

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.