

**GENERAL INFORMATION**

<b>Name of Library</b>	Natleigh College
<b>Address</b>	555 College Road
<b>City</b>	Collegeville
<b>State</b>	PA
<b>Zip Code +4 digit Extension</b>	19426
<b>Contact Name</b>	Leigh-Anne Yacovelli
<b>Contact Title</b>	Head Librarian
<b>Contact Phone</b>	555-555-5555
<b>Contact e-mail</b>	L.Yacovelli@natleigh.edu
<b>Federal Identification number of applicant</b>	00-8675-309-000-123
<b>U.S. Congressional District</b>	PA's 6 <sup>th</sup> District
<b>Type of Library</b>	Academic Library and Archives

**PARTNERS (required if there are partners in the proposed project)**

**Project partners are outside agencies or organizations who will be working in partnership on the project for the mutual benefit of the applicant and partner. Vendors under contract are not usually considered to be project partners.**

**List each partner's name and describe their roles in the project. Use a separate row for each partner.**

Partner Name	Role
Center for Presbyterian History	The Center for Presbyterian History is providing the Vass photos and the subject expertise. The archivist/subject specialist at the Center will be hired with grant money to assist in the creation of a local thesaurus, cataloging, and marketing.

**ABSTRACT**

**Briefly describe the purpose of the proposed Digitization project. Include the collection/item the project will digitize, and expected purchases and user outcomes/benefits. (Maximum 500 Characters)**

Digitization of items pertaining to the American Presbyterian Congo Mission during King Leopold II of Belgium's oppressive regime will enable people to access them regardless of location and time, provide for the discovery of the collections' contents to a broader audience, and contribute to the region's, the religious community's, and the world's historical memories. The grant will fund software at the college's library, the disbinding of a photo album, preservation-quality housing for the fragile collections, and staff.

**Total Cost of the Project:**

**Total LSTA Grant Funds requested (up to \$50,000):**

**Total Matching Funds (minimum dollar-for-dollar match for requests over \$30,000): \$0.00**

**NEEDS ASSESSMENT User Interest**

**Describe the extent of expected user interest in the materials you want to digitize. How was this interest ascertained? (Maximum 2,000 Characters)**

The college recognized that interest in a digitization project was multi-layered. Natleigh College's students use the missionary collections for primary source research. Several faculty members at the college, particularly those who teach anthropology and history, refer the students to the library's special collections and the Center's archives. Thus, student interest in these collections is high.

The college began as an African-American college in 1878, with ties to a local black Presbyterian congregation, which adds interest to this project by supporting the belief system of the people in the area. The college also receives numerous requests for information on the missionaries who attended the college. Sheppard's collection is one of the more highly requested given his status as one of the first African-American Presbyterian missionaries. This shows that there is strong support from outside the college community for the missionary collections, and digitizing the materials would improve access to them.

The fragility of the collections' items led to restrictions on the public's access, something neither the Center nor the college wished to continue. Digitizing the collections would enable the public to continue to access them without harming the original pieces. If digitization did not occur, then access to items in the collections would continue to decline.

**Describe the current accessibility limitations of the original resource. (Maximum 2,000 Characters)**

The collections are currently processed with folder-level finding aids. They do not offer detailed descriptions of the materials at the item-level. This limits users' understanding of the collections.

Accessibility limitations also include restricted access to a photo album in the Kellersberger Papers. Some photos in the album are no longer secured to the pages, creating the potential for damage, loss, or theft. The college restricts access to the original albums in the Sheppard collection due to brittle paper and fragile, fading photos. Photocopies of the restricted items are offered to researchers. However, the copies are an inferior quality. Digitizing the collections would provide far better quality copies.

The handling of the originals, even by those who are experienced and careful, facilitates damage, and speeds deterioration. This became readily apparent over the past year as workers attempted to assist an increasing number of researchers outside the area with copy requests. Digitization would provide a safe alternative. It will also enable the institutions to move the fragile originals safely into cold storage to diminish further image destabilization.

Other limitations to the collections' access include limitations on access into the buildings. The college allows visitor parking near the library. These reserved spots are limited in number and the length of time they can be occupied. The majority of library visitors must park across the campus in a visitor-only parking lot since the reserved spaces fill up quickly. Access to the special collections department is limited to the days when the college is open, and the public room closes earlier than the library. Visitors to the Center use on-street metered spaces or a pay-to-use lot a block away. The Center's hours of operation are six hours during the day for four days a week. Digitization of the collections will enable people to access them regardless of location and time.

### **Content Value**

**Describe what makes this material unique in format and/or content of the material, and why it should be digitized. If applicable, include details about how the materials in digital format might support K-16 education. (Maximum 2,000 Characters)**

The collections pull from primary source materials that document the 1898-1910 American Presbyterian Congo Mission, making them unique in format and content. The Sheppard papers contain photos of Sheppard and W. M. Morrison, another missionary, at a 1909 trial in Leopoldville charged with libeling the Kassai Rubber Company, as well as many photos of native people of the Bateke, Baluba, Bakuba, Zappo, and Zap tribes, and hymnbooks printed in Tshiluba and another unidentified language. The Vass photos present a wide range of topics, including the S.S. Samuel N. Lapsley I and II (ships that brought missionaries and supplies to stations in the Congo), mission stations and indigenous villages. They also capture examples of Belgian atrocities carried out on indigenous people. The Kellersberger oversized photo album contains his medical notes relating to diseases that plagued Congolese people and sketches of villages.

Digitization provides free access to all of the items by many researchers at the same time. It reduces physical handling of fragile items, and frees staff to plan new projects and work on preservation and maintenance of the collections. Researchers of the history of the Congo region during European colonialism, native fashions and living communities, and the experience of U.S. missionaries, would benefit from the digitization of these items and possibly lead to someone identifying the hymnbook language.

It supports K-16 education by providing teachers assigning student research projects free, immediate access to materials so students can explore using primary sources under the teacher's direction while still in class. It also increases the possibility of students returning to the digital site for research in other classes, or going to the physical collections to learn more. Library and Center staff will work with teachers to spread awareness of the collections, which in turn will make students more aware of their local history as it relates to world history.

**Describe the material's capacity for enhanced access through digitization, indicating how metadata, and possible full-text searching through the application of OCR, will reveal the content to the user. (Maximum 2,000 Characters)**

Physical access to many of the items in the collection is currently restricted currently due to their fragile nature. The hymnbooks would be scanned using ABBYY OCR software, which allows the contents of the scanned hymnbooks to be uploaded into CONTENTdm as XML files. The language used in the hymns is Tshiluba and an unknown foreign language. Since the hymns are well-known, the material would enable students of both African history and African languages to learn words in Tshiluba. Internet search engines would find the CONTENTdm's XML files, leading to potential recognition of the unidentified language in the hymnbooks as traffic increases from free online access to the digitized works.

Cataloging the collection-level records using MARC format would allow the school and the Center to enter the collection into their online catalogs, so people not familiar with the collection will find the items when searching for similar topics or names. CONTENTdm provides for the upload of metadata to WorldCat, which would join the contents of the digital collection to materials located in other places that are related by topic or names defined through the use of Library of Congress Subject Headings as a controlled vocabulary. This federated search capability would enable more researchers to find the collection.

Dublin Core (DC) metadata's item-level descriptions in CONTENTdm further improves the chance a researcher unfamiliar with the collection would find and retrieve items in the collection. Users will be able to search by the 16 core components of the DC Metadata Element Set required by AccessPA, which are title; creator; subject; description; date; publisher (repository information); contributor; type; format; identifier; rights; source; language; relation; coverage; and audience. Users will be able to tag materials, which staff will approve and add to the controlled vocabulary when appropriate.

**Provide details of searching for commercial vendors who provide the same material/titles, including the presence or lack of such entities. If title is offered by a commercial vendor, is the proposed material in addition to material offered by the vendor (e.g. earlier and later dates of a newspaper), and what value is added for the public in digital access to these materials when combined with the commercial material? (Maximum 2,000 Characters)**

No known commercial vendor provides information on this topic. Print and e-books cover the political and economic history of the region and some specific people, as noted in a Google search, but none of the three collections' creators. A non-profit website, Presbyterian Mission Agency, has a paragraph on a webpage about this topic, but the remainder of the website is about its current missions and goals. A 2013 blog, part of PACSCL Hidden Collections Processing Project, provides information and links to the Vass finding aid held at the Center.

Several searches were made to prove the originality of the digital collection. A search in CONTENTdm was run to rule out whether similar collections existed. Nothing came back for "Congo." "Belgium" brought up a lantern photograph collection that focused on flax crops in Russia. "Colony" brought up a Sierra Leone collection, but its topics included the abolition movement. "Africa" brought up topics that included various types of wildlife; musical instruments; World War II and related topics; a cities and buildings database; maps; and a collection about the Canary Islands. "Presbyterian" brought up a digitized notebook with stories, photographs, and letters that belonged to Dr. Macpherson (1879-1966), but it did not contain any

topics on missionaries or Africa. “Missionaries” retrieved collections that focused on Japan, China, India, Native North America and Canada, by Vikings, Jesuits and Catholics, but not Africa or Presbyterian.

A broader search at Internet Archive of “Congo colony” brought up modern radio broadcasts, documents in other languages, and an outline of the Congo’s 10-year development plan in 1950. “Africa missionaries” and “Presbyterian missionaries” brought up many different media types. Narrowing the searches to text resulted in biographies and books on specific people, but none of the three collections’ creators, and no other related topical media.

### **Copyright**

**Has copyright clearance been obtained or is the material in the public domain?** Yes, deeds of gifts are attached.

### **Provide proof of copyright clearance, search of copyright and/or public domain status. (Maximum 1,000 Characters)**

Vass’ grandson gave the Center Vass’ collection by deed of gift. The school received the Kellersberger’s collection from his daughter by deed of gift. Sheppard’s widow and W. M. Morrison, a missionary in some of Sheppard’s photos, gave the school the contents of the Sheppard Papers. All deeds of gift assigned “all literary and other rights” to the institutions.

Sheppard died 1927, Morrison died 1939, Vass died 1942, and Kellersberger died 1934. They took photos for personal enjoyment not publishing. The photos are outside the applicable 70-year copyright rule. The hymnbooks were published and distributed in the US for the missionaries. They are bound, printed material with authors and dates of publication unknown. Public distribution makes them published works. It is assumed the dates used, pre-1923, make them public domain. Even if published 1923-1977, they are public domain for no copyright notice. The hymns in the books have been printed worldwide for centuries and are public domain.

### **PLANNING**

**Describe the planning process involved in developing the project. Include previous experience and experts consulted. (Maximum 2,000 Characters)**

An examination of the Center’s visitor slips showed an increase in requests to view its collections. Student visitors to the Center remained steady over the past two years. However, public visitors increased from an average of 10 per month in 2010 to 15 in 2012, which now included students under the age of 18. The increase in requests to view the missionary collections correlated with the Center’s decision to make its database available on its website. Since the Center referred researchers to the school library’s special collections, an examination of the library’s database of visitors confirmed a reciprocal visitor increase.

The need to provide access to popular materials now restricted helped prioritize which items to digitize. Materials from the American Presbyterian Congo Mission during the oppressive regime of King Leopold II of Belgium (1865-1909) were chosen from three separate physical collections held at the two partnering institutions. The printed text items in the collections were selected

over handwritten documents because they could easily be made searchable. The popular photo albums will be taken apart for scanning, but the pages will be scanned intact to allow researchers to see the notes on the margins of the pages with the photos as the creators made them. The Philadelphia Conservation Center for Art and Historic Artifacts was consulted as to whether items could withstand the stress placed on them during digitization, and conservation handling afterwards. Ownership of the materials was confirmed by examining the deeds of gift. Since digitizing items in the public domain was preferred, copyright was determined to verify the chosen items could be considered as such. Model releases were deemed unnecessary since the Congolese in the photos are dead, and it would create an unreasonable hardship to track down any living relatives. The photos do not depict any religious activities, nor will they cause any embarrassment to people or governments.

### TECHNICAL VIABILITY

Items to be digitized – indicate quantity of each of the following items:

You will need to re-key this information into the online application because this section cannot be copied and pasted.

	<b>Number of titles, issues or pieces</b>	<b>Number of pages (if applicable)</b>
Books		
Black and white photos		
Color photos	220 (grayscale and monochrome photos that need to be scanned in RGB mode)	
Prints		
Maps		
Journals	2 hymnbooks (50 leaves, double-sided)	<b>200 pages</b>
Pamphlets		
Albums	3 albums 1.) 167 leaves, double-sided 2.) 20 leaves, double-sided 3.) 150 leaves, double-sided	<b>674 pages</b>
Other – Please describe in next question		

**If you chose other in the question above, please describe the type of material to be digitized, including the suitability of the material for digital presentation. (Maximum 200 Characters)**

N/A

**Describe how the materials are suitable for online use. (Maximum 1,000 Characters)**

The materials range in size from approximately 4x5 to 11x15 and can be accommodated by the college's oversize flatbed scanner, EPSON Expression 11000XL, which has a scanning area of 12x17. The collection is composed of monochrome and grayscale photographic images, and printed and handwritten text. The proposed scanning resolution, 24-bit RGB, with a resolution ranging from 400 dpi (larger photos and text) to 600 dpi (small photos) is in line with FADGI's minimum recommendations while allowing for manageable file sizes and significant enlargement in print reproductions. These scanning specifications allow for significant magnification when viewed online, allowing users to zoom in to see details that may be illegible on the original item. Metadata will provide item-level access, enhancing search and retrieval and providing the ability to collate items. The ability to edit images digitally by adjusting brightness or contrast will significantly improve their legibility for users.

**Describe any physical challenges you expect to face in scanning the original, such as brittleness of pages or tightness of binding (which could potentially cause guttering, or an unreadable darkened edge). (Maximum 6,000 Characters)**

The Eugene R. Kellersberger photo album is currently bound and in poor condition. Natleigh College has obtained an estimate from a local conservation institute of approximately \$2,500 to disbind, repair, and rehouse the album. The spine of the book is severely compromised and no longer functions properly; the sewing structure and linen tapes have failed or are failing. The majority of the album's leaves are a poor quality paperboard and have become brittle and stiff. Many leaves have fractured at stress points near the binding and some have come completely loose. A few (less than ten) photos have become loose from these attachments and would need to be reattached prior to scanning. There are five loose elements which appear to have no corresponding location and would need to be scanned separately. After the album is disbound, each leaf, which is 11x15 can be scanned separately on a flatbed scanner. Items will not be removed from the paperboard in order to preserve the context of the photographs, handwritten notes, sketches, and ephemera that were organized in a specific way by Dr. Kellersberger.

The Kellersberger album also contains various attachments which would need to fold out. Folded items would require multiple scans of a single album page. The fold-outs have been carefully assessed to ensure that they will not extend beyond the dimensions of the Epson flatbed scanner. Upon completion of conservation work, each album page will be housed in a polyester film L-sleeve and housed in a custom made, buffered, acid-free, lignin-free, alkaline buffered box. The album will be housed in a climate controlled archives and access will only be provided through the digital surrogates, in effect ensuring the preservation of the original album. The need to preserve the context of each album page and the enormity of the cost to remove and preserve each item individually, led to the decision to keep album pages intact. Though the acidic nature of the paperboard is irreversible, the continued degradation of the album will be slowed in its climate-controlled environment.

The two albums from the William H. Sheppard collection were previously disbound, but the paper is also brittle, and there are several loose items. Both albums have pages ranging in size from 5x7 to 9x 12, and would fit easily on a flatbed scanner. Individual photographs range in size from 4x5 to 8x10, and are attached to pages with photo corners or glue. There is only one

photograph on each album page. Whenever items are not glued to the paperboard, they will be removed and housed in a polyester L-sleeve and buffered acid-free folder.

The Epson Expression scanner is recommended by conservators at the Philadelphia Conservation Center for Art and Historic Artifacts as a scanner that limits light and heat exposure to the items being scanned.

Loose photographs from the Vass collection range in size from 4x5 to 8x10. Some photographs are curling and would need to be handled carefully to prevent damage to the photographic emulsion. Proper handling techniques, such as those recommended by the Library of Congress *Preservation Guidelines for Digitizing Library Materials* will be strictly enforced throughout the scanning process, requiring extreme care when handling brittle pages and curled photographs. There will be adequate space provided to stage materials, cotton gloves will be used when handling photographs, items will be protecting during transport, and materials will be promptly covered or returned to their containers to limit light exposure.

Some photographs have issues related to deterioration, such as fading, silver mirroring, and yellowing, as well as issues related to improper exposure. Though preservation master files will remain without any digital manipulation or retouching in order to capture the original content of the historic photographs, in some cases derivative versions of the image may need some degree of adjustment in order to be more useful for users. Adjustments will only be made to ensure images are legible and will not be an attempt to restore the image to an unknown original state or alter their contents.

The two hymnbooks are currently bound, however, the binding and covers are loose from the text block and there are some pages that are completely loose from the stitching. Pages easily open without causing further damage. Due to the fragility of the item and the loose elements, careful handling is required.

The value, rarity, and fragility of the items to be scanned makes the best argument against outsourcing digitization and metadata to a commercial vendor. Even if a vendor with preservation and conservation expertise was chosen, the transportation of the materials poses a substantial threat. By digitizing the materials in-house, Natleigh Library's archivist will have complete control over handling and digitization procedures. In addition, library and archives staff will be able to relate the valuable hands-on expertise gained from this large-scale project to future digitization projects as well.

**If materials to be scanned are bound, will they be disbound?** Yes and no. One item will be disbound, two items will not be disbound.

**If not, tell how the proposed scanning equipment is appropriate for use with bound materials. (Maximum 1,000 Characters)**

The College archivist, who has a certification from the Rutgers University Preservation Management Institute, has evaluated the condition of the hymnbooks. Though they are bound, the covers are loose from the text block. The stitching does not prevent the books from opening,



and pressing the pages on top of the flat bed scanner will not cause further damage. The college's oversize flatbed scanner (EPSON Expression 11000XL) has a scanning area of 12.2" x 17.2". It will easily accommodate the hymnbooks, which measure 5.5" x 7".

**Scanning specifications. You will need to re-key this information into the online application because this section cannot be copied and pasted.**

	<b>Scanner Type (flatbed, overhead, etc.)</b>	<b>dpi proposed</b>
Black & white text or line drawings	Flatbed	400 dpi, with OCR
Color	Flatbed	400 to 600 dpi
Grayscale		

**Digital Files. You will need to re-key this information into the online application because this section cannot be copied and pasted.**

	<b>File type</b>	<b>Estimated file size</b>
Master files	TIFF	
Web display files	JPEG	
Web format for scanned text	ASCII encoded with XML	

**Provide the total estimated storage capacity required (in bytes). (Maximum 50 Characters)**

**Will your agency host the server for Internet access? No**

**Will you use the Access PA Digital Repository to host? Yes**

For Digital Repository eligibility see: <http://www.accesspadigital.org/>

**How will the master files be archived? You will need to re-key this information into the online application because this section cannot be copied and pasted.**

	<b>Type of storage (disk, online)</b>	<b>Storage Location</b>
Digitized materials	CONTENTdm, online	Access PA server

**What institution will archive the resulting digital masters and any master microfilm that will be produced? (Maximum 50 Characters)**

Natleigh College (6TB Network-attached server)

**What plans are in place to ensure the long term preservation of master images and master microfilm, including funding, after the grant period ends? (Maximum 4,000 Characters)**

Natleigh College is currently an Access PA member and will submit the digital collection to the Access PA digital repository, which is maintained on the Access PA server. They provide the CONTENTdm license in perpetuity, archive images, back up data, and provide copies of data to

members. CONTENTdm's Digital Archive provides a controlled operating environment dedicated to the ongoing managed storage of digital content. CONTENTdm stores data in XML, a standard markup language that is the most common tool for data transmissions, allowing easy migration between all kinds of applications. This will ensure data can easily be migrated in the event of a CONTENTdm upgrade or if the institution decides to move data to a new digital repository.

The use of standard and non-proprietary image formats, TIFF for master files and JPEG for web derivatives, will also ensure that data will continue to be useable and migratable in the future. TIFF masters will be saved as uncompressed files to ensure that they continue to be lossless. JPEG, while lossy, is also a standard format that can be transmitted quickly on the internet and opened in all browsers and operating systems.

Over the past three years, Natleigh College has made a commitment to digital projects by dedicating a portion of the budget, starting at \$5,000 in the first year and increased to \$10,000 in year three, to digitization goals, and has raised additional funding through grants and private foundations. Funding was obtained from the William Penn Foundation and private donors to create a scanning workroom, which included the purchase of two dedicated computers, two Epson large format scanners, Silverfast scanning software, color management software, and ABBYY OCR software, two licenses for an Adobe Design suite, a mirrored 6 TB network-attached server system, and two new staff positions: one part-time Cataloger/Metadata Specialist, and one full-time Digitization Technician. The college will continue to seek and match funding from government and private grant agencies and foundations to add to this collection and fund future projects. Funds will be set aside to replace outdated equipment and storage devices, purchase larger storage capacities, and renew digital production staff positions.

The college will also use its own server to back up preservation master TIFF files. The library currently maintains a VMware virtual network and server with a 6 TB network-attached server used specifically to back up preservation master images. The 6 TB server is duplicated at the college's sister campus located in Arizona for disaster recovery purposes. The Center for Presbyterian History staff will be provided a log in and password to access the project images stored on the 6 TB NAS. Master files on these servers will be locked so that they can only be deleted by designated library systems staff. The Cataloger/Metadata Specialist will create a checksum file for all master TIFF files using Exact File software and will save the checksum file in the same folder as the images on the NAS. Checksums generate a unique identification number for each image file stored within a folder. The checksum file can be automatically checked against the TIFF files at any time to ensure there has been no data degradation.

The online environment is maintained with high-speed network connections, and computer equipment is stored in temperature and humidity controlled environments, with emergency power generators to ensure the continuity of service in the event of power failures. Additionally, automated back-up tape systems are maintained and stored off-site.

**Metadata records for cataloging and indexing. You will need to re-key this information online.**

	Type	# of records	Description
MARC	MARC21	1	Collection-level
Non-MARC	Dublin Core	896	Item-level
Non-MARC			
Non-MARC			
Non-MARC			

**Describe how you will create your metadata. Will records be created at the item, title or collection level, or some combination thereof? (Maximum 6,000 Characters)**

The library will use a combination of collection-level and item-level cataloging. A MARC21 collection-level record will be created for the digital collection as a whole and added to the college library's online catalog, the Center for Presbyterian History's online catalog, submitted to OCLC for addition to WorldCat, and will be available through the Access PA federated catalog. The WorldCat Digital Collection Gateway, offered by CONTENTdm, provides a self-service tool for uploading metadata to WorldCat at no additional charge.

An item-level record will be created for each page of the photograph albums, each loose photograph, and each hymnbook using a CONTENTdm specific variation of Dublin Core (DC) extended. The Sheppard photo albums contain one image per page, while the Kellersberger album contains various elements on each page. This adds a degree of complication to metadata creation for the Kellersberger album and may require more detail in each item-level record within this set. The DC Metadata Element Set required by Access PA defines 16 core components, which will be used for this project: title; creator; subject; description; date; publisher (repository information); contributor; type; format; identifier; rights; source; language; relation; coverage; and audience. These fields will be used as outlined by Access PA in their *Access Pennsylvania Digital Repository Guidelines* (p.4).

Structural metadata will be used to document relationships within and among objects; for example, within the photo albums and hymnbooks, thus enabling users to navigate the pages which comprise them. Since this can only be accomplished to a small degree with the Dublin Core *relation* element, METS will be used to supplement Dublin Core. The required structural map element in METS will be utilized to enable users to easily understand the relationships among materials from the same album or hymnbook and navigate logically between them.

As part of the scanning workflow, some administrative and descriptive metadata will also be embedded with the master TIFF files by utilizing Adobe's metadata templates (IPTC core), which are readable across operating systems (Mac/Windows). Embedding this information early in the workflow will ensure that the metadata is inherited by all derivative versions of the master file. Adobe Bridge allows this metadata to be applied to image files in batches by using a template, which is a time efficient practice. The following metadata elements will be used: creator; title; source (for name of institution); copyright notice; and rights usage terms. Adobe metadata automatically embeds technical metadata related to the file properties, such as file format; file size; pixel dimensions; resolution; bit-depth; color mode; and color profile.

Library of Congress Name and Subject Headings will be used whenever possible to describe materials. Subject and name headings have already been established for these collections, since they have been processed and have finding aids. If a name heading needs to be created, the College Cataloger/Metadata Specialist will create one using AACR2 as a guide and add it to the local authority file. With the expectation that LCSH will not meet all of the needs of this collection, a supplemental local thesaurus will be created in order to consistently describe materials related to tropical diseases and medicine and African native customs.

ABBYY OCR software will be utilized when scanning the two hymnbooks, which contain both printed and handwritten text. ABBYY automatically converts text to ASCII (American Standard Code for Information Interchange) and the tab-delimited ASCII file can be imported into CONTENTdm and exported to XML.

**How many staff will be dedicated to this project (if they will only devote part of their time to it, how many hours per week)?**

**Natleigh College:**

- 1 Archivist/Project Manager, with specialization in preservation (Mary Smith); 15 to 20 hours per week
- 1 Head Librarian (Leigh-Anne Yacovelli); 10 to 15 hours per week
- 1 Cataloger/Metadata Specialist (Lisa Davidson); Lisa is currently half-time and will be grant funded in order to work 30 to 40 hours per week
- 2 part-time student volunteers will assist in metadata creation; 20 hours each per week
- 1 full-time scanning technician (Jane Johnson); 25 to 30 hours per week
- 2 part-time scanning technicians (to be determined); grant funded in order to work a total of 40 hours per week

**Center for Presbyterian History:**

- 1 Archivist/Subject Specialist who will assist the cataloger with the creation of a local controlled vocabulary and marketing/promotion of the collection (Natalie Shilstut); 15 hours per week, grant funded

**IMPLEMENTATION PLAN**

**Describe the activities that will happen related to scanning, quality review of output, metadata creation, site design, data loading, etc. Outline the timetable by quarter of the activities that have been allotted for each step, including quality control, site testing, feedback evaluation, and any revisions that will occur due to these processes. (Maximum 2,000 Characters)**

**January - March 2014**

- Conservation work to disbind the Kellersberger photo-album begins
- Hire two part-time scanning technicians and two part-time student volunteers

- Create local controlled vocabulary to describe subjects related to tropical medicine and native customs
- Define file-naming conventions for master and derivative files
- Create IPTC core metadata template and Dublin Core/METS metadata template in Excel
- Begin establishing project workflow

#### **April - June 2014**

- Calibrate scanning equipment and monitor; define color profile
- Pilot project to test workflow and procedures: Select differing materials that are a good representation of the collection as a whole in order to test the effectiveness of the scanning specifications, file-naming convention, and metadata workflow. Make adjustments to workflow as needed.
- Define quality control procedures
- Begin scanning

#### **July - September 2014**

- Scanning and quality control continues
- Batch processing to embed IPTC core metadata and DC/METS metadata creation in Excel
- Convert master TIFFs to JPEG derivative; archive TIFFS on NAS and create checksum files
- Begin import of image files and metadata into CONTENTdm
- Make alterations to CONTENTdm user interface to improve access and usability
- Add MARC collection-level records to catalogs and WorldCat

#### **October - December 2014**

- Test usability and effectiveness of site with sample group of college students
- Create visit counter and online survey
- Promote launch of digital collection on both institution's websites; social media (Facebook, Twitter, Pinterest, History Pin); listservs; the College and CPH's online and paper newsletters and blogs
- Begin preparation of educational session/lecture
- Evaluate response and make improvements

**Tell which staff will be responsible for each task, indicating extent of time committed, and relevant experience if possible. Include the level (Librarian, Library Assistant, student worker, volunteer, etc.) of each staff member and evaluate their suitability for the task assigned. If tasks are to be outsourced, indicate when this will happen, the duration of the work, and which project staff will be responsible for working with the vendors. (Maximum 2,000 Characters)**

Mary Smith, Archivist/Project Manager, MLIS, certificate from Rutgers University Preservation Management Institute (15-20 hours/ week): Communicate with Conservation Center on

Kellersberger album conservation work; supervise scanning technicians to ensure proper handling; quality control.

Leigh-Anne Yacovelli, Head Librarian, MSLS, PhD (10-15 hours/week): Supervise the Cataloger/Metadata Specialist; assist in creation of local thesaurus.

Lisa Davidson, Cataloger/Metadata Specialist, MLIS, Digital Archives Specialist certified by SAA (part-time employee working 20 hours/week, grant funded for additional 10-20 hours): Develop local thesaurus; create Excel metadata template; create checksums; archive images on NAS; upload to CONTENTdm; supervise student volunteers.

Student volunteers (20 hours/week): 2 current MSLS graduate or undergraduate students in African studies department hired to enter metadata in Excel.

Jane Johnson, Digitization Technician, BA History (25-30 hours/week): Has extensive experience with scanning historic materials, Adobe software, and ABBYY software acquired through internships at the LOC and NARA; will define file naming convention and scanning workflow; scan materials and apply IPTC core metadata

Scanning Technicians (20 hours/week, grant-funded): 2 technicians with previous experience in scanning historic materials will be sought.

Natalie Shilstut, Archivist/Subject Specialist, MSLS, PhD, certified archivist, expert in topic of the Presbyterian American Congo Mission (part-time Center employee working 15 hours/week, grant-funded for additional 15 hours): Assist Cataloger/Metadata Specialist and Head Librarian with creation of a local thesaurus; assist in promotion of digital collection by hosting educational session and lecturing on American Presbyterian Congo Mission.

Edward Bennett, IT Specialist (5-10 hours/week, on-call for IT issues): manages the virtual network and College website; available to attend meetings and on-call for IT support.

**Describe how the digitized materials will be cataloged and submitted to the Access PA Database. Will it be cataloged in MARC, Dublin Core simple, Dublin Core extended (the schema for projects in the Access PA Digital Repository), or another schema. If the items will not be cataloged in MARC, name the schema and describe its suitability for the materials, including any elements (fields) that will be added, outside the standard set. (Maximum 2,000 Characters)**

A MARC21 collection-level record will be created for the digital collection and added to both the school's and the Center's online catalogs, submitted to OCLC for addition to WorldCat, and made available through the Access PA federated catalog. The WorldCat Digital Collection Gateway, offered by CONTENTdm, provides a self-service tool for uploading metadata to WorldCat at no additional charge.

An item-level record will be created for each page of the photograph albums, each loose photograph, and the hymnbooks using a CONTENTdm specific variation of Dublin Core

extended. The DC Metadata Element Set required by Access PA defines 16 core components, which will be used for this project: title; creator; subject; description; date; publisher (repository information); contributor; type; format; identifier; rights; source; language; relation; coverage; and audience. These fields will be used as outlined by Access PA in their *Access Pennsylvania Digital Repository Guidelines* (p.4).

Structural metadata will be used to complement DC in order to document relationships within and among objects; for example, within the photo albums and hymnbooks to enable users to navigate the pages. Since this can only be accomplished to a small degree with the DC *relation* element, METS will be used as a supplement. The required structural map element in METS will be utilized to enable users to understand the relationships among materials from the same album or hymnbook and navigate logically between them.

As part of the scanning workflow, administrative and descriptive metadata will be embedded with the master TIFF files by utilizing Adobe's metadata templates (IPTC core), which are readable across operating systems. Embedding this information early in the workflow will ensure that the metadata is inherited by all derivative versions of the master file. The following metadata elements will be used: creator; title; source (for name of institution); copyright notice; and rights usage terms.

## **PROJECT EVALUATION**

**Describe how you will measure achievement of the project's goals and the overall impact of the project on your library users. You can use quantitative (statistics) or qualitative (survey, focus groups, etc.) methods to gather evaluative data, or a combination. Commonwealth Libraries will accept proposals that include a plan and budget for an outside consultant to conduct testing or evaluation. (Maximum 10,000 Characters)**

The project's goals for the digitization of certain unique and restricted items from the three individual archival collections held at two separate locations are to:

- provide access at any time for people located anywhere in the world
- increase visitor foot traffic at both sites by 25% over the next year\*
- reduce the physical handling of fragile items
- increase student awareness of the collections.

\*It was established before the project began that visitors to the physical collections at both sites increased when the Center made its database available online. The numbers showed a 50% increase over two years in student visitors, which now also included students under 18 years old.

During the last three months of the project, staff will:

- test usability and effectiveness of site with a sample group of college students
- create an online survey
- promote launch of digital collection on both institution's websites; social media (Facebook, Twitter, Pinterest, History Pin); listservs; the college and Center's online and paper newsletters and blogs
- evaluate responses and make improvements

The purposes of the above-checklisted activities are to build the structure by which the project evaluations will be based.

During the first month after going live, staff will assess internal operations by evaluating duties to determine if changes are needed, and if so, make adjustments to the workflow to remain efficient. Any additional preservation actions not foreseen in the preparation of the project, but that came to light during the implementation of it, can be handled at this point. Monthly reviews of the digital collection's site to ensure there are no dead links, and the thumbnails and content match, should continue. As results from other evaluations over the following months are examined, additional revision to job duties and collection maintenance will occur.

Evaluating of the project's success will rely heavily on CONTENTdm's reports feature. Many other popular site analytic software, such as Google Analytics and Piwik, require access to a website's source code to integrate the analyzing software or cost additional fees. CONTENTdm's report features will be adequate for our needs. Event tracking is typically used to monitor downloads or clicks on social media icons to share or like an image. This shows that the visitors to the site were engaged in an activity and not stumbling upon something then moving on to another page. Time spent on each page is another way to monitor visitor activity. Neither of these reporting methods appear to be supported by CONTENTdm. However, CONTENTdm provides reports on views by collection and item; top searches; monthly, daily, and hourly statistics; top referring URLs; errors that visitors encountered on the collection site; user statistics such as IP addresses, browsers used, and countries of origin.

The benchmark for studying the results of the online collection will be based on the highest number of visitors during the first month. The goal for access will be revised once that number is determined to gauge the success of the project.

Physical counting of visitors by desk staff at both sites will continue to answer questions pertaining to physical access of collections.

Questions to ask for the evaluation of the project's success include:

- How many used the collections when they were just physical collections kept at two different places?
  - Monthly statistical tracking of the number of visitors from CONTENTdm reports.
- How many physically and remotely accessed them during first three months after active marketing?
  - Monthly statistical tracking of the number of visitors from CONTENTdm reports.
- How many physically and remotely accessed them at month nine when all active marketing has ended and only passive marketing such as through classrooms with teachers encouraging use of site?
  - Monthly statistical tracking of the number of visitors from CONTENTdm reports.
- Was there a specific time of day users were most likely to access the collection, and can they access them in person at that time?
  - Hourly statistical tracking of the number of visitors from CONTENTdm reports compared to the library and Center's hours of operation can determine if the digitization of the collection enabled people to access the materials at their convenience during closed hours.



- Where did remote visitors come from, or alternately, why did they come to the collection's site?
  - Examine CONTENTdm reports for the referring website urls to determine why visitors came, e.g., linked from the school library could mean student research; linked from the Center's website could mean other research; linked from social media pages could mean general interest; link from Internet search engine can show possible related sites.
- What were visitors looking for?
  - Examine CONTENTdm reports of visitor clicking patterns to determine if people browsed the collections or jumped to items after searching by keywords.
- Were visitors successful in finding what they wanted in the digital collection?
  - Use CONTENTdm's report on the pages per visit to measure the degree in which people stayed on the site searching, as can a report on successful links from keyword searching in the collection.
- Were visitors repeat visitors, and if so, were they looking for the same thing or searching collection for something else?
  - Sort by IP address in CONTENTdm reports to see if there were repeat visitors, and compare the repeated IP addresses to the keyword searching report.
- How much time has staff spent retrieving physical items before and after digitization?
  - Measure staff time spent retrieving items for in-person requests and distance requests for the digitized collections, and compare to time spent doing similar activities for other popular physical collections.

Ongoing surveys will measure awareness of the digital collection. The librarians at the school, and the front desk clerk at the Center, can ask patrons about whether they ever accessed the collection physically or virtually, and record the answers. They will also ask visitors to fill out a standard paper survey if they have visited the collections, or take the survey with them if they have not accessed, but would like to do so, virtually.

Email surveys to students, faculty, alumni, and Center members at the third month after going live will direct people to the collection and request feedback on ease of use in searching the collection, layout, images, and content. A question will be included to determine if they ever visited the digital collection. SurveyMonkey will be used. It provides a simple, free survey. "At-the-source" surveys, such as KISSinsights (now Qualaroo), charge a fee to use. Cold surveys will be just as effective as at-the-source because people will be directed to immediately go to the site, so their experience will be fresh. This can also be a form of marketing in that some people may not be aware of the digital collection.

An impact survey at two months and at six months after the collection goes live to students, faculty, alumni, and Center members will evaluate their feelings regarding the digitized collection, and their feedback on the need for future digitization of the remaining portions of the collections or the digitization of other collections. This will work well in conjunction with a comparison to staff time spent retrieving items for in-person requests and distance requests for the digitized collections to other popular physical collections.

A mini-conference, which gives historical insight into content and information on the planning and implementation of the digitization project, will be held for researchers interested in the content immediately after the collection goes live. A question and answer session will follow, treating the attendees as a focus group, to measure their thoughts on what was missed in the project’s processes, what about the site they did not understand, and if the choice of materials was adequate or if more needs to be added.

Digital humanities conferences and workshops will be attended by staff to benchmark what other places are doing or have done. They will be encouraged to discuss their experience with the project in order to get feedback on what other archivists and digitization experts feel about the project’s methods and content. Marketing in this fashion will end within six to seven months after going live.

**BUDGET NARRATIVE**

**List each item/service, etc., that will need to be purchased to implement the project successfully. Identify the source of the funds using the drop down menu in column four. Use a separate row for each item/service. Every item/service must be listed in the table below and in the Budget Detail section at the end of the application. Be specific in terms of different types of purchases needed for the project, by making sure that different types of similar items are broken down so that the reviewers will see how each item will be used to satisfy the needs described in the Needs Assessment for this project.**

Item/Service	Description of Item/Service purpose for the project	Cost	Grant, Local or Other Funds, of N/A
Staff position	Lisa Davidson, part-time Cataloger/Metadata Specialist: 10-20 additional hours per week, no benefits		Grant funding
Staff position	2 part-time scanning technicians: 20 hours per week each position, no benefits		Grant funding
Staff position	Natalie Shilstut, Archivist at the Center for Presbyterian History: 15 hours per week to work at Natleigh College		Grant funding
Conservation work (disbinding and rehousing)	Kellersberger photo-album conservation work and rehousing in archival housings	\$2500	Grant funding

Software	CONTENTdm hosting charge/subscription fee (annual)*		Grant funding first year
Equipment	Computer station: desktop computer and monitor	\$1000	Natleigh College
Software	2 Adobe Design Suite licenses	\$1198	Grant funding

\*Natleigh College will pay hosting and subscription fees from the Digitization projects budget after the first year.

**If Other was chosen for any of the items/services above, explain the source of the fund for each. Maximum 500 Characters.**

N/A

**Describe how the program will be sustained after the grant year. Include details of digital master and possible microfilm master storage and preservation. Be specific as to how much money will be needed and source of future funding to sustain the project. (Maximum 2,000 Characters)**

The CONTENTdm license is provided in perpetuity by Access Pennsylvania. CONTENTdm's Digital Archive provides a controlled operating environment dedicated to the ongoing managed storage of digital content. Natleigh College will pay an annual hosting/subscription fee of \$XXXX after the first year, and this money will be taken from the digital projects budget.

The use of standard and non-proprietary image formats will also ensure that data will continue to be useable and migratable in the future. TIFF masters will be saved as uncompressed files to ensure that they continue to be lossless. JPEG, while lossy, is also a standard format that can be transmitted quickly on the internet and opened in all browsers and operating systems. To ensure against data degradation the Cataloger/Metadata Specialist will create a checksum file for all master TIFF files using Exact File software and will save the checksum file in the same folder as the images on the NAS.

Natleigh College dedicates \$10,000 annually to digitization goals and will continue to seek and match funding from government and private grant agencies and foundations. The digital projects

budget will be used to replace outdated equipment and storage devices, purchase larger storage capacities, and renew digital production staff positions.

The College will use its 6 TB network attached server to back up preservation master TIFF files. The server is duplicated at the college's sister campus located in Arizona. Master files on these servers will be locked so that they can only be deleted by designated library systems staff.

The online environment is maintained with high-speed network connections, and computer equipment is stored in temperature and humidity controlled environments, with emergency power generators to ensure the continuity of service in the event of power failures. Automated back-up tape systems are maintained and stored off-site.

### **BUDGET DETAIL**

**A drop down menu to select the prepopulated budget expenditure categories will be provided in column one. A response may be drafted to complete columns two through four. Within the online application, choose the appropriate Budget Expenditure item from the drop down menu for both requested and local funds. Note: This information will need to be re-keyed into the online application because this section cannot be copied and pasted.**

Broad Budget Expenditure Category	Description	Amount of each one	Quantity
Staff position	Lisa Davidson, part-time Cataloger/Metadata Specialist: 10-20 additional hours per week, no benefits		Grant funding
Staff position	2 part-time scanning technicians: 20 hours per week each position, no benefits		Grant funding
Staff position	Natalie Shilstut, Archivist at the Center for Presbyterian History: 15 hours per week to work at Natleigh College		Grant funding
Conservation work (disbinding and rehousing)	Kellersberger photo-album conservation work and rehousing in archival housings	\$2500	Grant funding

Software	CONTENTdm hosting charge/subscription fee (annual)*		Grant funding first year
Equipment	Computer station: desktop computer and monitor	\$1000	Natleigh College
Software	2 Adobe Design Suite licenses	\$1198	Grant funding