



NDSR Project: Preservation of Scientific Research and Collection Datasets at the American Museum of Natural History

Goal Summary

To obtain a broad overview of the extent and status of American Museum of Natural History (AMNH) digital assets pertaining to Science. A survey will be developed to measure and describe scientific digital assets resulting in a metric to predict ongoing data curation needs as a baseline to review options for long-term digital preservation. It will also identify existing practices and policies for integrated data access and management at the AMNH.

Specific Objectives

To develop and implement a survey of existing digital assets in AMNH Science Departments that include the Research Library and other administrative units administered by AMNH Science. Among the data to be gathered will be the current storage location, uses, and administrative management of the assets, the status of associated metadata, the software applications used to access the data, and the file format along with a metric for storage requirements and associated costs specific to that format, as well as projections for growth. The survey will also include notes about the present lifecycle of each digital asset and the elements needed for the development of an institutional appraisal policy to determine the duration of data retention and which data will be selected for long term preservation.

To research the issues surrounding the preservation of scientific databases like those used by AMNH Science, e.g., current work in searchable long-term preservation of scientific data through semantic web representations.

To chart a comparison of the various options for long-term digital preservation including cloud and collaborative possibilities coupled with broad estimates of present and anticipated costs.

All of the above will be combined into a report as a basis to review the status of AMNH scientific data that may be used as a model for similar research-based natural science museums.

Timeframe & Deliverables

Overall – 9 months

Months 1 through 4 — The resident will be introduced to the primary project team and to key personnel within the Museum. After initial planning discussions,

the resident and mentors will meet with the IT department to determine the overall technological landscape. Initial group meetings will be held with members of each Scientific Department. The survey will be in development during this initial process and ready to be used, and modified as needed, during individual interviews with the AMNH curators and relevant scientific staff in charge of collections (approximately 50 individuals). In addition to identifying data formats, the amount of data in each format, and current storage and backup schedules, discussions will be held about how the data is used and how long it is useful, to distinguish working data from final data, to determine how to evaluate and choose data for long term digital preservation. These parameters will vary based on the processes of different scientific disciplines and individual scientists.

Deliverable: Draft of detailed survey of AMNH scientific digital assets, subject to revision as the project proceeds, and beyond. Guidelines for evaluation of policies for: appraisal for determining the length of retention of digital assets and data management plans for federally funded grant proposals.

Months 5 through 6 — Research on current methodologies regarding the preservation of scientific databases. Based on information and context gained from the initial survey and interviews, the resident will research and document current methodologies that address the management of scientific and collection databases.

Deliverable: A report that contextualizes AMNH scientific datasets within the larger digital preservation community. General and some specific recommendations will be made, including identification of issues that might require further evaluation.

Months 7 through 8 — A comparison of long-term preservation solutions. Based on the resident's accumulated knowledge of the institution and its priorities, s/he will review various options for long term digital preservation including cloud and collaborative opportunities, incorporating current research reports and developments in the field of digital preservation.

Deliverable: A comparison of possible long term preservation options for AMNH based on the digital assets survey which will include current cost estimates and a five year projection.

Month 9 — Final Report compiled and submitted.

Deliverable: The final report will be a compilation of all previous deliverables into a comprehensive document in a form, outlined as a resource for ongoing digital preservation planning at the AMNH including guidelines for local best practices for digital asset management and preservation. It should be noted that the structure of the report will be complete but allow for easy identification of areas where more research is needed, either institutionally or within the broader digital preservation community.

*Resources
Required*

2 mentors (primary mentor, Barbara Mathé, MSLS, Museum Archivist and Head of Library Special Collections, and Scott Schaefer, Ph.D., Associate Dean of Science for Collections), 1 resident

Access to AMNH staff in Science Departments including library, archives and museum (LAM) collection staff and AMNH IT staff.

An office and a laptop computer will be provided in the Library for the resident.

Context

Science at the AMNH administratively incorporates 5 scientific divisions, their researchers, students, fellows, associates and collection staff. Collections are also included in the Research Library, the Center for Biodiversity & Conservation, the Sackler Institute for Comparative Genomics, which includes the Ambrose Monell Cryogenic Collection (AMCC). The Science Computer Cluster Facility is a major resource used by museum research scientists, postdoctoral fellows, graduate and undergraduate students whose work relies heavily on high-end capability computing in areas of biology, genomics, astrophysics, and anthropology. The Microscopic Imaging Facility contains state-of-the-art imaging instruments and image analysis software for use across departments. Other sections under Science administration include the Hayden Planetarium, the Southwestern Research Station, the Richard Gilder Graduate School, an innovative Ph.D. program in comparative biology, and Natural Science Collections Conservation which oversees the conservation of the biological specimens in collections and on exhibit. The data in these sections will overlap with—or be minimal as compared to—the Science divisions and computational and imaging facilities.

AMNH scientists are at the forefront of developing and utilizing cutting-edge approaches in computing paradigms to address problems of broad application in the biological and physical sciences. For instance, researchers in Invertebrate Zoology have developed and implemented phylogenetic algorithms used by scientists around the world. While those in Astrophysics, in collaborations with scientists world-wide, use high-resolution numerical simulation techniques in research and also in the Hayden Planetarium. Anthropological research employs an anthro informatic approach, utilizing computational and phylogenetic tools to build on databases of human kinship variants. The NSF recently funded a 10 year effort to create a shared portal for digital data and images of biological collections. iDigBio <https://www.idigbio.org/> promises to integrate access to information about collections held across the country, but the preservation of assets in the collaboration is dependent upon its individual partners. AMNH is a federally-funded partner in this initiative. AMNH Science has few peers in the museum community in terms of the breadth and scale of its digital collections; however, it is similar to many research museums in lacking a unified strategy to manage its research data. Except for efforts at the Smithsonian, no other major museum has delved into this issue. At the same time, funders, including the NSF and NIH, are increasingly requiring researchers to produce data management plans (DMPs) that

describe how the results of grant-funded research will be preserved. Research universities have responded with a variety of solutions but this project will articulate the challenges of research data management in a museum-specific environment.

Required Knowledge and Skills for Resident

This position will involve interactions with scientists across many disciplines. No individual can be expected to have a comprehensive understanding of these many and varied areas of research but an interest in science, intellectual curiosity, and the willingness to respond to the challenge of determining different individuals' methodologies and needs is essential for this project. Specific requirements include: A graduate degree in Library/Information Science, Archival Studies, Information Science or equivalent.

- Deep interest /understanding of the issues pertaining to management of digital resources
- Awareness of/ability to research current relevant efforts in digital preservation
- Strong organizational skills and the ability to shift focus from big picture infrastructure issues to the details of a unique database design
- Ability to communicate clearly and effectively in writing, meetings and interviews and to explain complex technical ideas in a simple and concise way to others with less technical knowledge.
- Ability to work with others as part of a team
- Facility with MS Office, particularly MS Excel

Preferred Knowledge or Experience

- Knowledge or experience in current and emerging best practices, tools, principles and standards for digital preservation and curation
- Familiarity with descriptive, technical and preservation metadata standard application and use
- Conceptual familiarity with XML encoding and Linked Open Data within the semantic web
- Understanding of database technologies and server environments
- Knowledge of survey and information gathering technologies
- Knowledge of digitization guidelines and parameters (color space, file formats, resolution)
- Conceptual/practical knowledge concerning the cost environment of managing and preserving digital assets
- Experience in a digital library, archive or related heritage environment

CARNEGIE HALL

NDSR Project: Carnegie Hall Digital Archives Project: Born-Digital Asset Management and Preservation Policies

Goal Summary

To create and implement policies, procedures and best practices for Carnegie Hall's Digital Archives Project, a multi-year initiative to preserve and digitize archival collections and manage born-digital assets by: 1) Developing policies and procedures for digital preservation; 2) Implementing best practices and workflows for creation and management of born-digital assets; 3) Configuring and testing the new Digital Asset Management System. This project will lay the foundation for the Carnegie Hall Archives to have the tools to manage, preserve and sustain the Digital Archives.

Specific Objectives

To conduct interviews with internal stakeholders in departments that produce digital content to evaluate current workflows and digital usage and create detailed inventories of existing born-digital assets.

To use inventories of born-digital assets to inform requirements and recommendations for long-term preservation and sustainability of digital files. Using information gathered from interviews, create a Digital Preservation and Sustainability document for the policies, procedures, best practices, and workflows for management of digital files including selection and acquisition policies for born-digital assets.

Digital Asset Management System configuration and integration activities including generating workflows for ingest, cataloging, delivery, testing, feedback, and applying taxonomy to born-digital materials.

To gain intellectual control of Carnegie Hall's born-digital materials and better understand how various departments are creating digital assets; to develop and streamline production workflows for creating and ingesting digital assets into the DAMS; and finally, to document and test new policies and guidelines related to digital asset management and digital preservation.

Timeframe & Deliverables

Overall — 9 months

Months 1 through 3 — Orientation, Interviews, Survey, Information Gathering and Inventory Strategy

Acclimation to the Archives Department and Digital Archives Project to date; identify key departments and staff responsible for creating digital content; craft a strategy for the interview and inventory process which will involve informational interviews with staff, gather hard drives from internal departments and external contractors of Carnegie Hall, exploring departmental directories to uncover unknown assets, and dealing with existing born-digital materials in the Archives; survey and analysis of current digital workflows—find out how staff acquire, store, manage and add metadata to born-digital assets. Work with and guide departments in evaluating and selecting valuable assets for long-term preservation.

Months 4 through 6 — Analysis of Workflows, Review Inventory; Select Assets and Gather Metadata for Ingest; Draft policies

Using the detailed inventories and information from the interviews and survey, create a workflow for ingesting born-digital assets into the DAMS; make recommendations on how to improve file naming and other pre-ingest workflows; prepare selected born-digital assets and gather the descriptive metadata by working with staff; draft preservation policies related to ingest, access, and disaster planning.

Months 7 through 9 — First Ingest of Born-Digital Assets; Finalize Policies and Procedures

Review of the descriptive metadata for selected assets; ingest the first batch of born-digital assets; test file ingest and metadata for legacy and born-digital assets; refine policies and workflows related to DAMS acquisitions and ingest; document and provide feedback on ingest, cataloging, and management of digital assets. Prepare final deliverables.

Deliverables:

- Detailed inventory of existing born-digital assets at Carnegie Hall including, but not limited to, digital content, types, file formats and storage locations.
- Tested and documented workflows and best practices for the preservation and management of born-digital assets
- Policy document for Digital Preservation and Sustainability

*Resources
Required*

2 mentors (Miwa Yokoyama, Kathleen Sabogal), 1 resident

Resident will have a workstation and computer in the Carnegie Hall Archives. The resident will also participate in staff meetings and be integrated into the work environment of Carnegie Hall.

Resident will interact with and have access to relevant Carnegie Hall staff including IT, Interactive Services, e-Strategy, Public Relations and WMI (the Weill Music Institute).

Resident will also be able to interact with AudioVisual Preservation Solutions (AVPS), Carnegie Hall's consultants for the Digital Archives Project, and Orange Logic, creators of Cortex, Carnegie Hall's new Digital Asset Management System software.

Context

For close to 125 years, the story of Carnegie Hall has been synonymous with the history of music and performance in America. The record of that history—including 124 years of concert programs; promotional flyers; select audio, video, and film recordings; photographs; autographs; musical manuscripts; correspondence and business records; and architectural drawings, objects, and other materials detailing the origins, activities, and growth of Carnegie Hall—is brought to life through the collections of Carnegie Hall's Archives and further extended through robust digital print, audio, and video content being developed today to support Carnegie Hall's extensive performance and music education activities.

In the past, many of Carnegie Hall's "legacy" materials have been accessible only to people who have paid an on-site visit to the Hall's Archives. Select historic items have been put on display to concertgoers and the public in Carnegie Hall's Rose Museum. Through the institution's Digital Archives Project, Carnegie Hall's highest-priority archival collections are now being digitized, creating an opportunity for them to be housed alongside recent contemporary "born-digital" content in an online repository serving audiences in a new and modern way—a resource for all who wish to see, hear, and explore the history of the world's greatest artists, performing on one of the world's greatest stages.

The Digital Archives Project (DAP) is a multi-year grant-funded initiative to digitize archival collections documenting Carnegie Hall's history, ensuring that they are preserved for future generations and made increasingly accessible to the public, both on-site in its Archives and online. A majority of the Hall's archival collections will be digitized through 2015. A new digital asset management system (DAMS) has been selected to house the recently-digitized legacy collections as well as the more ever-growing born-digital content actively produced by the Hall. The foremost purpose of the DAMS is to provide an institution-wide system to centralize storage and ease search and retrieval of all digital assets. As part of the project, select digital assets will be made available to the general public online via carnegiehall.org.

As we enter into the crucial phase of DAMS implementation, we need to focus on creating strategies and documentation for the management and sustainability of the digital archive. The project activities and deliverables are aligned with the mission of the NDSR program as it is Carnegie Hall's goal to preserve the integrity of our digital archives and provide access to future generations of researchers. The Digital Archives Project will change the way Carnegie Hall staff create and manage digital content by streamlining workflows and documenting new guidelines and procedures for digital preservation, with the ultimate goal of making more content

accessible to researchers. We believe this project will be a model for other cultural arts institutions since it is an institution-wide initiative to bring together the legacy archival materials and the newly created digital content into one DAMS for a broader, more effective and efficient way of working with and presenting digital content on our website.

Required Knowledge and Skills for Resident

The resident must have a Master's Degree in Library or Information Studies, Moving Image Archiving and Preservation, or Archival Studies from an accredited university.

The qualified resident will have the following knowledge and skills:

- Familiarity with trustworthy digital repository standards and guidelines, digital preservation principles and best practices
- Ability to communicate clearly and effectively in interdepartmental meetings, emails, and written reports
- Ability to work well independently as well as with a team
- Interest in digital archiving
- Experience using Mac and Windows computers
- Detail oriented with spreadsheets and data

Preferred Knowledge or Experience

The following skills are preferred but not required:

- Experience working with born-digital materials
- Experience with an enterprise Digital Asset Management System
- Knowledge of a wide range of digital print, audio, video, and image file formats
- Music or performing arts background



NDSR Project: Adapting Preservation Standards to Meet the Information Needs of Time-based Media Conservation

Goal Summary

The goal of this project is to use The Museum of Modern Art's new digital repository as a test-bed for serving the information needs, documentation practices, and preservation standards of media conservation in a manner compliant with international digital preservation standards. The project aims to produce deliverables that document workflows for physical media carriers and metadata, content standards, application profiles, ontologies and controlled vocabularies for actionable use in MoMA's digital repository. With this work, the repository and associated standards would be available for potential re-use by institutions stewarding similar collections.

Specific Objectives

To survey existing metadata standards/initiatives for the description of analog videotape and other tangible media carriers and their relationship to their digital masters.

To engage in the documentation, capture, and ingest of a set of collections materials.

Document gaps between currently implemented digital preservation standards and information needs within media conservation.

To provide recommendations for the application of existing standards, taxonomies, and controlled vocabularies that may bridge identified gaps.

To assist in developing taxonomies for breaking down the complex relationships between tangible media, dedicated equipment, and digital materials.

To research how MoMA's digital repository may facilitate preservation and access of born-digital conservation and artwork records produced by other areas of the Department of Conservation (i.e. sculpture, paintings, photography, paper).

Timeframe & Deliverables

Overall — 9 months

Months 1 through 3 — Orientation and research

- Introduction to the Museum, media conservation's responsibilities, and

the current workflows and systems for the cataloguing and preservation of MoMA's collection.

- Shadow Assistant Media Conservator on the capture and preservation of a subset of the Museum's single-channel analog videotape collection materials
- Test ingest and cataloging of single-channel works in digital repository, identifying gaps between employed standards and museum community's information needs and preservation requirements
- Research metadata standards that may meet gap

Months 4 through 6 — Hands-on testing and implementation pilots

- Adapt identified standards to application profile for MoMA's media conservation documentation needs
- In collaboration with Digital Repository Manager and mentor Ben Fino-Radin, design and test workflow that applies the recommended standards to a range of video and complex software-based works
- Begin documentation of proposal to Museum staff for the implementation of identified standards, application profiles, and workflows

Months 7 through 9 — Revision and formalization

- Consider how MoMA's digital repository may facilitate preservation and access of born-digital conservation records being produced by other areas of the conservation department (i.e. sculpture, paintings, photography, paper)
- Complete work on documenting the set of complex time-based media pieces within the collection, which include audio, film installations, video, and computer-based art
- Solicit feedback from mentors and MoMA staff on the identified workflow and methodologies
- Complete proposal for full implementation of identified strategy into media conservation's current workflow and present findings to key MoMA stakeholders

*Resources
Required*

2 mentors (Kate Lewis and Ben Fino-Radin), 1 Resident

MoMA will provide workspace and support to the resident in its midtown Manhattan offices. The resident will also work with members of MoMA's Media Working Group, which includes staff from the Conservation, Registrar, Curatorial, Exhibitions, Information Technology, and Audio Visual departments.

Context

MoMA has a long history of collecting and exhibiting time-based media art since its founding in 1929, as one of the first memory institutions of any kind to collect celluloid film. Time-based media art differs from other more "traditional" works in the collection in that it is dependent on technologies to be properly cared for and exhibited. The Museum's commitment to its historically significant and

expansive collection of works of this nature necessitated the creation of a media conservation team in 2007 to safeguard this material. In recent years, the team has grown considerably (from one media conservator to a group of three) to meet the needs and challenges presented by these artworks.

MoMA has necessarily begun to focus more attention on the care of its growing born-digital collection as artists have shifted significantly from working in the analog to the digital and the needs of analog works in the collection face various forms of obsolescence. Media conservation—collaborating with internal colleagues and other institutions grappling with similar questions about these collections (specifically through Matters in Media Art; a collaborative project between the New Art Trust and its partner museums: MoMA, the San Francisco Museum of Modern Art, and the Tate)—began developing the functional requirements for a digital repository for time-based media artworks. MoMA is now in the final stages of development of an OAIS-based digital repository that employs the best practices established by the broader digital preservation community in libraries and archives, but also integrates the concepts, practices and needs of time-based media conservation. This unique system will meet a significant need within the museum community, not only for effective digital preservation, but conservation management for complex time-based media artworks. To that end, MoMA specifically built the repository as opensourced software and will make it publicly available for sister institutions.

The Museum anticipates the repository will be deployed this autumn, which would align well with the start of an NDSR residency and present an exciting opportunity for an emerging digital stewardship professional. The resident would work collaboratively with two mentors—Media Conservator Kate Lewis and Digital Repository Manager Ben Fino-Radin—and other MoMA staff to help identify where there may be gaps between the employed standards and the use cases and needs of conservation. This work is essential to providing a more effective tool both for MoMA, as well as other institutions that wish to adopt the system.

Required Knowledge and Skills for Resident

The successful resident will have a graduate degree in Library and Information Science, Audio Visual Preservation, or equivalent. Additionally, the successful candidate will have the following:

Specialized Knowledge:

- Analog and digital media preservation
- General knowledge and understanding of analog and digital video and audio formats
- Familiarity and understanding of ISO 14721:2003 (OAIS)

General Knowledge:

- Art history / contemporary art

*Preferred Knowledge or
Experience*

- PREMIS
- METS
- BagIt
- Python or similar scripting language(s)

NDSR Project: Web Archive Management at the New York Art Resources Consortium (NYARC)

Goal Summary

To develop policies and best practices to support the consortium's web archive collection.

The Resident will master quality assurance (QA) protocols and interview interns, project staff, external web archivists, and Archive-It support specialists to compile an inventory of known QA issues and create an open, centralized online documentation of issues and procedures to support streamlined workflows.

To identify seeds requiring capture by Hanzo Archives, which will be used as an on-demand service for sites that cannot be adequately captured by Archive-It.

The Resident will craft metadata guidelines and input workflows for preservation metadata requirements for stewardship of web archive materials.

The Resident will manage the DuraCloud implementation and produce written documentation to administer the WARC (Web ARChive) archive, noting reporting capabilities and task intervals.

Specific Objectives

To implement sustainable workflows and best practices for NYARC's web archiving program in three specific areas:

- Quality assurance
- Preservation metadata
- Archival storage

Timeframe & Deliverables

Overall — 9 months

Months 1 through 3 — Orientation and Quality Assurance (QA)

Review NYARC web archiving project reports and workflow documentation published from 2010-present. Train in existing Archive-It procedures through direct work with staff and webinars. Begin QA work. Inventory QA issues through direct experience and interviews with web archiving interns, staff, and external partners. Work with Archive-It support specialists to verify categorization of fixable/non-fixable issues. Review The Bentley Historical Web Archives: Guidelines and Procedures (<http://tinyurl.com/mzt8v8e>) as a potential model. Consult with colleagues at Columbia University and the Folger

Shakespeare Library about local QA processes that relate to long-term preservation. Test and establish improved QA workflows that take some of the guesswork out of knowing what to look for, what is fixable, etc. Create QA guidelines to document procedures, categorize Archive-It capture constraints, and identify seed characteristics for Hanzo Archive captures. Attend Archive-It Partner Meeting in Montgomery, Alabama on November 18, 2014.

Deliverables: A detailed inventory of QA issues and guidelines for improved QA workflows hosted on the NYARC wiki (<http://wiki.nyarc.org>). Presentation of QA work at NYARC stakeholder meeting.

Months 4 through 6 — Metadata

Begin research of DuraCloud features. Evaluate preservation metadata requirements (administrative, technical and structural) to support long-term preservation. Survey metadata elements contained in WARC header, the local permissions database, and DuraCloud. Conduct gap analysis and prepare recommendations for best practices. Present survey findings and recommendations to project stakeholders and metadata consultant. Based on feedback, create preservation metadata guidelines and routines.

Deliverables: Presentation of preservation metadata survey and recommendations for the program. Written documentation of preservation metadata guidelines and routines. Submit paper proposal for Art Libraries Society of North (ARLIS/NA) annual conference.

Months 7 through 9 — Archival Storage and DuraCloud

Continue DuraCloud evaluation with focus on management of archived files in cloud storage. Identify tasks necessary to verify digital backups and for ongoing repository management. Implement DuraCloud's Archive-It backup feature and produce procedural document detailing management activities to ensure long-term sustainability of the archive.

Deliverables: Implementation of DuraCloud service. Tested and documented procedures for DuraCloud managed archival storage. Present project outcomes at ARLIS/NA annual conference Fort Worth, Texas (dates to be determined).

Resources Required

2 mentors (Duncan (primary) and Pregill), 1 Resident

Access to key NYARC staff and contacts at partner institutions, specifically Archive-It, DuraCloud, and Columbia University.

Access to Archive-It, DuraCloud, Basecamp project management site, and background documentation on the NYARC web archive.

A workstation with dual monitor set-up to conduct QA work and a phone to participate in conference calls.

Context

Recognizing that publication methods for art research materials, such as auction catalogs and catalogues raisonnés, were shifting from analog to digital, NYARC began to explore web archiving as a collection building strategy. With the support of the Mellon Foundation, the consortium conducted a study in 2012 entitled *Reframing Collections for a Digital Age: A Preparatory Study for Collecting and Preserving Web-based Art Research Material*. This study investigated publication trends, web archiving technologies, infrastructure, and workflows needed to support a NYARC web archive collection. Outcomes from that study indicated that, while the tipping point from print to digital is still on the horizon, our libraries needed to take action or risk developing significant gaps in the art historical record. This exploratory study led to the current two-year Mellon grant (2013-2015) for the NYARC libraries to actively expand its web archive collection and develop workflows for administering that collection. This activity has become mission critical for NYARC.

Web archiving is increasingly being viewed as a mainstream collection development activity by both archives and libraries. The Archive-It service has over 275 partners, of which over 50 are also OCLC Research Library Partnership members. However, many challenges exist to integrate web archiving into library practice in efficient and sustainable ways. This project's tri-part structure is directly targeted at finding implementable solutions for quality assurance, metadata, and archival storage.

In the absence of automated processes, verifying the completeness of a harvested website requires manual and laborious quality assurance tasks. Identifying patterns of capture pitfalls and creating guidelines to streamline processes are local necessities, but this documentation can also assist Archive-It in tool development. The Resident's work in this area will build on lessons learned during the Folger Shakespeare Library's previous NDSR project and ongoing work at Columbia University.

Identifying the adequate level of metadata to support discovery and long-term usability of digital collections is an open question. Metadata creation is often a manual task and not easily repurposed across systems. Decisions surrounding metadata models must strike a balance between available resources and requisite elements to ensure the renderability and authenticity of digital materials. NYARC is contracting with a metadata consultant for a descriptive metadata model for web archive materials. The Resident will advise on preservation metadata requirements to support the collection.

Storing a locally managed copy of web archive files outside of the Internet Archive was a recommended digital preservation strategy that came of the 2012 NYARC study. WARC file management is new territory for many libraries, including the

NYARC libraries. The new DuraCloud service integration with Archive-It will be used as the NYARC storage solution; however management functions for maintaining that storage need to be thoroughly investigated and integrated into current workflows. The Resident's digital preservation expertise will be essential to critically examine and lead the DuraCloud implementation.

In all three areas, this project's deliverables will be shared with the broader web archiving community and the Resident will participate in outreach activities to present outcomes of their investigation. As an active member of the web archiving and art library communities, NYARC is committed to collectively finding solutions to better manage hybrid research library collections. This project expects to demonstrate solutions for managing web archives that can move this endeavor forward.

² For reports on this study see "Readying for Reframing: Reports on Web Archiving," <http://www.nyarc.org/content/readying-reframing-reports-web-archiving>.

Required Knowledge and Skills for Resident

Candidates will have an ALA-accredited master's degree in Library and Information Science or a graduate degree in Computer Science, Information Technology, or associated field. Additional qualifications include:

- Experience within an archives, library, or museum setting.
- Demonstrable organizational, analytical, and problem solving skills.
- Ability to work both independently and collaboratively with other team members in a consortial environment
- Excellent interpersonal skills, with a demonstrated ability to communicate effectively, both orally and in writing.
- Knowledge of basic web technologies, such as HTML, CSS, and CMS.
- Knowledge of metadata standards: Dublin Core, MARC, METS, PREMIS
- Inquisitive nature, sense of humor, and an eye for detail.

Preferred Knowledge or Experience

- Experience with digital preservation, web archiving and/or metadata creation in a research library or comparable environment.
- Knowledge of current trends and emerging technologies in web archiving.
- Knowledge of art history.

NEW YORK UNIVERSITY LIBRARIES

NDSR Project: Access and Discovery of Born-Digital Archives

Goal Summary

To assist in the planning and development of the tools, infrastructure, policies and workflows it needs to make born-digital archival collections discoverable and accessible to scholarly research.

To engage in a series of immersive studies that encompass the entire born-digital collections lifecycle. In the course of the Residency, the Resident will develop a deep understanding of current digital archives acquisition and preservation methodologies, analyze current public service models, identify gaps in services models and research potential solutions for discovery and access of born-digital collections.

To research and assess how virtualized environments and emulated computer systems can be integrated into public service workflows. The resident will work closely with staff from the Fales Library and Special Collections to determine how it's students researchers and scholars needs can be better served in the future, both onsite and remotely.

Specific Objectives

To be trained in forensic disk imaging, forensic analysis techniques and archival processing methods for born-digital collections and to review, assess and update existing documentation.

To gather requirements from Fales Library and Special Collections staff and researchers and model potential workflows and access points for scholarly access to born-digital archives.

To research the technology infrastructure at NYU and peer institutions for tools, workflows and policies that could be applied to providing access to born-digital collections.

To research and assess how virtualization and emulation can be integrated within library infrastructure to provide access points to born-digital collections.

To participate in the prototyping a discovery interface for born-digital archives with Digital Library staff.

Timeframe & Deliverables

Overall — 9 months

Months 1 through 2 — Immersion in Digital Archives Technologies

The goal of the first month is to immerse the fellow in the current tools for managing born-digital collections implemented in the Digital Library. The fellow will learn how born-digital collection materials are acquired, accessioned, preserved, arranged and described.

The Resident will forensically image both fixed and removable media from born-digital collection; ingest digital collection content and metadata into institutional repository; assist in the arrangement and description (archival processing) of born-digital collections with archival staff; install, configure and document the use of emulators for accessing born-digital collection materials from obsolete and otherwise inaccessible formats; review, assess and update existing documentation for all acquisition and preservation workflows; assist in the acquisition of born-digital archival collection materials (particularly those not bound to media, i.e. e-mail, social feeds, websites, etc.) with curatorial staff from NYU Libraries.

Deliverables: Updated current documentation for born-digital acquisition and functionality of workflows related to the acquisition, accession and preservation of born-digital archives; Report outlining how computer emulation could be used to access born-digital archives from NYU Libraries' collections.

Months 3 through 6 — Analysis of Scholarly Research and Access Systems

In the second phase the Resident will work extensively with the curatorial staff and researchers. The Resident will also have regular meetings with technologists from both within and outside of NYU, to research existing and emerging technologies that could be used for providing access to born-digital collections.

The Resident will Work with curatorial staff and researchers from the Fales Library and Special Collections to gather requirements for in-person and remote access to the Exit Art Archives and Jeremy Blake Papers; Meet with and interview other archival repositories engaged in providing research access to born-digital archival collections; Work with NYU Libraries' technical services, IT, Data Services, Digital Scholarship Services, Digital Library, and the NYU Virtual Computer Lab, to research technologies being implemented or developed at NYU that could be applied to the access of born-digital archival records.

Deliverables: Draft recommendations for policies and workflows necessary for providing access services for born-digital archival collections; Presentation of findings to NYU Libraries' Special Collections Coordination Committee.

Months 7 through 9 — Prototyping and Testing

In the third phase the fellow will work with IT units in NYU Libraries to design and prototype a discovery system for born-digital archives. Using agile methodologies, the Resident will iteratively review and assess designs with staff and researchers from Fales.

The Resident will Design and prototype discovery models using data from forensic analysis and archival descriptions from the Exit Art and Blake collections; Review and assess designs and mockups iteratively with curatorial staff and researchers.

Deliverables: Finalized system designs, wireframes and mockups of discovery interface for born-digital archival collections; Report that summarizes Residents research and work done during the course of fellowship.

*Resources
Required*

Digital Library Mentor (Donald Mennerich), Fales Library Mentor (Lisa Darms)

Dedicated workspace in DLTS digital archives lab, access to imaging workstation, forensic software and legacy software and hardware systems

Access to staff from NYU Libraries' special collections departments, Digital Library, technical services, Data Services, Digital Scholarship Services and NYU Virtual Computer Lab

Access to peer institutions: Rockefeller Archives Center, Columbia University, Beinecke Rare Book and Manuscript Library, Yale University Library, MoMA, MITH, UNC SILS

Context

The management of born-digital records within an archival context has proven to be a significant challenge to archival repositories. In essence, all aspects of archival practice: acquisition, appraisal, arrangement and description, preservation, discovery and, in particular, access present new and unique problems to the stewardship of born-digital objects. Born-digital records now arrive at repositories in diverse (and often obsolete) physical formats. They are transferred through email and shared through cloud storage services with an ever-increasing frequency and in startlingly growing volumes. The ease with which we now create and share records with computing technologies can be deceiving when considering born-digital records in archival repositories. The ubiquitous problems of sheer scale, hardware and software obsolescence, and little-known file formats all illuminate the challenges posed to existing workflows, policies, tools and infrastructure. While libraries and archives have made great strides in several critical aspects of the management of born-digital archival collections, the tools necessary to provide scholarly access to them have, by and large, remained elusive.

NYU Libraries has committed itself to continuing its tradition of collecting, preserving and providing access to collections of significance in the sciences, arts and humanities. Reflecting the standards of the University, NYU Libraries is a recognized leader in defining the library of the 21st century, including: design responsive to the way today's students work; rich collections in every imaginable format from papyri to born-digital; moving image archiving and preservation; digital preservation; management and accessibility of complex multimedia archives; and global service delivery.

Hosting the resident will be a collaborative effort between the Fales Library and Special Collections and DLTS. The Fales Library and Special Collections, the primary rare book and manuscript repository for the arts and humanities at New York University, houses the Fales Collection of rare books and manuscripts in English and American literature, the Downtown Collection, the Food and Cookery Collection, the Riot Grrrl Collection, and the general Special Collections of the NYU Libraries, encompassing over 355,000 printed volumes, over 10,000 linear feet of archive and manuscript materials, and more than 89,000 media items. In recent years, the Fales Downtown Collection has received international attention, and is widely recognized as the most important special collection for scholars, curators and critics studying New York-based art practices from the 1960s to the present. In recent years, Fales has acquired two important archives as part of its Downtown Collection that comprise significant born-digital components: the Exit Art Archives, documenting a seminal Downtown alternative art space, and the Jeremy Blake papers, comprising the archive of a ground-breaking artist who worked primarily in digital mediums.

NYU's innovative Digital Library program has become a nationally recognized model in adapting new technology for teaching and research, enabling new modes of scholarly inquiry, and preserving collections for future scholars. A sampling of NYU's many digital library projects includes the Afghanistan Digital Library, which is reconstructing the first sixty years of Afghanistan's published cultural heritage, from 1871-1930; the Hemispheric Institute Digital Video Library, which brings together materials that have been available only in small, little-known archives that are inaccessible to scholars and that lack the resources to properly care for and provide access to these cultural video documents; the development of ArchivesSpace, a state of the art archival collections management system; archiving potentially at-risk political web sites; the Arabic Content Online project is making available 10,000 titles from across the Arab world; creating and publishing in the MediaCommons platform, an umbrella of several scholarly publications, primarily in Media Studies; working with researchers in NYU's Center for Urban Science and Progress to build large scale repositories for the manipulation and preservation of numerical data from New York City municipal agencies; working with the "Databrary" project in the Psychology Department to build services for storing, preserving, and sharing large video datasets; researching the possibilities for long-term preservation of scholarship created in contemporary web-based environments such as WordPress, Drupal, and Scalar, in collaboration with the Alliance for Networking Visual Culture; Towards Interoperable Preservation Repositories, an early partnership among NYU, Cornell, and the Florida Center for Library Automation, funded by a National Leadership Grant from the Institute of Museum and Library Services to explore the implications of distributed storage and stewardship among repositories within different institutions, using disparate technologies; the creation of a workflow and digital archive for the Public Broadcasting Service, a collaboration with WNET, WGBH, and the Library of Congress.

In the course of the Residency, the Resident will be exposed to all phases of digital

collections lifecycle management as they relate to born-digital archival collections. The Resident will collaborate with both technical and curatorial staff to research and evaluate how NYU Libraries can better provide access to its born-digital archival collections, both onsite and remote. The Resident will regularly attend meetings and participate in the Library's working group for access to born-digital records. The resident will work closely with staff and researchers from the Fales Library and Special Collections to examine its existing workflows for scholarly access to its collections and explore potential models for the Exit Art Foundation Archives and the Jeremy Blake Papers; collections that together comprise of over a half-million born-digital records. The Resident will meet with technical staff from across the University's IT departments to research and assess the current infrastructure and technologies being supported and developed. In particular, the Resident will focus on how virtualized environments and computer emulation can be integrated into the Libraries' services to support scholarly research. Through this research, the Resident will author a set of recommendations outlining the needed resources to better provide access to born-digital archival collections. In the final stage of the Residency the Resident will work with Digital Library staff to prototype a discovery interface that will combine data gathered both the forensic analysis and archival appraisal of the collection.

The Resident will work in the Elmer Holmes Bobst Library, NYU's flagship library that provides students and faculty members with access to the world's scholarship and serves as a center for the University community's intellectual life. The Resident will have a dedicated workspace in DLTS' recently designed digital archives lab, designed specifically for the acquisition, analysis and processing of born-digital archival collections. The lab space includes an imaging workstation equipped specialized hardware (write blockers, controller cards) and software packages (Forensic Toolkit), dedicated computers for emulation and analysis, and a collection of now-obsolete computers, peripherals and software that should enable access to all collections containing born-digital materials. The NDSR Resident will have access to NYU Libraries' many curators, archivists, preservation and technical staff engaged in the stewardship of digital content across the enterprise. Working in NYU's innovative environment will provide the experience and knowledge necessary for leadership in stewardship of digital collections in libraries and archives.

Required Knowledge and Skills for Resident

The successful resident will have a graduate degree in Library and Information Science with studies in archives management and have experience working on technical projects.

- Demonstrated experience with the arrangement and description of archival collections.
- Demonstrated experience with public service, reference and providing access to special collections.
- Familiarity through work or coursework with XML and digital library and archival metadata standards such as EAD, METS, MODS, Dublin Core and PREMIS.

*Preferred Knowledge or
Experience*

- Experience with web programming, HTML, CSS and Javascript.
- Experience with software development tools and agile methodologies.
- Familiarity with file type identification tools and migration strategies.
- Familiarity with computer emulation and virtual machines.
- Knowledge of computer forensic hardware, software and analysis techniques.
- Knowledge of digital preservation principles, tools and technologies.