whether texts or images, the foremost step is to represent contained information into any formats of surrogate records, including library catalogs. With the advent of digital imaging in pathology, visual findings related to the diagnoses of disease are increasingly captured and stored in digitized formats. However, the descriptions of these images are not always linked to clinical records, specimen preparation information, and demographic information. This is problematic because there is no standard for a "complete" set of image metadata, i.e., data about data. Moreover, microscope systems do not usually record clinically relevant information such as histologic grading, cells, genes, or other patient follow-up data with the images. Additionally, data sharing or even submission to a journal requires conversion to a simple two-dimensional format, leaving critical metadata disconnected and lost. Project activities in this early career development project involve four phases to collect, merge, create, describe, and evaluate a metadata set for pathologic images. The project team will assess four existing sources of potential metadata identified in preliminary studies and collect relevant data elements; however, since no single system currently provides all the data elements required to adequately describe pathologic images, this review will validate the strengths of disparate existing datasets, determine areas of overlap and duplication, and provide a foundation for the project team to collect, clean, and map potential candidate data elements. Separate files for individual data elements will be created and merged into a single file in Protégé, a free open-source ontology editor and knowledge-base framework. A series of focus group meetings and interviews with domain experts from pathology, ontology and library science, and imaging will be conducted to determine relevant test image descriptions and to construct a standard that can effectively represent imagery information contained in the set. This stage will allow the project team to review and finalize the merged metadata elements and their relationships as well as describe and select the most appropriate describable units for scanned images. Focus groups will also identify and finalize potential queries to be tested based on industry standards for pathologic imaging properties and evaluate data retrieval effectiveness. Expected project outcomes include significant translation of core concepts in information representation, i.e., cataloging, classification, authority and access control, subject analysis, arrangement and display, and vocabulary control which have been developed, standardized, and practiced in libraries, into new and emerging information management needs. In extending librarians' organizational knowledge and skills to a non-traditional collection, pathologic images, the proposed metadata framework offers an innovative model of librarianship to support a novel and emerging need given today's unique datasets in the field of biomedical imaging. Specifically, the project aims to create a metadata framework to better support pathologic imaging description through knowledge representation. This study will contribute significantly to the digital imaging field by merging these existing data standards into one integrated metadata standard to provide a seamless query tool among different pathologic imaging systems.

NARRATIVE

Assessment of Need

Libraries have long served as a central hub of information organization in that librarians play active professional roles in describing various information resources for the dual purposes of efficient knowledge management and information retrieval. Thus, library and information science professionals are uniquely positioned to bring a wealth of organizational knowledge and supporting skill sets to the development and management of non-traditional data collections, a novel and emerging need given the technology infrastructure that supports today's unique datasets. In particular, the core concepts of information representation (cataloging, classification, authority and access control, subject analysis, arrangement and display, and vocabulary control), which have been developed, standardized, and practiced in libraries have significant potential for translation, if appropriately adopted, into other new and emerging disciplines. A critical need for librarianship in medicine, for example, has emerged in medical image description and management and retrieval of affiliated information. For instance, visual findings in pathology are the core of its practice. Advances in biomedical imaging in pathology have meant that well-described digital images are increasingly valuable resources for pathologists, health care practitioners, and biomedical researchers. ***The major goal of the proposed study is to create a standardized metadata framework of pathologic images so that appropriately detailed descriptive information can be stored, retrieved, and shared.*** Within this context, an innovative model of librarianship will result from the activities proposed in this application—one that will support the information management and access needs of users of these non-traditional data collections.

Metadata, data about data, has been developed over the past few decades in many disciplines and has been used as a backbone for integration of various information sources and as a seamless query tool among different information systems (Lambrix et al., 2003). However, there are very few metadata tools for biomedical image libraries. Moreover, most of the biomedical image metadata projects only provide a basic framework rather than a comprehensive and complete data description (Sim, 2004). Clinically relevant information, such as pathologic findings, clinical history, specimen preparation information, and demographic information that is indexed with content, images, and image acquisition information will serve to create a powerful biomedical image database. Currently, there is no standard for a "complete" set of metadata, for example, how an image-capturing system (such as a digital microscope camera) should be described. Moreover, microscope systems do not usually record clinically relevant information such as histologic grading, cells, genes, or other patient follow-up data. In addition, no single unified storage file format currently exists for such data. Data sharing or even submission to a journal requires conversion to a simple two-dimensional format, like TIFF, leaving the critical metadata unsupported and lost. In addition, while most commercial image-capturing software includes sophisticated processing and analysis tools, most data require customized analysis solutions. Even for trained personnel, a fair amount of time is required to master the software tools.

A particularly difficult problem is data management. For example, most digital imaging applications now require quantitative analyses. There is currently no method to coherently manage analytical results along with the image data. Further, analytical results and image data are not linked to relevant biological and clinical data, which make complete information retrieval an unnecessarily complex process. Invariably the two are linked only by the person who performed the analysis. If that individual leaves the laboratory or even forgets how he/she generated a result, the link between image and result is broken. Nevertheless, this set of challenges is one the library and museum community has addressed in extensive efforts to standardize core sets of metadata for various information, including digitized images in various digital library projects for the past decade.

Previous studies on biomedical imaging and, specifically, data analysis and description, have resulted in a growing body of literature focused on investigations of biomedical imaging in general; however, there has been very little research into pathologic imaging in the context of a standardized metadata framework. The proposed study is designed to address these needs by development of a comprehensive new resource for the management of biomedical image data.

Audience and audience needs

The visual findings (which result in medical diagnosis) from biomedical images such as radiological films, microscope slides, and photographs are a crucial tool in biomedicine (Kim & Rasmussen, 2008; Kim & Gilbertson, 2007). In pathology, in particular, an important function of pathologists is the accurate documentation of morphological findings. This task is achieved through descriptive prose which carries with it inconsistent variations in vocabulary and form and the differential diagnoses of individual pathologists. Recent advances in digital imaging in pathology have led to innovative changes in the discipline with the result that digital imaging has become an integral part of the discipline. However, one of the few studies on information management and retrieval found that pathologists are only given limited access to the digitized images *because these images are rarely associated with meaningful imaging information* (Yagi & Gilbertson, 2005; Gilbertson, 2004). For instance, pathologists can retrieve images by a simple identifier such as a surgical pathology number, but this identifier is limited in use due to issues of patient privacy and confidentiality issues. Moreover, the surgical pathology numbers associated with pathologic images do not fully support meaningful access points, such as the nature of the diagnostic findings, identification of the anatomic region of interest, information about specimen preparation, or any relevant clinical history of the individual. To accelerate the efficient use of pathologic images in practice, research, and education, more complete and meaningful metadata sets are needed in a standardized format. This evidence strongly suggests the need for the skills and expertise of librarians to organize complex digitized images. The proposed study offers *exceptional potential* to address the problem of standardizing pathologic images, creating a new audience and client base for information specialists.

Potential data sources

Existing metadata schemes in medicine include some datasets; however, *no single system currently provides all the data elements required to adequately describe today's detailed and complex pathologic images*. Some describe textual information (e.g., Unified Medical Language System or UMLS); some focus on radiological images (e.g., Digital Imaging and Communications in Medicine or DICOM); some cover health information on the Web (e.g., Medical Metadata Core or MMC); some describe biospecimen preparation context (e.g., the web-based caTISSUE biospecimen bank); and some cover technical details rather than clinically relevant contexts (e.g., Open Microscopy Environment or OME). Comprehensive descriptions, developed according to the types of standards governing library and information science resources, should encompass more complex imaging information such as pathologic findings, clinical or demographic information, general image file properties, imaging acquisition information (e.g., capture devices, including microscopes and digital cameras), and biospecimen preparation, and these should be stored and ideally mapped consistently to each image. Such an approach would support practitioners in biomedical disciplines at currently unprecedented levels in biomedical imaging but in keeping with state-of-the-art practice in library and information science.

Due to the scope and coverage of the metadata schemes mentioned above, these data sources contain only partial imaging descriptions required for microscopic images. More importantly, the noted schemes overlap somewhat in terms of their contextual and structural coverage. OME is an open source software project to develop a database-driven system for the quantitative analysis of biological images. In the OME, the definition of image metadata is expanded to include all traditional image metadata around the data acquisition event (objective lens, detector settings, illumination system, etc.), and it also includes a definition of the biological and experimental systems that contribute to making the sample--the cells, genes, mutants, inhibitors, temperature, etc(Goldberg, 2005). However, OME does *not support any specific biosample-related data components* which describe the original source of the scanned images (e.g., microscope slides, etc.). For biosample-specific descriptions, caTISSUE Core is the widely accepted standard.

DICOM is a product of the American College of Radiology and the National Electrical Manufacturers Association to promote communication of digital image information (mainly radiology images), regardless of device manufacturer (DICOM, 1999). A subset to DICOM, the Visible Light Supplement has been introduced to support explanations of diagnostic imaging devices (endoscopes, microscopes, and cameras) that produce reflection or transmission of color photographic images (Bidgood et al., 1997). The Visible Light Supplement also specifies a new anatomic frame of reference that does not rely on a patient-based coordinate system, but describes orientation in terms of anatomic landmarks. The imaging procedures supported by the DICOM Visible Light Supplement include fiber-optic and rigid-scope endoscopy, light microscopy for anatomic pathology, surgical microscopy, and general anatomic photography (DICOM, 1999). Both DICOM and OME are quite comprehensive in terms of their imaging acquisition descriptions. However, their *technical complexity* is not appropriate for simple core metadata descriptions which can be supported by the National Library of Medicine's (NLM) Metadata set. The NLM Metadata is based on metadata terms maintained by the Dublin Core Metadata Initiative (DCMI) designed for use with electronic resources published by the NLM. To achieve *seamless interoperability and wider accessibility* to acquired microscopic images, pathology imaging communities and knowledge representation communities must collaborate to standardize metadata descriptions.

Impact

The proposed early career development study, aligned with the Laura Bush 21st Century Librarian Program to "spur new innovations in library service," has significant potential to translate core concepts in information representation, such as cataloging, classification, authority and access control, subject analysis, arrangement and display, and vocabulary control, which have been developed, standardized, and practiced in libraries into new and emerging information management needs. The proposed metadata framework will provide *an innovative model of librarianship* to support non-traditional data collections and description. Specifically, this project aims to create a metadata framework to better support pathologic imaging descriptions through an organized and systematic representation of knowledge. Additional impacts of the proposed study include: *advancing understanding of imaging metadata* that will benefit academic medical institutions and healthcare providers requiring digital imaging description as well as the library community supporting these information resources; and *establishing the foundation for academic liaison* between the library and information science and biomedical imagery communities. For instance, as a process of core data identification, two working projects at the University of Kentucky (UK) will adopt the expected findings into the development of human and mouse sample management systems to provide preliminary test cases for full utilization of project outcomes. Ultimately, the findings will *expand librarians' organizational knowledge and skills* to non-traditional collections, specifically pathologic images. In this sense, this study will bring greater opportunities for long-term interdisciplinary research between library and information science and biomedical informatics disciplines. Finally, knowledge management through application of organizational skills is a core expertise which is trained through the library profession. Therefore, supporting a standardized description of digitized pathologic images through library expertise will benefit biomedical researchers who deal with molecular analytic results. The proposed study will be an initial step to *providing clinically relevant, technically seamless, and meaningful descriptions* through an integrated metadata framework.

Diversity

Libraries have long served the information needs of broad and diverse communities across divergent disciplines; however, traditionally, healthcare professionals seeking biomedical images and their affiliated data elements remain underserved through library services directly supporting their specific imaging information requirements. The proposed study is an innovative initiative to bring the specialized expertise of the library community to bear on the specific needs of a relatively non-traditional client community. In advancing understanding of the application of core library science skills to a novel area of librarianship, this project has significant potential to broaden access of the biomedical imaging community and the practitioners who rely on it to expanded biomedical library resources. In addition, library patrons who have no medical background will be able to achieve easier access to digital images if the images are well-described in a standardized manner. Therefore, the proposed study will benefit diverse groups of underserved information seekers who have previously had no access to digitized biomedical images.

Project Design and Evaluation

The proposed study aims to develop a standardized metadata framework of pathologic images so that detailed and complex biomedical information can be stored, retrieved, and shared among various people and organizations. The findings of the study will be used to develop a clinically relevant, technically seamless, and meaningful linkage standard for describing microscopic imaging through a metadata framework.

Project Goals

Goal 1: To identify candidate data elements representing pathologic imaging information from four existing biomedical data sources. These include the NLM Metadata, caTISSUE Core, OME, and DICOM._The first goal is to identify existing data elements which will be used as source data elements in the proposed study. The PI will collect potential data elements from multiple data sources to be integrated into a standard metadata format.

Goal 2: To merge identified data elements representing pathologic imaging information (from Goal 1) into the Protégé metadata software. With the help of this sophisticated metadata tool, the proposed study will merge all the data elements collected into a single output file which will allow the PI to create mapped relationships among the different data sources.

Goal 3: To describe pathologic images by applying the newly created core metadata elements. In this phase, the PI will create metadata records for a collection of diverse test pathology images by applying the core metadata set as defined.

Goal 4: To evaluate the outcomes of the merged file against the user or end information requirements for using these pathologic images. A single merged file in the form of a standard metadata framework will be generated and evaluated by measuring frequently used relevance measurements such as precision, recall, and F1 measures.

Specific Activities to implement the project

Four phases of the proposed study will be conducted according to the goals indicated above. In the first phase, the proposed study will collect potential candidate data elements from four existing sources including NLM Metadata, OME, caTISSUE Core, and DICOM. Four separate files for individual data elements will be merged into a single merged Protégé file in the second phase. For the third phase, the merged data elements will be tested to identify whether individual data elements should be included in core metadata set or not. A group of experts will finalize a metadata framework including core metadata set for pathologic images through a series of focus group interviews. In the final phase, evaluation of retrieval effectiveness will be measured between a set of imaging queries and described pathology images. The first two phases are to develop a metadata framework for pathologic images and the remaining two phases are to evaluate whether the metadata developed can be met by imaging queries or not (see Figure 1).

In the first phase, a primary task to be conducted will be to *collect and merge candidate data sets*. The majority of the project activities in this period of time will be devoted to setting up hardware and software to merge multiple data sets from the existing data sources including NLM, OME, DICOM, and caTISSUE data sets. These data sources will be used based on previous studies completed recently by the PI (Kim, 2007; Kim, 2008). The NLM Metadata was chosen because it contains a basic level of description based on the DCMI set, widely accepted in various online communities as their description standard. However, the NLM Metadata lacks the components of imaging descriptions provided through the DICOM Visible Light supplement. The DICOM framework has been a leading standard in medical imaging focusing on radiological imaging. The Visible Light Supplement was added to cover pathologic imaging descriptions which contain extensive data components that describe technical aspects of imaging acquisition. However, clinically relevant and laboratory-oriented image processing (e.g., patient-specific and biosample-oriented information, etc.) are not included in the DICOM supplement. Therefore, the proposed study will also use the caTISSUE Core data elements in order to cover biosample preparation/processing and patient specific clinical and outcome data components. In addition, the OME framework will be included because it provides an open source program for microscopic imaging databases which can be further used to store and collected test pathology images. The OME also contains analytic data components relating individual imaging projects to various image analyses and biosamples.

<Figure 1> Overview of Study Design and Phases

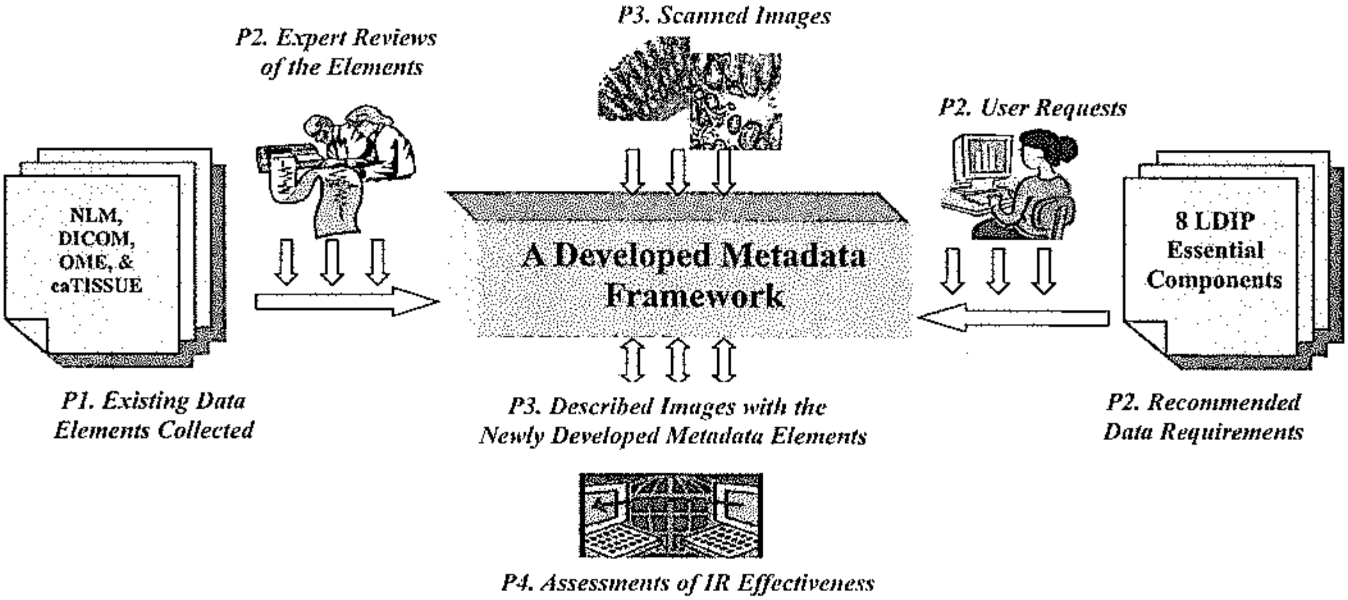

P2. Expert Reviews of the Elements
P3. Scanned Images
P2. User Requests
NLM, DICOM, OME, & caTISSUE
A Developed Metadata Framework
8 LDIP Essential Components
P1. Existing Data Elements Collected
P3. Described Images with the Newly Developed Metadata Elements
P2. Recommended Data Requirements
P4. Assessments of IR Effectiveness

Figure 1 illustrates the four study phases. Individual activities involved with the proposed study are fully outlined in the study timeline in the *Schedule of Completion* in a separate attached file.

Additionally, in this phase, the duplicate and overlapped as well as matched data elements among the collected data elements from these different resources will be individually imported into Protégé metadata editing software. In order to collect, clean, and map the collected data elements, two research assistants will be trained to manage the Protégé software in this phase. Expertise in software engineering and knowledge representation are required in this phase, and the proposed study will recruit and support two graduate research assistants from library and information science (LIS) and computer science (CS) at UK. The LIS student will extensively review individual data elements and enter them into the Protégé editing software for further analysis. This activity requires understanding of metadata representation. Thus, the PI will train students on how to appropriately create individual elements and their associations in Protégé. The CS student will be asked to customize the Protégé open source coding so that the collected data elements can be better displayed and mapped in Protégé. The research assistants and the PI will work cooperatively to acquire comprehensive data elements for pathologic descriptions in this phase.

The second period of the proposed study is to *create and describe the created metadata set* by comparing potential queries in the field of pathologic images. The potential queries to be tested will be developed in this phase based on recommendations from The Laboratory Digital Imaging Project (LDIP) and potential user groups in both pathology and imagery communities. The LDIP under the direction of the Association for Pathology Informatics recommends these critical components of the pathology image data exchange specification in Table 1 below.

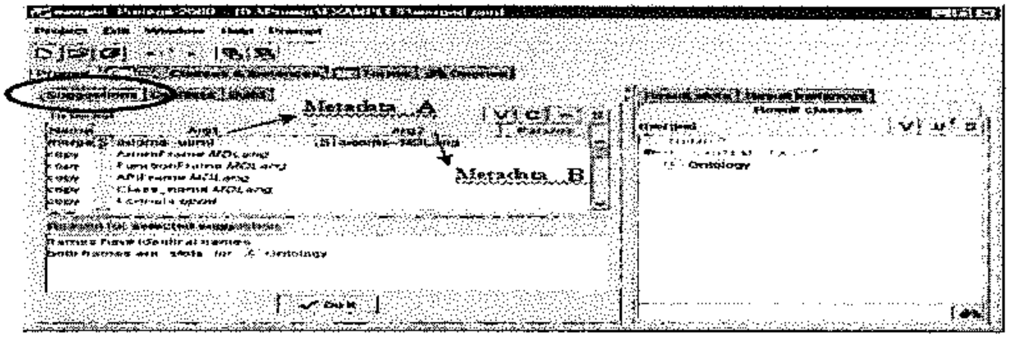
<Table 1> A Schematic Entry of an Original Source Project in Protégé 3.1.1.

Concept Name	Definition	Super-class	Sub-class	Sample value	Source Origin
Display-Options	…This specifies mapping of image channel components to RGB or greyscale colorspace with one byte per pixel per RGB channel….	Image	DisplayOptions/ Display, DisplayOptions/ID, DisplayOptions/ Zoom	RGB | Grey	OME
EmFilter	Emission filter manufacturer specification.	Filter	EmFilter/Type	LongPass | ShortPass | BandPass | MultiPass	OME

[Note: Concept Name indicates name of concept (data element name); Definition defines the concept; Super-class and sub-class represent hierarchical relationships (narrower/broader concept) in a collection of concepts; Sample Value is an example of actual data to be stored; and Source Origin is where the concept is taken from (including NLM, OME, DICOM, and caTISSUE)]

In this phase, the proposed study will merge all data elements (concepts) collected in Goal 1 from the four sources (NLM Metadata, caTISSUE, OME, and DICOM) by comparing the original source elements (source project) into a newly created working project (a merged project) based on suggested operations by the Protégé metadata editing program. When merging multiple metadata standards, the Protégé software automatically creates a list of suggested operations so that the user can perform an operation by choosing one of the suggestions or by specifying an operation directly (copying or deleting). The generation of suggestions is based on similarities in the concept names. Editing operations will be performed before the suggested operation by examining conflicts of four merged metadata sets and their creation of new operations into a merged project. A sample entry for a merged project can be found in Figure 2 below.

<Figure 2> A sample entry for a merged project in Protégé 3.1.1.



[Note: When a user is merging two metadata standards, Protégé software generates an initial list of suggestions shown in Suggestion Tab which is based on the similarities in class names. For example, it proposes to merge the two classes of data elements from a metadata A and a metadata B. This also provides the Reason for selected suggestion (lower left box).]

In addition, microscopic images will be scanned and stored in a Web-based image database that will be designed in this phase. Two CS students who have expertise in developing (or customizing) a large file image database will be recruited to write the database and do web programming (Oracle and Java-based platforms). The Web-based application to be developed will store the scanned images and provide an indexing aid feature that will capture the selected descriptors for individually stored images.

In the third phase, the project team will evaluate Protégé's suggested operations for the newly merged project. Quality will be evaluated by measuring precision and recall. Eight recommended requirements by the LDIP (see Table 3) will be used to pilot test Protégé's merging operations. In Table 2, the artificial results regarding precision and recall for two test cases, A and B, are shown. The Suggestion column reports the number of suggestions that were made. The Correct column shows how many of the suggestions were relevant while the Missing column show how many of the relevant suggestions were missing. In total there were 5 possible cases for merging in case A and 9 cases in case B (Correct plus Missing). Both A and B got perfect precision (1) and A recall (0.6) was higher than B (0.44). Six pairs of mapping suggestions will be analyzed among the four metadata sources.

<Table 2> Quality of Suggestions in Protégé 3.1.1

Test ID	Suggestions	Correct	Missing	Recall	Precision
A	3	3	2	0.6	1
B	4	4	5	0.44	1

At this stage, a series of focus group meetings and interviews will be conducted to determine relevant descriptions for test image descriptions. Domain experts from pathology (2 pathologists), ontology and library science (2 knowledge representation experts), and imagery (2 pathology imaging experts) will be asked to participate in the group interviews. The key component of this phase is to form a group of 6 domain experts that can exchange their expertise to construct a standard that can effectively represent imagery information contained in the set. Rigorous statistics will be required in this phase to assess inter/intra-indexer consistency among the different description results from the 6 domain experts. The experts will be recruited to perform the following activities: (1) review and finalize the merged metadata elements and their relationships; (2) describe and select the most appropriate describable units for the scanned microscopic images; (3) identify, review, and finalize potential queries to be tested based on LDIP's 8 recommended pathologic imaging properties; and (4) decide whether the described images with the newly created metadata framework match or not against individual queries. The proposed study will report the results of the discussion in a list of core metadata set with informative instructions for description.

In the final phase, *evaluation of retrieval effectiveness* will be tested between a set of imaging queries and described pathology images. A set of test queries will be collected from a group of domain experts to test retrieval effectiveness of the described images. The described images will be collected from two sources including the Department of Pathology and Laboratory Medicine at the University of Kentucky (Lexington, KY) and the Pathology Division at the National Surgical Adjuvant Breast and Bowel Project (NSABP, Pittsburgh, PA). The project director of the proposed study has a joint appointment in UK's School of Library Science and Department of Pathology and Laboratory Medicine. Images to be collected will be limited to breast cases from both gross and microscopic images. No personally identifiable and confidential information will be gathered in conjunction with images for the proposed study. UK IRB approval for exempt certification will be obtained for the process of acquiring scanned microscopic images for the proposed study. The collected images will be given to a group of domain experts along with a core metadata set so that the group can describe images for further analysis. The retrieval effectiveness will measure precision and recall between the images and the queries. In addition, F1 measures will be calculated to further analyze precision and recall. Since the recommended data properties of the LDIP contain core characteristics of microscopic images that are rather abstract, these will be further addressed by searchable user queries. Six domain experts and online user communities will be given the LDIP recommendations for their comments. Based on these comments and LDIP recommendations, the PI will write a set of test queries.

<Table 3> Laboratory Digital Imaging Project Recommendations (LDIP) for Digital Pathology Image Annotation

No.	Recommended elements for digital pathology image annotation
1	General file properties, such as who created the file, when the file was created, the purpose of the file, and any intellectual property rights and restrictions. This section may contain data elements that authenticate the file or its creator, and ensures the approved IRB/Privacy Board status of the file.
2	Binary object properties, such as the organization, structure or mathematical properties of the binary image(s), so-called image header data, technical image or image display descriptors, and either the binary object itself (rendered in ASCII base64) or with a pointer to a URL holding the binary image file.

3	Image capturing device information, specifying the microscope/camera and any other hardware devices contributing to the capture of the image.
4	Image acquisition information, such as device settings and physico/optical parameters related to the capture of the image, and calibration data or protocols.
5	Histologic features, such as staining information, or pointers to experimental protocols for the preparation of the image.
6	Specimen information, which may include the methods used to procure or prepare the specimen and pointers to specific specimen-related records in tissue databases or specimen repositories.
7	Pathologic information pertaining to the image, including diagnosis, or specific pathologic descriptions of defined regions of interest.
8	Clinical or demographic information related to the patient providing the specimen. This section can be provided with de-identified or encrypted data elements or with data intended to authenticate or otherwise ensure the confidentiality, privacy of the record or ensure compliance with federal regulations.

[Note: Source: API Working Group Session – Open Discussion of Pathology Digital Imaging Standards by Jules J. Berman, Ph.D., M.D. and Ulysses J. Balis, M.D. in APIII, Pittsburgh, PA, Wednesday, Oct 6, 2004 available at: http://www.ldip.org/ldip_jb.ppt]

Project Resources: Budget, Personnel, and Management Plan

This study is an interdisciplinary collaboration to support biomedical imaging through the direct application of core library science expertise to imaging datasets. Dr. Kim, the project director, brings direct experience in both disciplines to this project. She is also an experienced educator with excellent credentials in informatics who will be responsible for training the research assistants, each of whom will bring distinct programming and library science skills to the project. To ensure appropriate end-user outcomes, domain experts from pathology, ontology and library science, and imagery will participate in Goals 3 and 4. In particular, Dr. Cibull has particular expertise in managing biospecimen repositories and in collaborating on the development of related imaging datasets. Dr. Kim will oversee and participate in all facets of the project, coordinating work processes and monitoring progress. A study coordinator will assist Dr. Kim in managing communications with all team members and in scheduling project activities. Project consultants are experts in the field who bring specialized ancillary expertise in pathology, statistics, and information technology to ensure valid project outcomes.

Although it is not mandatory for the early career development of research category (according to IMLS guidelines), the University has agreed to cost share a portion of the project's expenses including travels ($6,000), a research assistant ($73,370), tuition ($26,200)and indirect costs ($40,952) for three years in the total amount of $146,522.

Dissemination

The results of the proposed study will be used for the following collaborative projects. The project director and her database team have been developed three working Web systems that are currently implemented in *two UK biorepositories (ukTISSUE and UKRMTB) and a Korean Lung Tissue Bank (KLTB-BIS, Seoul, Korea)*. The ukTISSUE is an Internet database that tracks biospecimens stored in the UK Biospecimen Core Tissue Bank (Director, Dr. Michael Cibull, MD). The National Cancer Institute at NIH is working on a system to connect biospecimen repositories nationwide. This project will provide a way to participate in the national biorepository information network and to become a partner and share local information sources (ukTISSUE). Most importantly, the expected results will be directly used to map all the described pathologic images and biosamples in these three systems so that images and biosamples across all three systems can be seamlessly retrieved. As a result, project outcomes will directly benefit the pathology community. In collaboration with Dr. Michael Cibull, MD, UK Professor of Pathology and Laboratory Medicine; Director of UK Surgical Pathology; and Director of the UK Markey Cancer Center Tissue Procurement Service, and his team, the proposed study will analyze essential data elements for the department's prototype image database. Currently, the Department of Pathology and Laboratory Medicine keeps digitized images from autopsy and surgical cases in the network drive with a simple file identifier in a shared network folder. The department will collaborate with the project director to identify a more advanced way to organize the images for use in training, case consulting, and archival purposes.

Secondly, results will be applied to the development of a mouse biorepository information system for describing its core data elements. In collaboration with Chemyong Ko, Ph.D., UK Center for Excellence in Reproductive Health Sciences, the PI will represent information in a formal metadata modeling tool set up with Protégé to describe mouse repository-specific information. The UKMRTB was recently implemented in Dr. Ko's lab and he is currently looking for *a way to be integrated into a human biosample databases* internally and nationally.

The outcomes of standardized metadata set describing pathologic images will be disseminated through four dissemination channels. First, the findings of the metadata standards will be applied to two in-house systems (developed by the project director and her team) that currently require annotation standards. Second, the outcomes will be submitted to academic journals such as JASIST or JAMIA for publication so that researchers who seek information about imaging description and its metadata application can be accessible via academic papers.

Third, there is a massive online user community which seeks information about biomedical imaging on a variety of aspects. These include library community interested in metadata standard for biomedical images. DCMI for crosswalk development accessible at: http://dublincore.org/groups/, Getty's AAT research site (Art and Architecture Thesaurus for imaging metadata standards) accessible at: http://www.getty.edu/research/conducting_research/standards/, and NLM (National Library of Medicine for medical imaging description through metadata) at: http://www.nlm.nih.gov/tsd/cataloging/metafilenew.html. In addition, four agencies which the proposed study will be used as existing data sources will be targeted to disseminate the findings that can be included as a part of metadata framework. These include OME (Open Microscopy Environment for database developers) at: http://lists.openmicroscopy.org.uk/mailman/listinfo/ome-users/, caTISSUE (cancer Bioinformatics Grid for biosample annotation) at https://list.nih.gov/archives/cabig_tbpt-l.html, DICOM (Digital Imaging in Communication in Medicine) at: http://medical.nema.org/, and NLM metadata development team accessible at: http://www.nlm.nih.gov/tsd/cataloging/metafilenew.html.

Lastly, the findings will be presented to various conferences and meetings such as ALISE (Association for Library and Information Science Education), ISBER (International Society for Biological and Environmental Repositories), and KBRIN/INBRE (Kentucky Biomedical Research Infrastructure Network). In addition, the findings will also be distributed as class exercise and project to library and information science students at the UK. PI of the proposed study teach two targeting classes including LIS602 (information storage and retrieval) and LIS639 (introduction to medical informatics) which will benefit to have a real working example of metadata set and description standard for learning exercise. All the findings of the study will be posted and updated at PI's homepage (http://www.uky.edu/~skim3) as well as College and School's news release on research activities for consistent access.

Sustainability

The project will continue to grow beyond the grant period through departmental support of metadata development and application activities. The UK School of Library and Information Science continuously supports faculty research activities, especially in the area of metadata applications and innovative technologies. The department will continue to support faculty research activities by providing research space in metadata lab and travel money to present the findings of the proposed study after the grant period. The developed metadata framework will be extended to an online retrieval system that allows automatic creation of XML encoded file to be integrated for other imaging applications. To extend the future application of the developed metadata framework, the PI will seek intramural and extramural grants to continuously extend the proposed study. PI will also plans to preserve the developed course contents of metadata exercises in an online tutorial format so that upcoming students who take LIS602 and LIS639 can practice at their own pace. All of the outcomes including mapped metadata set and source codes for XML conversion tool will be sharable through the PI's homepage.

References

- Bidgood, WD (1997). Documenting the information content of images, Proceeding of American Medical Informatics Association Annual Fall Symposium, 423-428.
- Bidgood, WD, Bray, B, Brown, N, Mori, AR, Spackman, KA, Golichowski, A, et al. (1999). Image acquisition context: Procedure description attributes for clinically relevant indexing and selective retrieval of biomedical images. Journal of American Medical Informatics Association, 6(1), 61-75.
- caTBPT (2005). caTISSUE, Retrieved March 20, 2005 from http://cabigcvs.nci.nih.gov/viewcvs/viewcvs.cgi/catissuecore/Use%20Case%20Document/caTISSUE_Core_Use_Case_Document_v2.0.doc.
- Digital Imaging and Communications in Medicine (DICOM): Supplement 15: Visible light image for endoscopy, microscopy, and photography, Retrieved October 12, 2004 from ftp://medical.nema.org/MEDICAL/Dicom/Final/sup15_ft.pdf.
- Gilbertson, JR, Gupta, R, Nie, Y, Patel, AA, & Becich, MJ. (2004). Automated Clinical Annotation of Tissue Bank Specimens, Medinfo, 607-610.
- Goldberg IG, et al. (2005). The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging, Genome Biology, 6:R47
- Kim, S. and Rasmussen, E. (2008). Characteristics of tissue-centric biomedical researchers using a survey and cluster analysis, Journal of American Society for Information Science and Technology. In Press.
- Kim, S. and Gilbertson J. (2007). Information requirements of cancer center researchers focusing on human biological samples and associated data. Information Processing & Management, 43 (5): 1383-1401.
- Krichel, T. (2001). A metadata framework to support scholarly communication, Proceeding of International Conference on Dublin Core and Metadata Applications, held on October 24-26, 2001, NII, Tokyo, Japan, 131-137.
- Lambrix, P, Habbouche, M, & Perez, M. (2003). Evaluation of ontology development tools for bioinformatics, Bioinformatics, 19(12): 1564–1571.
- Leong, FJWM & Anthony Leong, ASY. (2004). Digital imaging in pathology: theoretical and practical considerations and applications, Pathology, 36(3): 234–241.
- Lundin, M, Lundin, J, Helin, H & Isola, J. (2004). A digital atlas of breast histopathology: an application of web based virtual microscopy, Journal of Clinical Pathology, 57: 1288-1291.
- Malet, G. (1999). Enhancing Internet medicine document retrieval with search engines and controlled vocabularies, Journal of the American Medical Informatics Association 6:163-172.
- McEntire, R, Karp, P, Abernethy, N, Benton, D. (1999). An Evaluation of Ontology Exchange Languages for Bioinformatics, Retrieved November 12, 2006, from http://xml.coverpages.org/OntologyExchange.html.
- National Library of Medicine (NLM) (2004). NLM metadata schema, Retrieved December 1, 2004 from http://www.nlm.nih.gov/tsd/cataloging/metafilenew.html.
- Open Microscopy Eenvironments (2005). OME XML Schema, Retrieved March 1, 2005 from http://www.openmicroscopy.org/XMLschemas.
- Sim, I, Olasov, B, & Carini, S. (2004). An ontology of randomized controlled trials for evidence-based practice: content specification and evaluation using the competency decomposition method, Journal of Biomedical Informatics, 37: 108–119.
- Tulipano, PK, Millar, S, & Cimino1, JJ. (2003). Linking molecular imaging terminology to the gene ontology (Go), Pac Symp Biocomput, 613-23.
- Yagi, Y & Gilbertson, JR. (2005). Digital imaging in pathology: the case for Standardization, Journal of Telemedicine and Telecare, 11: 109–116.
- Zeng, ML & Chan, LM. (2004). Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems. Journal of American Society for Information Science and Technology, 55(5), 377-395.