

Using a Medical Digital Library for Education Purposes

Mara Nikolaidou
Libraries Computer Centre
University of Athens
Panepistimiopolis
15771 Athens, Greece
E-mail: mara@di.uoa.gr

Dimosthenis Anagnostopoulos
Dept. of Geography
Harokopio University of Athens
70 El. Venizelou Str.
17671 Athens Greece
E-mail: dimosthe@hua.gr

Michael Hatzopoulos
Dept. of Informatics
University of Athens
Panepistimiopolis
15771 Athens, Greece
E-mail: mike@di.uoa.gr

Abstract

Athens Medical School is one of the largest medical research institutions in Greece, where educational activities are combined with every-day practice in the University Hospital. The laboratories operating in the University Hospital produce a large amount of research material, mainly consisting of medical images and videos in digital format. In this paper, we present the Digital Library System built to support the collections of Athens Medical School laboratories. The system facilitates access through the Web to medical research material and on-line presentations/lessons. Laboratory collections have different requirements regarding digital material structure and metadata modelling. Thus, while designing the Digital Library architecture, the organisation and administration of dynamically created collections was explored. We introduced the term dynamic collection management to denote the support of automated collection definition and management within an integrated Digital Library environment. Laboratory staff is responsible for selecting the material to be added in the library, while they also participate in metadata creation. Furthermore, they use this material to create presentations or on-line tutorials, also stored in the Digital Library. To facilitate the aforementioned functionality the Digital Library system also supports a complex workflow environment.

1. Introduction

Athens Medical School (AMS) is one of the largest medical research institutions in Greece, where educational activities are combined with every-day practice in the University Hospital. The laboratories operating in the University Hospital produce a large amount of research material, mainly consisting of medical images and videos in digital format. This material should be categorized and used for educational purposes. Digital Library Systems facilitate the management of large collections of digital material and resources providing advanced accessing capabilities [5]. Medical applications may benefit from this kind of technology, since they require handling of large amounts of data, such as medical images or videos. In this paper, we present the Digital Library System built to maintain the material –mainly digital images and videos– produced by the laboratories of Athens Medical School (AMS DL). Laboratory staff is responsible for choosing the material that should be added in the Digital Library. The selection process is based on educational and research criteria. Although the digital material is part of patient medical files, it is characterised only by content-based properties that are meaningful in medical research. The system facilitates access through the Web to authorized researchers and

students. AMS DL was developed by the Libraries Computer Centre in cooperation with the Department of Informatics.

Typical Medical Image Archive systems, as those presented in [2] and [8], facilitate access to medical material based on web-based multi-tiered architectures. All the material added in the system is part of a single collection. No specialised research characteristics are taken into account and all digital objects are described by a core metadata set. The digital library of Athens Medical School must support a number of collections, each one fulfilling the educational requirements of a specific laboratory regarding digital material structure and metadata modelling. Furthermore, the number of collections supported is not predefined. Thus, while designing the Digital Library architecture, the organisation and administration of dynamically created collections was explored. We introduced the term *dynamic collection management* to denote the support of automated collection definition and management within an integrated Digital Library environment. In the medical archive systems previously discussed, the medical material is gathered and processed by a specialised unit before added in the system. The Digital Library environment built for Athens Medical School supports a more complex workflow, since the research material is added directly by the researcher, while he/she also participates in metadata creation. Furthermore, this material may be used to create presentation or on-line tutorials, also stored in the Digital Library.

2. Digital Library Architecture

The digital library architecture is based on a multi-tiered client-server model and is implemented using Java and *IBM Content Manager* platform [6]. The system consists of three main layers: *Collection Repository*, *Collection Management* and *Provided Services*. *Collection repository* facilitates storing and searching of digital objects and consists of IBM Content Manager components. The *Collection Management* layer facilitates the automated collection definition and administration within AMS DL environment. A *collection dictionary* is used to maintain collection-related information.

Collection Management layer consists of two main modules. *Repository Access* module is responsible for interacting with IBM Content Manager Platform using predefined APIs. This module is activated upon request whenever there is a need to store or retrieve data or metadata information to/from the Content Manager, but it cannot facilitate structuring and administrating a digital collection. This functionality is included in the *Collection Manager* module providing services to external clients. Thus, clients do not interact with the Content Manager platform and consequently have no knowledge of its existence. This ensures system modularity and extendibility and enables supporting different data and metadata models at data storing and data management levels. Repository access module is also responsible for DOI assignment and management.

Provided Services include Java clients, e.g. *AMS DL Administration*, *Processing and Cataloguing* and *Create Presentation/Lesson* applications, and servlets, as *Collection Search*, used to provide web access. Search results are presented using XML pages containing entity related information and links in the body part and the related metadata information in the header using RDF format [1].

3. Supported Workflow

Research material is produced by laboratory medical equipment in digital format and it mainly consists of TIFF images and videos of various high-analysis formats. Laboratory staff selects the material to be added in the Digital Library and forwards it to the Central Library of Health Sciences. The cataloguers must fill the metadata fields, process the digital object and add it in

AMS DL. Since the metadata characterising each image is strongly related with the specific medical area, cataloguers were unable to fill it. Thus, this is partially done by the laboratory scientific staff. The system provides workflow support to facilitate cooperation between researchers and specialised cataloguers, as shown in figure 1 using UML notation. Each medical object entering the workflow is characterised by a state. Only objects characterised as *Published* are actually available for viewing.

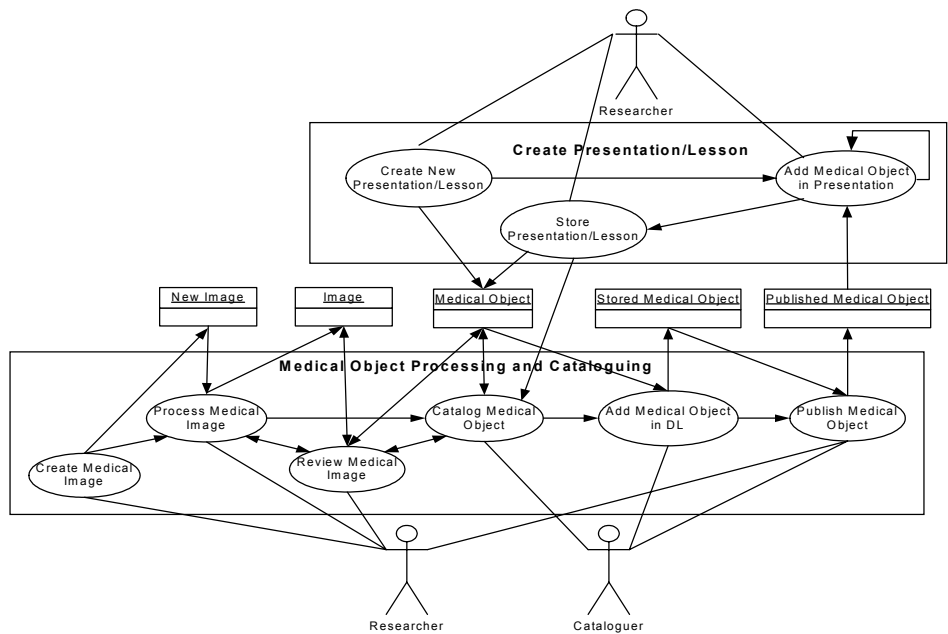


Figure 1. AMS DL supported workflows

One of the problems we faced during system employment was to integrate Medical Image creation in researcher's daily work. Selection of images takes place while the researcher is examining properly processed tissue samples with specialised microscopes capable of producing digital images. The digital image can be stored in a hard disk but, in order to add it in the library, the researcher must leave his position in the microscope and enter the *Processing and Cataloguing* application operating in the workstation next to it. This was not feasible due to researcher's workload. Thus, it was decided to store images during tissue examination, while review and characterisation of images are performed by researchers on a weekly basis. Researchers usually use the material stored within the Library to create presentation and on-line tutorials for medical students. The presentations are treated as composite medical objects stored within AMS DL consisting of sequences of images, notes and videos, as shown in figure 1. The workflow environment (*Create Presentation/Lesson* application) supports collaborative editing of simple presentations consisting of discrete steps of either presenting a medical object stored within the Library or writing simple comments.

4. Dynamic Collection Management

Since the number of supported collections is not predefined or static, two requirements are identified: a. the need to easily create new collections and b. the need to extend or modify collection characteristics. The term *dynamic collection management* is introduced to denote the support of automated collection definition and management within an integrated digital library environment. A *collection dictionary* is used to maintain collection related information.

The term *digital object* is used to denote material stored within the Digital Library. Digital objects are usually compound objects/documents consisting of *parts* of different medium type (e.g. text, image, sound, video), which are indexed by different tools. The digital objects belonging in the same collection are characterized by the same metadata set. Each collection is characterised by *general* metadata information, common in all collections, and *domain specific* metadata, useful for researchers in the specific domain. Thus, for each collection added in dictionary, the administrator must define the corresponding object structure and metadata scheme forming *collection description*. For either general or domain specific schemes four categories of metadata are maintain: *descriptive*, used to describe the material, *technical/preservational*, related to object/part type/format (e.g. image quality properties) and storing properties, *rights*, used for access control and *educational*, related to educational categorization (e.g. corresponding course or lecture). The metadata scheme used may be a standard one (e.g. DC), a variation of it or even a local one. Implementation properties, e.g. whether a field is bilingual, multi-valued or mandatory, are also included. The value type of each field is also recorded. Restricted value lists are also supported for specific fields.

Collection description can be derived from existing ones by extending the object structure and metadata model, e.g. a collection description can be defined as the descendant of an existing collection description, while additional object parts and metadata fields can also be defined. This feature allows flexibility during collection definition and facilitates collection description in a simplified manner. *Collection interoperability rules* may also be defined. This corresponds to the definition of mappings between the metadata characterizing each collection. Mappings can be partial (usually they are and concern general metadata information) and are useful when there is a need to search all the collections. The mappings introduced for a specific collection are valid for all its descendants as well.

In order to efficiently support dynamic collection management, AMS DL facilitates dynamic interface creation. The same interface is used for all collections, while screens presented to the user are dynamically created based on collection description. It is also multilingual, currently supporting Greek and English languages. The user is responsible for choosing the client's *locale* property. Searching and presentation activities are performed using the selected locale. Bilingual support increased interface complexity. To facilitate dynamic collection management AMS DL supports uniformly storing and accessing objects having different structure and being characterised by different metadata. Thus, the same services can be supported for different collections.

5. Data and Metadata Representation

Since the supported collections include medical material, we have decided to define a generic *Medical Collection* and use it as a prototype to create all laboratory specific collections. Medical collection consists from three sub-collections, *medical image collection*, which includes compound objects consisting of different analysis images (*medical image objects*), *medical video collection*, which include compound objects consisting of different analysis videos (*medical video objects*) and *presentation collection*, which consists of presentations edited by researcher using the material included in the two aforementioned collections (*presentation objects*) (figure 2). For each collection the object structure and metadata model must be defined. The metadata kept for both collections are similar, except the technical ones, which differ. The following parts are included in the *medical image/video objects*:

1. *Original image/video*. It is the original image/video produced in the Laboratory. It is of high quality; it cannot be efficiently transferred over the Web and should be strongly protected regarding copyright issues. Thus, access to it is restricted.

2. *Derivative image/video*: It is produced from the original usually in JPEG/MPEG format to be accessed through the Web. Access to it is restricted.
3. *Thumbnail Image*, to be shown in the Collection Search application.
4. *Description* in Greek and English (as the application should be bilingual).

The original image and the description are produced by the researcher, while all other formats are produced by the cataloguer during image processing.

The metadata scheme introduced to describe the *medical image/video objects* is based on Dublin Core, although it also supports customisations for medical material and specific educational metadata, such as *course* or *lecture*. Dublin Core is a widely adopted scheme, used in medical images archives [2] and health care applications [3,7]. The DC.Subject field was extended to support NML medical subject headings and local thesaurus schemes. The DC.Format field was extended to maintain information related with the image file characteristics. The Dublin Core Identifier field is used to store the Medical Image DOI produced automatically. Many DC fields and subfields, such as DC.Type, obtained default values.

The *Presentation Objects* consists of two parts: a *description* (in Greek and English) and a multi-valued part, named *presentation pages*, which contains links to objects of the Video and Image Collections. The technical metadata associated with this part contain information regarding its structure. The Dublin core *relation* field (*has_part* type) was used for its description.

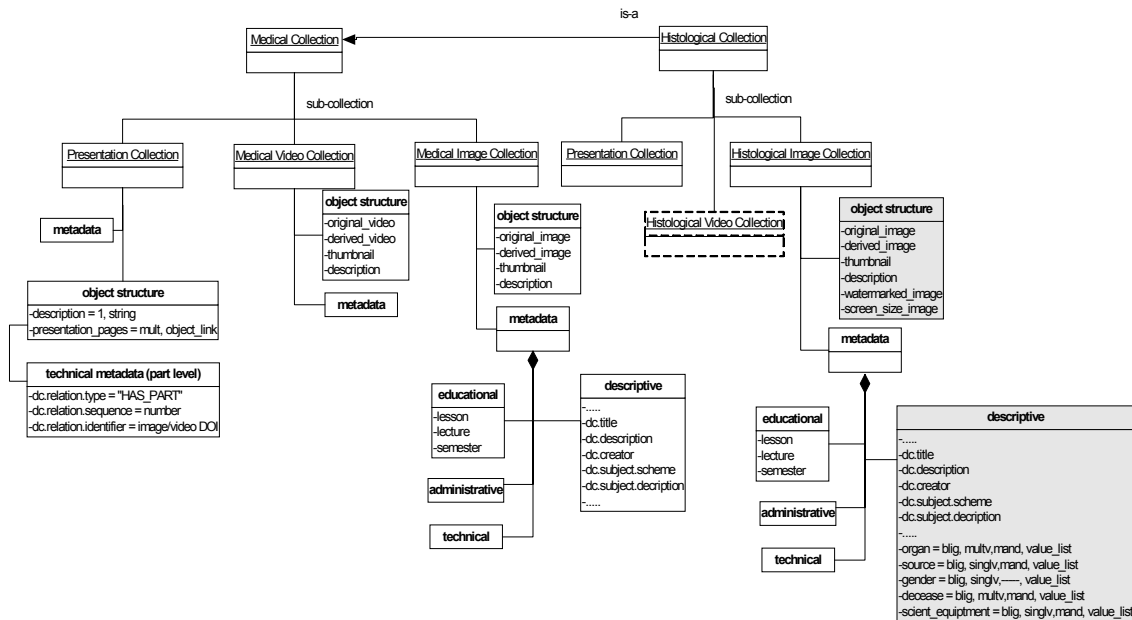


Figure 2. Medical and histological collection definition

The *Medical Collection* is practically empty, while all other collections are easily defined as its descendants by adding *collection-specific metadata fields* and extending the properties of medical image/video objects. *Collection-specific* metadata scheme can be defined using Dublin Core basic fields or extensions or even local fields.

As an example, we discuss the definition of the *Histological Collection*. As indicated in figure 2, the *Description* of *Histological Collection*, corresponding to the Laboratory of Histology, is derived from *Medical Collection Description*. There are no alterations in the *Presentation Collection* Description. Since the laboratory produce digital images, only the *Medical Video sub-collection* remains idle. The *Medical Image Collection* description was

extended. The *Medical Image Object structure* was extended by adding two new parts: *Watermarked Image*, produced from the derivative image watermarked using the symbols of the University and the corresponding Laboratory, and *Screen Size Image*, a medium-quality image produced from the derivative image to be easy shown through the Web. *Collection-specific metadata fields* are added in the descriptive metadata. These fields are considered as *local* (they are not Dublin Core fields), since they are useful only when searching the specific collection. Metadata description is accompanied by implementation properties, such as indicators of whether the field is bilingual or not, multi-valued or not, mandatory or not, and field type. While defining the Histological Collection, only additional features have to be described contributing to the simplification of collection definition process.

6. Conclusions

Athens Medical School Digital Library (AMS DL) facilitates access to medical research material –medical images, videos and on-line presentations - to researchers and students for educational purposes. Two strong requirements were imposed during system development: the organisation and administration of dynamically created collections and the support of advanced workflow capabilities. Medical material added in the library directly by researchers, may be used to create presentation or on-line tutorials also stored in the library.

We introduced the term dynamic collection management to denote automated collection management. In order to support dynamic collection management, a collection dictionary was implemented enabling collection definition in an advanced fashion. Collections are described in detail regarding their structure, supported metadata and the relationships between them. The definition and administration of composite objects as those belonging to the Presentation Collection proved to be straightforward and efficient. Derived collection definition enabled the description of all collections by extending the description of a simple one (medical collection) that provided basic capabilities. It contributed significantly to the simplification of the overall process, as metadata field definition is rather time-consuming, especially when dealing with fields with bilingual and predefined values.

References

- [1] Amann, B., Fundulaki, I., Scholl, M., “Integrating ontologies and thesauri for RDF schema creation and metadata querying”, *International Journal of Digital Libraries*, Vol 3, No 3, Springer-Verlag, 2000, pp. 221-236.
- [2] Bristol University, Technical Report: Bristol Biomedical Image Archive Case Study, 2000, available <http://www.brisbio.ac.uk>.
- [3] Davenport Robertson, W., Leadem, E.M., Dube, J., Greenberg, J., “Design and Implementation of the National Institute of Environmental Health Sciences Dublin Core Metadata Schema”, in *Proceedings of Int. Conf. On Dublin Core and Metadata Applications 2001*, National Institute of Informatics, 2001, pp. 193-199.
- [4] Duval, E., Hodgins, W., Sutton, S. Weibel, S., “Metadata Principles and Practicalities”, *D-Lib Magazine*, Vol 8, No 3, 2002, available <http://www.dlib.org/dlib/april02/weibel/04weibel.html>.
- [5] Fuhr, N., Hansen, P., Mabe, M., et. al., “Digital Libraries: A Generic Classification and Evaluation Scheme”, in *Proceedings of ECDL 2001*, LNCS 2163, Springer-Verlag, 2001, pp. 187-199.
- [6] IBM Corporation, *Content Manager Documentation – Planning and Installation Guide*, 2000.
- [7] Sakai Y., “Metadata for Evidence Based Medicine resources”, in *Proceedings of Int. Conf. On Dublin Core and Metadata Applications 2001*, National Institute of Informatics, 2001, pp. 81-85.
- [8] Suh, EB., Warach, S., Cheung, H., et.al., “A Web-based Medical Image Archive System”, in *Proceedings of the Medical Imaging 2002*, International Society for Optical Engineering, 2002, available http://cmag.cit.nih.gov/file/SPIE_MI_Paper.pdf