

Providing Web Access to Archival Photographs

Diana Lomarcan
Systems Librarian
Columbus State University
lomarcan_diana@colstate.edu

Reagan Grimsley
University Archivist
Columbus State University
grimsley_reagan@colstate.edu

Abstract

In 2002 the CSU Archives received a grant to digitize and preserve its collection of over 2,000 historic photographs of the Columbus area. The digitization project included the development of a web-accessible database containing descriptive information and thumbnail images of the photographs. The history, design, implementation, and future of the database and search system will be discussed.

Introduction

The CSU Archives holds over 2,000 historic photographs of Columbus and the Chattahoochee Valley area of Georgia. Most of these photographs are from the Historic Photographs Collection, but some are from other collections housed in the archives. While called the Historic Photographs Collection and being comprised of mostly photographs this collection also contains postcards, negatives, and slides. The term photographs will be used throughout this paper to refer to images of all formats in this collection.

For many years the only descriptive access to these images was provided by one printed list of images. This list, like the collection, was in accession order (the order in which the images were added to the collection) with no attempt to index or group the photographs by age, format, donor, or subject. The list was typewritten copy, many years old, and not frequently updated,

so many photographs had no descriptive information available at all.

What descriptive information was available was very brief. There had been no attempt to include provenance for the images, nor were usage restrictions clear. Physical details such as the format, size, or color of the images were also excluded.

Each image was given a unique identification number as it was added to the collection. For photographs housed in the Historic Photographs Collection the folder and photo numbers were used. Box and photo numbers were used if the photograph was housed in another collection. Because these identification numbers were assigned to the photographs in the order they were processed, photographs with no relation to each other were often housed in the same folder. Given this method of arrangement and the lack of indexed, organized, and up to date finding aids, it was very difficult to find any individual photograph in the collections.

Beginnings

In the fall of 2000 the librarians at CSU decided they had to develop a better system for finding historical images. Initial plans were to simply retype the existing finding aid into a Microsoft Word document and then add descriptions of the newer photographs. Given the size of the collection, however, this seemed cumbersome and minimally useful. While the internal "find" function in Word could provide a form of keyword searching it was

not considered effective or efficient. Something more flexible was needed.

Systems Librarian Diana Lomarcán recommended creating a database for this information. A database would allow more precise searching, easier scalability, and better consistency in data such as names and subjects. The original database was conceived of as a tool for library employees to use. There was no thought given to making this database available to the public.

A small relational database was created using Microsoft Access, the campus standard for databases. Lomarcán created the database to include the data in the original finding aid, then added fields for physical description, provenance, subject headings, dates, and photographers. This was done with the knowledge that the data, though not included in the original finding aid, might be very helpful to researchers and could be added at a later time. The database was designed as a relational database to provide simple control over repetitive data such as the names of people and places.

Using an Access database also provided the option to include scanned versions of the images. Including images in the database would give two major benefits. First, finding a particular image would be easier because the user could see a small version of it and would be able to determine if it was truly the desired image. Second, having digital versions of historic photographs helps preserve the physical artifacts. By using digital images to view or browse the photographs there is less physical wear on the originals. This is particularly important with historic photographs as they can become fragile relatively quickly.

Library staff typed the information from the finding aid into the database. This was done from the printed finding aid. Staff did not

pull the photographs to verify the data and did not add data into any of the additional fields at that time.

The project grows

In the fall of 2001 Reagan Grimsley joined the CSU Archives and, on asking about developing better access to the Historic Photographs Collection, was introduced to the Access database. The database was still in its unverified and incomplete form.

On reviewing the database Grimsley proposed expanding it. He wanted to verify the existing data and add the missing data and images. He also proposed that public access to the database be provided through the newly revised and expanded Archives web site. In making this database publicly available care had to be taken to ensure that no copyrights on the photographs were violated. Therefore, only photographs that were out of copyright or that CSU had copyright authority for were to have images included. In addition, to help forestall unauthorized use of these photographs by others, only small, low-resolution images would be placed on the web. Full-size, high-resolution preservation copies of the images would be stored on CDs/DVDs in the Archives and would be available only by request.

To achieve these goals, Grimsley applied for a grant from the Georgia Historical Records Advisory Board. This board is concerned with the preservation of primary materials relating to Georgia history. The proposed grant had two levels. The first level was to physically preserve the photographs. This included the arrangement and identification of each photograph and also provided acid-free folders and mylar sleeves to protect them. The second level of the grant was to provide access to collection. The proposed online database would fulfill the second part of the grant.

A previous grant from the Historic Chattahoochee Commission had been used to create a “digitization lab” in the Archives. This lab included a high end PC workstation with 200 GB hard drive storage space, CD/DVD burner, and a 19 inch monitor as well as an optical scanner with flatbed, film, and slide capability. Adobe Photoshop 7 image processing software was purchased to assist in cleaning up the scanned images. This lab provided the method for creating digital versions of the photographs, but did not provide web access.

For web access Lomarcan recommended using MySQL and PHP on a Linux server. According to research this was the de facto standard for creating web-enabled databases. Although the first database was in Access format, this format rapidly becomes unstable in high volume situations and requires the user to have full write access to the directory

that houses it. This is not desirable on a public web server.

CSU Computer Information & Networking Services (CINS) agreed to provide a small Linux box and assist in the installation of Linux, Apache web server, MySQL, and PHP. CINS staff also found and installed the PHPMyAdmin system to help administer the MySQL database. In the end, CINS personnel agreed to take over responsibility for the security and maintenance of the server.

Once the server was in place the process of converting the database began. The database structure was tweaked slightly to compensate for losing the ability to declare relationships as they were done in Access. Several tables were added to serve as connectors where 0 .. ∞ relationships were possible (figure 1).

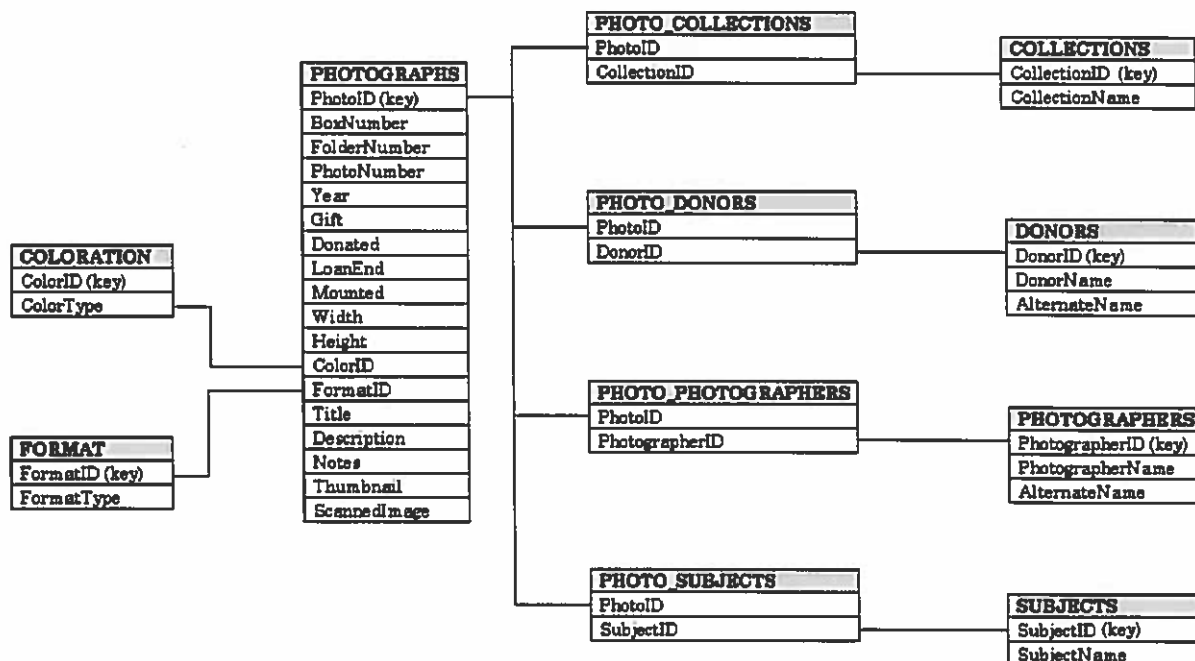


Figure 1: Entity Relationship Diagram

The data from the Access database was then uploaded into the MySQL database. This was done by exporting the data in delimited format and using the upload feature available in PHPMyAdmin. After the data was uploaded, Archives staff began verifying and completing the data.

The PHPMyAdmin interface is currently used for data entry. We are developing a secure web-based staff interface for database maintenance in the future. While PHPMyAdmin is fairly easy to use, it does presume a knowledge of SQL that we cannot guarantee in our future staff.

The original design of the database included a field for the thumbnail images of the photographs in the database itself. The idea behind including images in the database was that it would be easier to keep track of the images if they were housed with their descriptive data. This proved to be unworkable, however, so the design of the database was changed to include a field for the file name for each image. Images are now housed in a directory reserved for the images from the Historic Photographs Collections. These images conform to the naming conventions f###p## or b###p## where the #'s indicate the folder and photo or box and photo number pairs used to identify the original photographs.

The Public Face

The public front-end of the database is a simple set of HTML and PHP web pages. They consist of an HTML search form and a PHP page that runs the search then returns and formats the results. Server side includes and cascading style sheets are used to display the results in the same format used in the rest of the Archives web site.

The PHP code includes conditional statements that test for the presence and type of search terms. Numerical searches for

folder, photograph, or box number are performed as exact searches. Text searches on the title and description fields are performed using fuzzy matching.

Several loops retrieve the data from the database and set the display formatting. The description for each photograph is placed in an individual table. A conditional statement tests for the presence of an image and adds the appropriate table cell to hold the image if necessary. This was done to prevent empty image boxes from appearing where thumbnail images are not yet available. Because the images are all housed in one directory the only data needed from the database to retrieve the appropriate image is the file name. The PHP code includes the rest of the HTML tag.

A loop using variable variables creates the SQL queries used to retrieve the data from the 0 .. ∞ relationships. This loop creates queries that find all instances where a connection between the PhotoID (unique identifier for each photo) is connected to the IDs for collections, donors, photographers, and subjects. These queries are then looped individually to retrieve the appropriate values matching those ID keys.

The results display ends with a link that leads back to the search screen.

Project Results

The web-based version of the Historical Photographs Collection is up and in use. Currently there are very few photo images in the database, but Archives staff is continuing to add more. The search interface, though very primitive, has already proven to be successful for finding photographs requested by patrons.

The PHP used has turned out to be fairly portable. We have already added two other databases to the web. One is an Archival

database describing the contents of the Otis Burnham slide collection, a large collection of images taken over many years and several continents. The second database is for the Music Library, and is used to find information on the uncataloged portion of its sheet music collection. Both databases required very little change in the PHP code. Field names were changed for each database and some of the more complex result retrieval loops were removed because the two additional databases are not yet in relational format.

The success of this project has led to plans for the future. The primary plans for the Historical Photographs Collection involve data and searching. Data for each photo will be verified and completed and thumbnail images will be added where possible. More search functions will be added, such as the ability to search for multiple terms or over several fields at once. The ability to do exact left-anchored matching in text fields will be added to the current default fuzzy searching. Controlled vocabulary subject headings will be added to the database and made available through left-anchored searches. In addition, a system to break up large search results into groups of 10 or 15 per page will be added.

Even further in the future are plans to create similar web-searchable databases for other major collections including the Architectural Drawings Collection and the Columbus College Photographs Collection. Both collections are likely to include some form of linked graphics. An index database to the historical Industrial Index magazine collection in the Archives will be converted from its current format (dBase IV) to MySQL. We also hope to purchase a larger web server to support these databases and other Archival web projects such as EAD finding aids, streaming audio snippets of

oral histories, and scanned images of historical documents.

Thanks to

CSU Computer Information & Networking Systems – particularly Director Bill Johnson, for his willingness to provide hardware and to let us temporarily monopolize one of his staff, and Mack Ragan, for his willingness to jump in and learn Linux, MySQL, and PHP and then share his knowledge with us.

Vicky Adkins, who patiently retyped information for hundreds of photographs from the original typescript.

Georgia Historic Records Advisory Board for their financial support.

References

Apache Web Server: <http://httpd.apache.org/>

MySQL Database System:
<http://www.mysql.com/>

PHP: <http://www.php.net/>

PHPMyAdmin:
<http://www.phpmyadmin.net>

Welling, Luke, and Laura Thompson. PHP and MySQL Web Sevelopment. Indianapolis, Indiana: Sams Publishing, 2001.