The Database of the St. Catherine's Library Conservation Project in Sinai, Egypt

Athanasios Velios, Nicholas Pickwoad
Camberwell College of Arts, University of the Arts
London, UK

Abstract

The St. Catherine's monastery holds one of the most important collections of Byzantine and other early manuscripts, consisting of 3307 volumes. Most of them retain their original or early binding structures, few of which have not been studied in any detail. Camberwell College of Arts started a detailed survey of the condition of the manuscripts in 2001. Information about the book-binding structures and the condition in which they are preserved is stored in a database. In this paper we describe the principles of the database design, focusing on the need for constant updating of the database structure. We discuss the use and interpretation of drawings which allow recording of complex elements of book-binding. We also describe a fast, semi-automatic way for transferring our data from paper to the database. We finally examine the potential use of a global standard for Byzantine book-bindings and the possibility of an object-oriented database structure.

Figure 1. Interior of the St. Catherine's library

Introduction

The St. Catherine's monastery located in the desert of Sinai is the oldest active Christian monastery in the world. The monastery holds one of the most important collections of early manuscripts, consisting of 3307 volumes. The manuscript collection (Figure 1) has long been appreciated by scholars not only for its palaeographic value, but more recently also because of the variety of early book-binding structures that have survived in the remote and extremely dry conditions of the desert. In other similar collections, it is often the case that early Byzantine bindings have been replaced in more recent years by current binding structures, thus removing any evidence of the original Byzantine book-binding. This is not the case in the St. Catherine's library. For this reason, the collection is a unique and precious source of information for scholars studying early and especially Byzantine/Greek book-binding. However, the remote location of the monastery has until recently, prevented the extensive study of the collection.

Background to the St. Catherine's project

Camberwell College of Arts started a detailed survey of the condition of the manuscripts in 2001. The survey is dedicated to recording the book-binding structures and the condition in which every element of the structure is preserved. The recording is done at the library by specialised conservators. So far about two thirds of the collection has been surveyed. The manuscripts are recorded in great detail. Over 1000 fields are checked and if necessary, filled in per book. We are designing a database (St. Catherine's database) to store the recorded information for effective archiving and querying. The main functions of the database are to:

1. Prioritise conservation work at the St. Catherine's library in the future,
2. Allow us to estimate the resources and time needed to complete the conservation work,
3. Assist with future studies of Byzantine book-binding,
4. Act as a guide for future designs of databases for similar objects.

In addition, results are extracted from the database to be used for a number of projects. The first one is the boxing project, whereby certain fragile manuscripts will be stored in boxes for their protection. The boxes are produced according to the measured dimensions of each book. Information about the fragility and dimensions of each book is extracted from the database. Another project for which the database is proving useful is the manuscript digitization project (Brunner, 2004). During digitization, handling the book is necessary for each individual page to be aligned correctly opposite the camera. Occasionally the book structure is so fragile that it does not allow safe handling for photography. The St. Catherine's database provides data which describe how easy (or difficult) it is for a book to be handled for digitization and identifies problems in advance of actual digitization.

Previous examples of book-binding databases

Since research on Byzantine book-binding has been limited until now, there have been very few efforts to develop a database to store such specialised information. Researchers currently working in Mount Athos, Greece (Mount Athos Manuscripts Digital Library), are developing a database to store information relevant to Byzantine manuscripts. However, their database is orientated towards the bibliographic details of the manuscripts rather than their structure and condition. Major libraries such as the Library of Congress and the British Library have produced databases relevant to manuscript conservation and binding respectively, but to the authors' knowledge, none of them is directly relevant to Byzantine manuscripts and they tend to store very general information about the books. With our target being the development of a well-structured database for our detailed data, we had to consider a number of issues including that we did not always know in advance, what the data which we were about to record, would be. For example, the complex bookmark structures often observed on Byzantine manuscripts have not been extensively studied before. Therefore there was no existing database structure to store such information. The challenge facing our database design is for it to be built in a flexible way to allow unknown elements to be accommodated while the survey progresses.

Data collection

A record of the structure and condition of each manuscript is made at the library. Observations made regarding the manuscripts are recorded on a paper form which is specially designed to assist the recording process (Pickwoad, 2004). We chose to use paper as a means of recording because it is more flexible in comparison to a "strict" computer form. Modifications on the paper form are easily implemented on-site by the surveyors. The same modifications on the computer would involve specialised staff in a rather time-consuming process. Drawing is a helpful way to describe newly observed binding components. The use of paper is particularly beneficial in this case as surveyors are more confident with drawing on paper than on the computer screen. To automate the inputting process, as we will explain in the "Data inputting" section, the paper forms are digitized.

Digitization of paper forms

Here we briefly describe the digitization procedure which has been explained in detail in the past (Velios and Pickwoad, 2004). The paper forms are scanned in a resolution of 200 dots per inch?? (dpi) using a document scanner and are saved as compressed JPEG (Joint Photographic Experts Group file format) files alongside the database on the project's server. We use the JPEG format for disk space economy, as the total amount of pages to be scanned is more than 33.000. We are not concerned with the reduced quality of the compressed JPEG images as the scanned pages will not need to be reproduced in high quality in the future. Their quality is sufficient for inputting data, as we will describe in "Data inputting".

Considerations for our database design

The data accommodated on the paper form is grouped logically based on the components of Byzantine/Greek book-binding, starting from general ones (e.g. Endbands, Boards etc.) and continuing to the more specific. The database is designed along this principle, so that book conservation and paleography scholars can easily become familiar with it. Alterations to the paper form are inevitable during recording, as components not studied before are now observed. Our database is designed in such a way that alterations are absorbed within the current structure and no restructuring of the database is needed. Before presenting our adopted structure we will explain what alterations we had to consider.

Alterations of definitions

An interesting comparison between different versions of the paper form, before and after the alterations, can be found in Pickwoad (2004).

In a simplistic database design the alterations on the form can be problematic, as the definitions of components recorded change among the different versions thus demanding database restructuring. To emphasize the potential problem it is worth considering that the survey was scheduled to be completed over a period of 5 years whereas results from the database, such as information about the dimensions of books for the boxing project, were needed from the beginning. So the database needs to be functional while being under development and at the same time able to accommodate changing definitions.

If we try to categorize the alterations needed, we need to consider three different forms:

1. *Updating.* Available variations for a specific component have been enriched with new ones. An example is recording different types of endband attachment to the boards, where newly observed attachment types need to be added to the existing list. Figure 2 shows the part of the form where we record the type of board attachment. The old version of the form is on the left and the new one is on the right, which includes two more variations.

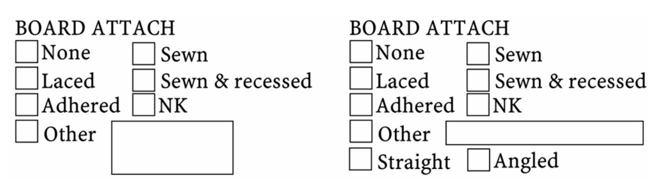

Figure 2. Example of changes in board attachment variations

2. *Generalization*. Components which were initially considered as unique variations, now need to become part of group of variations; hence their definition is generalized to include the new items. For example, endbands were assumed to be located always on top of the head and tail edges of a textblock. Their definition did not include any other possibility. However, when an endband was observed which was recessed in the textblock, this definition had to be updated. The updated definition includes a new property which defines whether the endband is recessed into the textblock or not. Figure 3 shows the part of the form where the endband is recorded. The old version of the form (left) shows the cross-section of the endband located on top of the textblock. The new version of the form (right) shows the endband recessed into the textblock.

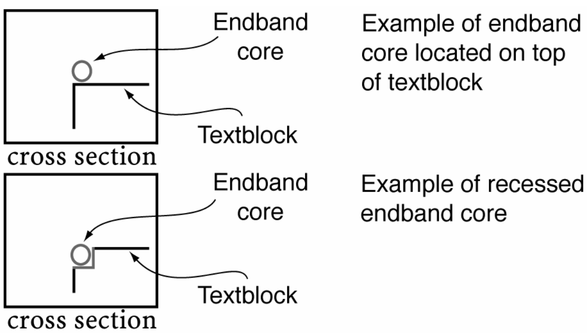

Figure 3. Example of generalized definition of endband

3. *Redefinition*. Newly observed variations of a component are too complex to be embedded into the existing structure by simply creating a new property, as it happens with generalization (described in point 2). Therefore, they are redesigned. An example of this is the way bookmarks are recorded. In the early versions of the form they were recorded within a text field, whereas in the current version, they are recorded through a series of properties on their type of attachment, decoration, material, color etc.

The above examples illustrate the need for the database to be flexible and be designed in such a way that every version of the changing definitions can be accommodated. In the next section ("Developing the database structure") we will describe how we are addressing this issue.

Multiple components and properties

It is often the case that multiple similar components are observed on a book. For example, two different endbands may sometimes be found on the same book. Also, there are many cases where specific properties of a component have more than one value, which is often the case with bookmarks having multiple colors. This introduces another level of complexity to the structure of the database as such relations ("many-to-many" relationships) need to be possible.

Data from drawings

The schematic drawings used in the survey hold a great deal of information which is useful for interpreting and categorizing the components they describe. We need to emphasize that this information cannot be used by the computer as it is not in any textual or numerical form. Automatic extraction of such data may be possible using highly specialized programming techniques; however, it is not our intention to develop such a tool. In the next section we will describe how the adopted database structure along with human interpretation can help to address this problem.

Developing the database structure

In our quest for the optimum way to build the database, we faced a dilemma on whether to use a relational or object-oriented database. We chose the relational model because of the availability of standard software tools for it. We recognize however, the potential of object-oriented databases and we briefly refer to them at the end of this section.

The structure of the St. Catherine's database closely reflects the structure of a Byzantine book. In general there are several *components* observed in a book (e.g. endband). These can be described by certain *properties* (e.g. material, color). The properties are relevant to the book-binding techniques and the preservation condition. They have several *variations* each (e.g. for material thread, silk). There are properties which are common to many components and therefore an attempt to classify them has been made as discussed next.

Database libraries – sub-libraries

We will describe the adopted structure by using an example. A common property for almost every component of the binding is the material. The range of materials observed varies. To keep this common information consistent, it has been stored in a separate database table which acts as a database reference library for every material observed. This database library is being enriched during the progress of data inputting. *Simple alterations* on the form are possible by updating the library with the newly observed variations.

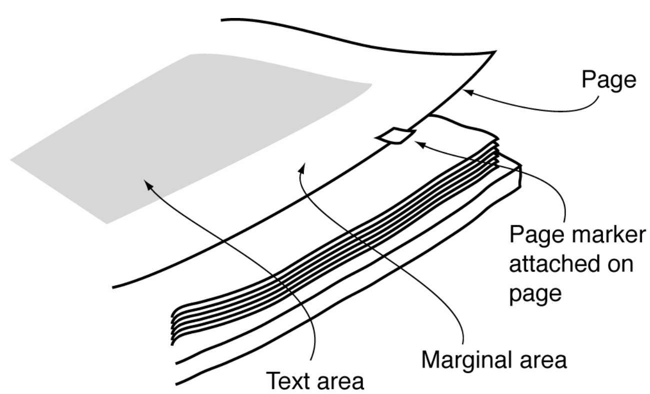

Figure 4. Example of page marker

However, not all materials apply to every component. For example page markers (Figure 4) are normally made of thin and flexible materials, such as leather or paper, as they are attached on the edge of the pages. It has never been observed and it would be unlikely to find rigid materials such as wood used for page markers. To avoid offering unnecessary options for recording the page markers' material, a materials sub-library has been compiled specifically for page markers, limiting the potential materials to the ones likely to be observed. Database libraries allow control over the updated variations of a property, whereas sub-libraries define the components that every variation is applied to. A graphical representation of this can be seen in Figure 5.

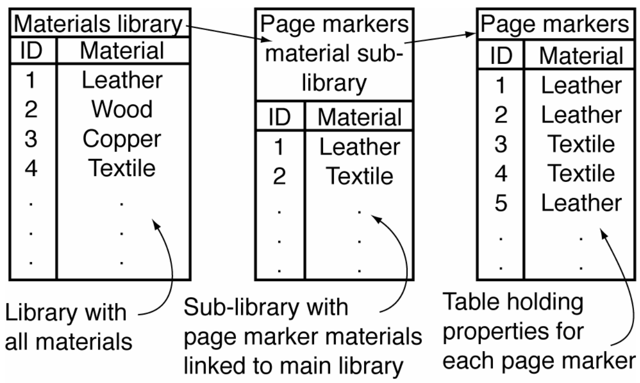

Figure 5. Example of using database libraries and sub-libraries

The same principle has been followed throughout the database whenever a list of variations for a component is needed. This allows for several reference libraries to be created covering almost the whole range of variations and components of a Byzantine book-binding structure. Occasionally these libraries have only one variation/record, as is the case with the endband location in the early version of the form which could only be located on top of the textblock. Although producing a database library for a single record may seem unnecessary, it is particularly helpful at the time where an unexpected additional variation is observed. Using this approach, *generalization alterations* of the form are possible without modifying the database structure, provided that the altered components depend on a database library for each of their properties.

Cases where *redefinition* of certain components is necessary cannot be tackled in our structure. There may be ways of automating such alterations but these would almost certainly make the database structure too complex and given that in our case such alterations are not frequent, it has been decided to incorporate them by manually restructuring the database.

Multiple components and properties

Multiple components or properties can be linked to a single book or component respectively using the adopted structure. An additional table needs to be introduced to store the many-to-many relationships. Again, the values of the individual properties are drawn from the different database libraries. Figure 6 shows how the structure described earlier can be modified to include multiple components.

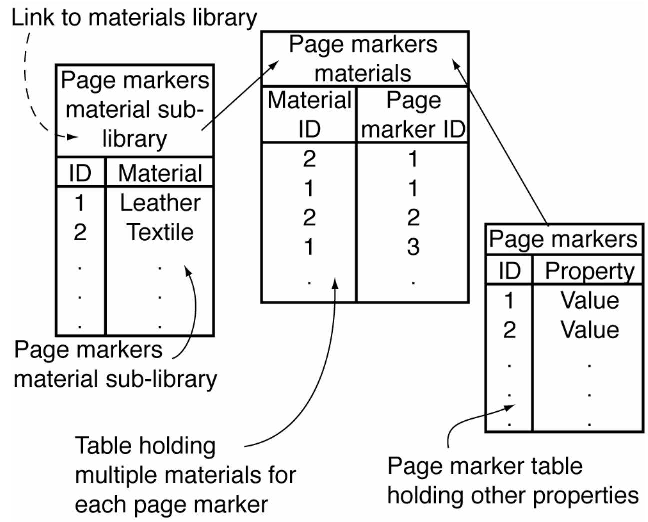

Figure 6. Example of multiple properties (many-to-many) relationships

Images

Modern database design software allows for storage of any kind of digital file through Binary Large OBject (BLOB) types, including images. In the St. Catherine's database, we have chosen not to use BLOBs. Instead we save the images on the hard disk as separate files and only a link to the file is stored in the database. This approach is marginally slower when it comes to the speed with which images in the query results are returned. However, its main advantage is the fact that the files are easily accessible through the operating system and not through the database, and thus can be quickly modified. Another important advantage is the fact that backup time is dramatically reduced, since the overall database size is kept small. Regular backups of the images are not necessary as they are not updated as often as the database.

As mentioned earlier, processing conservation-related data depicted in drawings is only possible after the drawings have been interpreted to simple numerical and textual forms. The schematic drawings are produced in a standardized way. An example of this standardization is shown in Figure 7 where the endleaves are drawn in an informative rather than naturalistic way. Such drawings are stored "as is" and are quickly interpreted by a user who converts them to numerical and textual values during data inputting. There are cases where newly discovered variations need to be drawn, but a predefined standard does not exist. The interpretation of these images has to be done case by case. We are planning to perform the individual interpretations after the survey has concluded, to allow for a complete view of the variations. At the end of the survey, our task will be to group and classify these variations and produce new libraries with the newly discovered features.

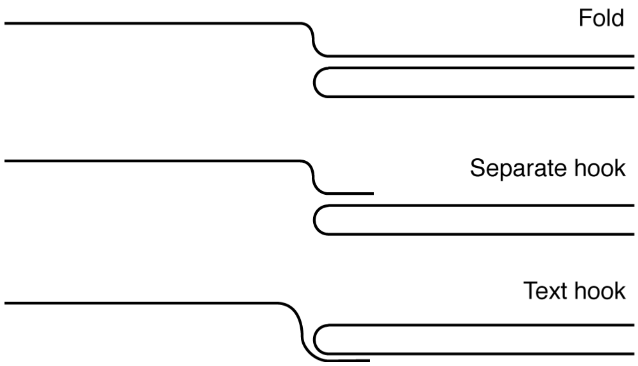

Figure 7. Example of standardized drawings of endleaves

Object-oriented approach

In the previous sections we mentioned the complexity of the database due to the large number of fields. As we will explain in "Data querying", this also affects the automatic querying tool devised for the interface.

Object-oriented databases have been an alternative to the relational model when the data structure is particularly complex (Kim, Nicolas and Nishio, 1989) The St. Catherine's database can be implemented using the object-oriented approach in a beneficial way. Many of the complexities introduced in the relational approach are due to the fact that different properties of, say, the potential materials are associated with the same component. For example, paper and parchment being different materials are characterized by completely different properties. However, they can both be part of the book's textblock. An object-oriented approach is ideal for a structure recording such complex data. In an object-oriented database both paper and parchment can be defined as objects which can be part of the textblock object, but at the same time have completely different properties.

Although there is a lot of discussion about the lack of standardization of the object-oriented database systems, our intention is to produce an object-oriented database to store Byzantine book-binding structure information. We have been experimenting with the open source tools offered by db4objects (www.db4objects.com). We are also interested in a comparison between the usability and speed of the relational and the object-oriented databases.

Database interface

Alongside the main relational database development, the production of a practical database interface has been one of our main concerns.

Data inputting

Transferring the data recorded on the form to the database is not an easy task. It is estimated that conventional database interfaces would allow a page of average complexity to be input in the database in about 2 min by a fairly experienced user. However, manual inputting increases the risk of error on the user's behalf. To address this problem and speed up the inputting procedure, a program has been developed for semi-automatic data inputting. The program projects the scanned image of a page onto the screen. It leads the user through the page by highlighting the individual fields sequentially. The program ignores the empty fields and only leads the user to where recorded data exists. The program is instructed to seek the information for every field at specific coordinates within the page, which have been measured in advance. It automatically identifies marked checkboxes and areas with hand-written notes. Data from checkboxes is transferred to the database automatically whereas handwritten notes need to be typed in. User confirmation is needed for inserting or updating each record. The total time for inputting data from pages of average complexity can be as low as a few seconds. The risk of inputting erroneous data is minimized, as records are inserted and updated automatically by the program.

Data querying

Returning results to specific queries is one of the main functions of the database. We have designed several tools to build database queries bearing in mind that the audience who will use the database is not familiar with database systems. We also considered the fact that the database structure will be constantly updated and therefore a flexible and customizable interface is needed.

Due to our previous familiarity with the Microsoft .NET programming environment, we have decided to use Active Server Pages (ASP) to build the database interface. The immediate advantages this gives are:

1. Production of an interface accessible through a web-browser, hence no specialized software or knowledge is necessary on the user side.
2. The database is developed in Microsoft SQL Server. ASP allows easy but sophisticated programming of MS SQL Server through specialized set of commands.

Querying the database through the interface is done in two steps. The first is to form the query as an SQL statement and the second to pass the query to a webpage which presents the data. To form the query the user has a number of options:

1. Select a predefined query from a list.
2. Build the query by choosing criteria through a set of dropdown lists.
3. Manually type in or edit an SQL statement.

The dropdown lists mentioned in point 2 is the most versatile way to build a query for the inexperienced user. Each dropdown list corresponds to a specific property of a component. The list holds all variations that can be found on the specific property. By choosing variations from different properties, the user can detect patterns about the frequency of any possibly related components among the manuscripts.

The complexity of the database often leads to queries with particularly long SQL statements. Despite the efforts to keep the database consistent, inevitably certain database elements are exceptions to the overall structure and can potentially lead to erroneous SQL statements when these are built automatically through the dropdown lists. Our efforts currently focus on further generalizing the automatic querying tool to successfully include every possible query without compromising ease of use.

When it is necessary for a query to return images among the results of a query, only part of a scanned page is usually needed. In these cases, an ASP script is called which opens the image file, crops the image to the requested area and returns the file as a new JPEG image. The information about the cropping boundaries for the image are stored in the database where the location of the small drawing within the larger page is defined by a set of x, y coordinates and the height and width of the area.

XML schema - SVG

XML (Extensible Markup Language), being strongly supported by the World Wide Web Consortium (W3C), is a major step towards standardizing the way information is presented. The field of Byzantine book-binding and conservation lacks a universal model for describing the binding structure and the condition of a book. The St. Catherine's database will be a pioneering source of information towards the standardization of such a model. Our intention is to work with experts in the field and use the database as a starting point for defining a model and a terminology for Byzantine book-binding through a global descriptive structure implemented through XML (XML schema). This will then allow comparisons to be made between books kept in different libraries around the world, which are not always physically accessible to scholars.

SVG (Scalable Vector Graphics) is a graphics format based on XML which has recently become a W3C recommendation. Our drawings are mostly linear, schematic depictions of variations of components, and can be easily implemented in the SVG format. The combination of the SVG format with a suitable XML schema would assist standardizing the way Byzantine bindings are described.

Conclusions

In this paper we discussed the development of the St. Catherine's conservation database. Our discussion focused on the problems arising from constantly updating the structure of the database. We explained how we addressed this problem by using reference database libraries and sub-libraries for each component of the binding. Following this approach we showed how additional variations to existing properties can be added and how generalized descriptions of binding components can be achieved without the need to restructure the database.

Another important discussion was the production of an innovative interface for transferring data from paper to the database, which visually leads the user to the fields that need attention. Our interface not only reduces the risk for user errors but also minimizes the time needed for inputting each page.

We explained the benefits of using drawings for recording complex data during the survey, which are that they are quick and descriptive.

Finally, we considered the usefulness of XML and SVG for standardizing the way Byzantine/Greek book-bindings are described and consequently ensure the long term preservation of our data.

References

1. Nicholas Pickwoad, The Condition Survey of the Manuscripts in the Monastery of Saint Catherine on Mount Sinai, The Paper Conservator, 28 (2004), pg. 33.
2. Athanasios Velios, Nicholas Pickwoad, Collecting Digital Data on Paper, 2nd International Conference of Museology - Technology for Cultural Heritage, Mitilini, Greece, (2004).
3. Earleen Brunner (Ed.), The monastery digitizing programme receives foundation assistance, Sinaiticus, St. Catherine Foundation, (2004), pg.4.
3. Won Kim, Jean-Marie Nicolas, Shojiro Nishio (Eds.), Deductive and Object-Oriented Databases, Proc. DOOD'89, Kyoto, Japan, (1989).

Biographies

Athanasios Velios received a degree in archaeological conservation from the Technological Educational Institute (TEI) of Athens, Greece, and a PhD from the Royal College of Arts in London, UK, in computer applications to conservation. He has previously been a lecturer at the TEI conservation course. He was recently appointed as a research fellow at Camberwell College of Arts, London, UK, mainly working on the development of the database for the St. Catherine's Conservation Project.

Dr. Nicholas Pickwoad trained with Roger Powell as a book conservator. He has been Advisor on Book Conservation to the National Trust of Great Britain and editor for The Paper Conservator journal. He taught book conservation at Columbia University Library School in New York and was Chief Conservator in the Harvard University Library. He is now the project leader for the St. Catherine's Monastery Library Project and also teaches courses on the history of European book-binding.