

A Project for the Management and the Publication of the University of Milan's Egyptological Archives

Glauco Mantegari¹ and Patrizia Piacentini²

¹Department of Computer Science, Systems and Communication
University of Milan Bicocca
Milan, Italy

²Department of Scienze dell'Antichità
University of Milan
Milan, Italy

glauco.mantegari@disco.unimib.it

Abstract

In 2005, an interdisciplinary project started that was aimed at improving the accessibility of the University of Milan's Egyptological Archives, as well as at experimenting with the correlation and retrieval of information by means of innovative modalities. This paper describes the data acquisition methodology and the development of a prototypal Web portal based on the use of Open Source technologies. The project is still in its initial stages with regard to the quantity of documents acquired up to now and the development of the data storage, management, and publication system. This first experience represents a starting point for new experimentation and for the integration of additional modules and mechanisms for the improvement of the navigation experience through the available data. In particular, it is possible to sketch a future architecture that will constitute the definition of a domain ontology.

1 Introduction

The project for the management and publication of the University of Milan's Egyptological Archives is born of the collaboration between the Chair of Egyptology of the Università degli Studi of Milan and the Department of Informatics, Systems and Communications of the Università degli Studi of Milan Bicocca. It has been developed thanks to a grant of the CARIPLO Bank Foundation.

The purpose of this interdisciplinary research is twofold: on the one hand, it aims to increase the accessibility of a sizeable quantity of archives of fundamental importance to the knowledge of Egypt and the Ancient Near East, as much for the specialists as for a public fascinated with the subject; and on the other hand, its goal is to furnish researchers in the computer science sector with an application domain and a set of data enabling them to experiment with the correlation and retrieval of information by means of innovative modalities.

In the following section (2), we will attempt to illustrate the nature and the characteristics of the archives, explicitly following the history of acquisitions in the last few years and describing the collection of material on which we have thus far concentrated the work of digitization. Section 3 is dedicated to the description of the digitization and data acquisition procedures. Section 4 examines the Web portal development for the management and diffusion of the digitized material. In the final section (5), we will discuss the acquired results and potential for future development as it relates to the definition and use of specific domain ontologies and innovative mechanisms for the structuring and retrieval of information.

2 The Archives

The Egyptology Library and Archives came into being in 1995, the year in which a policy for the acquisition of major private collections was inaugurated. These collections have significantly enhanced the first nucleus of volumes and documents that was established in the University between the 1950s and 1990s. The Masanes library was acquired in 1995, and it contains monographs and catalogues of exhibits and museums for basic research. In 1999 the Edel library (Piacentini 2006) was acquired. It is made up of over 15,000 titles of books, periodicals, and offprints, coming from at least eight libraries of famous scholars, collected by the great German Egyptologist Elmar Edel during more than 50 years of scientific activity. It contains, in addition, rare and unpublished documents from the archives of this scholar (letters, philological notes, ancient and new photographs, facsimiles of hieroglyphic inscriptions, etc.).

Subsequently, in 2001, a generous benefactor donated a substantial part of the Varille library to the University, consisting of around 400 volumes of great value, belonging to the French Egyptologist Alexandre Varille, or bought on the market of antiquities (Piacentini 2003). **In 2002**, the archives of the same Varille were bought, containing more than 40,000 photographs taken from the middle of the 19th to the middle of the 20th centuries, as well as excavation journals, philological notes, facsimiles of hieroglyphic inscriptions, and over 3,000 letters and miscellaneous documents. These resources allow us to reconstruct the history of Egyptology and the relationship between Egypt and Europe in the first 50 years of the 20th century. These archives also contain those of Victor Loret, the Professor of Varille, formed of multiple materials: travel diaries; manuscripts relating to the

Museum of Boulaq, the first public collection of Egyptian antiquities in Egypt; specific studies in history, philology, and archaeology as well as natural sciences and musicology; unpublished works on the excavations at Saqqara or in the Valley of the Kings (Piacentini and Orsenigo 2005); and a rich array of correspondence, photographs, squeezes, and watercolors.

In 2004, the Mariette-Brugsch Archives were acquired. They include, among other things, 135 unpublished letters sent to Heinrich Brugsch mainly from Auguste Mariette, the French Egyptologist who founded the Antiquities Service in Egypt. The same year, the antiquarian bookseller *Ars libri* in Boston offered to the University a collection of papers belonging to the English Egyptologist George Fraser. In 2006, the same bookseller made the gift of many volumes and offprints from the library of Erich Lüddeckens, containing a great number of egyptological documents, letters, and other material. Finally, some months ago, the University bought the Archives of the Danish Egyptologist Wolja Erichsen.

The wealth of such a collection housed at the University of Milan is obvious. It is also a tremendously complex task to digitize and archive the material; the economic and structural resources at the disposal of the project were and still are insufficient to allow us to think of undertaking immediately a truly comprehensive operation. For these reasons, from the beginning, we had to set priorities and propose viable goals. Thus we have concentrated the first phase of work on the section of photographs taken by Alexandre Varille during his travels in Egypt and North Africa and his visits to the principal Egyptian collections, both public and private, in France and Europe. This portion of the collection is homogeneous and of great scientific value. In fact, Varille, organized the approximately 40,000 photographs he took in a most meticulous fashion. For many of these, the negative and print still exist, and for others the glass plate. The photographs relate to the main Egyptian archaeological sites, including monuments and objects that were subsequently destroyed or scattered far and wide: of great interest are the images of the Temple of Karnak, the total photographic

rendering of the walls of around 60 Theban tombs, as well as the unique excavation photos taken at Saqqara and in the Theban area.

3 The Digitization Process

The work of digitization, recognition, and classification of subjects portrayed in these photos is only at the initial stage and is being carried out by graduate Egyptology students or by postgraduate students of the Specialization School in Archaeology (*Scuola di Specializzazione in Archeologia*). The digitization procedure (Figure 1) requires the use of an A3 scanner for the acquisition of the photographic plates and of an A4 scanner for the acquisition of traditional negatives. The images, scanned at a resolution of 600 dpi, then undergo photo-retouching procedures effected by an expert in the treatment of digital images and are then archived in precise progressive sequence in two different formats: a master copy in non-compressed TIFF format, and a consultation copy in JPEG format. The resulting files are preserved on both external hard disk and on DVD back-up.

Each image is catalogued using a database created with Microsoft Access. Choosing this product was a result of two main considerations: on the one hand, Access is easy to use and provides simple tools supporting the development of forms for data access, insertion, and modification; on the other hand, it meets the needs of students, the majority of whom own laptops operating in Windows with Microsoft Office by guaranteeing them easy installation and use of the database.

The database offers a main internal filter with several dropdown lists containing pre-defined sub-menus that support a uniform insertion of data where possible. In order to further facilitate the use of the database, several specific functions have been developed in Visual Basic, such as the dynamic loading of an image preview or of the search filters.

The experience, nevertheless, has brought to light several differences in the quality of work carried out by the

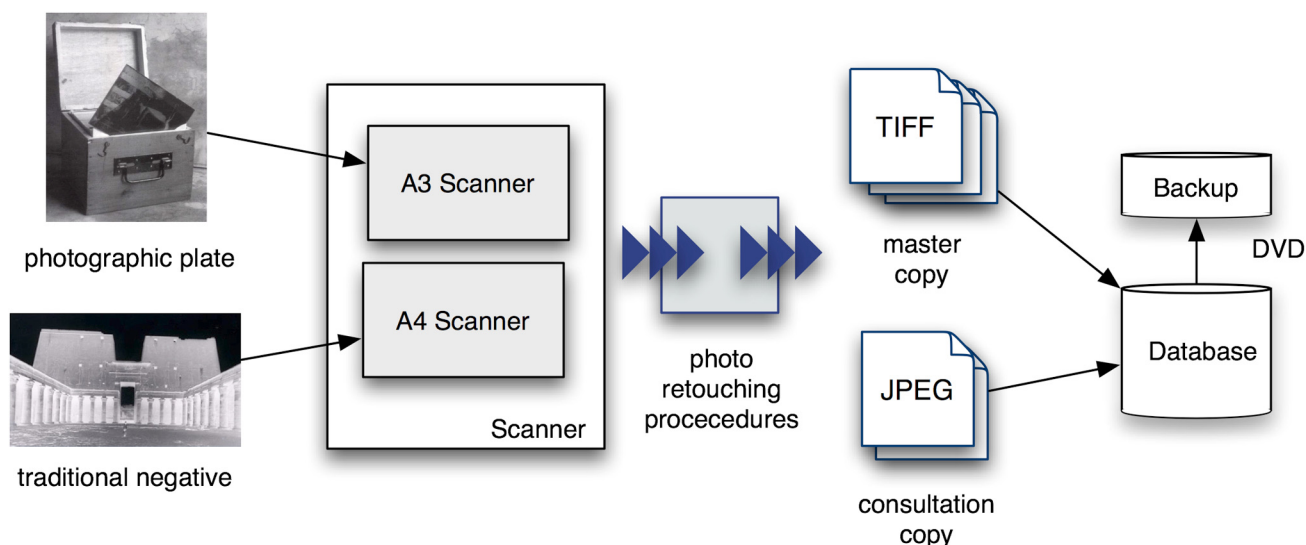


Figure 1. A diagram of the digitization procedure.

students. Such differences are attributable, in the majority of cases, to the dissimilar preparation of each student, which can sometimes cause problems in the quality of the insertion of data. Thus, a key focus of the future development of the project necessitates a thorough analysis of this problem. This will inevitably bring about an improvement in the organizational and technical aspects of the project. As such, it constitutes a work in progress at the stage of definition and implementation, and therefore cannot be discussed here in detail.

4 The Web Portal

All the images scanned and the information describing them are archived on a server of the Università degli Studi di Milan, site of the portal for the project (<http://users.unimi.it/egitto>).

The portal comprises two main areas, one for public access and one with limited access. The public access area contains several pages related to the history of the acquisitions and to the concise biographies of the three major scholars to whom the archives originally belonged (Edel, Varille, and Loret). Others will be added in the future. The limited access area, which can be accessed by means of a username and password assigned upon request by the library administrators, supports the consultation of progressively filed digital material. This restricted access control mechanism is necessary in order to safeguard copyright on the material, which is at present unpublished for the most part.

The portal has been created entirely by means of Open Source technologies, which were employed mainly because they offer the possibility of modifying the application code, thus allowing us to fully personalize the applications according to our needs. These technologies are also sufficiently developed at this time and so readily accessible as to be widely adopted in a growing number of analogous projects and systems. In particular, we have used the well-known LAMP architecture (Figure 2), an acronym derived from the software names that comprise it (Linux, Apache, MySQL, and PHP). The choice of this architecture guarantees optimum reliability in terms of the integration of components and the wealth of online documentation, which will

facilitate the finding of solutions to problems that may occur during system development.

The MySQL database combines excellent stability with unrivaled speed in data management compared to other Open Source and commercial database servers. In the database, only the data in the image files has been memorized, while the scansions of the images themselves have been memorized in the file-system of the server, using the unique code with which the image was catalogued, as the file identifier. This is a much more efficient solution with respect to the possibility of incorporating the images directly into the database, consistent with the specific functions offered by MySQL. Importing the new images is made so easy because it is possible to upload the images and the relative information at different times, as well. Furthermore, since, in this way, the number of operations the database has to carry out is diminished, the system responses are accelerated. It is important to realize that, in fact, our present implementation foresees the use of only one server upon which both the database and the Web server operate. This requires that we make every attempt not to overtax the resources of the system, which could result in a deterioration of its performance.

The structure of the archived data in the database is very simple: to each image is assigned all the information present in Access in the database, divided into four main categories: subject data, photographic re-take data, technical data, and file processor data.

Subject data consists of:

- The Inventory Number, or the unique identifier of the photo;
- Typology of the subject with predefined values (e.g., the site, the residential area, the temple, the necropolis, the tomb, the pyramid, the archaeological find, etc.);
- The photographed location;
- The period to which the represented subject belongs (e.g., Prehistoric, Pharaonic, Greco-Roman, Coptic, Islamic, or subsequent period);
- The dynasty to which the represented subject is attributed, active only in the case of the Pharaonic Period;
- Description, in free-text format;
- Bibliography on the subject;

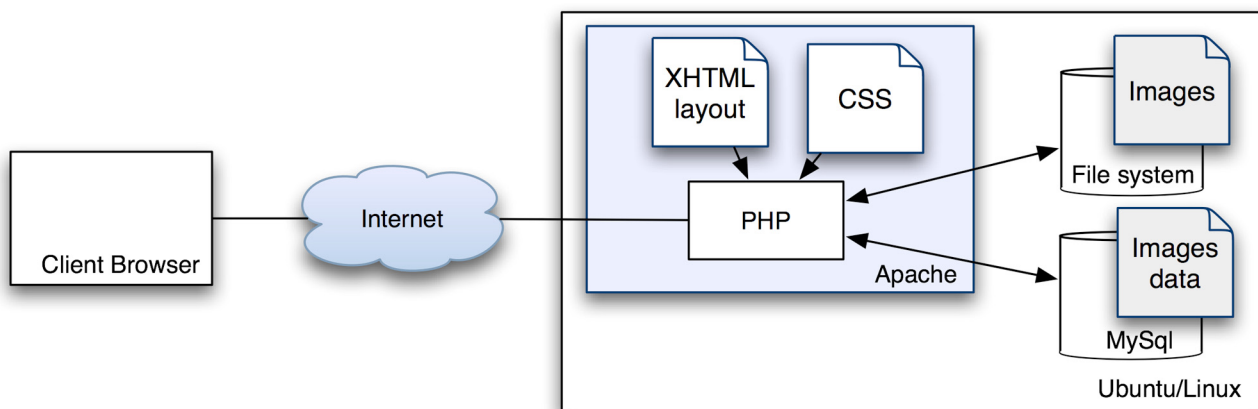


Figure 2. The Web portal architecture.

- Notes.
- Photographic re-take data consists of:
 - The name and surname of the photographer;
 - The date on which the photo was taken;
 - The method of acquisition, with pre-defined values (e.g., the Edel, Loret, or Varille archives, gift, deposit on loan, etc.).

Technical data consists of:

- Type of photographic mounting (photographic paper, post-card, photographic plate, negative);
- Photograph dimensions (height and width);
- State of preservation, with set standards of quality (e.g., illegible, deficient, sufficient, good, excellent).

File processor data consists of:

- Name and surname;
- Date of the initial processing of the file;
- Date of the file's last revision;
- Description of potential problems relative to its processing, what happens when the subject portrait is not easily identifiable or possible to describe.

This structure offers the advantage of being uncomplicated, thus facilitating the students in their task of cataloguing the material. The importation process from the database in Access to the database of the server is further facilitated because it not only ensures the exact correspondence in design between the two databases, but also prevents excessive rearranging of the stored data.

The importation process is made easy by means of an inter-exchange format. Among all the alternatives experimented with, the simplest format has been the Comma Separated Value (CSV), which consists of a simple text file in which all the field values are separated by commas. The Access database is exported in CSV, using for this purpose the special function made available by Access; then the CSV

file is imported within the MySQL database. The importation is carried out directly through the Web, by using a PHP script, which we have created to allow the user to select the CSV file from his own computer and to upload it onto the server, by means of a very simple and effective automatic procedure.

The photographic scansions, however, do not lend themselves to a direct importation via the Web because of their size, in terms of memory. The operation would be too taxing. So, in this case the copying of images is relegated to the DVD unit on the server and online, publication-size previews are made possible through a batch process.

Querying the database on the archived photos is made possible thanks to the use of a Web interface done with PHP. This interface (Figure 3) offers a mask that is composed of most of the fields with which the database is also composed, specifying values or ranges of values (as in the case of dynasties) in accordance with the necessities of the research. For example, it is possible to query about something that relates to all the images taken in the region of "Thebes" related to the Twentieth Dynasty.

The results are presented in thumbnail form (Figure 4) that can potentially be subdivided on more pages of 20 elements each. Each thumbnail carries beneath it a brief description, containing the inventory number and the locality portrayed.

Access to the complete file of the photograph (Figure 5) is gained by selecting the relevant thumbnail. Within the file, the image is shown in greater size, so that the elements may be visualized better. The hyperlinks, placed at the bottom of the file, ultimately allow the user to return to the results page with the thumbnails, or to carry out a new search by means of the appropriate mask.

While developing the portal we paid particular attention to the Web accessibility guidelines, such as those defined in the World Wide Web Consortium (W3C)

realm of activity, and especially in initiatives such as the Web Accessibility Initiative (WAI) (Caldwell et al. 2006). In Italy, moreover, the principles of accessibility are also recognized and promoted on the legislative level, specifically to public organization sites (see for example the so-called "Legge Stanca" of 2004 and subsequent decrees).

This requirement is very important because it guarantees a satisfactory use by the public, a goal in keeping with the basic future directions of the project. At the moment, our efforts to satisfy the main requirements for accessibility have been

Figure 3. The database query interface.



Numero di inventario: 114
Località: da Saqqara



Numero di inventario: 115
Località: da Saqqara



Numero di inventario: 133
Località: Saqqara (?)



Numero di inventario: 136
Località: da Saqqara



Numero di inventario: 140
Località: da Saqqara



Numero di inventario: 143
Località: Saqqara (?)



Numero di inventario: 145
Località: da Saqqara



Numero di inventario: 182
Località: da Saqqara



Numero di inventario: 184
Località: da Saqqara



Numero di inventario: 185
Località: da Saqqara



Numero di inventario: 186
Località: da Saqqara



Numero di inventario: 197
Località: Karnak



Numero di inventario: 198
Località: Karnak



Numero di inventario: 199
Località: Karnak



Numero di inventario: 200
Località: Karnak



Numero di inventario: 224
Località: da Karnak nord



Numero di inventario: 227
Località: da Qurnet Mura'i (Tebe ovest)



Numero di inventario: 231
Località: Nag' el-Medamud



Numero di inventario: 232
Località: Nag' el-Medamud



Numero di inventario: 233
Località: Nag' el-Medamud

Figure 4. An example of a list of results.

Dati soggetto

Numero d'inventario:	34
Tipologia:	tempio
Località:	File
Periodo:	greco-romano
Descrizione:	Colonnato occidentale del I cortile, sezione meridionale (PM G-L; 27-32), rilievi di Augusto e Tiberio offerenti davanti a teorie di dèi; in basso a destra, ingresso della scala sotterranea che conduce al nilometro.
Bibliografia:	ROBICHON-VARILLE 1937, pl. 150
Note:	Scatola 5. Cfr. 9, 10, 16, 17, 18, 28, 29, 30, 33, 174, 175.
Dinastia:	



Dati ripresa

Fotografo:	Fernand Bisson de la Roque
Data:	
Acquisizione:	Archivio Varille

Dati tecnici

Larghezza (mm):	179
Altezza (mm):	130
Bianco e nero:	S
Colori:	N
Presenza di negativo:	N
Stato di conservazione:	ottimo

Compilazione scheda

Compilatore:	Luigi Prada
Data compilazione:	22/01/2005
Data revisione:	22/01/2005
Problemi:	N
Descrizione problema:	

[Torna alla pagina dei risultati](#)

Figure 5. A complete file of a photograph.

focused on the creation of pages that clearly separate content from presentation. In particular, we have created pages in XHTML 1.0 Transitional code for indicating content, and the CSS 2.0 style sheets for portal layout. Thanks to other small precautions, we have attained the WAI standard of level "A" and it is our intention in time to also satisfy the necessary requirements to improve our levels of accessibility, ultimately seeking to reach level "AAA."

5 Conclusions and Future Developments

The project described here is still in its initial stages with regard to the quantity of documents acquired and the

development of the data storage, management, and publication system. However, this first experience, besides having brought us to the achievement of a completely functioning and functional version of the database and its own system of Web management, represents a starting point for new experimentation and for the integration of additional modules and mechanisms for the improvement of the navigation experience through the available data. In particular, these mechanisms are, on the one hand, focused on a refinement of the research functions (in a pull information access scheme, with the user actively querying the system), and on the other hand in terms of an enhancement of navigation paths through the autonomous suggestion of contents of interest by the system (in a push information access scheme,

with the system suggesting the visualization of additional contents).

We can tackle these developments in a particularly innovative manner by defining the ontology of a specific domain. This will represent a fundamental endeavor in the execution of the project and will involve experts in the field, connected with the Chair of Egyptology, and experts in knowledge representation, connected to the Laboratory of Artificial Intelligence. An ontologically-based photographic description can in fact allow us to execute content searches in a much more precise way than is currently possible, and above all will allow us to easily navigate between contents, not only thanks to predefined paths, such as those offered by form-based searches, but also in a much more similar way to what we are using to navigating the Web—using a structure based on dynamically set hyperlinks, for example on the basis of analogies or correlations between a currently visualized content and a suggested one.

To this end, we are thinking of using the well-known Web Ontology Language (OWL) (McGuinness and van Harmelen 2004) that, among other things, makes it possible to annotate the images themselves, in accordance with the design suggested by W3C (Van Ossenbruggen et al. 2006). The architectural plan that integrates these components could be the one presented in Figure 6.

An example may clarify the advantages that could be derived from ontologically represented content. Let us suppose that we have just executed a search for identifying photographs representing the Valley of the Kings, and that we have individuated the images of the tomb of Amenhotep II. Data associated with these images show that we are dealing with a tomb. To the tomb is associated a geographic position; the tomb itself may contain other important locations (corridors, rooms), which could also be part of a larger complex (in this case, the Valley of the Kings). To a tomb we may associate the persons buried in it, the collected remains, and the archaeologist who has discovered it. These people furthermore are conveniently associated to a specific role: for example, Amenhotep II was a king. Beyond

the name of the king himself and those of his parents and children, as well as the historical period in which he lived, we can correlate information such as, eventually, the length of his reign and the names of his predecessor and successor. If all this information refers to an ontology, we could easily visualize, for example, the tomb of the successor of Amenhotep II, or visualize all the other tombs discovered by Victor Loret, who is precisely the one who discovered the tomb of Amenhotep II. By associating the definition of a domain ontology to a hypertextual navigation of the contents, we could thus attain flexibility in the contents search itself, which would be very difficult, if not impossible, by implementing a more rigid structure based exclusively on relational data.

Acknowledgments

We would like to thank Christian Orsenigo and Luigi Prada for the supervision and co-ordination of the data acquisition phases; Andrea Bonomi and Stefano Bottaro for the development of the portal prototype, and Joyce Myerson for the support in the translation.

References Cited

- Caldwell, Ben, Slatin, John, and Vanderheiden, Gregg, eds. 2006. Web Content Accessibility Guidelines 2.0 (W3C Working Draft). Updated 27 April 2006. <http://www.w3.org/TR/WCAG20/> (Accessed 1 July 2006).
- McGuinness, Deborah L. and van Harmelen, Frank, eds. 2004. OWL Web Ontology Language Overview (W3C Recommendation). Updated 10 February 2006. <http://www.w3.org/TR/owl-features/> (Accessed 1 July 2006).

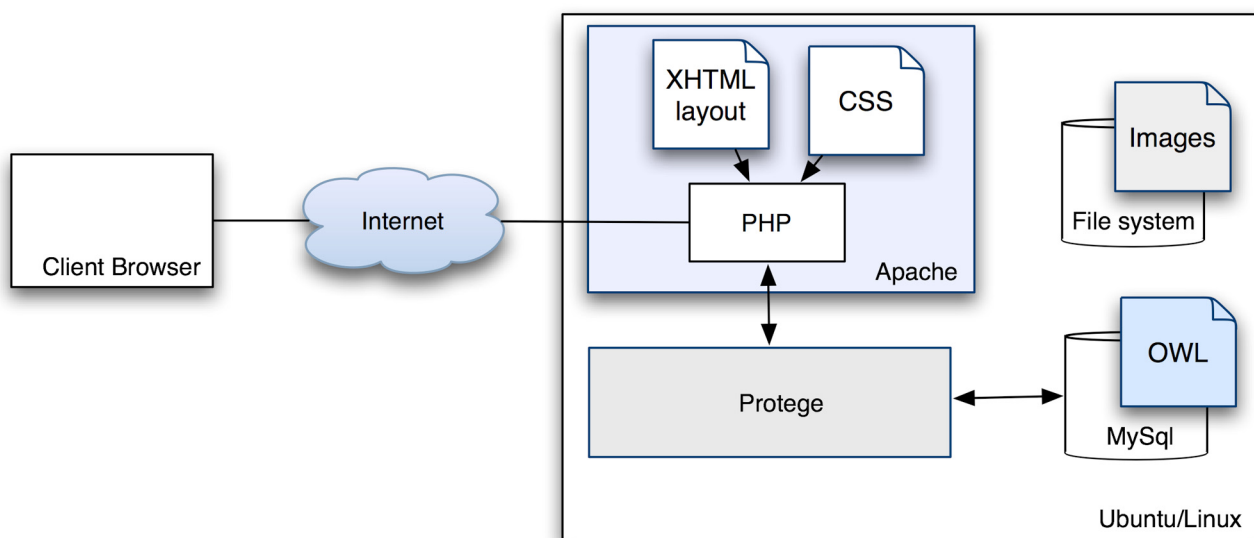


Figure 6. A hypothetical future architecture of the system.

Piacentini, Patrizia, ed. 2006. *Gli archivi egittologici dell'Università degli Studi di Milano. 1. Il fondo Elmar Edel*. Milano: LED, Il Filarete 230.

Piacentini, Patrizia. 2003. La biblioteca di Alexandre Varille e le prime fotografie aeree dell'Egitto. In, *Atti del VII Convegno Nazionale di Egittologia e Papirologia, Siracusa, November-December 2001*. Corrado Basile, Anna Di Natale (eds.), pp. 127-137. Quaderni del Museo del Papiro XI, Istituto Internazionale del Papiro.

Piacentini, Patrizia and Orsenigo, Christian. 2005. *La Valle dei Re riscoperta. I giornali di scavo di Victor Loret (1898-1899) e altri inediti*. Milano: Università degli Studi di Milano-SKIRA.

Van Ossenbruggen, and Jacco, Troncy, Raphaël, Stamou, Giorgos, Pan, and Jeff Z., eds. 2006. Image Annotation on the Semantic Web (W3C Working Draft). Updated 22 March 2006, <http://www.w3.org/TR/swbp-image-annotation/> (Accessed 1 July 2006).