Collecting and Archiving Modern Cultural Heritage

Judith Andrews; Smithsonian Institution National Museum of African American History and Culture; Washington, D.C.

Abstract

As digital collecting by museums, libraries, and archives has increased over recent years, the types and complexity of digital objects has also multiplied. The lessons learned and solutions created by the Digitization and Cataloging team at the Smithsonian's National Museum of African American History and Culture (NMAAHC) in acquiring, processing, cataloging and preserving these new types of digital collections can assist others in identifying processes and workflows to preserve and make accessible the ever-expanding amount of digital collections that will grow into tomorrow's digital cultural heritage.

Digital Cultural Heritage

By the end of the 20th century, the percentage of born-digital cultural materials had flourished. Many objects that were once created in a physical medium are now digital first. By continuing to focus on cultural heritage preservation as only preservation of the physical, we create a gap in preservation and potential for a future "dark age" in the cultural heritage sector. Already, the types of digital objects have expanded from simple files to complex, multi-file objects that may be dependent on proprietary devices or software. Attention to the preservation of digital cultural heritage must expand accordingly to expend the same resources on the preservation of digital cultural heritage as on the physical.

As a new museum, the Smithsonian Institution's National Museum of African American History and Culture began collecting and preserving Primary Digital Collection Objects (PDCOs) from the beginning of its collecting life. The term PDCO is used Smithsonian-wide to denote collection objects and archives that exist only in the digital sphere, rather than being a digital surrogate of a physical object in the collection. A PDCO may be born-digital or born-analog, where the file(s) started as a surrogate of a physical object, but the PDCO designation means only a digital version is in the collection. These early digital objects were simple, consisting mostly of digital image files and born-digital audiovisual material. As modern cultural heritage has shifted more to the digital world, the NMAAHC curatorial team began collecting not only higher numbers of digital objects, but also more complex ones. Digital preservation practitioners in the museum field define a simple digital object as "one object equals one file" that is not dependent on proprietary devices or software and complex digital objects as those where multiple files that create one object are dependent on multiple technical characteristics and may consist of proprietary file types.

The initial cataloging and digital processing guidelines at NMAAHC were created to accommodate simple PDCOs, exhibiting the tendency to think of digital objects as analogous to the physical objects being collected. This framework emerged to support the simple digital image files which are an ideal of the "one file equals one object" standard. As curatorial teams began to identify more complex digital objects for acquisition, the cataloging guidelines, preservation and access processing, and repository workflows were ill-equipped to manage these new digital collections. Many institutions are in a similar situation, as evidenced by the growing institutional conference presentations and sharing of born-digital processing workflows as well as the industry-wide conversations and collaborations among digital asset managers, digital archivists, librarians and others. Given the opportunities to collect ever more complex digital objects, we are presented with the responsibility to manage this modern cultural heritage with updated procedures that ensure preservation and intellectual control will last as long as the medium allows.

With its first full-time staff member dedicated to digital asset management and the acquisition of several complex digital objects between 2020 and 2023, NMAAHC seized the opportunity to update and expand acquisition, preservation, and cataloging standards and guidelines to include complex PDCOs. We approached preservation and descriptive cataloging of these objects as a learning process to ensure documentation, metadata, and preservation would serve complex digital objects the same way they serve the needs of simple digital objects.

Roles and Learning

Every individual in the collection acquisition process plays an important role in the intellectual control and preservation of PDCOs. The Cataloging and Digitization team, situated within the Office of Curatorial Affairs, was centrally placed to make the necessary standards updates and provide education on the industry's best practices of acquiring and preserving PDCOs. Multiple individuals played key roles in updating the processes and standards detailed in this paper. The cataloging team of Collection Information Specialist Emily Houf and Media Cataloger Bryan Miller worked closely on PDCO cataloging updates, conferring with Digital Asset Manager Judith Andrews when applicable. The Digital Asset Manager revised existing preservation workflows and guidelines in consultation with a standing committee made up of rights specialists, imaging specialists, and cataloging specialists she then created acquisition-related educational materials and held a series of informative presentations for the Office of Curatorial Affairs staff.

Learning the best practice standards of digital collecting was an important part of improving NMAAHC's PDCO collecting. The Digital Asset Manager created multiple documents that include technical considerations for collecting, acquisition guidance, a step-by-step process for receiving files, a one-page handout for easy reference and a pre-acquisition questionnaire. Each reference document provides information for increasing the Museum's knowledge of files prior to receipt and increased the possibility of the Museum receiving complete files and file packages. Along with providing educational presentations of the new process and guidelines, the Digital Asset Manager worked individually with curatorial staff to answer questions as new types and more complex digital objects were sought for acquisition.

The pre-acquisition questionnaire alone proved immeasurably useful as the start of the COVID-19 pandemic coincided with an almost 100% shift to internet transfer delivery from the delivery of files on optical disk or physical drive. The questionnaire ensured that the Museum was expecting digital files prior to their transfer from the donor. Knowing the files were expected allowed the Digital Asset Manager to manage the transfer of files to the Museum by creating Dropbox upload folders separately for each donor. With restrictions on which internet file transfer services can be used by a government entity, Dropbox is currently the best available to NMAAHC. Providing a single, central point for file transfer meant that the Digital Asset Manager controlled access to the original files and could create access copies for curatorial review, protecting the preservation files from corruption and unintended edits.

Case Studies

The three case studies detailed in this paper include a 3D model and photogrammetry data, formal publishing packages including supporting and process files from the cover artwork creator, and mobile-optimized PDCOs. Each case represents a different challenge to NMAAHC's original acquisition, cataloging/intellectual control, and preservation processes. These challenges brought into focus our handling of proprietary file formats, interfile reliance, preservation, file quality, and present and future access needs.

Case 1 – 3D Model and Photogrammetry Data

Figure 1. 2D orthographic rendering of 3D model of the Last Supper sculpture by Akili Ron Anderson, 2021.73.1. Original 1982 by Akili Ron Anderson, Digitally re-created 2021 by Joseph Aaron Campbell, Collection of the Smithsonian National Museum of African American History and Culture, original artwork © Akili Ron Anderson.

The Last Supper Sculpture is an artwork created by Akili Ron Anderson in 1982 under commission by the New Home Baptist Church in Washington, DC. This sculpture, depicting The Last Supper featuring a Black Messiah, was created as a mural on the wall behind the altar of the church. It is very large and measures 22' wide by 11'6" tall. The New Home Baptist Church eventually moved to a new location in Maryland and sold the building to The Church of Jesus Christ of Latter-Day Saints who covered the sculpture with drywall. The sculpture was discovered by the current owners of the building, the Studio Acting Conservatory (SAC) in 2019. When NMAAHC learned of the sculpture's discovery during building renovations, our curatorial and collections team began working with SAC leadership to acquire the artwork for the Museum's collection. The physical reality of removing the artwork intact from the wall, essentially removing the wall, was deemed unrealistic so the Museum and SAC turned to the possibility of creating a 3D model of the sculpture for acquisition.

Prior to photography and photogrammetry, the sculpture was treated by conservation professionals to repair and clean the artwork. On June 29, 2021, Joseph Aaron Campbell and Dontez Henderson, two Digital Imaging Specialists from NMAAHC, imaged the sculpture to create a 3D model through photogrammetry. This 3D model PDCO (2021.73.1) was then accessioned into the NMAAHC permanent collection as fourteen files consisting of a dense point cloud, master, render, and web model derivatives. Each derivative consisting of multiple files.

The 3D model accessioned into the collection presented a number of digital preservation hurdles. First, the raw photogrammetry data was not collected and though in the field of 3D modeling and photogrammetry, the possibility of re-rendering a model is ever present, museum industry standards expect an accessioned collection object to remain static. On this basis, we decided that the 3D model that was accessioned will be the only version in the collection. The photogrammetry data will be stored and preserved within the NMAAHC standard guidelines 3D models created as object surrogates, but will not be looked to for a re-render replacement of the collected 3D model.

Second, was how to preserve the model. The Smithsonian's Digital Asset Management System (DAMS) excels at bit-level preservation but does not have a robust packaging tool for files. The answer was to use the Library of Congress BagIt protocol to create a file manifest and zip the 3D model files into a single folder for digital preservation ingest. A 3D-specific repository is in production by the Smithsonian Digitization Program Office 3D Team and will be used for future storage of all NMAAHC 3D models- both surrogates and PDCOs- but the DAMS cold storage of a zipped Bag will always be used for PDCO 3D models. The DAMS is provided as a repository to all Smithsonian units by the Office of the Chief Information Officer (OCIO) and is the official digital repository of NMAAHC collection objects. The DAMS has previously been assessed under ISO 16363:2012 for digital repository trustworthiness and is currently undergoing an updated assessment.

Accurate descriptive cataloging and effective intellectual control presented new tests for the cataloging team. A compromise, made during the creation of NMAAHC's general 3D Guidelines, was to maintain the standard terminology of the 3D and technology sphere- using the word "MASTER" to denote the main and highest resolution model. NMAAHC's cataloging approach is very mindful of the implications of language rooted in the history of enslavement, but we accepted the 3D terminology to retain technical clarity.

The next challenge was how to best describe the object. The Museum collected a 3D model, but in order to accurately describe the digital object, the catalog record needed to also capture the details of the original, physical artwork. Not only did proper credit need to be given to the artist, but also the physical material and immense size needed to be conveyed without users mistaking the record as being for the sculpture instead of the 3D model. The cataloging team chose to include descriptions of both the files and the original artwork, listing both dimensions for the artwork as well as the digital file sizes of the model and derivatives. The sculpture's artist, Akili Ron Anderson, is listed as the primary creator while the 3D model maker is identified as the producer. The title of the work remains the same, but the addition of an object name denotes this version is a 3D model.

Finally, it was necessary to create a 2D access image to display in our collection database and as a representative image to be delivered online with the object record. To fill this need, our imaging specialist created a 2D orthographic rendering from the data. This file was handled with our standard object surrogate workflows, naming the file after the object number and ingest to DAMS for automated delivery through the Smithsonian Collection-DAMS Information System (CDIS) and Image Delivery Service (IDS) tools. The CDIS syncs assets in the Smithsonian DAMS with records in a unit's collection or archival database while the IDS is an automated delivery service for TIFF, JPEG, and PDF files. These tools work together to deliver images both internally and externally throughout the Smithsonian and are built and actively managed by the teams in OCIO.

Case 2- Publishing Packages

When the Museum's curator of design acquired formal publishing packages, one from ELLE DECOR magazine and another from the author and designer of a children's book we built on the knowledge learned from cataloging and preserving The Last Supper Sculpture 3D Model PDCO.

In the case of ELLE DECOR, the Museum accessioned the simple PDCO of the digital illustration with digital process sketches featured on the cover of the March 2021 magazine directly from the artist, Rachelle A. Baker (2021.47.1-.6). From ELLE DECOR (2021.85) we received the cover art design which included an Adobe InDesign file, font files, Photoshop and text file with a finalized PDF of the cover created by the magazine. The complex PDCO comprised the InDesign and its referenced font and artwork files. In a standard file configuration, InDesign references the font files and the artwork image file rather than embedding the data in the InDesign file. In the Smithsonian DAMS, this interdependency link is lost when the files are ingested. Like the 3D model we Bagged and zipped the InDesign package and used the PDF derivative delivered by the magazine as the visual access copy for the PDCO. Since this PDF was created by the donor and creator of the original complex PDCO, we could rely on the color space, sizing, and other technical benchmarks to meet the standards of the magazine. This is preferable to the Museum creating a TIFF, JPEG, or PDF derivative copy for access or future use.
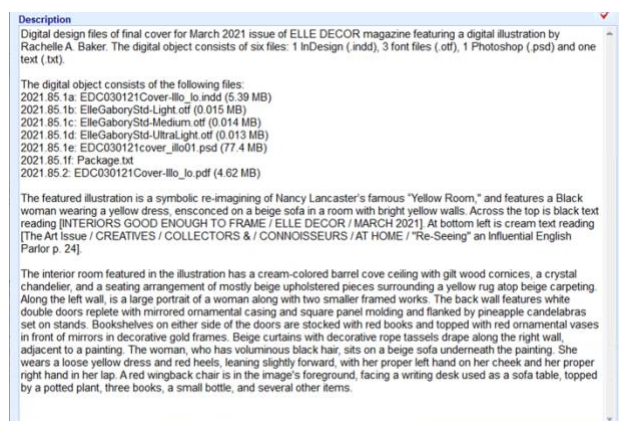

Figure 2. TMS collection database Description field detailing individual files that comprise the ELLE DECOR publishing package and content description, 2021.85.

For cataloging purposes, the PDF was not treated as its own standalone simple PDCO because it was a derivative export of the original InDesign cover design. The InDesign file package was cataloged similar to the 3D model, where the individual files are enumerated by file type and size, but the Description field focuses on the visual image they create which is more akin to traditional physical-object-centered cataloging of an object with content. The files themselves are considered components comprising the single object.

A second type of publishing package came to the Museum in late 2022. This accession (2023.9.1-.3a-g) consisted of a physical book, a set of 2 PDFs representing the interior and exterior of the book, and a Photoshop file with referenced JPEGs comprising the digital art that made up the pages of the book. Similar to the InDesign case, if the Photoshop file was opened without the referenced images in close digital proximity the file would not function properly. Unlike the ELLE DECOR case, the PDF files delivered by the author are not derivatives of the Photoshop file package. The PDFs include the text of the book whereas the Photoshop package only includes the art. As had become our standard by this point, the Photoshop package was Bagged and zipped for preservation.

Unlike the ELLE DECOR cover art design, the two PDF files were cataloged as two components of one object- the book- and the Photoshop package was cataloged as the artwork publication package of the book, creating two separate database records for the digital files in this collection.

Case 3- Mobile-Optimized PDCOs

The final case study is comprised of a set of emojis called "iDiversicons the World's First Diverse Emoji." This was the first set of diverse skin colored emojis to go to market in the app stores of mobile devices in 2013. This collection arrived at the Museum as five sets of themed emojis and one set of animated GIFs, totaling 989 files comprising 484 emojis and 21 GIFs. The donor provided the original Adobe Illustrator art files as well as a PDF derivative for each. As had become our complex PDCO standard, the preservation files were Bagged and zipped for collection-level ingest to the DAMS.

This collection presented a challenge in making the files accessible using the preservation and automated delivery systems available to the Museum. The collection was not item-level cataloged in the database so one additional hurdle was how to maintain intellectual control over the files in the DAMS through the CDIS sync. When very large numbers of files are synced to a single database record, it slows the loading time of the database so syncing every individual PDF derivative file to the single collection record was not an option. In order intellectually manage the files and link a representation of the collection to the database for access, we created contact sheets of the files by folder and CDIS synced those PDF contact sheets to the collection database record. An access copy of each PDF derivative remains in the accession folder for easy reference, but the contact sheets in the collection database provide a central place to visually find each individual file in the collection and identify the location of the preservation file in the repository.

These mobile-optimized files were small in size compared to most PDCOs collected by the Museum. This size differential was an opportunity for learning. As part of the curatorial education program on collecting PDCOs, the reference documents addressed different technical standards between born-digital and born-analog PDCOs and the variety of technical characteristics that may be present in potential PDCOs. NMAAHC holds to a higher technical standard for potential born-analog PDCOs than born-digital objects. This particular

collection of mobile-optimized files presented the opportunity to show that small file size does not equal a poor-quality file nor should the size of a file negatively impact curatorial collecting efforts. As NMAAHC collects more PDCOs that were created specifically for social media or other mobile and web purposes, the file size, type, and resolutions will organically vary and should be expected to be extremely different from a professional uncompressed TIFF digital photograph.
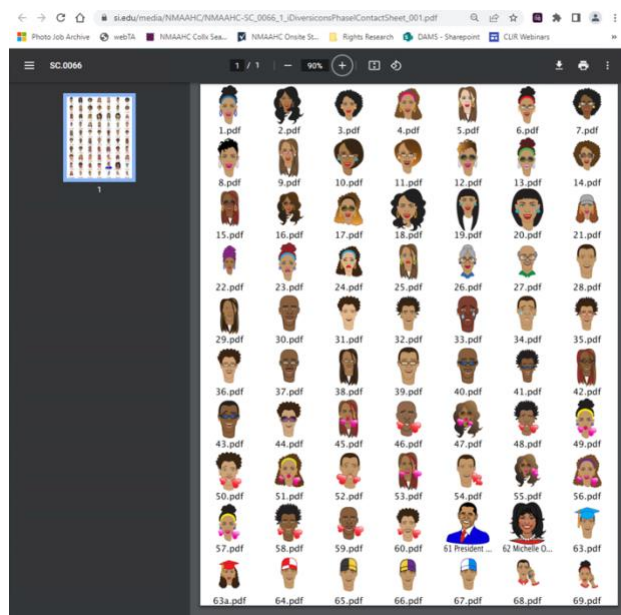

Figure 3. IDS-delivered contact sheet of the iDiversicons Phase I folder of emojis created for improved access and intellectual control (SC.0066).

Process and Standards Updates

These PDCO case studies allowed NMAAHC to improve acquisition, management, and care of our digital cultural heritage. Using these objects as initial models, the cataloging team added language specific to complex digital objects to the NMAAHC cataloging guidelines. Core cataloging data fields for dimensions and overall description are used to capture both digital file information and content description.

Our internal preservation processes required updates to allow for full preservation of the original complex PDCO in the DAMS while allowing easy access to a visual representation. The Smithsonian DAMS is the trusted digital repository for both NMAAHC PDCOs and digital surrogates and also the repository that serves images through automated tools like CDIS and IDS to systems across the Smithsonian through automated distribution. Our preservation process updates included updates to our embedded and DAMS-level metadata in order for the automated crosswalks to deliver the correct files and to mark the original preservation files as such. These updates did not require any changes to the technical infrastructure of the DAMS, instead we worked within the structure of the repository which was built to excel at the preservation of standalone files.

Finally, NMAAHC instituted a program of curatorial and staff education to provide basic technical knowledge about PDCOs. From teaching curators not to plug a flash drive into their own computer to learning how to request the most original and preservation-version files from a donor, the educational program informs curatorial staff on the best practices of digital collecting. This particular program is not limited to complex PDCOs, but has made a positive impact on digital collecting overall. We find that PDCOs acquired since the start of curatorial education arrive with richer metadata and are more easily processed through our preservation steps because we know more about the files and their needs. This program is ongoing and will likely continue evolving as new and varied PDCOs are created and catch the interest of our curatorial staff.

Looking Ahead

Cultural heritage will continue to become more digital and more complex as more material culture emerges digital first and NMAAHC will continue to collect digitally, preserving the modern, contemporary cultural heritage of African American communities. There is a possible future where technical infrastructure updates to repositories and digital processing spaces may need to take place, but by updating our standards and procedures we can reliably preserve and make accessible complex PDCOs today. One future change that is mere months away is the ability to utilize the Packrat 3D repository built by the Smithsonian's Digitization Program Office for preserving and delivering all of our 3D models. In the future, institutions may need to enlist more repositories like Packrat that can natively preserve, manage, and deliver for access complex packages of files. For now, NMAAHC learned that we can be nimble and respond to the changing needs of complex digital objects within our standards to acquire, catalog and preserve modern cultural heritage for future generations.

References

[1] Cataloging Standards and Guidelines, National Museum of African American History and Culture, Smithsonian Institution.

[2] 3D model of The Last Supper Sculpture by Akili Ron Anderson, http://n2t.net/ark:/65665/fd5cdad7379-011b-4ac3-a52a-003e98c3a371

[3] Cover design for ELLE DÉCOR magazine with illustration by Rachelle A. Baker, http://n2t.net/ark:/65665/fd5316ace11-7a13-404c-916f-f2b7f2716021

[4] "Toward anti-racist technical terminology." The Association for Computers and the Humanities, https://ach.org/toward-anti-racist-technical-terminology/

[5] M. Addis, "Minimum Viable Preservation." Digital Preservation Coalition Blog. (2018)

[6] A. Beking, Reformat Required: Addressing "Paper Minds" in Archiving through a New Collaborative Model for Digital Knowledge Transfer, Proc. iPRES, pg. 329. (2019)

[7] M.J. Bennett, "Evaluating the creation and preservation challenges of photogrammetry-based 3D models." UConn Published Works, 52 (2015).

[8] K.D. Bollacker, "Avoiding a Digital Dark Age." American Scientist, 98, 2 (2010).

[9] R. Erway, "Defining 'Born Digital'." OCLC Research. (2010)

[10] P. Falcao, C. Smith, S.D. Thomson, Preserving Complex Digital Objects Revisited, Proc. iPRES, pg.489. (2022)

[11] ISO 16363:2012, Space data and information transfer systems – Audit and certification of trustworthy digital repositories. https://www.iso.org/standard/56510.html

[12] Library of Congress, bagit-python, https://libraryofcongress.github.io/bagit-python/

[13] S. Rippington, Preserving Photogrammetry Outputs: A case study at the University of St Andrews Libraries and Museums, Proc. iPRES, pg. 466. (2022)

[14] C. Smith, P. Falcao, S.D. Thomson, Preserving Complex Digital Objects, Proc. iPRES, pg. 491. (2019)

[15] Smithsonian Institution, Digitization Program Office, Packrat, https://github.com/Smithsonian/dpo-packrat

[16] P. Wheatley, "File Format recommendations – I wouldn't say they are unacceptable, but I wouldn't recommend them either." Digital Preservation Coalition Blog. (2022)

[17] A. Whiteside, S. Abrams, S. Rogers, K. Koo, H. Kataw, Building for Tomorrow White Paper, (2021).

Author Biography

Judith Andrews is the Digital Asset Manager at the Smithsonian Institution National Museum of African American History and Culture. She holds a BA in Anthropology from the College of William and Mary (2006), an MA in Museum Anthropology from Arizona State University (2009), and the Digital Archives Specialist Certificate from the Society of American Archivists (2019). Her work focuses on digital preservation and making museum collections accessible to communities.