

**A Final Report to**  
**The Andrew W. Mellon Foundation**  
**Scholarly Communications Program**

**on**

**WGBH Media Library and Archives**

**Digital Library Initiative Phase Three:  
Participatory Cataloging**

*revised for public release*

March 31, 2015

---

Contact:  
Karen Cariani  
Director of Media Library and Archives  
karen\_cariani@wgbh.org  
WGBH  
One Guest Street  
Boston, Massachusetts 02135



## **Introduction**

WGBH's Media Library and Archives would like to thank the Andrew W. Mellon Foundation for its support of Phase Three of WGBH's Digital Library Initiative, the Participatory Cataloging Project. The Media Library and Archives (MLA) provides access to WGBH's public television and radio content dating back to 1947 onsite and, for selected materials, through its online repository, Open Vault. With much of our twentieth-century history, culture, and news recorded on audio, image, and video media, it is increasingly important for scholars to have easy access to multimedia archival collections. With the Mellon Foundation's support, we have shared with researchers and the public an assembled listing of programs contained in our archives, which has already attracted nearly 800 requests for more information. We have worked with scholars to test frameworks for accessing content and assembling metadata, and in conjunction with the Hydra community we have put into place an open-source architecture that lays the groundwork for future efforts to expand access to our holdings. Through these efforts, we have brought to light WGBH's rich collection, over 60 years of broadcast media.

## **Activities During the Reporting Period**

### ***Summary of the Project and Purpose of the Grant***

Phase Three of WGBH's Digital Library Initiative, the Participatory Cataloging Project, had the following objectives: (1) to publish and expose the catalog of the WGBH MLA to test its findability and use; (2) to invite scholarly contributions of metadata to the catalog for enrichment and increased discoverability; (3) to develop a workflow for providing offsite access to our materials; and (4) to work with scholars, digital humanities organizations, and digital publishing platforms to make our archival materials accessible in an open and sustainable manner.

### ***Progress Toward Expected Outcomes and Significant Accomplishments***

#### **Posting the Catalog**

On September 26, 2014, WGBH posted an inventory of its major program titles on Open Vault (<http://openvault.wgbh.org/>), opening to public view for the first time a collection of nearly 40,000 records. This list encompasses the television and radio archives, with a focus on broadcast programming. In conjunction, we added 140 digitized audio and video items selected by scholars. This newly available content represents over 63 hours of full programs and 19 hours of interviews. In total, Open Vault now has 2,244 videos, 94 audio clips, and 953 images available for public access. (See Appendix A for images of Open Vault.)

Posting the program catalog has generated excellent visibility for the archives. Researchers and scholars have long wanted to know what WGBH has in its collection and have requested more WGBH productions from our archives be available online. Previously, researchers contacted the MLA directly with inquiries about a program or series, or even to discover what the archives might have related to their work. Now a researcher can find a title that exists at WGBH via a Google search and see on Open Vault the same descriptive information as the archive staff.

Once visitors arrive at Open Vault, they are able to search the catalog by keyword or browse the series list. Already-digitized media are available for immediate access. With easy registration, scholars are able to flag records of materials they would like digitized or transcribed. (The

digitization workflow is discussed in further detail below.) In the six months since September 2014, 696 users registered for Open Vault. Ninety-seven percent signed up with an .edu email address, suggesting we are hitting our targeted academic audience. In this same timeframe, nearly 800 assets have been requested, underscoring the value of these resources.

Providing the WGBH catalog to the public proved to be a great technical challenge, but it has helped us move our infrastructure forward. The Open Vault website is now built on a Hydra technology stack that manages multiple components, including Fedora, an open-source repository platform; Solr, a search platform that indexes and accesses the data; and Blacklight, open-source software that allows the user to search and discover in Solr. To organize our metadata, we use PBCore XML, the standard for media materials that WGBH helped create and that has been adopted throughout the audio-visual archive community.

### **Working with Scholars**

A primary goal of this project has been to attract the scholarly community to the WGBH collection. At the same time, we tested opportunities to collect metadata from scholars' work and thereby enhance the catalog and its discoverability. With this aim in mind, we explored several approaches, including scholarly collections, survey solicitation, asset experts, and special collections.

To attract scholars to our archives and Participatory Cataloging project, MLA staff distributed information on the Open Vault blog, through message boards and email lists, and at academic conferences, archival professional association meetings, and digital humanities centers. With this publicity we sought to enlist user participation at all levels. (See Appendix B for examples.)

**Scholar Collections:** Our most successful strategy was the Scholar Collections, in which we invited researchers to create digital collections exploring archival assets on their topics of study. Scholars were paid for their work to provide an incentive for participation, with different amounts budgeted for different scholar categories. Most participants were associate professors and PhD candidates. Despite a higher rate of compensation, established scholars were difficult to engage. Fifteen scholars or associations created 18 Scholar Collections. The scholars' interests ranged from law to cinema studies to history and covered varied topics, including poetry, civil rights, children's television, cooking, and jazz. A full list of the scholars and descriptions of their collections is provided in Appendix C.

The Scholar Collections include an introduction, an academic essay, selected media, summaries and metadata for the media, links to related content, and a biography. The essay is the cornerstone of each Scholar Collection and refers to moving image or sound materials from the WGBH archives. The scholars were given the freedom to choose the essay's subject, audience, length, and style according to their interests and materials that WGBH had available. They also were able to select a limited amount of content for digitization and public access. As the catalog was not yet online, scholars worked with staff to select and access their media materials.

WGBH supplied the scholars with the metadata fields to collect as they examined materials for their collections. These metadata were added to the records on Open Vault. We made the decision to keep authoritative data under our control for the website—scholars did not have

direct access to adding metadata. To provide additional oversight, advisors from the scholars' specific fields reviewed and edited scholars' entire collections, including metadata. Fourteen advisors participated in this step of the project and affirmed the quality of the work.

The Scholar Collections created a mutually beneficial relationship between the participating scholars and the MLA. Researchers were able to create a digital resource collection using WGBH materials and to advance their scholarship, and WGBH was able to interact with the scholars and increase visibility of its archival assets.

To provide one example, Mark Garrett Cooper, Professor of Film and Media Studies at the University of South Carolina, created the Scholar Collection "[New Critical Television](#)." The collection examines I.A. Richards's *Sense of Poetry* (1957) and *Wrath of Achilles* (1958), WGBH's televised poetry lectures from the Harvard professor and prominent twentieth-century literary critic. Cooper has utilized his WGBH research on his blog with University of California, Davis, English professor John Marx, *Humanities After Hollywood*, in the post "[I.A. Richards's Failed MOOC](#)." He also reported that he uses material from this blog in his lectures as well as a planned book, currently titled *Mass Media U*. As a part of this project, Cooper has presented at two conferences along with the MLA staff. In March 2014 at the Society of Cinema Studies Conference in Seattle, along with Karen Cariani and others, he presented "The Televisual Archive: New Directions for Research and Access." In October 2014 at the Orphan Film Symposium in Amsterdam, Cooper presented "Metadata and the Future of the Humanities" and participated in a panel discussion, "New Research Networks for Obsolete Media."

Other approaches the MLA tested included:

**Survey Solicitation:** We developed and distributed surveys to gather metadata from visitors who request archival materials. (See Appendix E for survey examples.) Overall, we did not find this to be an effective way to solicit metadata, as most were not returned. It required extensive staff time to follow up, encourage completion, and incorporate any information into the catalog. It was deemed faster to have staff catalog the items themselves.

**Asset Expert:** We also offered scholars the opportunity to contribute extensive metadata for media they requested. The scholar would be credited on that resource page. Some scholars did generously share their notes and submit metadata. We heard from Carol Bunch Davis that her manuscript that includes discussion of a *Say Brother* episode has been accepted for publication. The metadata she contributed, an illustration of the information we requested, is included as Appendix D. Another notable asset expert is Charles Euchner, author of *Nobody Turn Me Around: A People's History of the 1963 March on Washington* (Beacon Press, 2010). In conjunction with the [March on Washington collection](#), which assembles on Open Vault radio coverage from that pivotal day, Euchner wrote a "[Listener's Guide](#)" for the first seven hours of broadcast and contributed it to the website. We hope to form additional relationships in this vein since such content expands the archive and further exposes the collection to Google searches.

**Special Collections:** As an additional model for scholarly contribution, WGBH solicited scholars who advised for past WGBH productions, and university partners to develop special collections around the programs they created. We built a collection around the work of longtime

WGBH producer Elizabeth Deane, capturing the monthly posts she has been writing for our electronic program guide, *Explore!*. In addition, the Boston Psychoanalytic Society and Institute contributed two collections: “[The Edwin G. Boring and Hanns Sachs Collection](#)“ and “[The Grete L. Bibring Collection](#).” Harvard University reached out to us to put online the *Facts of Medicine* series from the 1950s. And, in collaboration with the Museum of Fine Arts, Boston, we were able to provide access to the *Invitation to Art* series, which aired from the late 1950s through 1960.

### **Creating a Digitization Workflow**

A key aim of this project has been to provide digital access to our archives, saving scholars and other interested parties the difficulties of traveling to the WGBH office in Boston. With the new digitization workflow available through Open Vault, researchers will now be able to begin the process of requesting any needed archival material through a simple process online. There are costs associated with this work, however, which the requester would need to pay.

To begin the process, within a record on Open Vault, users are able to click a button to “Request Digitization” or “Request Transcript.” In response, the MLA staff informs the scholar of the process and costs associated. There is an initial fee to conduct rights research and consult with WGBH’s legal department about whether we can provide public access.

In parallel, we are working with our legal department alongside lawyers from the Library of Congress and Harvard’s Berkman Center for Internet and Society to consider how we can provide more categories of content for broad access, work that will benefit both Open Vault and the American Archive of Public Broadcasting. We also continue to explore the possibility of allowing access to right-restricted content behind a password for a limited time.

As noted above, nearly 800 assets have been requested. Three requesters have agreed to fund the digitization of their materials. The first program, already available online, is from WGBH’s 1970s debate series *The Advocates* and features audio from the episode [“Should The Law Permit Voluntary Euthanasia For The Terminally Ill?”](#)

The process of digitizing or transcribing this content also creates an opportunity for WGBH staff to add metadata on the resource to the Open Vault repository, making these materials more easily findable.

### **Working with Institutions**

WGBH has partnered with fellow institutions to make technical advancements that benefit both Open Vault and the field of media archives. We are a part of the Hydra community that discusses media management within the system, and our work in this space has allowed us to make connections with other organizations and projects, to our mutual benefit.

We have a close relationship with Stanford University and Indiana University, who have advised on our Hydra-related projects, and we are exploring the opportunity of providing content through Stanford’s “Spotlight,” which allows its scholars to create their own collections from a repository of digital materials. This feature may be incorporated into Open Vault in the future.

Our partnerships have proved helpful as the media archive community continues to move forward. By sharing our code, participating in the Hydra open-source community, and presenting Open Vault to the archival and library communities, WGBH has attracted extensive interest from Washington University and other public media stations with similar collections who want an access website “like Open Vault.” We plan to form new collaborative partnerships as we continue to develop both the field of moving image and audio archives and our new projects.

### **Sharing Our Results**

The WGBH MLA web developers have posted the code for Open Vault on GitHub at <https://github.com/afred/openvault>. GitHub is a website that facilitates the sharing of open-source software projects. Previous updates were posted to <https://launchpad.net/bvault> and <https://github.com/WGBH/openvault-core>.

We plan to communicate our results from this grant with the public through the Open Vault website, and we continue to share our work through symposium and conference presentations, as noted below.

### ***Plans and goals for the period subsequent to the grant***

While this grant is completed, work continues on Open Vault as we continue to refine and improve our website’s design, enhance our metadata, provide increased access for scholars and the public, and pursue new collaborations to advance the technical development of media archives.

It is clear that Open Vault has been valuable for scholars, who have eagerly engaged with our public-facing catalog. We are excited at the many requests for content we have already received and expect the inventory will continue to aid scholars as more discover it. WGBH continues to promote Open Vault through both Twitter (<https://twitter.com/wgbharchives>) and Facebook (<https://www.facebook.com/wgbharchives>). In addition, the Scholar Collections are also being promoted through social media platforms. The scholars themselves are allies in the promotion of their collections, which has translated into referral traffic to the Scholar Collections and WGBH’s media. The payment model resulted in the highest level of engagement, but going forward we will continue to capitalize on scholars’ interest in our collection by requesting any metadata or notes they are willing to contribute.

In addition, we have future plans for Open Vault that will allow us to post other valuable content from our archive beyond the series titles, such as outtakes. We also intend to continue updating the website to improve scholars’ experiences on Open Vault, benefitting from their feedback and usage. We have worked to make Open Vault responsive and user-friendly for tablets and smartphones, and we will continue to work to improve the design of the site and better visually manage the volume of data.

We also now have a venue integrated into Open Vault that allows approved scholars to create their own collections, subject to staff review. In the future it will be easy for us to allow new scholars to expand the existing Scholar Collections. We hope to include this feature within the American Archive of Public Broadcasting.

Beyond our collaborations with scholars, we are exploring additional crowdsourcing strategies. We have recently requested funding from the NEH to create a gaming model of crowdsourcing metadata tagging for the senior citizen community. We have also submitted a proposal to use crowdsourcing to correct inaccuracies in automated speech-to-text transcription. Both efforts would expose more data to search engines.

We continue to explore the best methodologies for assembling and regularizing metadata across our systems in line with the challenges we face. For example, we see great potential in employing Resource Description Framework (RDF) ontologies, which would allow a more flexible organization of data. In a recent hackathon, we explored the opportunity to leverage the European Broadcasting Union's existing EBUCore to bridge the gap, evolving PBCore's metadata framework so it can be used in RDF environments.

We are most excited at the opportunity to open to scholars 40,000 hours of new digital content through the American Archive of Public Broadcasting. This collection promises to expand our view of the history captured in 60 years of public media, opening a wealth of resources for researchers. The lessons of this project will inform our new work, and we look forward to keeping the Mellon Foundation updated on our progress in this area.

### **List/Description of Any Recent Evaluations, Publications, News Articles, or Other Materials Related to the Grant**

#### ***Evaluations***

Google Analytics have been collected since soon after the Open Vault website went live, and the statistics are evaluated regularly. Most researchers find content via Google searches, so we have improved our metadata to be more easily discoverable by search engines. We have also discovered visitors are finding the catalog through referral traffic from blogs and forums on various topics. Wikipedia also drives traffic, particularly from links in the articles on Vietnam.

#### ***Publications***

Our Digital Library Initiative team shared their work on this project at the following academic and professional conferences and networking events between 2011 and 2014:

- Open Repositories 2011 (June 2011)
- CURATECamp at Stanford University (August 2011)
- International Federation of Television Archives World Conference (September 2011)
- New England Archivists Fall Meeting at Dartmouth College (October 2011)
- THATCamp New England at Brandeis University (October 2011)
- Digital Libraries Federation Fall Forum in Baltimore (November 2011)
- Association of Moving Image Archivists Conference in Austin (November 2011)
- Society for Cinema and Media Studies Conference in Boston (March 2012)
- Society of American Archivists Annual Meeting (August 2012)
- Association of Moving Image Archivists (AMIA) in Richmond, VA (November 2013)
- New England Archivists Meeting in Portsmouth, NH (March 2014)
- Boston Area Collections Coalition in Boston, MA (April 2014)
- Hydra Connect in Cleveland, OH (September 2014)
- Orphan Film Symposium in Amsterdam (October 2014)

The scholars participating in the Scholar Collections also publicized Open Vault as they spoke on their topics and on digital media at conferences.

- American Historical Association in Washington, DC (January 2014)
  - Tracey Deutsch presented her paper, “Domestic Labor and New Cooking in the Age of Julia Child,” in the panel on “Cooking for the Masses? Culinary Celebrity, Gendered Labor, and Class in Mid-Twentieth-Century Argentina, Mexico, and the United States”
- 2014 Southwest Popular/American Culture Association Conference in Albuquerque, NM (February 2014)
  - Dana Polan presented “Postwar TV Moments: The Cooking Show as Cultural Mediation: The Case of Joyce Chen and Asian Cuisine.”
  - Daniella Perry presented “Television on Television Violence: Frontline’s Does TV Kill? and the Problem of Media Panics.”
  - Jim Cocola presented “Nine Poets Walk Into an Institute of Technology.”
- Northeastern History Graduate Student Association Graduate Student Conference in Boston, MA (March 2014)
  - Maureen Tannetta presented “Showing Status: WGBH and the President’s Commission on the Status of Women During the Kennedy Administration, 1961-1963” as part of the panel on “Preserving the Past Online: Digital Projects and Presentations.”
- Society of Cinema Studies Conference in Seattle, WA (March 2014)
  - Karen Cariani, Mark Cooper, and others presented “The Televisual Archive: New Directions for Research and Access.”
- Visual Resources Association in Milwaukee, WI (March 2014)
  - Amanda Sikarskie moderated and presented on the panel “Case Studies in Collaboration within Archival and Special Collection Environments.”
- Berkshire Conference on the History of Women in Toronto (May 2014)
  - Tracey Deutsch presented her paper, “The Julia Child Project: French Food, the Cold War, and the Labour of Cooking,” in the panel on “Women’s Labours and the Global Production of Consumer Culture in the 20th Century.”
- Orphan Film Symposium in Amsterdam (October 2014)
  - Mark Cooper presented “Metadata and the Future of the Humanities.”
  - Karen Cariani along with Mark Cooper and others participated in a panel discussion, “New Research Networks for Obsolete Media.”

The Scholar Collections scholars have begun to publish on topics using Open Vault assets as their sources.

- John Campopiano had an article, “Memory, Temporality, & Manifestations of Our Tech-Nostalgia,” published in the academic journal, *Preservation, Digital Technology and Culture* (PDT&C) (July 2014, Volume 43, Issue 3, Pages 75–85).
- R. Lisle Baker published the article “An Introduction to The Advocates and the Work of Professor Roger Fisher: A Prequel to the Program on Negotiation” on the *Harvard Negotiation Law Review* website (March 12, 2014).

### ***News Articles***

Open Vault has been featured in print and online as well as mentioned on television.

- “Time Capsule: An Hour with Steve Jobs in 1990” by Kasia Cieplak-Mayr Von Baldegg October 11, 2011, 1:49Spm ET *The Atlantic* (<http://www.theatlantic.com/technology/archive/2011/10/time-capsule-an-hour-with-steve-jobs-in-1990/246388/>)
- “Watch a Gay Marriage Debate from 1974 (Video)” by Tim Molloy June 3, 2014 @ 9:23am *The Wrap: Covering Hollywood* (<http://www.thewrap.com/watch-a-gay-marriage-debate-from-1974-video/>)
- WGBH’s *Greater Boston* makes frequent use of archival footage from Open Vault in their broadcasts and directs viewers to Open Vault.
- We also receive a large volume of traffic from blogs or forums on various topics. Wikipedia drives a great deal of traffic to Open Vault, particularly Wikipedia’s articles on Vietnam. The series *War and Peace in the Nuclear Age* and the complete interviews from the series *Vietnam: A Television History* are referenced in several Wikipedia articles. For an example, see [http://en.wikipedia.org/wiki/Archimedes\\_Patti](http://en.wikipedia.org/wiki/Archimedes_Patti).

### ***Other Materials***

- Appendix A: Open Vault
- Appendix B: Attracting Scholars
- Appendix C: About the Scholars & Scholar Collections
- Appendix D: Scholar-Submitted Metadata
- Appendix E: Survey Solicitation