DIGITIZATION PROPOSAL

Course: **LIS 4700, Introduction to Digitization**

New York's famous 369th regiment arrives home from France

Retrieved from http://www.archives.gov/education/lessons/369th-infantry

I. Description of Project

The Iowa State Historical Society intends to digitize a collection of letters created by American soldiers during WWI. Sent to their family members in Iowa, the correspondence reflects the various experiences of soldiers stationed in Europe in 1917 and 1918.

There are two main goals for this project. First, this project will provide digital and lasting access to popular materials that are increasingly fragile. Secondly, by digitizing the manuscripts, the historical society is engaging in digitization as preservation. Not only will this project capture the unique nature of the aging materials, but will extend the life of materials that will likely deteriorate with time.

Strategically, this collection is an excellent choice for a pilot digitization project. Teachers and students across Iowa frequently request copies of the materials and some staff even arrange field trips to the society. While getting students in the door is important, so is protecting this fragile collection. Digitization will reduce handling while at the same time create a new access point to our collection for students and teachers. Eventually, we will include in with digital collection lesson plans for educators. Additionally, we will work with state historians and faculty from Iowa colleges and universities to create content to complement the digital collection; such content will provide the historical context to guide researchers who access our content. By having an online presence, we hope to increase the number of annual visitors to the Iowa State Historical Society. We intend on sending our librarians and staff to various Iowa high schools on National History Day to introduce the digital collection to students and teachers.

II. Resources

To digitize the manuscripts, it is necessary that the historical society have proper scanning equipment. In order to complete the project in a timely manner and since the society typically works with two student interns, we recommend using two Epson Expression 10000XL scanners. The 11x17 inch scanning area, high optical resolution, and the high bit depth capabilities are ideal to capture the unique coloring of the manuscripts. These scanners also include transparency adapters that will inevitably come in handy as the society continues to digitize its collections. The scanners cost between $2,500 and $3,000 each (see http://www.epson.com for more details on cost and specifications).

To deliver the collection online, the historical society requires a content management system. We recommend CONTENTdm, a product of OCLC that is currently used by numerous cultural heritage organizations. CONTENTdm is equipped with many attractive features that will save the time of staff and interns. For example, once typed copies of the manuscripts are uploaded, the content management system can automatically perform optical character recognition (OCR) that allows the typed manuscripts to be easily searchable. Other content management systems such as Omeka do not perform this automated function and would require the purchase of Adobe Acrobat Pro software to do so. Further, CONTENTdm presents both the digitized, handwritten manuscripts and the typed files as compound digital objects; in other words, they will be presented as a unit, and users will not need to perform two searches for the related objects. Finally, with CONTENTdm, we will be able to upload batches of our collection items, and derivatives are automatically generated.

CONTENTdm can be installed on the historical society's current Microsoft Windows 2008 web server and is compatible with our Windows computer workstations. The historical society should have no problem hosting the management system since it has 9 GB of available RAM and 8 GB of available hard-disk space. This will also allow for the CONTENTdm website and project client. The society has Adobe Reader 8, a 256Kbps connection, a computer display resolution of 1024 x 768, an Internet connection for downloading .NET 3.5, and Microsoft Windows XP with SPD 32-bit.

CONTENTdm is ideal for users. We are able to integrate the existing EAD finding aid, so this information will not be lost. As far as viewing, users are able to zoom in and out, pan, or view the manuscripts full-browser. We feel that these viewing tools will provide an experience with our collections that users have not had when they physically handle the collection. Advanced search tools will allow for users to quickly find items. Since many educators will be accessing the collection, the ability to save items and export them for presentations is extremely valuable.

CONTENTdm's interactive interface allows for social tagging, user comments, and sharing via social media tools such as Facebook and Twitter. We hope that these tools will help build conversations amongst our users, and we foresee that the social media component might function as an independent marketing tool.

The WWI manuscripts collection is composed of 420 letters; on average, each letter is two pages, or two scanned files. Thus, the historical society can plan to digitize 840 items. We recommend that the society purchase CONTENTdm's Quick Start package that can store 3,000 items. This will give the society space for an additional 2,160 digital files for future projects. The Quick Start package is $4,200.00 plus the cost of the OCR extension (see http://www.oclc.org/contentdm/ for more details regarding pricing, package options, and adherence to standards such as OAI-PMH).

In order for the digitization project to be efficient and successful, the historical society must ensure that its staff has the appropriate skills and expertise to operate the scanners and work with the content management system. We will rely heavily on our digitization project specialist, Gina Schlesselman-Tarango, and propose continued professional development for both her and our metadata librarian. Schlesselman-Tarango will train our two interns in proper scanning techniques, provide other staff members with training when necessary, and create documentation of the digitization processes so that those in the future can repeat and improve upon her methods.

III. Standards and Technical Specifications

We recommend Dublin Core metadata standards, an internationally-adopted metadata schema that is supported by the Open Archives Initiative Protocol for Metadata Harvesting (OAI PMH). Dublin Core records are XML-based and can thus be migrated to new systems in the future to facilitate continued access to the collection. Since we will be creating compound objects, descriptive metadata will be provided at the volume or collection level rather than the individual level. As noted above, we intend to include in the records a link to the original EAD finding aid. For more information of Dublin Core, see http://www.dublincore.org.

To create the records, we will pull from a variety of controlled vocabularies, primarily those established by the Library of Congress such as subject headings and the thesauri for graphic materials and geographic names. Guidelines for indexing will be based largely on Resource Description and Access (RDA), but the metadata librarian will establish local rules that reflect the needs and language of our users. Finally, the metadata librarian will work with the historical society's cataloger to create the metadata, and there is discussion of recruiting an additional student intern to assist with this time-consuming task.

To best capture and preserve the nature of the manuscripts, strict scanning, conversion, and formatting requirements are to be followed.

Scanning requirements:

Document type: reflective
Mode: professional
Bit depth: 24
Auto exposure: photo
Dpi: will vary by size – capture 4,000 pixels across long dimension
Save as uncompressed .tif file using a prefix and a three-digit number, e.g.: war001.tif

Conversion and formatting requirements using Adobe Photoshop:

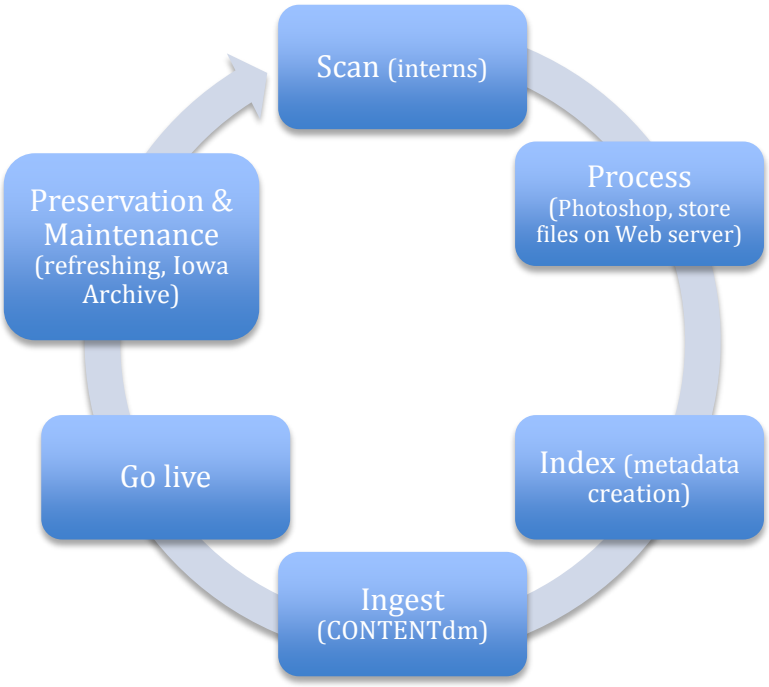
Create a service copy and save as an uncompressed .tif file; add s to original file name, e.g.: war001s.tif
Use service copy to create access image for web delivery: 72 dpi, 1024 pixel dimensions, save as .jpeg, add w to file name, e.g.: war001w.jpeg
Use service copy to create print copy: 300 ppi, adjust document size for printing, save as .jpeg, add p to file name, e.g.: war001p.jpeg
*Note: Schlesselman-Tarango will edit service copy in Photoshop before creating web and print derivatives

For more information on best digital practices, see the following:

California Digital Library. (2011). The CDL Guidelines for Digital Objects
Consortium of Academic and Research Libraries in Illinois. (2009). CARLI Guidelines for the Creation of Digital Collections: Digitization Best Practices for Images and Text

IV. Workflow Plan

The following chart illustrates basic steps to be following in the digitization project.



V. Sustainable Preservation Plan

Preserving digital master files is imperative in order for our manuscript collection to be maintained and thus accessible over time. First and foremost, we propose a program of refreshing, or copying two copies of each master file to new, up-to-date media, at least once every two years. This will ensure that our files are stored on a platform that is accessible and can be attended to if it is damaged or becomes corrupt. In the historical society's case, this means ensuring that our server continues to be technologically modern. Also, Microsoft participates in a continuous server backup program.

Secondly, we propose storing two copies of each master file in Iowa's shared cultural heritage repository, the Iowa Archive. The Iowa Archive[1] charges a five-year, flat rate participation fee of $2,000. If we initially agree to participate for 20 years, we will receive a 10 percent discount, paying $7,200 and saving $800. We recommend participating in the 20-year plan, as it guarantees a trusted backup at a very reasonable rate. Additionally, many of our colleagues at the Iowa State Historical Society are Iowa Archive board members, and we feel that using the archive as our repository is a smart organizational move.

All of our master digital files – those on our server and those in the Iowa Archive– will be stored as .tif files, as will any documentation we create regarding the digitization process. Additionally, preservation metadata (following the MIX – NISO Metadata for Images standards) will describe

[1] The Iowa Archive is a fictional digital repository.

compression, bit depth, file formats, and other information about scanning hardware and software.

VI. Project Timeline

We anticipate that completion of the digitization project will take 12 weeks. Below is a table detailing the tasks to be completed each week as well as the staff members involved in each task.

Week	Tasks	Staff
Week 1	Intern instruction and scanning	Schlesselman-Tarango and interns
Week 2	Scanning; processing	Interns; Schlesselman-Tarango
Week 3	Scanning; processing; indexing	Interns; Schlesselman-Tarango; metadata librarian, cataloger
Week 4	Final scanning; processing; indexing	Interns; Schlesselman-Tarango; metadata librarian, cataloger
Week 5	Processing; indexing; type manuscripts	Schlesselman-Tarango; metadata librarian, cataloger; interns
Week 6	Processing; indexing; type manuscripts	Schlesselman-Tarango; metadata librarian, cataloger; interns
Week 7	Processing; indexing; type manuscripts	Schlesselman-Tarango; metadata librarian, cataloger; interns
Week 8	Processing; indexing; type manuscripts	Schlesselman-Tarango; metadata librarian, cataloger; interns
Week 9	Final processing; indexing; final typing of manuscripts	Schlesselman-Tarango; metadata librarian, cataloger; interns
Week 10	Final indexing; ingesting	Metadata librarian, cataloger; all
Week 11	Trial run – make any necessary changes	All
Week 12	Go live	All

**Interns will be utilized throughout the process, but their primary roles will be in scanning the manuscripts and creating typed files for each letter.*

VII. Budget

The table below details the projected budget for the digitization project.

Item(s)/ Service	Cost	Total
Epson Expression 1000XL scanner (2)	~ $2,750 x 2 = $5,500	$5,500
CONTENTdm Quick Start package (1)	$4,200	$9,200
CONTENTdm OCR extension (1)	~ $1,000	$10,000
Iowa Archive 20-year participation fee (1)	$7,200	**$17,200**

VIII. Assessment Plan

To assess the effectiveness of our digitization project, we intend to develop relationships with a number of teachers in Iowa high schools. It is our aim that the opportunity to conduct formal and informal modes of assessment will spring from such relationships. For example, our librarians

and staff will not only interview and survey high school teachers about the usability of the collection and its corresponding lesson plans, but the society plans to host a bi-annual history and social studies education conference; this conference will provide rich insight into the ways in which we can improve the effectiveness of the online collection.

Since the Iowa State Historical Society will have a working relationship with the Iowa Archive, we will solicit feedback from its members regarding our collection. We will also survey experts in the digitization community, such as university professors, in hopes of gaining an understanding of the strengths and weaknesses of our digital collection. We firmly believe that others working in field will provide valuable suggestions as we continue to digitize items in our collections.