THE DIGITIZATION OF SONGS, PLAYS, AND MUSICALS, FROM

WORLD WAR II

7/9/01

L595/5312

Kopczynski

Introduction:
Recent years have seen an upsurge of interest in the World War II era with such movies as *Saving Private Ryan* and Tom Brokaw's book *The Greatest Generation*. A prominent feature of many of these cultural products has been a concern with greater realism in the portrayal of the feeling if not the substance of the era. The collection that our Digital Library Program plans to digitize can be placed at the intersection of both the feeling and the substance of the age. IU's Government Publications department is a Government Document Depository and is home to the entire collection of 110 War Department documents promoting the purchase of war bonds. Created between 1942 and 1945, this collection includes plays, musicals, songbooks and advertisements by soldiers and civilians. Some of these creators went on to great renown, such as Guys 'N' Dolls librettist Frank Loesser. Our mission is to increase awareness and to support research and educational experiences of the WWII era through digital reproductions and a searchable database with numerous access points. Our goal is to create a user-friendly digital library with advanced searching capabilities supported by well-planned information architecture that is suitable for present technologies and adaptable for those of the future. We anticipate a diverse community of users, including scholars, WWII aficionados, documentary filmmakers, photojournalists, genealogists and high school and college teachers. All materials involved in the project are either in the public domain—they were, after all, paid for and made by the public—or else will belong to the Digital Library Program and will be freely available to the public. The only issue approaching a copyright concern involves composers who went on to become famous, whose

1

estates will be contacted so that those portions of the collection will be handled with all due sensitivity and legality.

The collection itself is finite and collection development will largely cease at the completion of the project. Our position as a digital library program in an academic institution is fortunate, allowing us free server hosting and maintenance on IU's web server. The only upkeep on the project will fall within our normal range of responsibilities. Specifically, digital media storage, upkeep and updating will be the responsibility of our Web development manager. IU graduate students will perform bibliographic updates. The collection is neither very old nor found only in IU's Government Documents Depository—there are more than 1,300 such depositories nationwide-- so it is not a great burden to adequately transport or store. Its duplication in other depositories also means that it will be available through interlibrary loan for the duration of its absence from IU's Government Publications Department. Access is the heart of the project and, thanks to the lack of copyright restrictions, access to the World Wide Web will be the collection's only requirement of users. Technical requirements

will be minimal and advertisement and promotions will be maximized to reach the widest possible audience.

All images, text or graphic, will be stored as TIFFs and presented as GIFs, with the exception of OCR processed text. Text documents will be processed using OCR software and presented separately in a searchable format. All materials will be searchable and numerous access points will be provided. Each piece of sheet music will have a companion MP3 recording of voice and piano

2

renditions. Users will be introduced to various significant portions of the collection by scholarly annotations and resources for further study will be provided. The entire project will take no longer than 10 weeks thanks to the proximity of the collection, the existence of suitable technology and of suitable staff on campus and the previous experience of the DLP with projects of a similar nature.

Collection Description:

The Government Publication section of the Main Library at Indiana University—Bloomington is home to the World War II War Bond Promotions collection. This collection contains all 110 War Bond Promotions publications produced by the U.S. Government between the years of 1942-1945. These publications fall into four categories: plays, musicals, songbooks, and advertisements. Among the items of special historical significance in this collection are songs by noted musicologist, Charles Seeger, and a short previously unpublished one-act play by Tennessee Williams. The collection also contains the books for three musicals published by what is now known as the Government Printing Office. These hitherto unheard of musicals were written by Guys 'N' Dolls librettist, Frank Loesser, and 2001: Space Odyssey composer, Alex North, and were choreographed by folklorist, José Limon, all before they became celebrities. Within many of these publications are valuable photos, reproductions of newspaper articles, and letters. Although these documents are not rare or in need of restoration and preservation, we believe the organization of the documents into a single collection that we will make available and searchable

3

on the World Wide Web would be unique and would be valuable to great numbers of users.

Mission

One of the main purposes of digitizing the the World War II War Bond Promotions materials is to create awareness of the existence of this special collection of documents, not just in the academic community, but also among the general public. The collection is currently almost completely unknown, even among scholars in the field.

By digitizing the collection, we will make it searchable, which will allow users to interact with it more richly than if they were simply looking through the printed materials. Also, the digital version of the collection will permit users to actually listen to the musical pieces contained within the collection. We believe that this feature will increase its educational value significantly.

Given the recent increased interest in the World War II era, as illustrated

by the release of the movie "Pearl Harbor" and Tom Brokaw's book, The Greatest Generation, we believe that this is an opportune time to digitize this collection and help raise awareness about this important period in history. We hope that it will prove to be a valuable addition to other pre-existing digital libraries with World War II themes (a list of these other World War II related digital library sites can be found at
http://www.academicinfo.net/histww2library.html).

Goals

Our main goal is to create a digital library that is user-friendly and yet supports the kind of advanced searching capabilities that scholars and perhaps
4
lay people will find useful when conducting research on the collection. We hope to create a solid information architecture for the site. By separating the content from presentation, we expect to ease the process of changing the site's interface, which will most likely be necessary in the future as browser technologies change and new web markup language standards are endorsed by the Word Wide Web Consortium (http://www.w3.org). The separation of content from presentation will also make it possible to more easily port the collection to other media, such as CD-ROM or DVD-ROM, should this be desired at some point in the future.

Who is our audience?

We anticipate that this project will be useful to a variety of communities. Those interested in wartime songs—scholars (folklorists, ethnomusicologists, theater historians, and WWII historians) and WWII aficionados—will find these songs exciting because many of them were (a) composed by amateurs, and/or (b) written or adapted by children, and because the publication of songs of this type is rare. In addition, the three complete musicals—perhaps the most professional items in the collection--will pique an interest in theater historians, historical musicologists, and possibly professional and amateur theater producers looking for a new hit show in today's revival market.

We also believe this collection will prove valuable to those interested in wartime images. Documentary filmmakers and photojournalists, as well as academic researchers, often complain about the canonization of a particular core of images that happens to reside in the Smithsonian Archives. This collection
5
would make available visual perspectives of the World War II era to which photojournalists have not yet given a voice.

Those interested in genealogy might be interested in using this collection to find special information about their relatives and friends. This collection is interesting in that many of its materials link together specific names and places. There is, for example, a newspaper clipping of a woman winning a war song contest in Akron, OH. Family members may have only seen that clipping in the original newspaper, but may never have known that the U.S. government nationally recognized the incident.

High school and college teachers could make use of this material to present multiple perspectives on WWII. On one hand, professors could highlight the propagandistic, jingoistic nature of these materials—one gets a fresh insight

into how the U.S. government was promoting the war on every level. On the other hand, professors could highlight this period as a time when the government was encouraging great quantities of artistic expression from amateurs and semiprofessionals,
a time when there were numerous prizes and awards being given
to encourage creativity.

Copyright:
The War Department of the U.S. Government created all of the materials
in this collection, and as such, they all exist in the public domain; therefore, there are no known copyright issues. Pre-existing materials that the War Department used for propaganda were already in the public domain, and the government automatically gave public domain status to materials created by civilians and soldiers for the bond promotion campaign. Sound recordings made for
6
presentation on the Web will belong to the Digital Library Project, and we will make these freely available to the public. Although these materials are legally free from copyright restrictions, certain parties may not wish, for whatever reason, to see these works displayed on the Internet. We are concerned, namely, about the estates of the more famous composers and authors who have their works in this collection. Therefore, we will contact their estates to get a sense of how they would like us to handle this material.

Collection Maintenance and Development
This digitized collection will be hosted free of charge on IU's web server,
and therefore, any 'server upkeep' responsibilities will be theirs. Because of the finite quantity of data that we are digitizing, we do not anticipate significant development on this project after its completion; we foresee just two minor upkeep tasks. (1) We will keep copies of the project stored on backup media and will house this media in our facilities. It is possible that years to come, this media might become defective—the life-span of CD-Rs is still unknown. Therefore, we will be ready to transfer the files to other media when the time arises. Being a Digital Library Program, we have a member of staff that is responsible for the maintenance of all of our web projects. We will add the task of maintaining this project (i.e. making back-up copies of this project, updating media storage, etc.) to this staff member's workload, and s/he will periodically upkeep the project in conjunction with other maintenance tasks. (2) The bibliographies provided by our researchers will eventually grow out-of-date. We will periodically contact graduate students on our campus and ask them to update these bibliographies.
7
Acquisition Process and Handling Issues
All of these materials are stored in IU's Government Publications
department in its function as a U.S. Government Document Depository. This department has already given us permission to work with and electronically publish these materials, with the minor requirement that we give Web space to highlight their other collections. While we are working with these materials, they will not be available for use by library patrons. However, if patrons request these materials, they can acquire them through IU's interlibrary loan service from other

U.S. Government Document Depositories. These documents will not require extraordinary care for two reasons: (a) they are not ancient or even old, and the government published numerous copies of each item, and (b) because both the Government Publications Department and the Digital Library Project reside in the same building, there should not be any shipping or handling problems to arise when transporting them to our work area.

Access

This collection will be accessible through the WWW. Our aim is to disseminate these materials to as many users as possible, and thus, no fee or registration will be required to view our site. In order to make full use of this site, users will need Netscape Navigator or Internet Explorer 4.0 or higher, and RealPlayer. Because all documents included are in the public domain, anyone with access to the World Wide Web will have access to the collection.

For us, access also implies knowledge; users must know about the project before they can access it. Therefore, we plan to heavily promote the project in a variety of communities, so that anyone who might have an interest in the web site 8

will have heard of it. One idea that we have so far is to have the unveiling of the project coincide with a historic WWII date like Pearl Harbor or D-Day, just as IU's Hoagy Carmichael collection was unveiled on his birthday.

Features & Functionality:

We will digitally scan each page of an object in the TIFF format for archival purposes. We will then create smaller GIF files for use on the World Wide Web. Users will be able to flip through web pages as if they were flipping through the objects themselves, taking note of the formatting of the text, the organization of pictures, etc. We will create a digital table of contents for each page to facilitate easy navigation, and each page will contain hyperlinks to other pages within a given item. We consider this method of presentation more accessible than a PDF conversion, because of the limited dispersion of the plug-in that displays PDFs.

We will also convert pictures into GIFs, and will descriptively index similar to what we will do with the song lyrics. Index names like "boy with rifle pointed at mock Nazi," or "middle age woman happily buying war bonds in Ithaca, New York" will help users find the images they will be looking for. This part of the project will require a double phase of indexing. First, the appropriate historian will determine the most crucial information that s/he would convey about the content of each picture. Second, the metadata specialist would format the historian's captions into a mode suitable for further indexing.

We will also scan each object through an OCR application. The OCRproduced pages will allow us to construct a search engine that enables users to look for keywords of their choice within each object, or within the entire collection. 9

We will then format the OCR text and display on a text only version of the site; users will be able to download pages quickly for easy reading, and will have a second way to browse the material. We feel OCR is a preferable method of input to manual key entry because of many the unique text formatting issues. With

manual key entry, a typist would need to use extra time to determine a new margin layout for each given song or poem, while with OCR, the computer will be able to quickly accomplish this task.

The OCR processing will also allow us to extract materials from the objects and organize them in ways that are more meaningful; users probably will not be searching by object, but instead by the genre of material within the object. On the home page, for example, the user will be able to access an index and jump to a sub-index of all the song lyrics by title, song lyrics by first line (common in many song indexes), materials by year, materials by genre, or perhaps of composers and authors. Users who come to the site looking specifically for materials related to Frank Loesser will have the option of moving directly to our index rather than using the search engine.

A special component of our project will be the recordings that we will make of each piece of sheet music in the collection. Users will be able to hear what the music may have sounded like. This is particularly valuable in the many cases of song lyrics in the collection that bear the inscription, "Sung to the tune of…." Users today and in the future may not recognize those song references, and thus might feel alienated without the aid of the MP3s. All of the composers in this collection have arranged their music for voice and piano, and therefore their

10

songs will be relatively easy to record. We will convert these recordings to MP3 files, and users will be able to download them from our site. Users will find links to recordings next to song lyrics or items of sheet music, and there will be a separate index of each recording.

The collection will also include many research components that link it to a variety of academic communities, and to the public. Each researcher—an ethnomusicologist, a theatre and dance historian, and a WWII historian—will prepare an overview essay that situates the collection in a different historical context. The researchers will then annotate the songs, pictures, authors, and composers that they feel are noteworthy. These annotations will exist in an optional frame below each GIF and OCR page; users will decide if they would like to visit the site with or without a tour guide. Researchers will also provide a selected bibliography, discography, and/or filmography that direct(s) the users to relevant readings, recordings, and films.

Finally, there will be a page with links to other Digital Library projects, and to outside pages that have related materials. There will also be an 'about' page that encapsulates the mission and methods of the project, and that explains who and what is the nature of the Digital Library Project.

How It Will Be Done

The Digital Library Project already owns or has free access to much of the technical equipment needed to create this project. We have high-end digital scanners, software for scanning, for OCR, for web development, and for converting WAVs to MP3s, our own development server, a web server (IUB), and a recording studio (IU School of Music). Upgrades to this equipment are

11

automatically included in our annual budget, so if we need to upgrade this

collection in the future, we will certainly have the technology to allow this to happen.

We will begin our project by scanning each page of our collection as images into TIFF format. We will then rescan all photos and drawings as GIFs. This phase will take 2 weeks of full time work by an imaging technician. Our project supervisor will check to make sure that the scans are high quality enough for web presentation. After this phase, we will immediately send copies of the materials to the researchers, musicians, and front-end web developer. In this second phase, the musicians and engineer will record the music, the engineer will convert the audio files to MP3s, the researchers will compile essays, annotation, and the bibliographies, and the web developer will prepare the web design (i.e. graphics, layout, search engines, and overall thematic appearance). Simultaneously, a technician will be running OCR scans of the materials, and a part time graduate student will be proofreading the copies.

In the third phase of the project, we will send the annotations, audio files, images, and all other materials, back to the web designer for a period of final web development, and to the metadata specialist for indexing and cataloging. Toward the end of this phase, the Digital Library Project will design advertisements, and will get in touch with other libraries that have similar collections and let them know of the project.

Financial Plans

Below, is a table that outlines our budget for the project:

12

Activity Duration Cost

Scanning 2 weeks $11.50 x 80 hrs = $920 plus 25% benefits = $1150
OCR, extracting images and
XML markup
6 weeks $20.00 x 120 hrs = $2400 plus 25% benefits = $3000
Recording music 5 weeks $12.00 x 100 hrs = $1200 x 3 = $3600
Front-end web 4 weeks $20.00 x 160 hrs = $3200 plus 25% benefits = $4000
Back-end web 10 weeks $25.00 x 400 hrs = $10000 plus 25% benefits = $12500
Metadata 8 weeks $21.00 x 40 hrs = $840 plus 25% benefits = $1050
Project Supervisor 10 weeks $50,000 per year 10% = $5000
Researchers 8 weeks $50,000 per year 10% = $4000 x 3 = $12000

A word on the music fees: Indiana University—Bloomington has one of the more acclaimed music schools in the country. Any graduate student would be able to provide a professional level interpretation of the music of our collection. Being graduate students, however, we can ask them to record this music at a relatively low fee.

13

Project Phases/Timeline

Task Duration Start Finish Predecessors

Scanning 2w Wed 8/1/01 Tue 8/14/01
OCR, extract images, XML
markup
6w Wed 8/15/01 Tue 9/25/01 1

Record music 5w Wed 8/1/01 Tue 9/4/01
Front-end web development 4w Wed 8/1/01 Tue 8/28/01
Back-end web development 10w Wed 8/1/01 Tue 10/9/01
Metadata 8w Wed 8/15/01 Tue 10/9/01 1
Project Supervisor 10w Wed 8/1/01 Tue 10/9/01
Researchers 8w Wed 8/1/01 Tue 9/25/01
14
Conceptual Model
15