Journal of Contemporary Archival Studies

Volume 10 *2023* Article 11

2023

Data on Deck: A Case Study of a Historic Undersea Film and Video Digitization Project

Karen Urbec MLIS, CA
Woods Hole Oceanographic Institution, karen.urbec@whoi.edu

Audrey Mickle MLIS
Woods Hole Oceanographic Institution, amickle@whoi.edu

Lisa Raymond
Woods Hole Oceanographic Institution, lraymond@whoi.edu

Data on Deck: A Case Study of a Historic Undersea Film and Video Digitization Project

Cover Page Footnote

The authors would like to acknowledge the assistance of several key people during this project. We would like to thank Erika Fitzpatrick, WHOI's associate director for corporate and foundation relations, for her help with grant writing and administration; Brett Freiburger, WHOI archivist at the start of the project, who selected the films and videos for the project and established the library's relationship with our vendor, Paul Adams; and WHOI Repository Manager Deborah Roth and WHOI Systems Librarian Kirill Batyuk for their assistance and support throughout the project. We would also like to thank the JCAS editors and reviewers for their valuable input that improved this paper. This project was supported by a Recordings at Risk grant from the Council on Library and Information Resources (CLIR). The grant program is made possible by funding from The Andrew W. Mellon Foundation.

DATA ON DECK: A CASE STUDY OF A HISTORIC UNDERSEA FILM AND VIDEO DIGITIZATION PROJECT

Introduction

Digital preservation is not a new idea or effort. Since the early 1990s, librarians and archivists have worked to establish preservation standards and trusted digital repositories.[1] Our understanding of various media formats and knowledge of how best to manage them has grown, but we still struggle with obtaining the necessary resources, both financial and personnel, to migrate and preserve the aging media formats in our collections. According to the Heritage Health Information Survey, almost two-thirds of archives, libraries, and museums are maintaining digital collections or are preserving born-digital collections, but 73 percent of these institutions do not have a preservation plan or even an assessment of their digital collections.[2] Clearly, we all have work to do.

In 2021, the Woods Hole Oceanographic Institution (WHOI) Data Library and Archives (DLA), a part of the MBLWHOI (Marine Biological Laboratory Woods Hole Oceanographic Institution) Library, received a grant from the Council on Library and Information Resources (CLIR) Recordings at Risk program to catalog, digitize, preserve, and make available historic films and videos of *Alvin*, the first-of-its-kind manned submersible vessel that has shown us more of the deep ocean than had ever been possible and transformed our understanding of life on Earth. Librarians selected moving image media that document the design and fabrication of *Alvin*, and the early moving images that were recorded of the ocean and sea floor with the famed submarine.

These original moving image media have not been accessible to researchers and were far past the intended life cycle of their respective media formats. There was significant risk of complete image loss of these groundbreaking moving images if we attempted to view them. With the CLIR grant, we were able to migrate these significant moving images, to preserve them and also provide wider access to them. In this paper, we will share our experience of designing and implementing this project, we will document the historic importance of these moving image media, and we will briefly review the current literature on archival digitization and digital preservation practices. This was a challenging project, and we learned a lot. We hope that our lessons learned will help others to evaluate their collections and create plans and workflows that will optimize the digitization and preservation of other audiovisual collections that are near or past their expected lifecycles.

This paper begins with the history of the MBLWHOI Library and the importance of the *Alvin* submersible. Then, after a brief review of the pertinent literature, we will describe our methodology as we completed this project. We then will describe the particularly challenging aspects we faced and how we overcame them. Finally, we end with our outreach plans and future goals for these digital surrogates and for the audiovisual collection as a whole.

---

[1] Baucom, "A Brief History of Digital Preservation."
[2] Institute of Museum and Library Services, *Protecting America's Collections*.

History of the MBLWHOI Library and Historic Importance of *Alvin*

For over a century, the MBLWHOI Library has been the intellectual heart of the Woods Hole scientific community. The library is internationally recognized as defining current trends and practices in marine information sciences and bioinformatics. The Marine Biological Laboratory and the Woods Hole Oceanographic Institution maintain a vital partnership in the day-to-day operation of the MBLWHOI Library and provide services to the USGS Woods Hole Coastal and Marine Science Center, Sea Education Association, and Woodwell Climate Research Center. The Data Library and Archives (DLA) holds a diverse collection of administrative records, photographic prints, scientists' personal papers, moving images, historical scientific instruments, cruise data, ship logbooks, diaries, blueprints, and oral histories. Collections also include WHOI publications, maps, atlases, books, and technical reports.

The Marine Biological Laboratory was founded in 1888 and a thriving scientific community quickly grew around it. When Henry Bryant Bigelow, the noted oceanographer and first director of the WHOI, was looking for a location for a new, world-class research outpost in 1930, he chose the village of Woods Hole in part because there was already a notable library that could support scholarly research.[3]

One of the library's notable collections is the audiovisual collection, which documents the scientific work on and under the oceans and includes over 1,000 linear feet of 35mm, 8mm, and 16mm film reels and a variety of magnetic media including VHS, Beta, EIAJ, and Umatic cassettes.

The *Alvin* moving image media in this collection are the only images of one of the first manned submersible vessels that can reach the deep ocean floor. It was designed and built in 1963 and 1964, and during its tenure has allowed oceanographers to see the oceans and our planet from a truly unique perspective. This audiovisual collection at the DLA documents the fabrication of the vessel, initial shallow-water tests, and thousands of dives to the ocean floor. Most media are originals and are at significant risk of image loss and degradation due to the natural aging processes of the moving image media. The collection includes some duplicates that are past their lifecycle and are also at risk of complete image loss.

The Human-Occupied Vehicle (HOV) *Alvin* was built when it became clear, in the mid-1950s, that manned undersea vehicles were needed to further ocean science and research. Submersible vehicles, like the bathyscaphe *Trieste*, were being used in the 1950s, but they were large and unwieldy and Woods Hole scientists wanted a vessel that was more maneuverable and better suited to deep-ocean dives. Construction began in the early 1960s and the completed *Alvin*, named for WHOI scientist Allyn Vine, was delivered to Woods Hole on May 26, 1964, and commissioned on June 5th. *Alvin*'s seaworthiness was thoroughly tested in many shallow and tethered dives, and the vessel began deep water research in earnest in 1965.

[3] Bigelow, "Report of the Committee on Oceanography to the National Academy of Sciences," 153, 159.

Alvin has been used for research in several scientific disciplines, including biology, geology, chemistry, and ocean engineering. It was instrumental in the discovery of hydrothermal vents on the ocean floors in 1977, during dives near the Galapagos Islands in the Pacific Ocean. The abundance of exotic animals and plants close to the vents resulted in the conclusion that chemosynthesis, and not photosynthesis, is what supports life in this light-deprived region. *Alvin* was able to locate a hydrogen bomb lost off the coast of Spain in 1966 and was used in 1986 to explore the *Titanic* in the North Atlantic Ocean.

Notable events in *Alvin*'s history include when *Alvin* was attacked by a two-hundred-pound swordfish on July 6, 1967, during dive 202 on the ocean floor along the Blake Plateau. The eight-foot-long sword became wedged between *Alvin*'s sphere and fiberglass skin, just missing the vessel's electrical cables by a centimeter.[4] During dive 224 on September 24, 1967, *Alvin*'s mechanical arm was lost during operation, and then in October 1968, *Alvin* sank completely about one hundred miles south of Cape Cod when two lift-platform cables broke while lowering the vessel into the water.[5] The three scientists inside were able to escape with minimal injury, but *Alvin* rested on the ocean floor for nearly 10 months before being recovered.[6]

Alvin has been in use since its commissioning, except when it has undergone regular maintenance and has received necessary overhauls and improvements. Today, after completing the latest upgrades, the vessel can reach 99 percent of the ocean floor and achieve dive depths of 6,500 meters. Dives are now documented with video footage from multiple cameras. Currently, *Alvin* is equipped with six to nine exterior cameras, as well as several interior cameras, to aid the expedition of the in-sphere observers.[7]

Literature Review

What is digital preservation? Why do we do it?

Digital preservation is a challenge facing most archival institutions. It is expensive and time-consuming, and it requires detailed planning, technical skill, and the ability to adjust plans as new information is received and situations change. And yet it is something that we cannot avoid if we are to be responsible stewards of our collections. Digital preservation is the work that is done to preserve a digital object over time and ensure its accessibility to current and future users.[8] For it to be effective, intentional policies and workflows are needed to properly appraise, preserve, store,

---

[4] "*Alvin* Dive Log 202," Woods Hole Oceanographic Institution (WHOI), http://dsg.whoi.edu/divelog.nsf/c7e75e08f249295e85256812005364e2/627f0fff1713814c8525620b006c7ebd; Kaharl, *Water Baby*, 95–96.
[5] "*Alvin* Dive Log 224," WHOI, http://dsg.whoi.edu/divelog.nsf/c7e75e08f249295e85256812005364e2/7caf22a83e1247a68525620b006c7fa4; Kaharl, *Water Baby*, 114–17.
[6] Cullen, *Down to the Sea for Science*, 121.
[7] "HOV *Alvin*," WHOI, https://ndsf.whoi.edu/alvin/.
[8] Caplan, "Digital Preservation," 33.

describe, and migrate digital objects.[9] If we avoid this responsibility, we invite complete loss of irreplaceable data.[10]

While planning is required for success, librarians and archivists must be prepared to make changes quickly to respond to changing information as the project progresses. Preservation planning will always require "an iterative cycle of assessment, policy development and refinement, implementation, and maintenance" that balances an organization's current needs and expectations for future circumstances. [11]

Digitization of archival materials is also done to provide increased access to resources. Digital surrogates of fragile objects can be viewed repeatedly without damaging the originals, and these same surrogates can be accessed outside of the usual limitations of the reading room. Previously "hidden" collections can be seen widely when they are in a digital format.[12] Collaboration is necessary, in the form of workplace teams, partnerships with other institutions, or online collaboration via webinars and forums.[13]

Active management of digital objects is required because without it, digital authenticity and trustworthiness are not possible.[14] Active management includes the necessary financial resources, both of increased budgets and increased staffing.[15]

Digital preservation workflows

The best workflows begin before an object is accessioned into the archives and continue as long as the object is preserved. Though not always possible, an early intervention makes it more likely that an object will be preserved and also that the correct version of an object is preserved. It also makes it more likely that there will be sufficient context and metadata to properly interpret the object in the future.[16]

Digital workflows provide a roadmap to follow while processing collections, and also help future researchers understand the provenance of the objects. Clearly documented workflows also provide a clear audit history which creates transparency and authenticity.[17]

---

[9] Baucom, "Planning and Implementing," 5; Gracy and Kahn, "Preservation in the Digital Age," 32–33.
[10] Baucom, "Planning and Implementing," 5; De Stefano, "Moving Image Preservation," 122.
[11] Baucom, "Planning and Implementing," 5.
[12] Gracy and Kahn, "Preservation in the Digital Age," 30; Fantozzi et al., "Tape Music Archives," 236; Schottlaender, "Digital Preservation Imperative," 3.
[13] Spence et al., "University of Glasgow's Digital Preservation Journey," 4; Schottlaender, "Digital Preservation Imperative," 3.
[14] Schottlaender, "Digital Preservation Imperative," 2; Caplan, "Digital Preservation," 34, 36.
[15] Baucom, "Planning and Implementing, 5."
[16] Spence et al., "University of Glasgow's Digital Preservation Journey," 3.
[17] Baucom, "Creating Adaptable Digital Preservation Workflows," 5, 7; Baucom, "Planning and Implementing," 5.

Preservation challenges

There are many challenges to digital preservation. Funding is a persistent barrier, due to the considerable expense of migrating objects and the continuing costs of long-term digital storage and maintenance.[18] The time spent arranging and describing digital objects is also considerable and represents a cost in staffing levels and staff time. [19]

The fragility of moving image media is also a significant preservation challenge. Obsolete moving image media is extremely fragile and often cannot be played on the similarly obsolete equipment. [20] Fragile films need expert handling and may have only one playback left before they are lost forever.[21] Not surprisingly, expert assistance is expensive.

The management of digital resources is more complex than analog archival items.[22] It is clear to a user that all parts of a book belong together and relate to each other; it is even clear that items in a folder or box maintain a relationship with each other. But digital objects can easily lose their context when they are viewed online as individual items. Archivists and librarians must be vigilant in creating metadata that helps users understand the whole collection and the interrelatedness of the individual items so contextual understanding is not lost.

As digital media have become larger parts of archival and library collections, standards and practices have been developed to guide best practices for formatting, storage, backups, and the many details that must be addressed for conscientious stewardship. The articles we reviewed address many of these issues and agree on most major points. Case studies that speak to real-world challenges and learning curves were not as easy to find, and the authors hope this article will begin to fill that gap. Best practices and theoretical articles are helpful as learning opportunities and guideposts, but practitioners in the library and archives who are looking for practical and specific advice on workflows and ways to manage large digitization projects are also needed. This type of practical advice and honest accounts of challenges and lessons learned will help librarians and archivists make real-world decisions that are grounded in theory and also practical and achievable in a busy library environment.

In addition, the writers believe strongly that data and information are free and should be distributed as widely as possible. Underserved communities, at all ages and levels of education, often have more difficulty obtaining access to scientific resources than more privileged groups, and by digitizing these moving image records and placing them in an open-access repository, our intention is to ensure that any and all who wish to view, study, and learn from these resources will have the

[18] Smith, "Valuing Preservation," 7, 9; De Stefano, "Moving Image Preservation," 123; Gracy and Kahn, "Preservation in the Digital Age," 36.
[19] Spence et al., "University of Glasgow's Digital Preservation Journey," 5.
[20] Sorensen, "Analog Video," 7.
[21] De Stefano, "Moving Image Preservation," 122.
[22] Schottlaender, "Digital Preservation Imperative," 2.

opportunity to do so.[23] By digitizing them and making them available, it is our hope that barriers have been removed, or at least lowered, through our work in this project. The field of oceanography is notoriously privileged and too often does not include scientists and researchers who identify as women or who are people of color, and by making this historic and groundbreaking data available as widely as possible, we hope to encourage future oceanographers, whatever their gender identity or expression, race, ethnicity, sexual orientation, and/or socioeconomic background, to see possibilities for themselves and their careers in the opportunities this data represents.[24]

Digitization Project Methodology

Moving image media of *Alvin* were chosen for their historical and scientific importance, as well as the relative risk of media deterioration. As mentioned previously, the moving image media chosen for this project include 8-, 16-, and 35mm films and several magnetic media formats, mostly EIAJ and VHS. The diversity of formats presents a preservation challenge, and each requires careful, and unique, handling and treatments.

Motion picture film is a photographic medium with a laminate structure of a support layer (often polyester, cellulose ester, or cellulose nitrate) and an emulsion layer (or layers) on which the photographic image is captured.[25] Film can be viewed with a lightbox or the naked eye.[26] Magnetic media tapes consist of a support layer that is covered with a binder layer of suspended magnetic particles. There is often an antistatic black coating on the tapes, as well as a protective top coating. The support layer, since the 1960s, is polyester (usually polyethylene terephthalate) and the magnetic particles are usually ferric oxide or chromium dioxide.[27] They are machine-dependent records that require equipment that can play back the moving images.[28] This presents an additional archival preservation challenge as both the tapes and the machinery to view them will deteriorate over time.

To avoid any confusion, in this paper we will refer to 8-, 16-, and 35mm films as *films*; EIAJ, Beta, and VHS magnetic media tapes will be referred to as *videos*; when referring to digital versions of films or videos we will use the term *digital surrogates*. When referring to the project generally or to a variety of items collectively, we will use the term *moving image media*.

Moving image media selected for this project document *Alvin*'s fabrication, testing, and early dives in the early 1960s through the 1970s and until 1981. Since older media were more at risk of image degradation and loss, films and videos were selected chronologically. The maximum Recordings at Risk grant award is $50,000 and a budget was submitted with that in mind. Librarians worked with our vendor to determine how many items could be digitized, based on the budget, items, and

[23] Tsurusaki, "Framework for Diversity and Equity"; US Department of Education Office of Civil Rights, "Ensuring Students Have Equal Access"; National Center for Education Statistics, "Student Access to Digital Learning."
[24] Johri et al., "Pathways to Justice."
[25] Ritzenthaler, *Preserving Archives and Manuscripts*, 81.
[26] Ritzenthaler, *Preserving Archives and Manuscripts*, 167.
[27] Ritzenthaler, *Preserving Archives and Manuscripts*, 170.
[28] Ritzenthaler, *Preserving Archives and Manuscripts*, 167.

timeframe. Potential issues were anticipated and budgeted for, and so the final number of items digitized was relatively close to the original projection. A total of 433 items were sent in four batches to our vendor and after accounting for duplicates, especially in the early films, a total of 416 digital surrogates were created.

We worked closely with our vendor, Paul Adams of Mass Productions in Tewksbury, Massachusetts. He is an expert on obsolete media formats and digital migration. Adams is experienced in working with cultural heritage institutions for Recordings at Risk grants, and as a subject-matter expert in moving image media, we relied on him to help us evaluate our collection and develop our project plan.

Adams especially advised us on how many videos would require treatment before they could be played and digitized. When videos are deteriorating, as is the case with many types of magnetic media from the mid-twentieth century, they are at high risk of soft binder syndrome-sticky shed syndrome.[29] Sticky shed is when the binder layer of a video has deteriorated to the point that it is likely to adhere to a player's rollers and cause significant damage to the player and to the original video.[30] All image and sound may be lost when an unstable video is played in a machine without treatment to stabilize the binder layer. To mitigate this, the videos undergo thermal treatment where they are baked at extremely low temperatures which temporarily stabilizes the binder layer, allowing for successful playback.[31]

Once the items were on site at Mass Productions, Adams determined that fewer items required baking than were originally estimated. This made the migration process faster than anticipated and it freed up funds. Once CLIR approved the change to our grant, we were able to digitize 18 additional videos as a previously unplanned batch four.

The first batch of 138 moving image media was picked up at the library in September of 2021 by our vendor. WHOI librarians had prepared an Excel spreadsheet including the title, format, length of film in feet, color listed, sound, notes, and alternate ID. We also sent foil-backed labels printed with identifying numbers, which Adams and his staff affixed to the new, archival film canisters. Films were processed, and all films that were not identified as duplicates were digitized. All films with metal housings were rehoused in archival plastic containers to allow for increased air flow and to minimize the risks associated with deteriorating metal canisters.

Batch one was returned to the DLA in November 2021, and batch two was picked up. Batch two was returned in January 2022 and batches three and four were picked up; the last moving image media were returned to the library on March 21, 2022. Adams returned preservation and access digital surrogate files along with metadata exports of name, size, runtime container, resolution, codec, audio, and frames per second on hard drives provided by the library. Once the files were returned, metadata creation and processing of the digital surrogates was begun in earnest.

[29] Fantozzi et al., "Tape Music Archives," 236.
[30] Ritzenthaler, *Preserving Archives and Manuscripts*, 170.
[31] Fantozzi et al., "Tape Music Archives," 236–37.

Digitized videos were returned to the library on Transcend 4TB USB hard drives, with preservation copies of videos in MOV format and access copies in MP4 format. All digital surrogates were transferred to WHOI's server to create an additional copy that benefitted from the institution's tape backup schedules. Hard drives were placed in the library's climate-controlled film vault, to allow for optimal storage conditions of 50 to 55 degrees Fahrenheit and 30- to 40-percent relative humidity. Original films and videos were assigned a barcode and their shelf locations were recorded as they were returned to the film vault.

To provide access to the digital surrogates, the data librarian and archivist worked together to design a workflow and process that would make them as discoverable and accessible as possible. The first decision to be made was how to refer to each digital surrogate in the project. Titles were varied, as some used numbers or letters and others had narrative titles. Some of the numbers in the assigned titles referred to the recording number and some referred to the dive number, resulting in repeated early numbers. Also, since the digital surrogates would be described in both DSpace and ArchivesSpace, we needed a way to refer to each recording without confusion within and across platforms. We decided to use the barcode numbers that we affixed to the exterior canisters when they were returned to us after digitization to refer to the original media and digital surrogates in any iteration. In ArchivesSpace, both an analog record and digital record had to be created, so the analog records were assigned the barcode number and the digital records were assigned the barcode number with the prefix "DO" to indicate they were a digital object. In DSpace, records were assigned the barcode with the "DO" prefix. In this way, we can clearly identify each item without any confusion. By using the barcodes, we limited the opportunities for confusion as we processed the records.

Once we settled on a naming convention, we began a complete visual review of each digital surrogate. All librarians in the DLA helped with the review, but the data librarian and archivist performed most of it. Due to project and funding deadlines and the high volume, most were reviewed at double speed and two at a time. A total of 416 digital surrogates, amounting to 142.5 hours of moving images, were watched in 4 weeks. Early recordings are only a few minutes long and many of the later ones are over 30 minutes long; most are black and white ocean floor images, so anything anomalous was quickly discovered. Digital surrogates with sound were listened to in order to obtain more detailed metadata, since the scientists in the vessel often mentioned what they were seeing and doing during the dive. A handful included small segments of unrelated material such as television shows, soap operas, and commercials. The original versions were maintained for posterity, but the access copies were trimmed to remove these sections to avoid copyright conflicts when hosted online. During viewing, timestamps of interesting and engaging images were noted so they could be used as thumbnails in the repository.

The digital files generated from this grant were uploaded to WHOI's CoreSealTrust-certified, open-access repository, Woods Hole Open Access Server (WHOAS), by the data and repository librarians. WHOAS runs on the DSpace platform, which is an open source, Java-based application.[32]

[32] "About DSpace," DSpace, https://dspace.lyrasis.org/about/.

Because of the large size of the digital surrogates, all CLIR records needed to be uploaded in bulk on the back end of the DSpace application using the Simple Archive Format, also known as SAF.[33] We started this process by creating specially formatted spreadsheets with item metadata and transferring copies of all the files from one server to another. Metadata included title, description, and date or date range. For known dive material, dive number, cruise number, dates, principal investigator, depth, and purpose are included in the description.

Since our installation of DSpace does not automatically provide thumbnails for moving image formats, we created thumbnails to include in the upload. The data librarian used the command line tool FFmpeg to create thumbnails based on the timestamps previously identified during the comprehensive review.[34] To create the SAF packages for import, we used PySAF, a newer tool for bulk importing items into WHOAS.[35] PySAF is much more efficient than SAF builder, our previous tool for this process.[36] The finished "content" file then needed to be edited, using the "sed" text stream editor to reflect that the thumbnail should be placed in the "thumbnail" file bundle.[37]

Once the DSpace records were created and a digital handle was created, the archivist could create the digital object records in our open-source archive information management system, ArchivesSpace, using an export of the metadata from our repository collection.[38] Most films and videos were not previously cataloged in ArchivesSpace, and so records for both the analog items and the digital objects were created by the archivist. Because of the way ArchivesSpace displays metadata in the online module, most metadata was included in the analog record and not in the digital object record, even though some of the metadata refers to the digital instance. Metadata that was included in the records was title, date or date range (if known), location of analog item in the film vault, and film or video type and length. Digital object metadata included title and digital handle that linked back to the WHOAS record. Analog records also included dive data (when available), approximate dive location, dive depth, and scientific objectives of the dive, ship name and cruise number, and the sponsor of the dive. Analog records were tagged with Library of Congress authorized terms: *Alvin (submarine)*, and *Lulu (ship)*. Records were also given the local tags *Lulu (ship) Cruise [number]*, and *CLIR Recordings at Risk Grant*. Most data in the library is arranged according to the ship and cruise number, and so this information, when available, was quite important to better integrate these digital surrogates into existing archival and data collections. We opted to tag each record with the CLIR grant, so that as we continue to digitize the audiovisual collection, we will know which items were digitized with this funding.

---

33 "Importing and Exporting Items via Simple Archive Format," DSpace, https://wiki.lyrasis.org/display/DSDOC5x/Importing+and+Exporting+Items+via+Simple+Archive+Format.
34 "About FFmpeg," FFmpeg, https://ffmpeg.org/about.html.
35 PySAF, GitHub repository, https://github.com/cstarcher/pysaf.
36 SAF Builder, GitHub repository, https://github.com/DSpace-Labs/SAFBuilder.
37 "sed, a stream editor," manual, https://www.gnu.org/software/sed/manual/sed.html.
38 ArchivesSpace: https://archivesspace.org/.

Lessons Learned

This was a challenging project and there were many lessons learned. This was the first project of this size and complexity for management system the library, and so workflows had to be created entirely from scratch. Existing content applications required updates to meet the demands of this project. And there was a key staffing change in the middle of the project. Overall, the success of this project demonstrates the collegiality and close collaboration between the data librarian, archivist, repository manager, and systems librarian, all with the guidance and support of our director. Everyone played a significant role in seeing it through to success.

Prior to this project, the library had only digitized items on an as-needed basis. There was not a comprehensive digitization plan and there were few existing workflows. As outlined above, new workflows were created from scratch to accommodate this project. There were a few false starts and many brainstorming conversations to develop the best possible ways to approach managing and processing the data within the time constraints of the grant period. What we learned this time will help us to plan and manage future projects.

At the start of this project, the DLA had an archivist who identified the moving image media to be digitized and worked with the vendor to deliver the first batch of media. This archivist was instrumental in beginning the project, but between batches one and two, he moved on to another opportunity. The new institution archivist began just before batch two was sent and spent the first few weeks becoming familiar with this project as part of her orientation to a new workplace.

As soon as metadata creation started, it became clear that we would need to update our version of ArchivesSpace. We had been running version 3.0.2, but the latest version, 3.2.0, allows for more integrated batch uploading. The DLA systems librarian updated the application to the latest version on March 18, but it took longer than expected because of structural changes to the application's dependencies. After a few trials, the batch uploader created files without difficulty, however, due to time limitations, the records were deliberately spare to ensure we would complete at least a minimal processing of every digital surrogate before the grant period ended. We also expected to be able to batch edit records easily to add additional information later. Unfortunately, while there are some plug-ins that allow for batch record editing in ArchivesSpace, the ones we found only applied to very specific fields and did not match our needs for this project. They were also very difficult to alter without risking additional errors and complications. We decided to manually add the additional metadata. This manual edit included adding dive log data, dates, and subject headings to each analog record. In the future, we will ensure all records are as complete as possible when they are initially created, so the batch record creation tool can be used to its full effectiveness, and to minimize the need to add supplementary metadata later.

Another challenge was establishing a way to clearly identify each film and video, despite variable naming conventions over many years and based on the often-limited metadata written on film canisters. As we described, we opted to use the item barcode to clearly identify each item, but the barcodes were affixed when the originals were returned to the library after digitization, not before.

In future projects, barcodes will be added as items come off the shelf to provide us with a more systematic tracking mechanism. The barcode will be incorporated into the file-naming scheme to provide clarity and consistency. Since many of the films were rehoused as part of this project, it made sense to barcode the new housings upon return to the library, but in the future a temporary sticker can be used for items that will be rehoused so that barcodes can be used from the beginning to clearly identify all items.

We also have decided that we will establish and maintain a database for any future digitization projects. This will allow us to more easily track individual statuses, add new attributes throughout the process, and enforce better controlled vocabulary use. Using a database will help us easily expand on and track attributes and derivatives of our analog and digital objects.

With this project, we worked with Excel files in Microsoft OneDrive and SharePoint, in hopes that all would work within the same documents. Despite this, there was still confusion at several points about which document version was the most up-to-date and the most complete. Having experienced the steps of the digitization process in full, we will be better able to put in place data management practices and processes that will increase efficiency in the future.

Outreach and Future Plans

Outreach, as well as access and discoverability, was an important expectation of our grant planning after the completion of the digitization and cataloging process. We wrote a blog post for the History of Oceanography website to announce the completion of the project and shared the news with our WHOI colleagues through our institutional newsletter. [39] We also began to post weekly video clips on our social media channels to share the digital surrogates with the public. [40]

We have started collaborating with Digital Commonwealth, the Massachusetts hub of the Digital Public Library of America (DPLA), to have the repository records indexed and harvested this winter. This will make them available to the national DPLA database search. The DPLA reports that 1.5 million people used the portal in 2020.[41] During the grant period, on February 14th, the data librarian presented a poster about the project at the International Oceanographic Data and Information Exchange conference.[42] The Library plans to continue to digitize the collection, to limit the image loss of these unique and valuable resources, and to provide wider access to the collection as a whole.

There are many possibilities for this type of material, and we have already been in contact with film producers and others who will use this material to further science education and understanding to general audiences as well as more specialized and technical users.

[39] Urbec, "CLIR Recordings at Risk Digitization Project at MBLWHOI Library."
[40] MBLWHOI Library's Facebook page is at https://www.facebook.com/mblwhoilibrary.
[41] Williams, "How People Used DPLA Resources."
[42] Mickle et al., "Preserving Historical HOV *Alvin* Video Data."

There is still a large number of concerning at-risk *Alvin* dive videos from the 1980s and 90s. We have learned that some of the items' formats are particularly likely to be quite deteriorated. Our goal is to upgrade our digital archiving workflow while we continue to promote the moving images that we have already digitized with this grant. As we refine our digital archiving workflow, we also plan to write a preservation plan for our audiovisual collection, to document our current holdings and practices, and to create a comprehensive plan for future digitization efforts.

While many aspects of our workflow were helpful, we were able to clearly identify areas for improvement and illustrate where more investment is needed; since this project we have installed Archivematica and are learning to use it, to automate and streamline future electronic records processing and management. It also impressed upon our team the need to budget for the proper maintenance of newly digitized materials. More staff hours will also be needed to handle additional reference questions for these materials. Digitization is an expansion of the collection and requires additional resources to maintain.

Our hope is that what we have learned during this project has prepared us to embark on a much larger project to fully digitize our audiovisual collection. This will require significant funding and support; we look forward to working with our development partners to continue rescuing this priceless data and history, so we can make it as widely available as possible.

Conclusions

This is one of the larger and more complicated projects our small staff has completed in recent memory. Film and tape migration is a costly and time-consuming endeavor. The success and outreach of this moving image media as both historical footage and as scientific data are important contributions to the information that is currently easily available to researchers. We hope the success of this project will make it easier to secure funds for continued migration. The project was both stressful and rewarding, but we are relieved to have completed it successfully and are optimistic about the ongoing use of these resources and the benefits that they will bring.

Overall, this project has shown that the complexity of these media formats and the digital migration process requires a team-based approach to fully anticipate and manage all aspects of the process. No one person could effectively manage a project like this, since the required expertise is so varied and specialized. Any institution looking to begin a project like this one should keep that in mind and develop their plans with input from a diverse team within the library and archive. We look forward to continuing to migrate our audiovisual collection, in light of the lessons we have learned here, and in order to further provide access to these scientific and historic treasures.

AC-57.Film5.mp4
In this still image from film 5, *Alvin* crew are preparing for a test dive in Woods Hole Harbor on August 4, 1964. This tethered dive would reach a depth of 12 meters.

AC-57.Film78-4.mp4
In this still image from film 78, *Alvin* is seen between R/V *Lulu*'s hulls on May 17, 1965. *Alvin* is preparing for trials near Port Canaveral and the Bahamas and reached a depth of 8 meters.

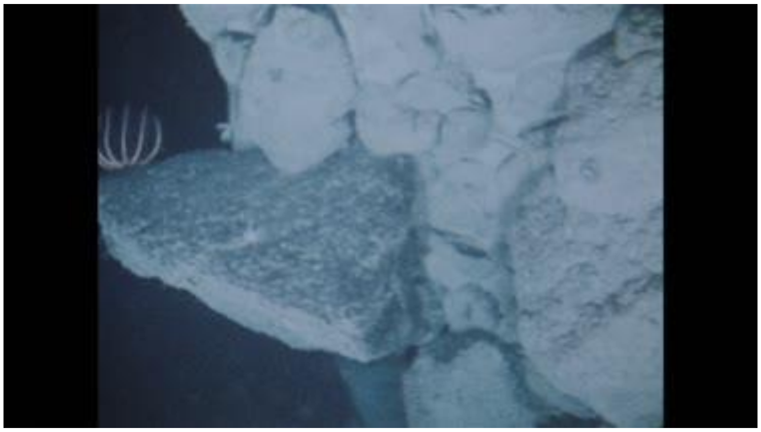
AC-57.Film126B.mp4

In this still image from dive 126, *Alvin* is being launched via crane into waters off Cartagena, Spain, to participate in a bomb search on March 12, 1966. After a U.S. B-52 bomber and KC-135 tanker collided in midair on January 17, 1966, over Palomares, Spain, four hydrogen bombs were lost. Three were found on land but one had fallen into the sea. On February 11, *Alvin* joined the team that searched for the missing bomb; it was recovered on April 7.[43]

AC-57.173.mp4

This still image from dive 173 shows areas of the sea floor in the Tongue of the Ocean, a large undersea canyon that is 40 miles wide and over 80 miles long.[44] This dive reached a depth of 1,372 meters on September 9, 1966.

[43] Kaharl, *Water Baby*, 65–81.
[44] Kaharl, *Water Baby*, 87.



AC-57.862-869.mp4

In this still image from dives 862–869, a rocky portion of the ocean is visible. These dives, in several locations near the Bahamas, were conducted from November 26 to December 4, 1978.

Bibliography

Baucom, Erin. “A Brief History of Digital Preservation.” In *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom. Chicago: ALA Editions, 2019. https://scholarworks.umt.edu/ml_pubs/31.

Baucom, Erin. “Planning and Implementing a Sustainable Digital Preservation Program.” *Library Technology Reports* 55, no. 6 (2019). https://doi.org/10.5860/ltr.55n6.

Baucom, Erin. “Creating Adaptable Digital Preservation Workflows.” *Library Technology Reports* 57, no. 4 (2021). https://doi.org/10.5860/ltr.57n4.

Bigelow, Henry B. “Report of the Committee on Oceanography to the National Academy of Sciences.” The Henry Bryant Bigelow Manuscript Collection, Woods Hole Oceanographic Institution Data Library and Archives, 1930.

Caplan, Priscilla. “Digital Preservation: Standing the Test of Time.” Florida Center for Library Automation, Gainesville, FL, 2007.

Cullen, Vicky. *Down to the Sea for Science: 75 Years of Ocean Research, Education, and Exploration at the Woods Hole Oceanographic Institution*. Woods Hole, MA: Woods Hole Oceanographic Institution, 2005.

De Stefano, Paula. “Moving Image Preservation in Libraries.” *Library Trends* 5, no. 1 (2003): 118–32.

“Ensuring Students Have Equal Access to Educational Resources without Regard to Race, Color, or National Origin.” US Department of Education Office of Civil Rights, fact sheet, October 1, 2014. https://www2.ed.gov/about/offices/list/ocr/docs/dcl-factsheet-resourcecomp-201410.pdf.

Fantozzi, Carlo, Federica Bressan, Niccolò Pretto, and Sergio Canazza. “Tape Music Archives: From Preservation to Access.” *International Journal on Digital Libraries* 18 (2017): 233–49. https://doi.org/10.1007/s00799-017-0208-8.

Gracy, Karen F., and Miriam B. Kahn. "Preservation in the Digital Age: A Review of Preservation Literature, 2009–10." *Library Resources and Technical Services* 56, no. 1 (2011): 25–43. https://journals.ala.org/index.php/lrts/article/view/5495/6745.

The Institute of Museum and Library Services. *Protecting America's Collections: Results from the Heritage Health Information Survey*. Washington, D.C.: The Institute, 2019. https://www.imls.gov/sites/default/files/publications/documents/imls-hhis-report.pdf.

Johri, Shaili, Maria Carnevale, Lindsay Porter, Zivian Anna, Melina Kourantidou, Erin L. Meyer, Jessica Seevers, and Rachel A. Skubel. "Pathways to Justice, Equity, Diversity, and Inclusion in Marine Science and Conservation." *Frontiers in Marine Science* 8 (2021). https://doi.org/10.3389/fmars.2021.696180.

Kaharl, Victoria. *Water Baby: The Story of Alvin*. New York City: Oxford University Press, 1990.

Mickle, Audrey, Lisa Raymond, Karen Urbec, Brett Freiburger. "Preserving Historical HOV *Alvin* Video Data for Science and Society." International Ocean Data Conference poster session, 2022. https://oceandataconference.org/files/92_OceanDataConference2022.pdf.

Ritzenthaler, Mary Lynn. *Preserving Archives and Manuscripts*. Archival Fundamentals Series II. 2nd ed. Chicago: Society of American Archivists, 2010.

Schottlaender, Brian E. C. "The Digital Preservation Imperative: An Ecosystem View." *Library Resources and Technical Services* 58, no. 1 (2014): 2–3. https://doi.org/10.5860/lrts.58n1.2.

Smith, Abby. "Valuing Preservation." *Library Trends* 56, no. 1 (2007): 4–25.

Sorensen, Lauren. "Analog Video in Moving Image Archives and Conservation: Infrastructures of Knowledge from Production to Preservation." *KULA: Knowledge Creation, Dissemination, and Preservation Studies* 2, no. 1 (2018). https://doi.org/10.5334/kula.32.

Spence, Alison, Valerie McCutcheon, Matt Mahon. "Case Study: The University of Glasgow's Digital Preservation Journey 2017–2019," *Insights* 32, no. 10 (2019): 1–9. https://doi.org/10.1629/uksg.461.

"Student Access to Digital Learning Outside the Classroom," National Center for Education Statistics. https://nces.ed.gov/pubs2017/2017098/index.asp.

Tsurusaki, Blakely (lead author). "A Framework for Diversity and Equity in K–12 Science Educational Media," 2020, report prepared for GBH Education's *Bringing the Universe to America's Classroom*.

Urbec, Karen. "CLIR Recordings at Risk Digitization Project at MBLWHOI Library." *The History of Oceanography* (blog). International Commission on the History of Oceanography. September 29, 2022. https://oceansciencehistory.com/2022/09/29/clir-recordings-at-risk-digitization-project-at-mblwhoi-library/.

Williams, Kathleen. "How People Used DPLA Resources in 2020." Digital Public Library of America. December 16, 2020. https://dp.la/news/how-people-used-dpla-resources-in-2020.