

SELECTED PAPERS

Cleared for Takeoff: Digitizing the World's "Most Experienced Airline" and Creation of an Aviation Portal in DPLA

Gabriella Williams, University of Miami; Jacqueline Wachholz, Duke University; Leah Tams, Duke University; Scott Williams, Digital Public Library of America; Adriana J. Millares, HistoryMiami Museum

The creation and launch of a subject-based portal, "Cleared for Takeoff: Explore Commercial Aviation," which features Pan American World Airways materials digitized from collections at the University of Miami, HistoryMiami Museum, and Duke University, alongside commercial aviation resources from the Digital Public Library of America's (DPLA) partner network, provides an important tool for researchers. The portal aims to enable students, teachers, scholars, other researchers, and aviation enthusiasts to easily discover and build connections across aviation collections nationwide. It includes an interactive Pan Am timeline exhibition, as well as a primary source set (PSS) and classroom lesson plan for instructors. The records, advertisements, and artifacts that were digitized and the primary sources that were carefully curated demonstrate the profound and ubiquitous impact Pan Am had on the world, as well as its substantial influence on globalization and modernization. Issues such as feminism and gender roles, diversity and racism, colonialism and environmentalism are among some of the overarching themes that have been fleshed out through this project. In 2018, these four institutions were awarded a Digitizing Hidden Collections grant from CLIR to expose these important collections and contextualize them alongside other important commercial aviation records from other institutions nationwide. The grant partners discuss the challenges of creating such an innovative research tool, the lessons learned, and potential avenues for future sustainability of the project.

Institutions

The University of Miami Libraries (UML) comprise the main library located on the Coral Gables campus, as well as five subject libraries. Founded in 1978, UML's special collections are vast and include an impressive array of scholarly resources, comprising over 10,000 linear feet of materials. Over 15 percent of that linear footage is taken up by Pan Am alone, and that collection is also supplemented by a number of aviation-related adjacent collections such as Eastern Airlines. Recently, UML also became home to the Jay I. Kislak Collection of the Early Americas, Exploration & Navigation, an eminent cumulation of materials that document the history of the Western Hemisphere and North/South poles. These two collections actually complement each other nicely, given Pan Am's history of pioneering air routes throughout the world.

UML has also successfully leveraged outstanding distinctive collections through a robust digital initiatives program. UML digital collections now comprise 36 exhibits and over 130 collections and are continually growing. Of those, over 46 are from Special Collections, and the Pan Am digital collection receives the most traffic. UML also hosts HistoryMiami Museum's first digital collection.

HistoryMiami Museum (HMM) is located in downtown Miami, Florida. It was founded as an association in 1940 and has been operating as a museum since 1962. HMM is the largest regional history museum in South Florida and their mission is to safeguard and share Miami stories of the communities, individuals, places, and events found in the city. The museum houses over 20,000 objects that help tell these stories. HMM also maintains an aviation subcollection that contains over 1,500 items, including artifacts from Pan American World Airways as well as Eastern Airlines and National Airlines. The museum's Pan Am collection was donated to them by the Pan American Historical Foundation and contains over 800 cataloged objects. Currently, HMM has three Pan Am- and/or aviation-themed exhibitions, including one that is on display at Miami International Airport in the J Terminal.

Duke University's John W. Hartman Center for Sales, Advertising & Marketing History is an intellectual center within the David M. Rubenstein Rare Book & Manuscript Library. The Library acquired the J. Walter Thompson Company Archives in 1987, and to this day this collection continues to be the largest advertising agency archives in the world. The Pan Am account files came with the J. Walter Thompson collection. Duke established the Hartman Center in 1992 and has continued to proactively document the history and process of advertising from the industry perspective, collecting records from ad agencies, trade associations, and people who have worked in advertising in a wide variety of roles. The Hartman Center is also focused on documenting outdoor advertising, direct marketing, and the experiences of women and people of color in advertising.

Pan Am History

From the time of its inception in 1927, Pan American World Airways was America's preeminent, and for a long time only, international airline, establishing air routes throughout the world and achieving many aviation records, including being the first airline to cross both the Pacific and Atlantic oceans. In fact, it was the closest to a nationalized airline the US has ever had and played an integral role in WWII and other military operations. Pan Am was well known for its luxury and frills, something that is quite evident in Pan Am's marketing and advertisements, and the culture and impact of Pan Am can still be felt today. Pan Am connected the world through increasingly larger and faster aircraft, and it is easy to see how the airline truly "shrank the world," as the company was fond of saying, and brought society closer together.

At the time of the airline's bankruptcy, the Pan Am archive contained over 85,000 boxes of material. Pan Am was fastidious in archiving its nearly-seven-decade history. After Pan Am dissolved in 1991, the University of Miami Special Collections received all the paper, photographic, and audiovisual records. Subsequently, HistoryMiami Museum and the Smithsonian Institution received many of the cultural artifacts and 3D objects. Duke University had already acquired Pan Am's marketing materials through their acquisition of ad agency archives. All three institutions had experienced some of the same organizational, preservation, and discoverability issues with their respective Pan Am collections; thus there was ample opportunity to collaborate on future projects.

Digitizing Pan Am—Part 1

UML has received two digitization grants for the Pan Am Archives. The first digitization project began in 2017 with a grant from the National Historical Publications & Records Commission (NHPRC). This project built on a previous processing grant that UML was awarded, also from NHPRC, to enhance access and use. Although the processing grant made the physical collection more accessible, the collection still was not available online, so digitization was the obvious next step.

The NHPRC grant was an 18-month project with UML acting alone to digitize the Printed Materials Series, which is the most used series among the 19 subject groupings in the collection. A total of 60 boxes were outsourced for digitization, as well as 10 boxes scanned in house. UML also designed three thematic digital exhibits that were added to their [Pan Am site](#). Additionally, all of the digitized material was harvested into DPLA to broaden access and discovery. This project laid the groundwork for the CLIR grant project.

Collaboration

Upon successful completion of the NHPRC grant, UML staff knew they wanted to supplement their digital collection with materials found in other repositories elsewhere. HistoryMiami Museum became a natural choice since it is local and has the cultural artifacts and 3D objects. Duke University has notable Pan Am advertisement holdings that complement UML's Pan Am records. Since UML had already worked with DPLA to harvest the existing Pan Am digital collection from the previous grant, we used that as a catalyst to jumpstart our grant proposal and continue our work with them. So, the four institutions together pursued a Digitizing Hidden Collections grant from CLIR. Discussion about a potential collaboration began in 2015, and the partners agreed to submit a proposal for the 2017 CLIR Digitizing Hidden Collections grant cycle. The application made it to the second round but was ultimately turned down. In early 2018, UML and Duke had a lengthy discussion with CLIR grant staff about how to improve the application and reapplied for the 2018 grant cycle. The revised proposal scaled back UML's contribution by 30,900 images and shortened the grant period from three years down to two. The requested amount also was reduced from \$448,000 to \$275,000, and the team refocused the grant narrative to highlight globalization of air travel with Pan Am and its globally diverse destinations.

CLIR Grant

For this project, UML digitized 69 boxes of Pan Am records, which were outsourced to Backstage Library Works for scanning. HistoryMiami contributed over 500 artifacts, which UML photographed. UML is also hosting the museum's new digital collection since the museum does not have the capacity to do so on their own. Additionally, Duke contributed over 6,600 advertisements that they digitized through Backstage. Bringing all three collections together, we then created a commercial aviation-themed subject portal in DPLA, along with a digital exhibit, primary source set, and a classroom teaching guide for educators.

Digitizing Pan Am—Part 2

For the CLIR grant project, it is important to note that UML took a different approach to scanning the records and assigning metadata to the digital assets. Unlike the previous NHPRC grant,

UML's materials were processed at the folder level because they have a mixture of documents, correspondence, reports, photographs, and other items that would have required an excessive amount of time if the items were scanned individually. In order to save time, UML took a More Product/Less Process approach and treated each folder as a digital object, using minimal folder-level metadata rather than describing each discrete item. Since all objects have a corresponding text file (created through OCR), it was more efficient to rely on the system's search capabilities to discover individual items. Additionally, UML developed sponsor statements that were included in the metadata to sort the digital holdings by grant, which subsequently helped us with the digital infrastructure of the portal.

Object Digitization

UML coordinated with HistoryMiami Museum to photograph all of the objects the museum contributed to the project. The digital productions technicians at UML handled all of the photography of the 500 artifacts that were captured for HistoryMiami. Dr. Drew Wofford, an aviation historian who worked for the University of Miami Libraries and also volunteers for the museum, chose the items from among over 800 cataloged items. Wofford selected the items based on potential historical significance and uniqueness; these items included a wide array of things such as uniforms, dinnerware and silverware, model airplanes, luggage, jewelry and pins, and more. Wofford also went to great lengths to provide detailed descriptions of each item. The biggest challenge for UML was integrating the Museum's metadata into the library's system, which presented several hurdles in terms of crosswalking and converting their associated metadata and enriching it to comply with library standards. Additionally, many objects were multipiece items, for instance a toiletry kit, which added an additional layer of complexity with regard to item capture and description.

Scan Totals

During the first digitization project, UML had a total of 113,684 pages, of which about 30 percent was scanned in house. The totals for the CLIR grant are fairly comparable, but with much less in-house production. The UML Digital Production Lab also photographed 545 cultural objects and scanned an additional 945 oversized materials and vendor corrections. Together, both digitization grants included 230,666 images in the collection in total, representing almost 7,100 distinct digital assets. This is by far UML's biggest digital collection to date, though it is only about 10 to 12 percent of the entire Pan Am archive. Because the Pan Am archive is UML's most used collection, it was important to maximize the collection's accessibility and supplement it with more digitized materials from other institutions that could fill in some contextual gaps.

Advertising Contribution

Duke University's Hartman Center drew from several primary collections in its repository. The J. Walter Thompson ad agency (JWT) had the Pan Am account from 1942 to 1974. Four different collections within the JWT archives were tapped to supply Pan Am advertising for the CLIR grant: the Domestic (US-based), Frankfurt, International, and Competitive advertising collections all hold related print advertisements. In addition, the Wells Rich Greene ad agency had the Pan Am account from 1982 to 1988, and that collection also resides in the Hartman Center.

Leah Tams was hired as Duke's Pan Am CLIR grant intern, and she selected ads based on preset criteria. Because the ads were being outsourced to Backstage Library Works for scanning, the selected images had to be able to fit on their scanners, so no oversized ads were included. Duke also elected not to send advertisements printed on newsprint due to the fragile structural nature of the paper. Similarly, magazine ads that required attention from the preservation department before scanning were excluded, and Tams was trained by the preservation department to identify preservation issues. Hartman Center staff wanted to capture the truly international nature of Pan Am while also being as inclusive as possible. As a result, ad selection was not limited by language—over 15 languages are represented by the ads in Duke's advertising contributions to the project.

Advertising Metadata

The Hartman Center has already launched at least seven previous web projects focusing on advertising, so Duke already had a good understanding about what metadata is most useful to describe print ads. Duke commonly includes date, company name, product name, tagline, and format. For this project, it was also important to include two additional fields that would support the unique information found in Pan Am print ads. Firstly, any travel destination mentioned in a print ad was noted in the location metadata field. Some ads may promote just one route or location, but others may include quite a few. In one 1946 print ad promoting Pan Am's Clipper Express shipping service 42 different destinations were noted. All of those locations were included in the metadata for that print ad, and now researchers can click on any one of those cities in the metadata to see other ads that include the same location. For instance, there are 122 results for Guatemala as a Pan Am destination.

Similarly, it was important to include aircraft make and models when noted in Pan Am print ads. Aviation scholars and enthusiasts will be particularly interested in being able to search by plane type. A 1957 print ad celebrating the thirtieth anniversary of the Clipper features 17 different plane models that were used over the course of those first 30 years of service. Researchers can now click on "DC-4" to see that there are 23 results that describe that particular aircraft.

Finding Related Content in DPLA

The materials from UML, HMM, and Duke specifically digitized and selected for inclusion in the DPLA commercial aviation portal are supplemented by thousands of related records from institutions across the DPLA network. DPLA was able to leverage an in-house system originally built to identify related records for the [Black Women's Suffrage](#) portal and apply the same approach to Pan Am and commercial aviation materials. The first step was executed by the Pan Am metadata team at UML and Duke. They compiled a list of hundreds of search terms that they suspected would return related records in the DPLA corpus. The queries were rough approximations of what the team thought would be relevant but were intentionally overly inclusive. These queries were then reviewed and refined through a collaborative and iterative process with DPLA staff.

Building Queries

After the Pan Am metadata team worked together to develop an initial list of terms and queries, DPLA constructed a Google Sheet for the team members to review and provide feedback on the efficacy of the tags being applied to the DPLA corpus. One section listed each specific query/tag combination (e.g., "aviation_boeing_307"). These would link to the query results in the DPLA

portal so the review team could quickly navigate from the spreadsheet to the search results. The team paged through results and where appropriate, logged changes for DPLA staff to make. For example, the team identified that the query for ‘boeing 307’ was pulling in nonrelevant results and that the query should be wrapped in double quotes. DPLA then tweaked the queries and reevaluated those changes on a subsequent iteration.

Review & Refinement

The final results of modifying the specific example of ‘boeing 307’ reduced the number of related items from more than 6,000 to 17. After the first iteration, the changes fell into the following four broad categories.

Global Exclusions

These were terms that the team identified as wholly irrelevant to the records we were trying to include. If any item records contained one of these terms, it was automatically excluded from the set of related records.

Modify Existing Queries

While reviewing the list of queries, the team found that existing queries needed to be modified by either adding an exclusion term to the query or changing the specific query terms. An example of adding an exclusion term would be an original search term of ‘boeing 737’ that needed exclusion terms of ‘military’ and ‘department of defense’ so that the records returned were limited to commercial aviation.

New Queries

The team also discovered that the initial list of search terms was incomplete and added new queries to pull in additional related records.

Investigations

These were the most time-intensive problems the team tackled. One example of problems we encountered was a query that appeared to be working correctly but was not returning a specific record which we expected. Another example was identifying and fixing specific edge cases where we needed to pull in individual records that could not be included using term queries.

Item Aggregation

By the end of this process, the team had developed approximately 350 individual queries to identify specific related content as well as to exclude nonpertinent content (e.g., military or noncommercial aviation). Moving forward, when new relevant content is added to DPLA, it will automatically be tagged by one of these queries and added to the DPLA commercial aviation portal. For example, DPLA added a Pan American World Airways poster from the David Rumsey Map Collection to the aviation portal because its title was a match for the “pan_american_world_airways” tag query. DPLA continues to automatically add this kind of related content to the aviation portal on a weekly basis as they update and add records to the corpus. Most of the process was completed by summer 2021, and by October 2021, when the project was initially complete, the portal included roughly 31,000 items. Because of the automation process and DPLA partners continually adding new content, we are pleased that the portal has now grown to include approximately 41,000 items as of October 2022.

Portal Overview

The **portal** has three main parts to explore: the homepage, the associated primary source set, and the interactive timeline. In bringing so much collection material together from so many different institutions, we wanted to make sure that there were multiple points of entry for people who want to explore the materials. On the homepage at the center there is a search box, allowing site visitors to search using keywords and phrases. Additionally, there are also some preselected topics that we determined would be of high interest to visitors. These preselected topics are represented by the tiles on the bottom half of the screen and include:

- Aircraft
- Airports
- Advertising
- Pan American World Airways
- Pilots and Flight Attendants
- Early Commercial Aviation

In addition to those main entry points, there is also an option at the top of the portal homepage to “Browse by Partner.” This option allows visitors to see all of the portal material aggregated from a particular institution that matches the query parameters set up by DPLA. As of October 2022, over 450 institutions have material in the commercial aviation portal.

Another key feature of the portal is the associated primary source set, or PSS. The goal of the PSS, whose landing page is on [a] slide [in our presentation], was to bring together a much smaller group of carefully curated material from the three grant partners, provide some basic context, and then provide discussion questions and classroom activities for students (primarily K–12). Each question highlights items from the three grant partners, putting them in conversation with each other and encouraging students to critically analyze them.

The final aspect of the commercial aviation portal is the interactive Pan Am timeline. The timeline can be accessed from the homepage of the portal by clicking on the “Pan Am Timeline” link at the top. Much like the PSS, the goal for the timeline was to exhibit material from the three grant partners and put items in conversation with each other while also showing them in a relatively digestible format. The timeline begins with the founding of Pan Am in 1927 and ends with its merger with Delta in 1991. Visitors can navigate the timeline by using directional arrows, clicking on a tile at the bottom, and/or dragging the timeline itself to progress. Each slide features at least one item from Miami, Duke, or HistoryMiami. It was a challenge to select which items would be featured in the timeline and PSS, especially with tens of thousands of items to choose from our three institutions. It took many hours of planning, discussion, and evaluation to arrive at our final product, but they do a wonderful job of bringing material together from disparate institutions in a meaningful way.

Project Challenges

With any grant or digital project there are some challenges or hurdles to overcome. COVID-19 brought two particular setbacks for this project. UML staff experienced a furlough during 2020, so it was difficult for them to work together in a timely fashion on the project. Duke was about 75 percent finished on item selection and metadata creation when the March 2020 closure occurred.

The Duke library reopened with a skeleton crew in July 2020, and the Pan Am intern was able to extend her time and finish the selection process, albeit about four months behind schedule. After Duke's selected ads were sent to the vendor for scanning there was another issue. Backstage Library Works suffered a data breach attack in February 2021, just weeks before the Duke scans and originals were slated to be returned. Duke's files were going through the quality control check at the time. Backstage had to stop work on all projects and make sure their work was recoverable. Their recovery and careful checks on the work for Duke took months, so the final files weren't returned until late May 2021.

Finally, the last notable hurdle for this grant project related to the launching of an online exhibit in the DPLA commercial aviation portal. Once all files and metadata were uploaded into the portal, the final task to complete the project was to create an online exhibit and primary source set with a teaching guide, both of which were noted as deliverables in the grant proposal. Unfortunately, in April 2021 DPLA informed the grant partners that their exhibit creation tool was being updated and could not be used during the time left to complete the grant project. DPLA also did not have an estimated delivery date for a new tool. The curation team could and did move forward with creating the primary source set and teaching guide, but needed to come up with a workable substitute for an exhibit. Leah Tams, who returned to Duke to help with the exhibit and teaching guide, suggested an illustrated timeline for Pan Am using Northwestern University's Knight Lab web tool [Timeline JS](#). Using scanned items from all three institutions found in the DPLA site, Tams built the timeline and it became a very successful and logical substitute for an online exhibit. ●

Author Bios

GABRIELLA WILLIAMS is the digital projects librarian at University of Miami Libraries (UML) and has been working on grant-related digitization projects since she began her employment there in 2017. She holds a BA in political science from the University of Iowa and an MLIS from Florida State University. In addition to her ten years' experience at Embry Riddle Aeronautical University, Gabriella offers a strong background in commercial aviation history, having served as the project manager for UML's first digitization grant from the National Historical Publications & Records Commission. She also served as the project manager and primary investigator for their second digitization grant awarded from the Council on Library and Information Resources.

JACQUELINE WACHHOLZ is the director of the Hartman Center for Sales, Advertising & Marketing History, part of the David M. Rubenstein Rare Book & Manuscript Library at Duke University. Before becoming director in 2004, she was the reference archivist for the Hartman Center. She was the illustration editor for the *Advertising Age Encyclopedia of Advertising*, published by Fitzroy Dearborn in 2002. Prior to coming to Duke, she was an archivist at The History Factory and worked with corporate collections from Sears, Prudential, Goldman Sachs, and GATX. She holds an MLIS from The University of Texas at Austin and a BA in Art History from the University of Minnesota.

LEAH TAMS is the accessions coordinator at Duke University's David M. Rubenstein Rare Book & Manuscript Library. She holds a BA in history and mathematics from the University of Mary Washington and an MSLS from the University of North Carolina at Chapel Hill. She has worked on numerous digitization and public history projects and has previously held positions at the Smithsonian Institution Archives, the National Museum of American History, and the Rubenstein Library's John W. Hartman Center for Sales, Advertising & Marketing History.

SCOTT WILLIAMS (non-presenting contributor) is a senior software engineer at the Digital Public Library of America (DPLA). Before coming to DPLA, Scott worked with museum data at the Yale University Art Gallery and the University of Pennsylvania Museum of Archaeology and Anthropology. He holds degrees in history and computer science from Miami University of Ohio.

ADRIANA J. MILLARES (non-presenting contributor) is the object collection curator at HistoryMiami Museum. She holds a BA in Art History from Southern Methodist University and a Museum Studies Certificate from Florida International University. Adriana started as an intern with HistoryMiami during her last year as an undergraduate, before beginning full time in 2005. She is now responsible for the care and management of over 30,000 collection items. As a museum professional, Adriana has dedicated her career to building, preserving, and providing access to the museum's collection.