Madden Library Digitization Plan

1. Introduction

1.1. The Vision

The Henry Madden Library participates in the creation and maintenance of the emerging global digital library by digitizing and sharing electronic information.

1.2. Goal

The goal of Henry Madden Library's digitization projects is to create, maintain, and provide access to electronic resources which support the scholarly and informational needs of the Fresno State community; and also to make these resources available to the wider community, thereby contributing to the educational, cultural, and economic well being of our region and beyond. The library will pursue its vision by expanding and enhancing access to published and unpublished materials of potential interest to the academic community and general public, especially those unique to Fresno State and its partner institutions.

1.3. Statement of Purpose

Digitization is the process by which an electronic representation of a tangible item is created. Transferring materials from tangible to digital format protects the original item from damage while providing public access to materials that might otherwise be restricted. The Henry Madden Library Digitization Plan has been developed in accordance with the objectives of the Library’s Strategic Plan in order to facilitate the accession and digitization of selected materials and collections. The Digitization Plan is intended to provide a framework for selecting, cataloging, digitizing, digitally storing, and providing electronic access to resources in the Library and partner collections.

1.4. Why Digitize?

There is a wealth of information in older printed materials and in special collections documents such as letters, photos, postcards, etc. However, many of these documents are in poor condition, and are too fragile for frequent use. It is important to capture a digital copy of these works before they deteriorate completely. Researchers can use a digitized version for most purposes, saving wear and tear on the original. By digitizing the unique documents in special collections and archives, we also make them available to a far wider audience. Researchers no longer have to travel to the place where the document is held; the document can come to them. People who would never be allowed to handle rare documents - schoolchildren, college students, casual researchers, hobbyists - can actually use these historical artifacts in their studies.

1.5. Objectives
1.5.1. Preservation

The Library seeks to preserve valuable items and collections by creating digital copies. These copies can be used in lieu of the original item. Many of the materials to be digitized are in a deteriorating state. The library will perform all necessary repairs to the original materials before beginning digitization. Preservation of the original is our primary concern, and we will take every precaution to protect the originals from damage. While digitization of fragile materials can prevent wear and tear on the original and can thus act as a preservation tool, it is in no way a substitute for the original material.

1.5.2. Access

Digitization of valuable items provides enhanced access to materials that might otherwise be unavailable for public use. The Library also seeks to digitize materials of great interest and items whose present format obstructs ease-of-use. Through the Library’s discovery systems and other methods, users will have the ability to access digitized information both on site and remotely.

2. Digitization Process

Digitization is a long and complicated process. There are many steps involved. Every project is different, but the four basic stages include:

Stage 1. Select Material
Stage 2. Convert normal text into electronic text
Stage 3. Format electronic text for the Internet
Stage 4. Create website for access and navigation

2.1. Selection

The following criteria will be taken into consideration when evaluating projects for selection:

- Uniqueness: Items that are unique or rare should be considered high priority candidates for digitization, as the likelihood of duplicating previous digitization efforts will be low, and digitizing these materials will greatly increase the public’s access to them as they are not widely available.
- Historical Significance: Items of significant historical value should be given high priority.
- Regional Significance: Items documenting the California Central Valley should be given high priority, given our mandate to serve this region.
- Pedagogical value: Items of significant educational value to students at Fresno State and beyond should be given high priority, given Fresno State’s educational mission.
● Past use or potential use by patrons: Items known to have high use should be given high priority, as should items known to have a high potential for use once they are digitized. For example, items with a specifically identified and interested audience.
● Context: Only identified photographs should be digitized.
● Size/Format: The size and/or format of the item to be digitized may limit the library’s ability to digitize it given the capabilities of our scanners. In some cases, it may be advisable to outsource the scanning of such materials.
● Fragility: Fragility may either increase or decrease an item’s potential for digitization. Fragile items that are able to be digitized with minimal damage should be given high priority, as digitization will reduce handling of the originals. However, some items may be so fragile that they cannot be digitized without significant damage; these should not be scanned without first stabilizing the item.

To determine eligibility for selection materials should fulfill the following criteria:

● Meet the research needs of faculty, students, and scholars within and beyond the CSU Fresno community. In assessing what material meets the needs of our constituency, consideration should be given to the scholarly content of the material; the uniqueness of the material; and the demand for the material.
● Benefit from increased access and should contribute to the Library's service and collection development missions. Materials that are difficult to access in their original formats or that would benefit from increased speed or depth of access via electronic delivery formats should be given priority.
● Have clear ownership and copyright clearance. Before a digitization project is undertaken, the Library needs to secure sound legal advice about the ownership and rights to reproduce or publish materials electronically.
● Be of interest to potential partners. Materials that would be of interest to campus and outside partners, both collaborators on the content and potential sources of funding and other support, should be given strong consideration.

Also, before selecting materials, consideration for their preservation is made from the following perspectives:

● a) Items should not be digitized wherein the scanning process is detrimental to the item itself;
● b) Items that receive heavy patron use and are quickly deteriorating should be selected for imaging in order to preserve the original. Although data migration is an ongoing concern, digital editions will not be considered preservation quality reformatting for original editions until technological issues are resolved and standards are widely accepted.

A specific checklist of attributes, access, infrastructure and preservation concerns are included in the "Suggested Collections/Materials to be Digitized" form, available on the Library's web site. The Collection Development Committee will make decisions as to which suggested materials will be chosen for digitization. Established collection development criteria and policies will be utilized. Selection for digitization requires that materials have enduring value and be available in
a sufficient number or quantity that they form a significant and unique research corpus. Further, the decision to digitize must take into account many factors, as evidenced by the criteria on online form.

In selecting materials, the Library will actively seek out partners, both collaborators on specific projects and supporting partners to supply funding or technical assistance. In addition, the Library respects cultural traditions of different ethnic and racial groups in preparing its digital collections; consultation with tribal or other interested organizations will be conducted prior to digitizing potentially sensitive materials.

2.1.1. Copyright

Securing copyright permission is an overriding concern with all projects. The most immediate problem involving copyright and digitization is identifying what collections or parts of collections can be legally mounted on our web server. The rigor of establishing copyright clearance is not grounds for automatic dismissal of potential projects; however, ease of establishing permission will influence the priority of projects. Digitization projects with clear rights or easily obtained rights should be undertaken first. While these projects are undertaken, rights can be sought for subsequent projects.

2.1.2. Selection for Institutional Repository

The FSDR accepts all kinds of scholarly materials and content including pre-prints, previously published material (if permitted by publisher policy), working papers, technical reports, presentations, data sets, and instructional or creative works. Content relevant to the University’s history such as speeches, interviews, reports, and campus events are also collected.

Guidelines for submission include:

- The content must be produced, submitted, or sponsored by Fresno State personnel.
- The content must be related to scholarly, creative, educational, or administrative activities at Fresno State.
- The content must be in digital form or in a form that can be digitized.
- The content should be ready for distribution and not in draft form.
- The author/owner must be willing and able to grant Fresno State the right to preserve and distribute the content via FSDR.

2.1.3. Selection for CONTENTdm/Digital Management system

- Must be owned by the Henry Madden Library or from a partner institution
- Must be funded by donor(s), a grant, the Library or a partner institution
- Must have staff dedicated to finishing the scanning and metadata
- Focus is on whole or partial collections, not single items
General principles: materials must be sensitive to privacy/confidentiality and right to publicity concerns. Oral histories must be transcribed (and captioned, if video files). Embargoes on materials should have an end date (otherwise, why keep it?)

2.2. To Scan or Re-key? (and Scan Once Method)

The condition of the materials will determine how they are converted to electronic form. Very fragile materials, anything printed before 1940, and any manuscripts will have to be re-typed, because the optical character recognition ("OCR") software used to convert a scanned image to text will be unable to recognize the textual characters. We use an overhead scanning device that is less damaging to books than a flatbed scanner. If the print is clear enough to OCR, the documents will be scanned, OCR'd, and saved as text files. Whether scanned and OCR'd or re-keyed, all text will be proofread. Our goal is 99.95% accuracy.

2.3. Step 3 & 4: Methods of Access to Digital Content

Digital content will be made available online in formats appropriate for online delivery, which may differ from formats used for preservation (e.g. lower resolution images, streaming video rather than downloadable files).

Metadata for digital material will be added to the library catalog at at least the collection level, with item-level metadata where appropriate.

3. Metadata

3.1. Descriptive Metadata

Descriptive Metadata refers to the elements required for discovery and identification of a resource. It describes bibliographic and physical attributes of the material while associating access points. These points are structured in one or more standardized and published schemes such as Dublin Core or MARC.

3.1.1. Element sets/schemas

- Digital collections of the Henry Madden Library may use and accept any descriptive metadata standard to describe objects in the collection. In instances where metadata is not already provided, Dublin Core is recommended as the minimum standard due to its operational use in CONTENTdm and DSpace and as a requirement (in unqualified form) for data exchange using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Minimally accepted descriptive elements include an identifier, title/caption, creator, type, and genre. Metadata already provided will be stored in Dublin Core format. Data dictionaries and/or templates may be required on an individual project
basis. Encoded Archival Description (EAD) is used for additional finding aids as needed. Some data may be crosswalked from other standards into Dublin Core.

### 3.1.2. Vocabularies

- Subject terms will vary by collection but will preference Library of Congress Subject Headings (LCSH), LC Thesaurus for Graphic Materials (TGM), Art & Architecture Thesaurus (AAT), ProQuest Dissertations & Theses (PQDT) Subject Categories, and locally created controlled vocabulary lists. Subject keywords or tagging may be added in addition to standardized subject terms.
- Name terms will vary by collection but will preference the Library of Congress Name Authority File (LCNAF), Getty Thesaurus of Geographic Names (TGN), and locally created controlled vocabulary lists.
- Language standards accepted include MARC Code List of Languages (terms or codes) and ISO 639-1.
- Date standards accepted include WC3- DTF and Library of Congress Subject Headings (LCSH) for named periods.

### 3.1.3. Content standards

- Provided or locally created data dictionaries may be used to create or enhance descriptive metadata for a collection. Other accepted data content standards include Anglo-American Cataloging Rules (AACR2), Resource Description and Access (RDA), Cataloging Cultural Objects (CCO), and Describing Archives: A Content Standard (DACS).

### 3.1.4. Data format

- CONTENTdm and DSpace export data in Dublin Core XML schema. EAD DTD is used for finding aids.

### 3.2. Administrative Metadata

Administrative Metadata includes rights management, preservation information, and technical details of a digital object's creation.

#### 3.2.1. Rights management

Rights and copyright clearance will be investigated and documented for every digitization project on an individual basis. This information may be stored in a Dublin Core rights or other descriptive field, in a MARC description field or note,
negotiated as part of a grant, or handled by individual departments and stored according to their own record keeping procedures.

3.2.2. Technical details

- CONTENTdm and DSpace capture and maintain minimal information about the original file such as size, digitization dates, and file format. Specific technical details related to scanning procedures (bit depth, resolution, compression, speed, etc.) can be found under “Digital Imaging Requirements”. Individual projects may necessitate changes in these requirements to be documented and kept with project plans.

3.3. Structural Metadata

Structural Metadata describes relationships between materials including links to related files, internal structures of documents, and how compound objects are put together.

- Organization of files relies heavily on proper file naming conventions appropriated at the local level and dependent on the library’s individual role in the project. Additionally, CONTENTdm and DSpace provide options for implying structural metadata such as hierarchy and grouping of files, pages and views, and types of files present. These systems also rely on searching through Optical Character Recognition (OCR) versioning.

4. Digital Imaging Requirements & Specifications

Digitization specifications have been developed for different material types to provide a method of evaluating the quality of images produced, to estimate the data storage for access files (on-line) and master files (offline), and to assist in maintaining a certain level of quality control. Inherent differences in document types dictate different approaches to scanning. Due to the desire to image documents in a consistent manner, all scanning is done in 8-bit grayscale or 24-bit color.

4.1. Material Specifications

Specifications have been developed for the following types of materials.

4.1.1. Textual Documents:

Media formats provided for scanning include original records and manuscripts, bound volumes, or microfilm. The scanning resolution for the master files of 400 dpi for smaller documents was selected to be compatible with OCR software. The lower scanning resolution of 300 dpi for larger documents was selected to be of
reproduction quality and to save file storage space. Archival copies are scanned to TIFF or PDF.

4.1.2. Photographs:

Media formats provided for scanning include B&W and color photographic prints and transparencies. The digital master files are produced at 400-600 dpi based on the size of the image format. Images are scanned to TIFF for archival copy.

4.1.3. Maps and Oversized Records:

Media formats provided for scanning include original and photographic copy. The scanning resolution for the master files is between 600-1200 dpi based on the size of the documents. The lower scanning resolution of 600 dpi for larger documents was selected to be of reproduction quality and to save as archival copy.

4.2. Graphic Records:

Line drawings, artistic illustrations, and other similar records. Media formats to be provided for scanning include original records. Smaller graphic records are scanned in the same manner as photographs and larger graphic records are scanned in the same manner as maps, plans, and oversized records.

4.3. MASTER IMAGE FILES:

Pixel Depth: 8 bit grayscale or 24 bit color, RGB mode.

Resolution and Image Size

4.3.1. Textual Documents and Maps Records:

Original textual documents are scanned at 300-400 dpi depending on size and condition; examples: 8” x 12” at 400 dpi; 16”x 20” at 300 dpi. Original maps smaller than or equal to 12”x12” are scanned at 1200 dpi; maps larger than 20” x 20” are scanned at 400-600 dpi effective resolution. Examples: maps- 9” x 9” at 1200 dpi; 24” x 20” at 600 dpi; 40” x 120” at 400 dpi.

4.3.2. Photographs and Graphic Records:

Scanning resolution adjusted to produce files with pixel arrays of 6000 pixels across the long image dimension by the proportional number of pixels for the specific photo format, or at 400-600 dpi, depending on the scanning requirements. Scanner shall provide true optical resolution of at least 6000 pixels across the long
dimension of the image. Interpolating to a higher resolution from a lower resolution scan shall not be permitted.

File Format:
Uncompressed TIFF or PDF.

4.4. File naming conventions:
A file naming scheme will be established prior to capture and will be determined by the project specifications.

4.5. Storage:
Production master image files will be stored on hard drive systems with data redundancy, such as RAID drives. Additionally, another set of images with metadata will be stored on an LTO-6 tape format, and will be archived on Amazon's Glacier long term archival solution.

4.6. Backups

Regular backups of the images onto tape from the RAID drives will be performed with checksums stored alongside image files. Regular backups of local archives will be copied over to Amazon Glacier for long term archival.

5. Digitization Projects Summary

5.1. New Projects
- California State Lands Commission Maps & Aerials
- Fresno State Herbarium
- Music Recitals
- Athletics Videos
- Rod House Aerial Negatives Collection
- Hirasuna Papers Digitized Collection
- StoryCorps interviews
- Valley Foodways
- Central California War Veterans Oral History Collection
- Britto Club
- Hirasuna family
- Walter E. Pollock
- Violet de Cristoforo

5.2. Ongoing Projects
- Maps and Aerial Photography
Digitized aerial photographs of California and maps from Fresno State and partner collections. These materials are also spatially indexed and discoverable in our Map and Aerial Locator Tool (MALT)

- Garabedian Collection
  - The V.E. Petrucci Library’s collection of Viticulture and Enology Department documents and images, including proceedings, slides and other research materials, pamphlets, and Fresno State wine labels.
- CSU Japanese American Digitization Project
- Water Archives Oral Histories

5.3. Accomplished Projects

- Collegen
  - The Henry Madden Library houses the complete collection of The Collegian, Fresno State’s student-run newspaper, on microfilm going back to 1922. In a three-year project the Library digitized the entire microfilm collection from 1922-1998 and it is now available online.

- Yearbooks
  - This collection of Fresno State yearbooks stands as a treasure trove of photographs and information and memories, and contains yearbooks from 1911 to 1991 (with a gap from 1972-1975).

- The Academy
- Arnie Nixon Center Cloth Books
- Collegian
- Dissertations and Theses
- Fashion Plate Collection
- Fresno Historical Society Hutchinson Collection
- Hammer Field Collection
- Harry Pidgeon Photograph Collection
- Henry Madden Library Realia
- Hye Sharzhoom Newspaper
- Insight
- La Voz de Aztlan
- Leon S. Peters Legacy Collection
- Leon S. Peters Papers
- Politi Inscriptions
- Radio Bilingue
- Retrospective Theses
- San Joaquin Valley Japanese Americans in WWII
- Theater Arts Garment Collection
- Topolobampo Collection
- University Archives Photograph
- Woodward Family Collection
- Woodward Local History Collection
- World's Fair Collection
- Yearbook Collection