

George M. Aldridge Memorial Collection
Digitization Project

SJSU

Spring 2017

Group 1:

Group Member 1

Group Member 2

Group Member 3

Mary Beth Romo

George M. Aldridge Memorial Collection

Digitization Project

Introduction

The George M. Aldridge Memorial Collection consists of 35 documents from Mr. Aldridge's personal collection of letters, autographs, and other historical paperwork and paraphernalia from notable Americans of the past. The late Mr. Aldridge was a history buff and an avid collector of such items, developing a sizeable array of historical and culturally significant items relating to the economic history and financial development of the United States. These items are written or signed by key figures involved with the early development of the American economy such as United States Secretaries of the Treasury. These documents are currently owned by Jean and Rafael B., and in the possession of Mónica B., Mr. Aldridge's granddaughter, for use in this digitization project.

Planning for Digitization

The overall theme of the collection is "Letters signed by United States Secretaries of the Treasury and other key figures involved in the early history of U.S. finance." This theme was easily determined from the collection material itself. Group 1 decided on this collection from Mónica's grandfather because of the interesting nature of the material and its genuine historical value. The 35 items we digitized are handwritten letters, written on paper, by historical figures and are dated between 1795 and 1894. It was exciting to discover that the collection contained material signed by prominent Americans of the past. A letter signed by Alexander Hamilton, for example, was especially interesting due to his current position in modern popular culture and art.

The main goal of this project is to provide access to digital surrogates of the George Aldridge Memorial Collection. The needs of users must therefore be considered. The potential

users of this collection are historians, writers and scholars interested in the United States Treasury Department, the historical figures involved in the letters, or 18th to 19th Century Americana. Other potential users would be individuals with a personal or professional interest in the subject or content of the letters, such as history buffs. Mr. Aldridge was a thorough individual who saved correspondence between himself and historians or authors requesting copies to be made of the personal and business letters for their study. Some of these individuals were conducting research for academic papers, while others were writing biographies or histories of the Treasury Department of the United States. There is clearly a real-world application for this collection and therefore an obvious need for these items to be digitized and accessible. The user needs would require enough metadata to ensure a high level of findability when searching. There would need to be details about who wrote the letter and to whom they were writing, the date of the letter, descriptions of the content of the letter or transcripts of the letter, high quality scans of each letter for close scrutiny of the content, and related information such as subjects headings, dates of service, and appointing President if applicable.

Selection

The selection process began with a closer look at the items to determine their exact nature, and how many were suitable for including in our digital collection. Mónica took pictures of all the documents for all group members to have a better look at their content and appearance. Additional items included with the documents consisted of identifying name cards that organized the documents in storage folders, information cards from the dealer Walter R. Benjamin Autographs, and transcripts of the documents, although not all had information cards and/or transcripts. Eliminating simple autographs without a corresponding letter or document, and letters simply requesting an autograph that had no historical value, left us with 50 letters and

official government documents signed by persons who served as Secretaries of the Treasury or who were closely associated with the Bank of the United States. Focusing only on items from before the Lincoln administration, the count was reduced to 35 documents written and signed by key players in the founding of the financial structure of the United States Government. This included a document written by Nicholas Biddle. Although Biddle was not a Secretary of the Treasury, as the last president of the Second Bank of the United States he was of great significance in the founding of the financial backbone of the United States, and was worth including in our collection.

For the sake of consistency, we changed our criteria to only include handwritten letters, and to exclude other types of documents. This required us to expand to items beyond the Lincoln administration. We also considered the quality and preservation of the original documents since they are aged and in deteriorating condition. One document was eliminated due to fragility of its binding which would have made it difficult if not impossible to scan with a flatbed. The final 35 documents we hope are suitable for the historians and researchers of our intended audience.

Copyright considerations

Intellectual property issues must be evaluated before digitizing a collection of analog materials. To avoid litigation and penalties, it is necessary to establish that works are in the public domain, or to secure permissions from copyright owners. Although the materials included in the George M. Aldridge Memorial Collection are dated between the years 1795 and 1894, it should not be assumed that they are in the public domain. In a real-world situation, the public domain status of these items should be ascertained. There are three conditions that would satisfy public domain status.

If these letters were registered or published before 1923, they are in the public domain. Copyright status should be explored by searching the records and catalogs of the U.S. Copyright Office. Alternatively, the U.S. Copyright Office could be requested to conduct a search for these materials (U.S. Copyright Office, 2013. p. 1).

If the items in the collection were unregistered and unpublished, they would be in the public domain 70 years after the death of the author. The last death date of an author is 1905, which means that if the letters were unregistered and unpublished, they are in the public domain.

If the materials were created as part of the duties of a US government employee, they are in the public domain. "The U.S. Copyright Act clearly states that copyright protection in the United States is not available for any work of the federal government" (Copyrightlaws.com, 2017). Group 1 has established whether the materials were created *during* the author's service in the government. However, it would require an expert, and is beyond the scope of this project, to determine whether many of the letters were written as *part of their duties*.

If public domain status cannot be established, digitization of materials would require permissions from any copyright holders. An institution in a real-world situation would need to research the existence of copyright ownership with due diligence. Keeping clear documentation of attempts to gain permissions is recommended in order to demonstrate good faith (De Stephano, 2000, pp. 11-13).

An institution may decide to take a calculated risk, and digitize these materials without copyright permissions as *fair use* (HATII & NINCH, 2003, p. 47-48). Title 17 of the United States Code, Section 107: Limitations on exclusive rights: Fair use (17 USC § 107) allows *fair use* of copyrighted works, which usually include "not-for-profit educational use and private research and study" (p. 65). Group 1 believes that these materials are likely unpublished, and that

the risk factor in digitizing this collection is low. However, establishing this definitively would require significant time and funding.

The rights status of the George Aldridge Memorial Collection is stated "clearly in the metadata associated with the digital resource" (HATII & NINCH, 2003, p. 62). The copyright statement for this project is as follows:

The materials in the George Aldridge Memorial Collection are made available for noncommercial, personal, research, or educational use only. We believe this constitutes a 'fair use' of any copyrighted material as provided for in section 107 of the US Copyright Law. The collection is owned by Jean and Rafael Bonilla. Collection metadata are for descriptive purposes only, and not definitive indications of the creators, copyright status, or copyright owners of these materials. The collection owners have no information on the status of copyright ownership in the collection, but we believe the materials to be in the public domain. The responsibility for securing any necessary copyright permissions to publish documents, texts, and images in this collection rests solely on the user. For more information regarding these materials, please contact the source of acquisition: Walter R. Benjamin Autographs, Inc. (518-263-4133).

Technical Production

Bit-depth and Resolution

The manuscripts in the George M. Aldridge Memorial Collection have soft-edged characters produced in ink, by hand, on reflective paper. The dynamic range, color cast, details, and physical condition are important characteristics to consider for digital capture, and influenced our choices of bit-depth, resolution, and scanning equipment.

The dynamic range of the manuscripts is low to medium, meaning that there is a relatively low contrast between the text and the background (Kenney, 2000, p. 37). Although the

colors in these manuscripts are flat, and the palette is limited, they convey important information regarding the age and condition of the documents. "Unintended color" has appeared in these documents through age and use (p. 42). "Some media, such as graphite pencil, early inks, and the gold leaf used to illuminate manuscripts, are highly reflective and more difficult to capture than more absorbing media, such as common printing inks or watercolors" (p. 37). Many of the documents in this collection are yellow with age and early ink, which could be reflective in nature, was used. The stains, markings, and smearing, require tonal distinctions to preserve the document's appearance (p. 33). Colors also reveal clues about the medium and support, which may be of interest to scholars and researchers. Group 1 chose to scan the documents at 24 bit-depth, as recommended by the "Minimum Digitization Capture Requirements" (ALCTS, 2013). "Most institutions have chosen to scan manuscripts in grayscale or color to maintain the softer-edged features characteristic of handwriting or impact printing (e.g., typewriting)" (Kenney, 2000, p. 35). The specific setting used during scanning was, "Original Type: Full Color Txt/Photo; RGB" (3 channels, 8 bits per channel= 24 bits).

Manuscript details involving legibility influenced the choice of resolution for digital capture. Details of manuscripts are measured in terms of the width of the finest stroke (Kenney, 2000, p. 44). The finest line or stroke should be covered by at least two pixels. The ALCTS (2013) recommends a minimum resolution of 400 ppi for manuscripts and Group 1 found that 400 dpi was adequate in maintaining the necessary detail for display. "Other projects have concluded that increasing resolution improves quality appreciably up to a certain point—between 400 and 600 dpi (grayscale or color)—increasing resolution beyond that must be weighed against rising costs or larger file sizes" (Kenney, 2000, p. 47). The specific setting used during scanning was, "Resolution: 400 dpi."

Scanning Equipment Used

A RICOH MP 3054 copier/scanner was used to scan all items in the collection. The specifications of this equipment are detailed below.

Scanner Specifications

| | |
|-----------------------------------|---|
| Scanning Speed (B&W & Full Color) | 200/300 dpi: 79 ipm (Letter) 200/300 dpi: 80 ipm (A4) |
| Scanning Resolution | Color & B/W: Up to 600 dpi TWAIN: Up to 1200 dpi |
| Platen Size | Up to 11" x 17" |
| Connection Type | Standard: 10Base-T/100/1000Base T Ethernet Optional: IEEE 802.11a/b/g/n Wireless LAN |
| Standard Features | Embedded Scan-to-Email/Folder/URL, TWAIN and Color Scanning (Copyfaxes, 2017) |

Problems arising during scanning process

The RICOH MP 3054 copier/scanner does not connect directly to a computer so as a result, documents were scanned onto a memory device and then transferred to a computer. This process, which did not result in any loss of quality, was time consuming. A real-world, large-scale project would be more efficient with a direct connection.

Protection of the original document is paramount to a successful project. "To assess documents, characterize any potential harm to the originals, such as strain on bindings, light sensitivity, brittle paper, or fragile glass plates" (Kenney, 2000, p. 33). The collection is

very old, and these paper products become brittle with time. Folded items were problematic during the scanning process. Letters that contain text on the interior were carefully unfolded for scanning. Folded documents with no content on the interior were scanned in their folded state to avoid causing strain and damage. This process provided an adequate digital surrogate of the original manuscripts, and did not result in any loss of important information. This method of digitizing documents in the current folded state is a reasonable solution to ensure preservation. The dimensions of these ‘unopened’ documents were recorded in their folded state.

Image production

The main goals of the George M. Aldridge Memorial Collection digitization project are to provide access to digital materials for scholarly research and to preserve aging documents. High quality detailed color reproductions are therefore required. The file format for archival masters of this project must "preserve the resolution, bit-depth, color support, and metadata of a very rich image file" (Kenney, 2000, p. 51). The TIFF format fulfills these requirements, and is the best choice for this project. Although new solutions continue to be explored, TIFF "remains the de facto standard" (p. 51). Group 1 scanned all original documents in TIFF format.

Although the TIFF format produces high quality images, files are large, expensive to store, and cumbersome to deliver. Large files are inappropriate for users who "expect fast retrieval, acceptable quality, and added functionality" (Kenney, 2000, p. 55). Group 1 was guided by the standard practice of “storing the master file uncompressed to facilitate digital preservation and limiting compression to access versions” (p. 54). Compressed JPEG files are of a lower quality than TIFFs due to the loss of information during the compression process. Nevertheless, JPEG files offer an acceptable balance between image quality and speed of

delivery. Photoshop CS6 software was used to compress the TIFF files into JPEG format, and to rotate the images to the desired display.

Metadata

Metadata for the George M. Aldridge Memorial Collection organizes and describes the collection, facilitates discovery, provides identification, and supports archiving and preservation. Record fields include descriptive, administrative, and technical metadata. Structural metadata will not be required for the purposes of this digitization project. (Guenther & Radebaugh, 2004, p. 1)

Descriptive metadata chosen for this collection identifies and describes each item and aids in discovery. Twelve of the fifteen basic elements established by the Dublin Core Metadata Initiative (2014) are used, including Identifier, Title, Subject, Creator, Description, Contributors, Date, Type, Format, Language, Rights, and Source. In addition to these fields, specific descriptive properties of this collection were identified by Group 1 to create additional metadata fields. The Group decided to refer to the unique aspects of this collection, and to anticipate user needs and expectations. "The best tools to locate a resource are those tailored to the user's knowledge and context" (Lagoze & Payette, 2000, p. 88). Members of the intended user group are US history scholars and enthusiasts. Group 1 decided to include a field for the Creator's Position in the US Treasury Department because this property unites the items in the collection and defines one of the main criteria for inclusion. The Appointing US President and the Dates of Service in the US Treasury Department provide related and useful descriptive information. Physical Dimensions and Condition of the documents were also considered essential descriptive information due to the age of the documents. Each document is accompanied by a Transcript, which the Group agreed was important since some manuscripts are difficult to read.

Technical and administrative metadata are "associated with the functions of administering digital resources, and [ensure] their long-term preservation" (Lagoze & Payette, 2000, p. 95). Rights, which is one of the 15 basic Dublin Core Metadata elements, is considered an administrative property. The Rights field contains information about intellectual property rights and use conditions. Digitization Specifications of the scanning software and hardware used to create the digital image, resolution, bit depth, and file type are important for purposes of preservation and integrity. All this technical information is contained in one field. The metadata will also identify who created the record for administrative purposes.

Below is a list of the 19 metadata fields used for the George M. Aldridge Memorial Collection, along with explanations of how we are using them and how we developed the data.

1. **Identifier** *descriptive*: This is a unique number assigned to each item to identify the item and distinguish it from other items. Items are numbered 1-35.
2. **Record Creator** (Lastname, Firstname) *administrative*: This identifies which member of Group 1 created the metadata record.
3. **Title** (Letter to Lastname, Firstname) *descriptive*: All documents in our collection are signed manuscript letters. Group 1 decided that the Title field should indicate that the item is a letter, and to whom the letter is addressed. (The "Creator" field describes who wrote the letter.)
4. **Subjects** (LC Political Science authorized subject heading; LC US History authorized subject heading/s) *descriptive*: We decided to use Library of Congress classification to describe the subjects of each item. After browsing LC authorized subject headings, Group 1 chose Political Science and US History as the main categories. US History is further specified by the dates of the Creator's service in the US Treasury Department.

5. **Creator** (LC authorized form of name) *descriptive*: This field identifies the US Treasury Department Secretary or treasury official who signed the document. Group 1 decided that this key feature unifies items in the collection.
6. **Creator's Position in US Treasury Department** *descriptive*: With the exception of Nicholas Biddle, who served before the existence of the US Treasury Department, all Creators acted as US Treasury Secretary. Group 1 decided to include the Biddle document because of the important role he played in US Finance.
7. **Appointing US President** (LC authorized form of name) *descriptive*: Group 1 decided that users may wish to search and locate documents based on the US President who was in office during the tenure of the Creator's service in the US Treasury Department. This finding aid may be useful for research purposes.
8. **Dates of service in US Treasury Department** (YYYY - YYYY) *descriptive*: Group 1 decided to include a field indicating the Creator's dates of service in the US Treasury Department. We felt this was an important detail, since some documents were produced outside the scope of the Creator's role in the US Treasury Department.
9. **Description** *descriptive*: This field is where the Record Creator will describe the object. This information will be a general description, accessible to the full range of users.
10. **Date** (YYYY-MM-DD) *descriptive*: This is a basic field describing when the object was created. The items in the collection all include a date in the original text.
11. **Type** (DCMI Type vocabulary) *descriptive*: This field will use controlled vocabulary to indicate the type of object. These items are all considered Text.
12. **Format** *descriptive*: This describes the physical medium of the object. Our documents are manuscripts.

13. **Language** (ISO 639-2 Code) *descriptive*: Because the objects in the collection are all text documents, we wanted to include a field to indicate that they are in English.
14. **Rights** *administrative*: This field contains our copyright statement for the collection.
15. **Digitization Specifications** *administrative*: Scanning hardware and software, bit-depth, resolution, and file type.
16. **Physical dimensions** *descriptive*: This information describes the dimensions of the paper the letter is written on.
17. **Physical Condition** *descriptive*: Group 1 decided to include a description of the condition of the objects due to their age. One term will be chosen from a list of three possibilities.
18. **Transcript** *descriptive*: Some documents are supported by a transcript obtained at the time of acquisition. Group 1 decided that transcripts will be uploaded to this field exactly as they appear in the original without attempting to make corrections. Errors may reflect errors in the original transcript.
19. **Source** *descriptive*: Group 1 is using this field to indicate from whom George Aldridge acquired the original objects. In most cases, this is Walter R. Benjamin Autographs, Inc.

Preservation

Preservation of the physical materials constitutes a portion of this digitization project. Digital access will reduce handling of these old and fragile materials. Items in the George Aldridge Memorial Collection are between 123 and 222 years old and some are brittle, torn, stained, and worn. Group 1 discussed and acknowledged the need for careful treatment during the scanning process to reduce the risk of further damage to original items. Documents were thus

scanned without the use of cotton gloves to ensure proper dexterity and nimble handling; hands were kept clean and free of moisturizer and oils.

The digital file format is crucial for preservation. Group 1 chose the TIFF format for digital images. "The digital surrogate is a form of preservation (although not a substitute for any other form of preservation) and must itself be preserved to ensure future access and use" (HATII & NINCH, 2003, p. 198). The master files will remain in the possession of the collection owners, along with the original materials. Correct storage of preservation files "and an environmentally controlled location will help to optimize their stability and protect them from loss," (HATII & NINCH, 2003, p. 201). The collection owners will begin this year to examine possible locations to donate these letters so this proper storage and preservation can be secured.

Other issues

Group 1, whose members are scattered across thousands of miles and separated by the Atlantic Ocean, faced logistical challenges in the design and implementation of this digitization project. The physical items were all in the possession of a single group member, which resulted in an uneven distribution of tasks. At the onset of the project, this group member provided access to a set of crude digital photos to the rest of the group via Google Docs, however, proper physical description required physical access to the materials. Scanning and uploading images to CONTENTdm also fell to the owner of the materials. The remaining group members attempted to compensate in other areas. For optimum quality control, access to originals by all planners would be preferable. While Group 1 agrees that the quality of the final project was not diminished, we recognize that the project would have benefitted from group input in all aspects of planning and digitizing. In hindsight, it might have been better to choose a collection to which all members could have contributed items. On the other hand, we unanimously agree that the

value of the George M. Aldridge Memorial Collection is worth any possible limitations occurring from the location of the materials.

Ideally, the full text of the manuscripts in this collection would be searchable. Several of the items in the George Aldridge Memorial Collection are not accompanied by a transcript or an information card. Group 1 decided to include these materials anyway, regardless of the lack of documentation, due to their value as part of a cohesive collection, which covers the US Constitutional period through the late nineteenth century. The lack of transcripts, presents challenges to users who may experience frustration while attempting to decipher aged handwritten manuscripts. For a fuller understanding of this material, scholars and handwriting experts would be needed to decipher and interpret the handwriting and apply Optical Character Recognition (OCR). Although these functions are far beyond the scope of current capabilities, transcripts for many of the materials are included in the metadata, and provide foundational information for future iterations of this project.

Conclusion

Based on analyses of various issues related to the George Aldridge Memorial Collection, Group 1 has planned and executed an online collection of digital images. Decisions regarding theme, potential users and their needs, copyrights, selection, technical and image production, metadata, and other issues were discussed remotely and agreed on by all group members. This small-scale project features 35 digital images of historically significant manuscripts, along with supporting metadata. A user group of US history scholars and enthusiasts will benefit from online access to these cultural heritage resources. Increased access to this collection contributes to its value for the purposes of education, research, and preservation.

This collection is dedicated to the memory of George Martin Aldridge, who was an avid collector and enthusiast of US History. Jean and Rafael B., current owners of the collection, wish to honor his legacy by inspiring a new generation of history scholars, experts, collectors, and enthusiasts. The George M. Aldridge Memorial Collection liberates cultural heritage materials from the bonds of ink and paper into the infinite realm of cyberspace.

References

- Association for Library Collections & Technical Services (2013, June). "Minimum Digitization Capture Requirements." Retrieved from <http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>
- BCR (2008). "BCR's CD Digital Imaging Best Practices Version 2.0." Copyfaxes. (2017). Retrieved from <http://copyfaxes.com/product/6768/Ricoh-Aficio-MP-3054-BandW-MultiFunction-Print>
- Copyrightlaws.com (2017, February). "Copyright laws in U.S. Government Works." Retrieved from http://www.copyrightlaws.com/us/copyright-laws-in-u-s-government-works/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+Copyrightlawscom+%28copyrightlaws.com%29
- De Stefano, P. (2000). Selection for digital conversion. In A.R. Kenney and O.Y. Rieger (Eds.), *Moving Theory into Practice* (pp. 11-23). Mountain View, CA: Research Libraries Group.
- Dublin Core Metadata Initiative (2016, June 14). "Dublin Core Metadata Element Set." Retrieved from <http://dublincore.org/documents/dces/>
- Guenther, R. and Radebaugh, J. (2004). "Understanding Metadata." Bethesda, MD: National Information Standards Organization.
- Humanities Advanced Technology and Information Institute (HATII), University Glasgow and the National Initiative for a Networked Cultural Heritage (NINCH) (2003). *NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*.

Lagoze, C. and Payette, S. (2000). Metadata: Principles, practices, and challenges. In A.R. Kenney and O.Y. Rieger (Eds.), *Moving Theory into Practice* (pp. 24-60). Mountain View, CA: Research Libraries Group.

Kenney, A.R. (2000). Digital benchmarking for conversion and access. In A.R. Kenney and O.Y. Rieger (Eds.), *Moving Theory into Practice* (pp. 24-60). Mountain View, CA: Research Libraries Group.

United States Copyright Office (2013). "How to investigate copyright status." Retrieved from <https://www.copyright.gov/circs/circ22.pdf>