

OHRAB Grant Mid-Project Report

August 31, 2013

Organization: The University of Akron, University Libraries, Archival Services

Project Title: Digitization of the Daniel Guggenheim Airship Institute Technical Reports

Project Director: S. Victor Fleischer

Telephone: 330-972-7670

E-mail: archives@uakron.edu

Summary of Project Activities and Accomplishments:

The purpose of this project is to digitize for preservation purposes and to make available online 182 technical reports from the Daniel Guggenheim Airship Institute. So far, everything is going smoothly and is proceeding on time. During the first month of the grant, the technical reports were located, pulled, and packed in shipping boxes with padding to protect them during transport by the Senior Archives Associate with the assistance of Archival Services student assistants. The reports were shipped via United Parcel Service (UPS) to The Image Collective (TIC) in Somerville, Massachusetts. The reports arrived safely and were not damaged during transportation.

Digitization of most of the technical reports is almost completed and the reports should be returned to The University of Akron Archival Services in the next few weeks. This is about three weeks past the timeline listed in the grant application, but will not affect the completion of the project. The vendor did encounter some issues scanning the reports, which caused this brief delay. The difficulties encountered were due to the fact that some of the pages in the reports are on onion skin paper with bled type, which required some adjustments in the scanning process, namely bumping up the contrast slightly. The vendor did this so the type will be more legible and so the PDF files could be OCR'd and searched. The other issue that was discovered during scanning is that some of the reports have fold outs or "tip-outs," which required a different procedure. The vendor solved this problem by skipping these pages and then scanning them later at a different resolution with an oversized scanner and then added them back into the PDF file, which is a standard procedure. Again, these unforeseen circumstances have slowed the process slightly, but will not affect the project's completion.

Digitization standards based on "Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images" produced by the National Archives and Records Administration (NARA) have been used during the digitization of the reports. The reports are being scanned at 300 DPI (dots per inch) uncompressed TIFF images for the master files. When the reports arrived at the vendor, the Head of Archival Services and the vendor discussed the file naming conventions, which have followed NARA

recommendations and include the report number followed by the document format extension. Dots in the report numbers were replaced with underscores. For instance, the uncompressed TIFF file for report number 02.01 is 02_01.TIFF and the PDF file is 02_01.PDF. The scanned reports have not been cropped or skewed by the scanning technician and reflect accurately the original image. Derivative PDF access files have also been created for each report by the vendor from each TIFF master file and will be used for online viewing.

The vendor has been using Optical Character Recognition (OCR) so the documents are fully searchable. The vendor has been backing up production master image files and derivatives on external hard drive systems with a level of data redundancy based on NARA recommendations. One external hard drive has been purchased and will be used to transport the master TIFFs and the derivative PDFs from the vendor to The University of Akron. Originally, an additional set of digital master file images with metadata stored on an open standard tape format, such as LTO (Linear Tape-Open), was supposed to be maintained offsite at the OCLC facilities in Dublin, Ohio. However, as we have found out throughout this and other digitization projects, OCLC does not store the high resolution master files or allow us to upload them to CONTENTdm. Instead, these files will be stored onsite on Archival Services' designated server, which is housed at The University of Akron Computer Center and maintained by Information Technology staff. This actually is a better option as it allows us quick and easy access to these digital files.

When the digitization is complete, the vendor will re-pack the boxes to protect them using the original packaging and ship them back to Akron. While TIC performs quality control onsite, another round of quality control and quality assurance will be conducted on all the digitized reports by Archival Services student assistants under the guidance of the Senior Archives Associate. This work will occur when the files arrive back in the archives, probably in early to mid-September, to ensure the digitization and metadata adhere to the image standards using the Archival Services training manual "Quality Control and Assurance Procedures for Scanning and Metadata Creation." Any issues will be resolved by rescanning the image or making adjustments using Adobe Photoshop (e.g., image rotation) or Adobe Acrobat Professional (e.g., pages out of order).

After they are trained by the Special Collections Cataloger in late September or early October, two Archival Services student assistants will create item level qualified Dublin Core metadata for all the reports. The student assistants will enter metadata for each technical report from late-September or early October to mid-December following procedures outlined in a metadata manual that will be created by the Special Collections Cataloger in coordination with the Head of Archival Services. To help facilitate topical searching, as part of its metadata each technical report will have a short description and one or more subject headings assigned to it. For subject heading controlled vocabulary, we will use names and place terms from the Library of Congress Name Authority File. Keyword indexed metadata fields will include author, title, subjects, description, date issued, and identifier. The Senior Archives Associate will perform quality and assurance control on the metadata. All PDF files and metadata will be uploaded to The University of Akron's Institutional Repository by mid-December.

While the digitization is a few weeks behind schedule, the creation of metadata is far ahead of schedule and is mostly completed. Since much of the metadata already existed as the finding aid includes an item-level description of each individual report Archival Services personnel used this information to quickly populate the metadata import spreadsheet. They copied and pasted this information from the finding aid, which included author, title, date, and report number. They then pasted this information into a Microsoft Word file and manipulated the data by putting returns after each report description and tabs between each set of information. They then copied and pasted this information into a Microsoft Excel spreadsheet. The returns placed each report's metadata on a different row in the spreadsheet and the tabs put the data into separate columns. This made metadata creation for the reports much quicker than creating it from scratch. Other information such as collection, institution, institutional contact, and the like were placed into the spreadsheet in the appropriate column and copied down to populate all of the remaining cells.

The only metadata that needs completed, which will be done by student assistants under the direction of the Senior Archives Associate once the reports return, is the item description and the subject headings. Each report has a summary or abstract of the report on the first page. Therefore, once the scanned images are received by Archival Services, the student assistants will open each PDF file, copy this information off the PDF copy, and paste it into the item description on the spreadsheet for each of the 182 reports. They will also use the abstract to help them select key words and subject headings. Once that is completed, the spreadsheet and corresponding PDF files will be uploaded to CONTENTdm by the Senior Archives Associate.

While we are waiting for the project to be completed to publicize the results to our community through various venues, Archival Services has already done some publicity to announce the grant. First, a news story was written by the Project Director and Archives Associate Senior and posted on Archival Services' website (please see http://www.uakron.edu/libraries/archives/about/archival-news_detail.dot?id=7b04c508-40fa-4b3f-a927-ac6bd65c34c7). This article also appears in the news ticker on the University Libraries' homepage and on the Library News & Announcements page on the University Libraries' website (please see <http://www.uakron.edu/libraries/> and <http://www.uakron.edu/libraries/about/news/>, respectively). We also submitted a news article to the Society of Ohio Archivists to be placed in the news section of the next issue of their newsletter, the *Ohio Archivist*. Additional publicity will be done in the coming months and as the project is completed.