COPYRIGHT PROTECTION & EXPLOITATION OF DIGITAL CULTURAL HERITAGE

Dimitrios K. Tsolis, George K. Tsolis, Emmanouil G. Karatzas, Dimitrios A. Koutsomitropoulos and Theodore S. Papatheodorou

{dkt, gkt, ekaratzas, tsp}@hpclab.ceid.upatras.gr
HPCLAB, Department of Computer Engineering and Informatics, University of Patras

ABSTRACT

The main issue addressed in this paper is the design and implementation of a framework for the copyright protection and exploitation of Digital Cultural Heritage. The framework is consisting of an Advanced Digital Image Repository, which offers specialized services and a Dedicated User Interface for the protection and management of the Intellectual Property Rights of digitized material. In addition, another main research area of this contribution is the implementation of a Web Based Library, supported by advanced technologies, for the proper presentation of the digital cultural content. The work described in this contribution focuses on digitized material of the Hellenic Cultural Heritage. It has been carried out as part of the PRAXITELIS Project, which was presented during various conferences ([8], [9], [10]). This project is focusing on countering "electronic theft" and protecting the copyright of digital surrogates of photographs of artifacts, monuments and sites.

1. INTRODUCTION

The revolution of information technology is changing access to information in fundamental ways. Increasing amounts of information are available in digital form, networks interconnect computers around the world and the World Wide Web provides a framework for access to information. Yet the same technologies that provide enhanced access also raise important issues concerning intellectual property and proper presentation of sensitive cultural topics of priceless educational value.

The evolution of technology is challenging the status quo of Intellectual Property (IP) protection and management in many ways. In recent years we have seen the exploration of many technical mechanisms intended to protect IP in digital form, along with attempts to develop commercial products and services based on those mechanisms [1]. These mechanisms include interchangeable metadata to characterize the digital object and its IP Rights (IPR), unique global identifiers for the digital objects, watermarking techniques to mark, detect and prove the existence of the copyright of a digital image, audio and video, encrypted and secure data transfer through networks, etc. This section is focusing briefly on describing the general concepts, the results so far and the continuation of the implementation of an information system, which incorporates all those mechanisms.

The general concept of the integrated information system for management, protection and exploitation of digitized photographs of the Hellenic Cultural Heritage is the design and implementation of a modular system that embodies the following characteristics and tools [8]:
- Content Creation / Digitization tools.
- Digital Repositories that support:
  - Metadata for the management of data about the digital image's description and copyright.
  - Global standards for the unique identification of the digital images.
  - Watermarking for the IPR management and protection.
- On – Line Distribution and Payment System.
  The primary target of the current phase is the implementation of an Advanced Digital Image Repository, which, except for the above capabilities (metadata, unique identifiers and watermarking), provides specialized services and user interfaces for the efficient and resourceful retrieval, including search by image content. In addition, this phase equally focuses on the creation of a Web Based Library, structured upon the Digital Image Repository, for the proper presentation of digitized material of the Hellenic Cultural Heritage.

2. ADVANCED DIGITAL IMAGE REPOSITORY

The design and implementation of the Advanced Digital Image Repository (Figure 1) for this information system is a very important task. A repository fully capable to serve as a platform to this system should incorporate the following basic characteristics:

- Scalable. That is, the capability of gradually improving its performance in accordance to the platform and the number of computing resources used. The information system is a distributed one. Independent repositories already exist throughout Greece and are installed in the Ephorates and other authoritative Agencies of the Hellenic Ministry of Culture. The integration of these independent installations requires a scalable and modular repository.
- Support for multimedia and advanced hypermedia applications for digital image, audio, animation and video.
- Internet oriented, providing secure and easy access to data and metadata via the Internet and corporate Intranets. The Hellenic Ministry of Culture already has a corporate Intranet (designed and implemented by the High Performance Information Systems Laboratory) and, as a result, the integration of the digital repository to this already existing infrastructure is considered necessary.
- Support for applications and interfaces for data and metadata manipulation. These applications do not only
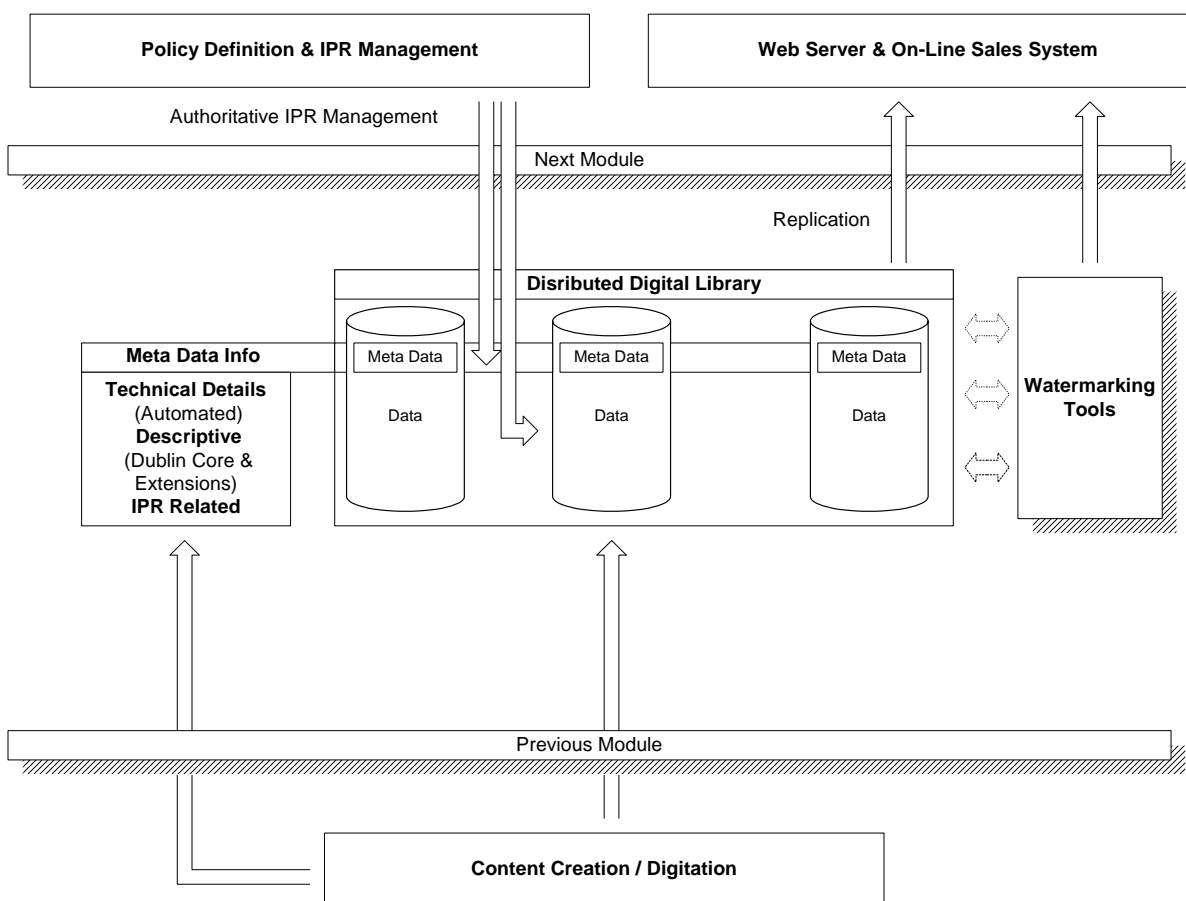

Figure 1: Digital Image Repository - Structure

include the built-in mechanisms of the database management system, but also user interfaces developed and customized in a way that they meet the needs and requirements of this project.
- Usability and effective management. Ease of implementation, configuration and use. Open to programmers and developers for its customization.
- Global access from distributed locations, supported by technologies for protected transactions and data transfer, replication and backing-up, data encryption and decryption.
- Open to new technologies and programming languages like JAVA and XML.
- Scheduled synchronization of distributed database instances in a national network of Digital Image Repositories. The central Advanced Digital Image Repository is being implemented in the premises of the Hellenic Ministry of Culture. Other databases will be distributed locally all over Greece, especially to locations of major cultural importance.

Except for these basic characteristics, metadata for digital images, unique identifiers and watermarking methods are embedded into the Advanced Digital Image Repository. The Advanced Digital Image Repository is implemented using the IBM DB2 Universal Database (Version 7.1, Fix Pack 3) with the AIV (Audio, Image and Video) and XML Extenders [12].

3. METADATA – DIGITAL IMAGES AND IPR MANAGEMENT

The specific tables and fields, which are used for the repository of the digital images, were selected in the basis of the next requirements:
- The user requirements of the Hellenic Ministry of Culture.
- The international standards for describing, characterizing and identifying digital objects.
- The international standards for managing and protecting the Intellectual Property Rights.
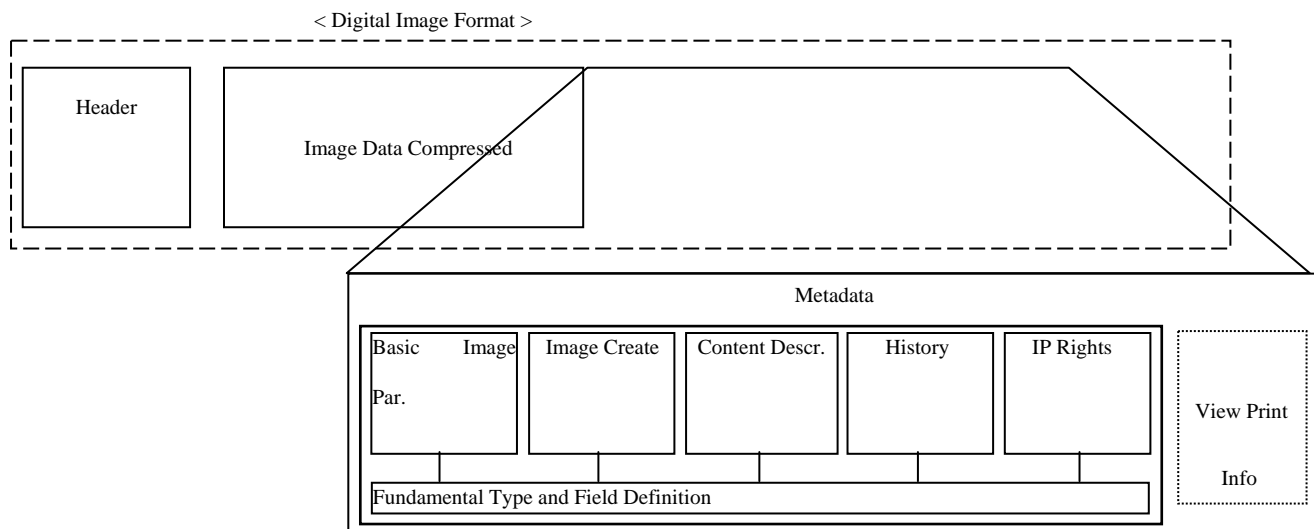

Figure 2: DIG 35 Metadata for Digital Images

The need of adopting the international standards is essential, especially for applications aiming cultural content exchange and on-line payment. Following these guidelines, the result is an open information system ready to be integrated into an expanding global, multi-language and multi-national network were the digital objects and the metadata coupled with them are uniquely identifiable and interchangeable.

The DIG 35 Specification "Metadata for Digital Images", Version 1.1 [3] (Figure 2) has a very important role in the selection of fields and tables, as far as the metadata for the digital images are concerned. This metadata standard is already being widely used from simple end-user devices to worldwide networks. The database structure has a special focus on metadata for the Intellectual Property Rights management. Consequently, the definition of sets of metadata, which are tightly coupled with the high quality digital surrogates of objects of the Hellenic cultural heritage, was a significant research area for this project. In particular, these sets are divided in five major sectors:
- Technical metadata.
- Image creation metadata.
- History metadata.

- Content description metadata .
- IPR related metadata.

Technical metadata are related to the image parameters, such as the image format, the image size, the compression method, and the color information. The image creation metadata include general information concerning the creation of the digital image. This information involves the time and date of the creation, the name of the creator, and information about the capturing device. The content description metadata contain descriptive information about the location, the participants (objects of people) and the capture time and date. The history metadata are necessary so as to identify and record the processing steps that might have been applied to a digital surrogate. This may help to avoid any further processing steps, and to identify independent objects in a composition of digital pictures. This set of metadata contains information whether or not a digital image is cropped, rotated, retouched, or suffered a color adjustment. The most important metadata set is the one related to Intellectual Property Rights management:

This table could be characterized as the lifeblood of the information system in the terms of Intellectual Property Rights protection and management (Fig. 3).
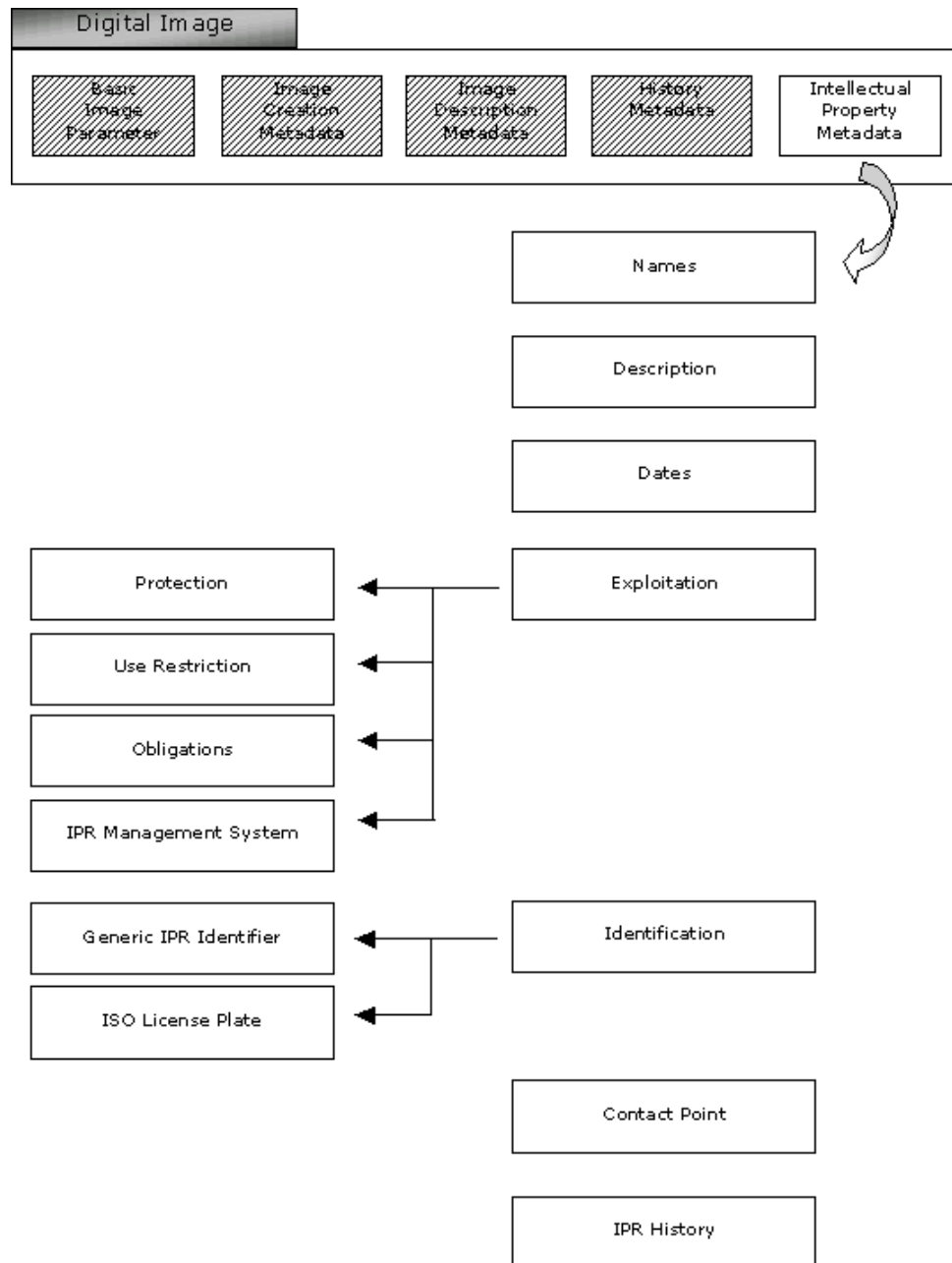
Figure 3: Intellectual Property Metadata Table

The Digital Image Repository is implemented in accordance with the described metadata sets. The metadata sets are incorporated through tables, fields, triggers and views in the Database. The Digital Image Repository will be distributed in regional Ephorates and Authoritative Agencies of the Hellenic Ministry of Culture.

4. WATERMARKING AND THE IMAGE REPOSITORY

Invisible watermarks are a proposed solution for the protection of Intellectual Property Rights and for dealing with the problem of illegal reproduction of content of multimedia objects [2]. Numerous watermarking schemes have been proposed and implemented and the performance evaluation has resulted in the selection of the best watermarking tools for this information system [4]. These tools have been obtained in SDK (Software Development Kit) format and are being automatically embedded in the Digital Image Repository using technologies that the database management system provides. The goal is the watermarks to be embedded in the digital images when the digital image is being stored in the

central repository. This general strategic planning is focusing on:

- Distributed management of the watermarking process and scheduled synchronization of the regional databases. The watermarks should be embedded in the digital image wherever a database exists and then the images should be transferred to the central Advanced Digital Image Repository in the premises of the Hellenic Ministry of Culture.
- The watermarks should be embedded only into the high quality digital surrogates of photographs of objects of the Hellenic Cultural Heritage.
- The watermarks should carry two ID numbers. The first number, will be common to all digital images, and will identify the Hellenic Ministry of Culture as the copyright holder of the digital image being processed. The second number will identify the instance of the transaction, containing the digital image and the persons that are involved to that transaction. This number is a reference that points to specific IPR related information in the database. This information is stored in the table described above, which relates to the IPR metadata.
- The watermark should be robust to attacks and detectable over Internet oriented environments. This is ensured in the basis of the exhaustive evaluation, which has been conducted [9].

This watermarking schema is being implemented via the creation of Dynamic Link Libraries (DLLs) that are available with the Dedicated User Interface. Watermarks are not the only means of dealing with the critical problem of copyright management and protection of digital surrogates of objects of the Hellenic Cultural Heritage, but are highly recommended for the detection of illegal copies and for the embedding of information identifying the copyright holder.

5. DEDICATED USER INTERFACE FOR THE REPOSITORY

The images are managed by archaeologists using a Dedicated User Interface. The archaeologists are members of the corporate Intranet already implemented for the Hellenic Ministry of Culture (HMC). The repository is designed to incorporate the user permissions and administrative rights of the HMC's corporate Intranet. The HMC's corporate Intranet is designed and implemented by the High Performance Information Systems Laboratory and, as a result, the integration of the digital repository to this already existing infrastructure is taken into account.

The selected GUI development tool is Microsoft Visual Basic 6.0. Custom-developed dynamic libraries (DLLs) are extensively being used for the interconnection of the GUI with the database system. The capabilities supplied to the end user are briefly presented here:

- Import and export high quality digital images. Search for digital images and metadata and management of image metadata.
- Watermarking: The watermark is embedded transparently into the digital image. An insertion of a digital surrogate into the central database causes the automatic embedding of the appropriate digital watermark. During this process, two kinds of information are incorporated into the digital image. A standard copyright ID number that actually declares Hellenic Ministry of Culture as the copyright owner of the digital image and the unique digital object identifier of the image. The user has the capability of initiating the detection of a watermark.
- The application under development creates XML documents containing information from the advanced digital image repository. For this project the usage of XML [15] is recommended as the standard metadata interchange format [9]. The primary reasons are, XML is widely adopted as an Internet-based language. XML is highly extensible and device independent. In addition, XML is language and application independent. The XML documents are produced by user-defined SQL queries. The XML documents are based on an XML schema, which derives from the XML schema of the DIG35 Metadata Standard. The XML schema is modified in accordance to the specifications of the Hellenic Ministry of Culture. The output files are automatically populated with XML Template Files (Figure 3). These template files interconnect the database schema and the XML schema. The presentation of the results is accomplished through the transformation of XML files to HTML, using Microsoft XSLT Processor or the IBM XML Lightweight Extractor (XLE) [16].

```
<?XML VERSION="1.0" ENCODING="UTF-8"?>
<!DOCTYPE METADATA SYSTEM "DIG35.DTD">
<METADATA TYPE = "SINGLE" TIMESTAMP = "VALUE"
XMLNS = "HTTP://WWW.DIGITALIMAGING.ORG/DIG35/1.0/XML">
  <BASIC_IMAGE_PARAM>
    <BASIC_IMAGE_INFO>
      <FILE_FORMAT>
        <FORMAT_TYPE>
           [MMDBSYS.FORMAT(BASICIMAGE.IMAGE)]
        </FORMAT_TYPE>
      </FILE_FORMAT>
      <IMAGE_ID>
        <UID>
           [BASICIMAGE.DOI]
        </UID>
      </IMAGE_ID>
      <IMAGE_SIZE>
        <WIDTH>
           [MMDBSYS.WIDTH(BASICIMAGE.IMAGE)]
        </WIDTH>
        <HEIGHT>
           [MMDBSYS.HEIGHT(BASICIMAGE.IMAGE)]
        </HEIGHT>
      </IMAGE_SIZE>
    </BASIC_IMAGE_INFO>
  </BASIC_IMAGE_PARAM>
</METADATA>
```

Figure 3: Sample XML Template File

8. WEB BASED IMAGE LIBRARY

The Web Based Image Library is structured upon a Digital Image Repository. This digital image library is not only the gate of the information system to the world, but also provides numerous basic and advanced services. The basic features include:

- News and announcements dynamically exported by the announcements database of the Hellenic Ministry of Culture (HMC). This MS-Access database is maintained by trained staff of the HMC and collects all the news and announcements of the authoritative agencies of the ministry.
- Information and helpful documents.
- Online help that depends on the web page the user is exploring. The online help ranges from simple text to animated examples of how the system works.
- Except for the basic features the advanced web based digital image library implements advanced services for the user. These services include:
- Registration forms collecting user data that are stored directly to the Digital Image Repository. This data will be the basis for statistics and reports and the future extensions of the system.
- Advanced methods for searching the digital multimedia content. The search methods consist of:
  - o Search for metadata and free text. The user is expected to fill the required fields. The fields are referencing specific database columns and produce dynamic web pages. The user is able to search the database using relational constraints (AND, OR) and free text queries.
  - o Query by Image Content (QBIC). This advanced feature allows the user to send a query to the database in terms of color and/or layout. The layout search provides for queries that are created by the user and represent basic shapes (circles, rectangular,…) filled with the preferred color. For example a user could search for an image shaped like a yellow disc and the query will result to several images representing the sun.

- Zoom-in collections are those that give the flexibility to the Internet user to zoom in certain images. These collections are firstly presented by zoom-able thumbnails and some basic information.
- Copyright notices and protection. The digital images are protected by digital watermarks that identify the HMC as the copyright holder of the digital content.
- A thematic based catalogue that organizes the digital content in categories. The user is able to refine his search using the content categories.
- Collections and Selected Cultural Presentations. Selected digital collections are created and enriched with accompanying information. They refer to specific cultural topics of common national and international interest.

8. 1 Implementation Principles
The most advanced web technologies were exhaustively tested and evaluated. The largest part of the web site interface is built on pure web technologies, avoiding impressive but high bandwidth technologies like Flash or Quick Time VR. The principles of human computer interaction and usability were taken under high consideration.

Dynamic web pages, for the search results, are created with PHP. The descriptive pages of the artifacts (containing a large image and detailed information) are created using Java technologies implemented as UDF functions in the database. The news pages are generated, managed and updated using JSP and Java technologies, powered by an Application Server. Zoom in collections are presented through Java applets.

The database repository of the digital images and their metadata is implemented using the IBM DB2 Universal Database, with the assistance of the IBM DB2 XML and AIV Extenders. The system provides support for the streamlined implementation of the aforementioned XML schema and the advanced services for searching and retrieving digital images. QBIC searches are created by Java applets with the support with the AIV extenders.

All the above implementation decisions were taken after a thorough evaluation of the technologies that could be used for each case. A special consideration was taken for the decision of the use of Dynamic Pages (generated on the fly each time the user asks for them) and Static Pages Dynamically Generated (when data are added or updated). A very important factor in the technologies evaluation was the interoperability, scalability and security of the end product (at least at the server side).

8.2 Implementation Paradigm
The Query by Image Content search engine is an advanced feature that allows the user to explore new ways of accessing the database content. Three main services are implemented:
- Query by Image predominant color and histogram.
- Query by Image layout.
- Query the database using a user specified digital image.

These advanced search tools are developed through the combination of web technologies such as PHP and Java Applets, embedded SQL, C++ and Dynamic Link Libraries (DLLs), all presented with simple HTML and XML forms.

For example the Query by Image Layout is considered (Figure 4). The user is able to draw his query in terms of colored rectangular and circles. Considering that the Web Server is installed in a different platform than the Database Server, in order to improve performance and interoperability, a Query by Image Layout involves the following steps:

- The user draws his query with the support of a Java Applet that provides for basic shapes and colors.
- A temporary file is created in the Web Server with PHP graphic libraries and commands, which is representing the user-driven query.
- A COM object is instantiated by PHP. The COM object is an ActiveX Control, which is a system registered Dynamic Link Library (DLL). The COM object is created with C++. The main functions of the COM object is the connection with the database, the necessary file format conversions of the temporary image (so as the transparency colors to be preserved), the import of the image in the database and the disconnection from the database. The image is stored in the database externally and as a result a temporary file is created in the Database Server. The COM object is also compiled as .so (shared object) file for UNIX platforms.
- An SQL statement is used for fetching a result set using success indicators and percentage of similarity, deleting the temporary fields and tuples.
- PHP is used to format and present the result-set in HTML tables and to delete all the Web Server's temporary images.

- The QBIC Layout search is a multi-user search tool.

```
<HTML>

    <? PHP

      $SQL = "SELECT imageid, mmdbsys.qbscorefromstr

      ('texture file=<SERVER, \"$tmp$Serverfile\">', IMAGE)

      from prax.basicimage order by 2 asc");

      $TemporaryFile;
      $TemporaryFileNames;
      $DeletionRoutines;


      $instance = com(QueryImage.Query);
      DLL instantiation
      Connect with Database
      File Format Conversion
      Import Image to Database
      Disconnect from Database
    ?>




    <!APPLET>
            LayoutSearch.class
            ColorSelection.class
            ColorSlider.class
            ColorSquare.class
            ColorPercentageGuage
            BrightnessSlider.class
            ImageButton.class
            ImageCanvas.class
    </APPLET>



</HTML>
```

Figure 4: Query by Image Layout – Structure

9. INDICATIVE SCENARIO

The main purpose of the indicative scenario is to present an overview of the system's everyday use. It is assumed that:
- Training of the necessary personnel of the authoritative agencies of the HMC is completed successfully.
- The system is installed in three selected organizations supervised by the HMC.
- Certain time for adjustment and performance optimization has passed.

In everyday use the system supports:
- The distributed and simultaneous insertion, deletion and management of digital surrogates in the selected organizations.
- Automated and imperceptible embedment of watermarks for the copyright protection of the digital content. The copyright notice "COPYRIGHT HMC" is incorporated in the images as a watermark.
- The insertion of more than twenty high quality digital images per day and per organization.
- The overall daily number of digital images is sixty.
- Once a month and when the HMC's network is not overloaded, these digital images are safely transferred through the network to the central database server.
- Low quality, watermarked copies are automatically presented through the "PRAXITELIS" Web Site. The Internet user is able to query the database by image content and layout.

The above scenario is based on the assumption that a trained user in every organization will be using the system five hours per day.

10. CONCLUSION - FUTURE EXTENSIONS

The general system architecture is based on the effective combination of several techniques and new technologies. Web technologies, databases, copyright management and protection techniques were exhaustively evaluated and studied. The main innovation of the system is that it is a real application that serves the digitization, management and protection of digital cultural heritage in an interoperable and distributed way.

The future extension of the system should aim at the enhancement of its educational purpose and the implementation of value-added services for the digitized material.

Firstly the web-based digital collection should be multilingual. Until today the content is presented only in Greek and this is an important drawback. The site should be translated at least in English, Italian, German, French, Spanish and Portuguese.

Enhancement of the educational purpose of the Web-based digital collection could be achieved by implementing on-line educational games and contests based on the digital content. In addition, a user should be able to create his personal digital collection based on common facts and characteristics of the cultural objects. For example, a user wants to create a collection consisting of ancient plates that refer to Ulysses and his imaginary journey. The user already has the ability to search for plates (layout search with a circle as a query) and refine his search by querying for metadata related to Ulysses. The user should as well have the ability to store his queries and selection of results for future use.

Implementation of value-added services for the exploitation of the digitized material is supported by the sustainability of this system. These services could aim specifically to the financial exploitation of the content. A system for online sales of digital images has already been pre-designed. In addition, a B2B system dedicated to museums and other cultural organizations is under consideration. Furthermore, electronic editions and CD-ROMs should be supported.

An online content management system is also under consideration. A Java-based user interface will assist archaeologists and photographers in managing digital images over the web.

11. REFERENCES

[1] ACM Publications, "Intellectual Property in the Age of Universal Access", 1999.

[2] Computer Science and Telecommunications Board, National Research Council, "The Digital Dilemma – Intellectual Property in the Information Age", National Academy Press, 1999.

[3] Digital Imaging Group, "Inc: DIG35 Specification - Metadata for Digital Images. Version 1.0", August 30, 2000.

[4] Fabien A. P. Petitcolas, Ross J. Anderson, "Evaluation of copyright marking systems".

[5] F. Bartolini, G. Bini, V. Cappellini, A. Fringuelli, G. Meucci, A. Piva, M. Barni, "Enforcement of Copyright Laws for Multimedia Through Blind, Detectable, Reversible Watermarking".

[6] F. Bartolini, R. Cardelli, V. Capellini, A. De Rosa, A. Piva, "Digital Watermarking: A Solution to Electronic Copyright Management Systems Requirements", WWW9 Culture Track, Amsterdam 2000.

[7] G. E. Kalamaras, D. K. Tsolis, G. K. Tsolis and T. S. Papatheodorou: Evaluation of Watermarking Techniques and Implementation of a Metadata Schema, for the Protection and Management of the Copyrights and Intellectual Property of Digitized Material of the Hellenic Cultural Heritage. Cultural Heritage Networks Hypermedia and Medici Framework Day, 16th - 19th September, 2000, Milano (available on-line at http://www.cultheritage.polimi.it/events/benire00/html/speakers/tsolis.htm).

[8] G. E. Kalamaras, G. K. Tsolis, D. K. Tsolis, P. A. Zafiris and T. S. Papatheodorou: Intergrated Environment for the Management, Protection and Exploitation of Digitized Material of the Hellenic Cultural Heritage. WWW9 - Culture

Track, May 16th - 19th, 2000, Amsterdam (on-line at http://www.medicif.org/dig_library/StateArt/Museums_Case_Study/Tsolis/tsolis.html).

[9] G. K. Tsolis, D. K. Tsolis and T.S Papatheodorou, "A watermarking environment and a metadata digital image repository for the protection and management of digital images of the hellenic cultural heritage", International Conference on Image Processing 2001 (ICIP 2001), Image Processing and Cultural Heritage, Thessaloniki, Greece, October 2001.

[10] G. K. Tsolis, D. K. Tsolis and T.S Papatheodorou, "An Advanced Digital Image Repository and an XML-based application for Intellectual Property Protection and Management of Digital Objects", WebNet 2001, Orlando, October 2001.

[11] George Voyatzis, Ioannis Pitas, "Protecting Digital-Image Copyrights: A Framework", IEEE Computer Graphics and Applications, pp. 18-24, January/February 1999.

[12] IBM, "IBM DB2 Digital Library", http://www.ibm.com.

[13] M. Kutter, and F.A.P Petitcolas, "A fair benchmark for image watermarking systems", Electronic Imaging '99, Security and Watermarking of Multimedia Contents, Sans Jose, CA, USA, 25-27 January 1999.

[14] N. Nikolaidis, I. Pitas, "Digital Image Watermarking: an Overview".

[15] The Apache Software Foundation. (2001). *The Apache XML Project*. http://xml.apache.org/.

[16] The IBM AlphaWorks. (2001). *XML Access Service Lightweight Extractor (XLE).* http://www.alphaworks.ibm.com/tech/xle.