

THE WGBH MEDIA LIBRARY AND ARCHIVES DIGITAL LIBRARY PROJECT

Phase II
Developing a Public Television Content
Delivery System for Academic Institutions

FUNDED BY
THE ANDREW W. MELLON FOUNDATION

March 2010

Principle Investigators

Karen Cariani
Peter Pinch

Project Staff

Chris Beer, Web Developer
Courtney Michael, Project Manager
Mayo Todorovic, Designer



CONTENTS

EXECUTIVE SUMMARY.....	1
INTRODUCTION	2
PROTOTYPING PROCESS	4
SELECTING THE SCHOLARS	4
INITIAL MEETINGS AND TASKS FOR SCHOLARS	5
SCHOLARS’ INITIAL REQUIREMENTS	7
WGBH’S CHALLENGE: HOW TO MEET SCHOLAR REQUIREMENTS	9
OPEN VAULT RESEARCH (OPENVAULTRESEARCH.WGBH.ORG)	11
PROTOTYPE CATALOGING.....	12
PROTOTYPE DESIGN.....	14
HOMEPAGE	14
THE RECORD PAGE.....	15
SEARCH.....	23
MY PAGE	26
THE RELATIONSHIP MAP	26
PROTOTYPE TECHNOLOGY	29
TECHNOLOGY EVALUATION	32
FEEDBACK AND REVISIONS	32
SCHOLAR FEEDBACK.....	32
USER TESTING	33
MIT “VISUALIZING CULTURES” CLASS.....	33
SUSTAINABILITY	34
OUR CONCLUSIONS REGARDING SUSTAINABILITY	36
DISSEMINATION	38
CONCLUSION	41
LESSONS LEARNED.....	42
IMPACT.....	42
FUTURE PLANS AND OPPORTUNITIES	45
OUR THANKS TO THE ANDREW W. MELLON FOUNDATION	45
BIBLIOGRAPHY.....	47
WORKS CITED.....	47
WORKS CONSULTED.....	47
APPENDICES.....	50

EXECUTIVE SUMMARY

WGBH seeks to provide access to an unparalleled collection of over fifty years of public television history through the development of a Digital Library for higher education use. In Phase II, *Developing a Public Television Content Delivery System for Academic Institutions*, WGBH (1) utilized scholarly input to design and pilot an online media archive content delivery system for research and classroom use and (2) developed a preliminary business plan for future sustainability and growth. The models constructed through this work were designed to be replicable by archives, libraries, historical societies, and public broadcasters who are also interested in making their moving image and audio collections available to higher education.

In collaboration with selected scholars, WGBH staff worked to create a framework for the Digital Library's future implementation by assessing researchers' needs and incorporating their input into the project's design. To accomplish this, we conducted the following activities:

- Consulted with three scholars in a shared field of research (history/political science) to thoroughly understand their needs and the potential uses of the WGBH collection.
- Collaborated with these scholars to create a metadata structure directed toward their needs while being compliant with appropriate standards.
- Worked with these scholars to determine which items from the archive would be most useful and examined rights issues/challenges to providing those materials.
- Researched, developed, and piloted a distribution/content management system using open-source software.
- Analyzed the cost of expanding the delivery system and providing content on an ongoing basis.
- Worked with Ithaka to research business models, explore internal policies that can support the system, and develop a preliminary business plan.

As a result of this work, we developed a prototype aligned with WGBH's Open Vault (openvault.wgbh.org), a digital repository of WGBH archival content. This prototype web service, Open Vault Research, has been populated with a limited collection of rights-cleared research materials and features a search interface designed to provide access to visual content. In consultation with Ithaka, we also crafted a preliminary business plan to inform Phase III implementation, laying the groundwork for the construction of an enduring online resource for scholarly use.

More broadly, we have created a model system using open-source software that others can adopt and adapt for their own use. Our deep knowledge of the types of assets preserved in media archives as well as our in-depth exploration of scholarly and pedagogical needs in higher education have allowed us to build a user-friendly framework for exposing rich collections of visual content suitable for research and classroom needs. We have shared and will continue to share both our software and conceptual models for the project with other institutions, and we will also bring this knowledge to inform WGBH's involvement in the American Archive project. As a result, we hope that our work will fuel scholarly and classroom interest and use of moving image and audio media as primary resources for research and teaching.

INTRODUCTION

As the keepers of historical materials, media archivists find themselves at a turning point. With the exception of the field of film and media studies, established humanities and social science scholarship has yet to broadly recognize film and television archives as important sources. At the same time, our culture and history are increasingly recorded in multimedia and “born digital” formats. Further, in an environment shaped by YouTube and Hulu, today’s young “digital natives” take web video for granted.

Just over the horizon, an emerging generation of researchers already expects archival content delivered online and on demand, and moving image archivists must be prepared. As the current model of intensive, text-based research gives way to newer modes of cross-media, interdisciplinary research and analysis, moving image archives must quickly find a way to make their materials accessible or else face rapid obsolescence. In this emerging paradigm, if an archive’s moving image content is not on the web, it may as well not exist.

WGBH, Boston’s PBS Station, produces over one-third of nationally broadcast PBS television programming. The WGBH Archives holds over 300,000 hours of moving image and sound content, as well as over 8,000 linear feet of related documentation and still images, represented in 745,000 catalog records. Not only do researchers find finished documentary films from our flagship productions (*FRONTLINE*, *NOVA*, and *American Experience*), but, more importantly for most researchers, we retain all of the production elements that go into the making of these films.

In order to meet the challenge of providing access to this unparalleled content, WGBH has launched a multi-phase Digital Library project to make its fifty-year collection of television history available to scholars and educators. In Phase I, the *Assessment for Scholarly Use*, our work demonstrated that the WGBH Archives holds materials of high relevance and value for scholarly researchers. Ultimately Phase III will mark the implementation stage of the Digital Library, as we will scale up our resource collections and open them to scholarly access while putting in place a sustainable business plan for future access and growth.

This report on Phase II details our collaboration with scholars on developing a prototype model in order to determine how to most effectively facilitate discovery and use of our materials and how to best deliver them over the Web. Through this process, we worked to address and provide workable solutions to challenges faced by all moving image archives that endeavor to serve scholarly researchers. These challenges include:

Exposure to archival collections. Moving image or media archives have traditionally been more difficult to access than paper-based archives because of obsolete formats and playback technologies. Consequently, lack of exposure to an archive’s holdings is a key concern. Add to this copyright barriers, some scholars’ wary or dismissive attitudes towards the use of media as a source, and the relative newness of history captured as moving images, and we find that media archives are only beginning to be recognized as key resources in the research arena. Our work consequently is motivated by a desire to

further the status and importance of media archives as valued repositories for scholarly research.

Access to the materials. The standards for cataloging media materials have traditionally mimicked bibliographic approaches by listing completed films with titles and authors rather than itemizing the raw elements that go into the finished product. While traditional paper archives generally summarize collections at the collection and folder level and less frequently delve down to the item level, raw footage and film elements must be cataloged at the item or even sub-item level to be accessible to researchers. Unfortunately, cataloging at these levels is not only extremely labor intensive, it requires direct access to the materials via film and video transfers, digitization, or the use of rare and aging playback equipment. Faced with what may seem an overwhelming task, many moving image archives find themselves struggling to make their collections accessible in ways that are meaningful and user-friendly.

Technical challenges. Even once materials have been cataloged, digitized, and stored in a repository or database, archivists lack the tools to manage and expose their data to the world. Few affordable and scalable applications and utilities exist to help a media archive provide access to its repositories. Similarly, there are few common or broadly applicable policies and frameworks for legal issues, vocabulary development, or information discovery. The digital media files themselves also pose problems, as the file size of preservation-quality video makes redundant copies of video expensive but still necessary in a volatile landscape of evolving compression formats and delivery needs. As a consequence, open source tools, widely accepted standards, and transparent, community-vetted policies are necessary for the field to move forward into the new age of digital scholarship.

As a result of our work in these areas, we successfully achieved the outcomes we proposed at the project's outset: (1) we have created a model system for providing online access to the MLA's archival media materials for scholars and educators; (2) we have built a prototype content delivery system through open-source software with applicability to other projects; (3) we have developed a preliminary business plan for the implementation of the Digital Library, including a review of the existing landscape in the repository field, models for long-term sustainability, and cost projections for expansion; and (4) we have assembled knowledge of the resources and metadata requirements that scholars find most important when they draw upon the collected media assets assembled through our history programming.

This project has also allowed us to explore collaborations with other institutions working with media materials. Through many of our dissemination efforts, we were able to learn from and interact with other Mellon grantees and compare notes in the area of developing digital systems for archives and scholarly research. With the generous support of this grant WGBH has been able to share its knowledge of digital media with these institutions, explore the use of open source solutions for digital media management, and truly research our target audience's needs. While we have found existing projects and solutions with similar goals and methods, we are excited to move media archives forward by layering new access tools and open source technologies on top of our extensive user research.

PROTOTYPING PROCESS

SELECTING THE SCHOLARS

We began this project by recruiting scholars to guide our work. We were looking for experienced, well-established faculty who would be willing and available to put in the time necessary to advise us. In addition, we hoped to find advisors who had some experience with, or at least a desire for, using media in their work. Finally, we hoped to align the strengths of our collection with our scholars' interests as much as possible.

We ultimately selected three scholars for this project: Dr. James Blight of Brown University's Watson Institute for International Studies, Dr. John Dower of MIT, and Dr. Peter Winn of Tufts University. Each brought considerable academic experience and each hoped to use media in unique ways to aid their current research. All also had previous relationships with WGBH, whether as a project advisors or expert interviewees for documentary films, and therefore were somewhat familiar with the types of materials available in our collection.

The research interests of these scholars (political science and history) correlated well with the results of our Phase I inventory, where we discovered that the WGBH Archives have great potential value for humanities and social science scholars. Beyond these research areas, however, scholars from diverse fields have also shown great interest in our collections for reasons we had not anticipated. For example, participants in user studies for this project expressed excitement about our science content and the possibility of researching the history of science through its portrayal on television over the past fifty years. Additionally, in our work with Columbia University on the Institute for Museum and Library Services (IMLS)-funded *Vietnam Project*, educators affiliated with the Center for New Media in Teaching and Learning have used WGBH archival materials from *Vietnam: A Television History* to support undergraduate and graduate instruction in journalism, education, and the Vietnamese language as well as history.

Below please find further details on the participating scholars.

Dr. James Blight

***Research Professor and Adjunct Associate Professor
Watson Institute for International Studies, Brown University***

In partnership with his wife Dr. Janet Lang, Dr. James Blight has pioneered a method termed critical oral history, through which he brings together in one room historians, former decision makers, and documents in order to revisit and study the past. He has employed this method successfully in conferences exploring the Cuban Missile Crisis and is currently directing a similar venture to examine the history of Iran prior to the Persian Gulf War. In addition to using this process for his own research, he wants to train others in the practice of critical oral history so that they can conduct their own similar conferences.

In preparation for a conference in December 2008, Drs. Blight and Lang searched the WGBH archives for materials on the Iran-Iraq War and the Iranian Revolution, events that did not register deeply in the American popular consciousness. They were also interested in materials on Jimmy Carter and the Iran hostage crisis of 1979-1980, and they were excited to mine the archives for non-American, non-Western views of these events in the Middle East. In addition to

using archival media for pure research documentation, they sought to use the media to get their critical oral history group in the right frame of mind before embarking on their sessions, creating a “time machine” to take people back so that they could better recall the era’s events and moods.

Dr. John Dower

Ford International Professor of History, Massachusetts Institute of Technology

Dr. John Dower studies twentieth-century United States and Asian history, race in the international context, and war as culture. He is very much visually oriented and a pioneer of a method of inquiry termed “image-driven scholarship.” Along with Dr. Shigero Miyagawa (MIT linguist and Kochi-Manjiro Professor of Japanese Language and Culture), Dr. Dower co-leads “Visualizing Cultures,” a web project that curates collections of historical images and analyzes the contemporary visual culture these collections both reflect and create.

Dr. Dower joined our Digital Library project because he was interested in archival sources on the Sino-Japanese wars beginning around 1931 through the Asia Pacific War (i.e. the Pacific theater of World War II), from which he hoped to draw materials for “Visualizing Cultures” as well as for his forthcoming book, *Cultures of War*. He also wanted to explore our collections around China and Vietnam in general along with any historical images – still or moving – of Asia in our collection.

Dr. Peter Winn

Professor of History, Tufts University

Dr. Peter Winn is a scholar of Latin America whose research interests currently include issues of history and memory, particularly around dictatorships and human rights violations in Latin America. One issue that he explores is the multiple ways in which media affects popular memory and historical narratives. He sees documentary film production as another form of writing history; like historians, media makers select the evidence or stories they include or leave out, and these selections reflect contemporary biases and perspectives. Consequently, through their work, they write history in the context of their time.

In the early 1990s, Dr. Winn served an advisor on WGBH’s *Americas* television series and authored the companion book *Americas: The Changing Face of Latin America and the Caribbean*. He joined our current project because he wanted to use materials collected for that series and others not only as archival sources, but also for the purpose of examining how the history has been interpreted, both in the original source material and then by media producers like WGBH. Because of a specific interest in El Salvador, Dr. Winn wanted to focus his participation in this project by using materials from three of our series – *Americas*, *Mexico*, and FRONTLINE’s *Crisis in Central America* – in order to examine how these three different productions treated the history of El Salvador over time.

INITIAL MEETINGS AND TASKS FOR SCHOLARS

At our initial meetings with the scholars, we sought to tease out answers to the following questions:

- What is your area of research?
- What materials are you interested in using from the WGBH archives?

- What information do you need from the archives in order to use an item?
- How would you like to use media (and other items) you find in the archives
- How do you use media in general?

Mapping the Metadata Requirements

In order to determine what kind of information our scholars would need to best utilize our archival content, we chose a few tapes from the collection relevant to the scholars' fields and screened the tapes with no introduction. We then asked the scholars what they would need to know about these materials in order to use them in their work. In this way, we hoped to discover how much and what kind of metadata is required by the scholarly researcher when using media materials. (This exercise had the additional effect of spurring the scholars' interest in the collection.)

Our scholars agreed that they wanted as much metadata as possible. For example, for an interview they wanted to know not only who was on the screen and why they were there, but also the production for which the film was shot, when and where it was filmed, who was behind the camera, and the prompts or questions that sparked the interviewee's response. In other words, the scholars were not about to take the speaker's words at face value but rather wanted to know as much contextual information as possible in order to accurately interpret the content and thus treat it as a valid historical source.

The amount of metadata requested by our scholars was a lesson for the MLA. The challenge posed by their requests would prove especially demanding because while our internal catalog contained some of the information, it is more often implied rather than explicit. As a consequence, much of this metadata would need to be newly added to the records and/or redescribed by using searchable fields or terms. Throughout the project, we have consistently found that the cost of assembling the metadata has exceeded all other costs, both in terms of time and full-time-equivalent staffing.

Assembling the Prototype Collection

Based on their research interests, we next sent our scholars lists of materials in the archives we thought would be of interest to them. We asked them to narrow the list and help us select some materials for inclusion in our prototype project, Open Vault Research. This process took several rounds and proved to be a "chicken and egg" problem – while we sought to make the collection accessible by testing a select group of media materials, our advisors could not select media materials until they could see the collection. They needed access to the big picture in order to pinpoint relevant content, but we needed them to pinpoint relevant content to test whether the big picture was worth accessing.

To work around this problem, we experimented with sending print-outs of series and program lists and other data extracted from our Filemaker-based physical asset catalog. We also provided transcripts of programs in order to gauge their interest. This process (akin to the "normal" process scholars encounter when visiting an archive) was a bit frustrating for our advisors, who were excited to dive into media but first had to wade through multiple PDF and Excel lists.

One positive effect, however, was that our scholars realized how sparse some of our cataloging is and offered suggestions for making materials accessible. Dr. Dower in particular encouraged us to publish what data we have and make the materials available alongside a boilerplate disclaimer which could note that we lack contextual information about many assets or warn where our limited data is incomplete. This also encouraged an idea we had hoped to explore in the course of this project: adding user-generated metadata to our records. Dr. Dower himself was conflicted by this idea – excited that users in Taiwan, for example, might be able to help identify images in our collection from the early history of Formosa, yet wary of the diverse motives of many online contributions. From this and other conversations, we determined to explore the idea of accepting user-generated metadata but refraining from incorporating it into our authoritative catalog until vetted. We intend to take Dr. Dower’s advice on future cataloging projects, most notably our planned Boston Local News Archive.

In the end, for the purposes of the prototype, we made educated guesses about which media materials to include. For Dr. Dower, we focused on materials gathered for the production of *Tug of War: The Story of Taiwan* (1998), including stills and footage from the Taiwan Military Agency not readily available outside of Taiwan. For Dr. Winn’s study of Latin and South American human rights abuses, we pulled materials from the *Americas* series (1993) and FRONTLINE’s *Crisis in Central America* series (1985), which included interviews with victims of Pinochet in Chile, FMLN guerillas in the Peruvian countryside, and the children of the Disappeared in Argentina. Both FRONTLINE and *American Experience* archives provided sources for Dr. Blight and Dr. Lang’s research on the Iranian revolution and events leading up to the hostage crisis, as well as materials from *War and Peace in the Nuclear Age* (1989). For their research, we included interviews and historical news footage gathered for documentaries on Jimmy Carter (1989, 2002) and Ronald Reagan (1998).

In addition we added several other assets from the archive to test how we might display different formats. For example, in the case of materials with an accompanying shot log, we experimented with how we could sync the shot log to the media or add older programs without transcripts.

A further issue in adding content to the prototype was rights clearances. While we could provide free access to our scholars without having to clear rights, we also wanted to provide public access to the prototype after the project’s end. Consequently it was necessary to pay for the necessary work to clear the rights of the media selected for inclusion.

SCHOLARS’ INITIAL REQUIREMENTS

Though they shared similar fields of inquiry, our scholars’ research methods and questions differed significantly. Dr. Blight’s use of media as a “time machine” called for compelling, never-before-seen images; Dr. Dower’s analysis of visual materials meant the provenance of the materials was key; Dr. Winn’s desire to compare and contrast historical accounts across time and place required diverse voices from the archive.

Despite these differences, however, all three shared basic needs for their analyses of archival media: raw materials, transcripts, search tools for speedy scanning, the ability to take notes on video (especially on smaller segments or points in time), rich metadata, and citation/embedding functionality.

Raw Materials

First and foremost, our scholars wanted full access to entire interviews and other content. They recognized that while 30 seconds of a particular interview might be cut into a final documentary, the full ninety-minute conversation lies hidden in the WGBH vault. A film producer might choose one neat soundbite in order to move his or her narrative along while ignoring the nuanced messages of the interviewee. Our scholars were also interested in all the “ums” and “ahs” in order to understand the full meaning of the speaker’s words. Untapped, “never before seen” conversations such as these are particularly exciting opportunities for discovery.

Transcripts

Early on in our conversations, it was clear that transcripts of the materials were extremely useful for our scholars. They attributed this to their “text-based” mode of analysis, i.e. the fact that they are used to skimming large amounts of text. Moreover, as they commonly cite textual materials in their work, they felt more comfortable dealing with the transcript rather than the video.

Search and Scan Tools

The need to manage time and the volume of resources were constant threads throughout our discussions. While our scholars wanted access to full-length outtakes, interviews, and footage, they also wanted to pinpoint quickly the sections most relevant to their research. Because video is not easily “skimmable” or “searchable” like text is, all have had the experience of spending hours watching the screen only to find that there is nothing useful for their work. In the end, we were able to exploit the accompanying transcript to make this a more fruitful process for our scholars.

Note-Taking

An added functionality our scholars encouraged us to adopt was a note-taking interface. They wanted to take notes on a particular moment in time and then be able to come back to that moment easily. Akin to writing in the margins of a text, they wanted to annotate their archival video source and be able to revisit the source material and their notes later on in their process. We were charged with making this process easier and more seamless than the low-tech ‘pause and pencil’ method – especially with web video where the controls are not as intuitive.

Metadata

Relevant content is the most basic need for the scholarly researcher, and contextual information about the authenticity and provenance of a source is necessary for the researcher to make best use of it. This need was laid out in our “metadata brainstorm,” when we asked the scholars what kinds of information they would need about a particular item in order to use it in their work. As discussed above, they requested rich metadata, including information about locations, dates, production purposes, filmmaker backgrounds, *et cetera*.

Citation/Embedding Functionality

Finally, our scholars wanted to integrate the materials into their scholarly works. At a most basic level, this means being able to cite the materials. At a more complex level, they envisioned someday embedding the materials into presentations and papers. While embedding was conceptually possible given the technology underlying our prototype, none of our scholars needed it for this project and we came up against vendor lock-ins that added to the difficulty.

However, our parallel IMLS-funded *Vietnam* project has addressed this challenge in different ways through the work of Columbia's Center for New Media Teaching and Learning. We hope to continue exploring this issue as rights and technologies permit and to promote moving images as viable objects in the evolving world of scholarly communication.

WGBH'S CHALLENGE: HOW TO MEET SCHOLAR REQUIREMENTS

About the WGBH Collections

As a media archive, our needs are very different from a library, yet we also do not fall into traditional archival frameworks. For example, we do not use traditional finding aids as access tools. We do arrange our materials as archivists in the sense that we manage series and collections of materials whose provenance is of utmost importance, but we tend to describe or catalog our holdings as librarians, at the item and even "sub-item" or "shot" level. This level of description is necessary for two main reasons: 1) because our collections and our primary users (filmmakers) are so visually oriented; and 2) because the materials exist on non-human-readable formats. The catalog record is, therefore, very much a surrogate for our materials.

Access Challenges for Scholars

From conversations with our advisors, as well as from experience with visiting researchers, we recognize that media archives are challenging venues for scholars. In the specific case of WGBH, our rate of production (and our corresponding scale of acquisition) makes for an extremely robust collection, and yet we recognize that our internal interfaces to the data also make external access complex. For example, our MARS Filemaker physical access database tool is extremely customizable for internal use, but it is hard for scholars to learn and use efficiently. While the WGBH MLA has experimented with offering traditional finding aids online, their hierarchical, text-based, and largely linear formats do not well suit our media collections. Internally we have solved this challenge by creating item and shot-level databases for film production, but using these databases requires a steep learning curve and significant investment of time that is frequently prohibitive to external researchers.

In recognition of this difficulty, we have already taken steps to provide user-friendly online access to selected archival content as funding permits. In 2006, the WGBH MLA launched Open Vault (<http://openvault.wgbh.org>), a searchable and browseable site that highlights four television series in our archives: *New Television Workshop*, *Say Brother*, *The Ten O'Clock News*, and *War and Peace in the Nuclear Age*. This site includes video clips and transcripts from the elements that went into the making of these series and is aimed primarily at educators and the general public. As an extension of the archives online, it has also greatly increased our collections' visibility to scholarly researchers. Over the last year, Open Vault had over 200,000 unique visitors and about 1 million page views with steady visitor growth. In addition, calls to our archive department for on-site research have steadily increased by approximately 37% over the last four years since Open Vault was launched.

Consequently, through the Phase II prototype, we sought not only to transition our patrons from analog to digital approaches, but to enhance and facilitate their use of our archives by developing innovative web and catalog technologies. In doing so, we have sought to develop new structures and processes that ameliorate the following barriers to common use:

Awareness and Access: Beginning at the most basic level, there has been lack of awareness in the scholarly community that we exist. Our Open Vault website has gone far to alleviate this problem. However, continued development of online resources and marketing is crucial to increase awareness of our collections' value. Without our online tools, visiting scholars must find the time and resources to support travel to WGBH. Once they arrive, they have a "traditional" experience of speaking with our archivist, searching through our database, ordering materials, and viewing them on site at our facility. However, there are additional points along this route which may pose difficulties.

MARS: Because this homegrown Filemaker Pro library management database is populated by production assistants, MARS materials are not cataloged with outside researchers in mind. Media Library and Archives staff oversee and groom the cataloging to minimal standards, but we simply do not have the resources to fully catalog our materials. As a consequence, MARS does not resemble a library catalog or a finding aid and therefore requires time and training to use. This is not to say that scholars cannot find useful materials in our collections, they just need staff support at present.

Dated Formats: Many of our materials are on older formats for which we have no playback equipment. As a consequence, these materials need forward migration to current formats in order to be viewed, which requires funding. This forces scholars to invest research funds in resources of unknown value to their work, and they cannot necessarily afford the cost or the time necessary for the transfer of this content. They need materials on demand and in useable formats.

Cost/Benefit Commitment of Time and Resources: Unlike text, video content is not easy to scan or skim. Also, unlike text, it is not easy to photocopy or scan to digital copies. Archival media material requires time and patience and, given the low level of cataloging, it is very possible that an item that required significant time to scroll through and view may not contain any relevant content. While this is true of all research materials, given the time investment necessary to access archival moving images, scholars must conduct a cost-benefit analysis. Sadly, without better tools and metadata the costs appear to outweigh the benefits for many potential researchers.

Usage Restrictions: Many times, scholars will want a copy of the material for their own use. Most of these times, we are unable to make copies for home use due to copyright and distribution restrictions, making it difficult for them to revisit their sources as they further develop their arguments. Additionally, in the very near future we expect that scholars will want to embed pieces of archival media into their digital scholarship, a use that was not foreseen in the legal agreements that covered our archived content at the time of their production. While the Digital Library project has helped WGBH push the envelope on internal copyright and distribution barriers to scholarly use of our materials, it remains an ongoing challenge that WGBH, and our users, must face.

Community Standards for Scholarship: A long legacy of text-based scholarship and negative attitudes toward the use of media as primary sources weighs heavily on our users. Although practices in higher education are rapidly evolving, many scholars still frown on the use of video

and audio in academic research and publication. Debates over the authenticity of oral histories, over the provenance of news feeds, over the biases of filmmakers – all of these enter the dialogue when a scholar turns to archival moving image and sound. Some scholars, therefore, remain wary of our resources, preferring to exclusively consult government documents and personal manuscript collections. It is our hope that this project and other efforts towards facilitating the use of archival media collections will go far to convince the academy that scholars of the latter 20th century must look to moving images and sound if they want to paint a full picture of recent history.

OPEN VAULT RESEARCH (OPENVAULTRESEARCH.WGBH.ORG)

By building upon the Open Vault foundation, we have developed Open Vault Research, a repository-based prototype site specifically designed for scholarly use that overcomes many of the above outlined barriers to use, bringing together the searchable functionality of an electronic finding aid and the multi-dimensional orientation necessary to fully understand our collections.

Because we found that the scholarly research audience has very different needs from those of the general public and educators, we greatly adapted the parameters of the original Open Vault. At the most basic level, scholars *want it all*, they want open and unmediated access rather than a curated or guided user experience. They also want to see an entire collection in order to ensure that their research is thorough, and they want to view full-length archival videos, not just pre-selected representative clips. In short, they want to do the data mining and curation work for themselves. Our prototype seeks to balance scholars' needs for undiluted and immediate access to a diversity of raw materials with elegant and intuitive design and tools to facilitate their use of the collection.

We therefore chose to focus our prototype design on solving the following challenges:

- Metadata – providing the rich information about the content required by our scholarly users in a manner more detailed than what we currently have and less “shaped” or editorially influenced.
- Navigation – helping our scholars find relevant materials quickly and orienting them to the collection contents through the search process rather than through finding aids or transcripts.
- Search and Synchronize – allowing users to scroll through and search media without having to view or listen to the entire piece, creating the media equivalent of “speed reading” or scanning texts.
- Connection – allowing users to navigate through and across collections by revealing similar and related assets “around” an individual asset, thus providing ways for serendipitous discovery of other relevant materials.
- Rights – clearing archival materials for public access and use, a process that currently requires time-consuming and costly processes.
- Reuse – enabling users to use our materials in their scholarly products, taking materials or their surrogates out of the repository and integrating them into their own digital context.

The Open Vault Research prototype is now available at openvaultresearch.wgbh.org, where users can register at no cost, browse rights-cleared assets within the collection, and test the functionality of the interface's design.

PROTOTYPE CATALOGING

Without thorough item level description at the very least, we have found our media assets are of relatively little use. At the same time, without collection level context, they lose much of their value to scholars. To exploit the meaning and contents of individual assets in a media archive, or any archive for that matter, the researcher must be able to see a larger picture. For our prototype, we attempted to draw this picture by illustrating selected metadata-driven relationships between assets.

To organize our descriptive metadata, we turned to PBCore. This metadata standard developed by the public broadcasting community offers a robust system and controlled vocabulary, which are essential in creating reliable discovery interfaces for scholars and others. Further, its schema for describing moving image and ancillary materials is significantly more granular and specific than other options.

An added benefit is that it aligns with existing WGBH systems, which already store similar metadata. WGBH has an internal Digital Asset Management (DAM) system from which we export the XML files, convert them to PBCore, and groom them as necessary. Consequently, we assembled our descriptive metadata within a PBCore XML document, one document per digital asset.

As PBCore is a relatively new and underutilized standard for describing media, we have found some areas for improvement and have adapted and supplemented it. In the midst of our work, we were pleased to discover the development of a European analog to PBCore, EBUCore, from which we borrowed the date elements to fill in gaps in PBCore. We also created a Dublin Core alternative for increased compatibility with existing tools and interfaces (e.g. Zotero, OAI-PMH, etc).

For the purposes of visualizing the collection and the relationships between assets, the most important elements in the PBCore document are coverage (spatial), creator, contributor, and subject. Within the PBCore schema, contributor is defined rather loosely and sometimes conflicts with the idea of “people as subject” cataloging. For example, in an interview with Jimmy Carter’s National Security Advisor Zbigniew Brzezinski, should Carter be classified as a contributor-subject or is he simply a subject? These challenges and our recommendations are now informing the PBCore 2.0 project led by WGBH and funded by the Corporation for Public Broadcasting in service of the American Archive Initiative.

Although the PBCore document serves as a high level summary of an archival asset, the substantive details which greatly enrich the cataloging come from the asset’s sub-item or content level description. For our prototype, we encoded the transcripts of each asset in TEI (Text Encoding Initiative) XML. We also experimented with encoding shot logs in the International Press Telecommunications Council’s NewsML standard.

Using the “Transcriptions of Speech” profile of TEI, we tagged people, places, dates, and subjects in the transcripts where applicable. We made use of the Library of Congress Authorities and Subject Headings as our cataloging authority wherever possible, and, as a back-up, we made

employed Wikipedia terms as outlined in conversations we encountered around the emerging Semantic Web. For geographic terms, we utilized the Getty Thesaurus of Geographic Names. Appendix A provides a sample from an encoded TEI transcript.

To make this information useful, we needed to add it to a search index. To do so, we created a master PBCore record that combined basic descriptive information with technical metadata and the terms from the TEI document using XSLT documents and Fedora relationships. We mapped each keyword type to an appropriate PBCore element.

The image below in Figure 1 illustrates these relationships.

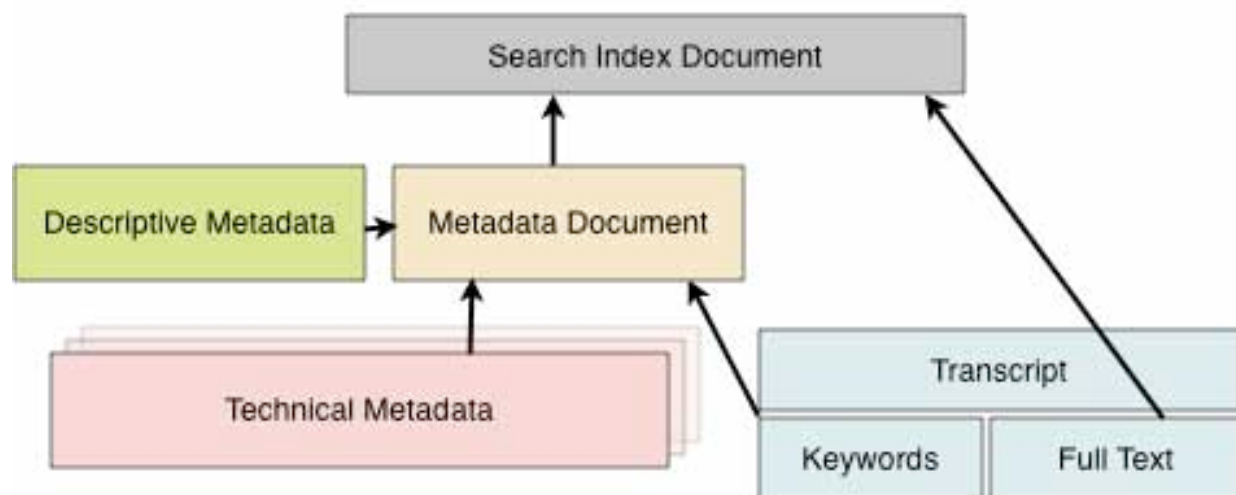


Figure 1. Metadata aggregation for search indexing

We also preserved the keyword URLs for future use, with the idea that future linked data applications will be able to take advantage of them and that data providers will offer a usable API to access additional contextual information. This, we hope, will provide a valuable entry point to the content for semantic web-enabled technologies and also allow us to use inferencing to provide broader and related terms in our browsing and search interfaces.

As a result of our in-depth cataloging, we are able to do some very practical things like improve our search results and provide relevant search facets. Perhaps more interestingly, we can also enable new and creative mash-ups of the data. These mashups provide a method for “serendipitous discovery” of our materials and also encourage researchers to explore our material in a less structured manner.

PROTOTYPE DESIGN

Below please find a discussion of the Open Vault Research web interface and components.

HOMEPAGE



Figure 2. Open Vault Research homepage

Our homepage builds upon the appeal of the openvault.wgbh.org “mosaic” design and includes a simplified menu and search. Our experience with the existing Open Vault site has been that most users never see this page, instead arriving on the site on a record page, either via a public search engine (such as Google) or a shared link. As a consequence, for this Open Vault Research prototype, we chose to focus our efforts elsewhere.

However, we did add increased metadata to the mosaic to make it more useful and enticing for scholarly researchers. Upon rollover, users can learn more about each tile and click to go to the record page for that asset. In addition, we added the link to series and collections as we felt this would be a familiar way for many researchers to begin their experience of the WGBH Archives.

THE RECORD PAGE

Interview with Robert McFarlane, 1996 (Part 1)

SHARE SAVE CITE

ANNOTATE TAG

00:00:09 / 00:18:33

Annotations:

PRIVATE (4)

- important stuff by Courtney Michael, Oct 30, 2009 [00:01:11 - 00:03:46]
- this interesting by Courtney Michael, Jun 12, 2009 [00:02:14 - 00:05:18]
- Guarded by Courtney Michael, Apr 16, 2009 [00:00:00 - 00:18:33]
- last straw by Courtney Michael, Apr 08, 2009 [00:04:41 - 00:06:09]

PUBLIC (3)

- McFarlane on defense build up by kim_ducharme@wgbh.org, May 20, 2009 [00:04:46 - 00:05:42]
- McFarlane on Iran by Courtney Michael, May 07, 2009
- Reagan as superficial by Courtney Michael, Apr 08, 2009 [00:00:35 - 00:01:23]

Transcript

PRINT TRANSCRIPT Search Transcript

Part 1

Hoyt: Interview with Robert C. McFarlane, Mr. McFarlane was National Security Advisor from October, 1983 until November, 1985. Interviewer: Austin Hoyt - Thursday, June 6, 1996, Washington, D.C.

Hoyt: Mr. McFarlane, how will Ronald Reagan be remembered? What's his important legacy?

McFarlane: I believe President Reagan will be a giant figure in history. For mostly, for his achieving what none of his predecessors had, the collapse of Marxism and the Soviet Union, the reduction of nuclear weapons, which neither Nixon nor Kissinger could achieve. He'll be remembered as a foreign policy President.

Hoyt: What's the most important misconception about him?

McFarlane: the misconception of Reagan is in --

-- I believe the characterization of Ronald Reagan as superficial is the leading myth.

He did focus on important things but he focused on them in enormous depth -- take Russia and--

-- before the Geneva Summit. I've never seen a President -- Nixon, Ford, Carter, absorb -- digest -- and master such a depth of information about the agenda and the other party, Gorbachev. Reagan was the master of that material and of that Summit.

Hoyt: What do you think the source of his appeal was to the American people?

McFarlane: I think the source of Ronald Reagan's appeal to Americans was his ability to inspire confidence, optimism, a sense that life is going to get better and that you can achieve great things if you work at it.

Ah, I used to dismiss that dimension of being the President, and I used to think it was more important for Presidents to know substance -- how to solve a problem of housing or healthcare. But, just as important, and I think probably more so, is the ability to inspire confidence, hope, and Ronald Reagan could do that like no other President.

Hoyt: How did his policies toward the Soviet Union differ from his predecessors?

McFarlane: President Reagan's policies were dramatically different from all of his predecessors, with regard to the Soviet Union. Detent, for example, rested on the premise that although we, in(?) the Soviet Union, might disagree, there was no prospect that we could prevail -- that because the Soviets had so much wealth, and oil, and gas and metal, that their inefficient system could survive, just by living off capital.

New search:

See this record in relation to others.

RELATED:

- Interview with Robert McFarlane, 1996 (Part 2)
- Interview with Robert McFarlane, 1996 (Part 3)

SOURCE:

Element: Interview with Robert McFarlane, 1996

Series: American Experience

Program: Reagan: An American Crusade

(c) 1996 WGBH Educational Foundation

DESCRIPTION:

Element Description: Robert McFarlane was the National Security Advisor from October, 1983 until November, 1985. He discusses the legacy of President Ronald Reagan, including Reagan's thoughts and actions surrounding the Strategic Defense Initiative, the Soviet Union, the Iran-Contra Affair, and American diplomacy in general.

Type: Interview

Topics: Soviet Union, National Security Advisor (United States), Marxism, Nuclear weapons, Antinuclear movement, Nuclear freeze, MGM-31 Pershing, SS-20 Missile, Strategic Defense Initiative, Korean Air Lines Incident, 1983, Armageddon, Communist party, Sound of music (Motion picture), Williamsburg Economic Summit (1983), Iran-Contra Affair, 1985-1990, Geopolitics, Iran Hostage Crisis, 1979-1981, Air Force One (Presidential aircraft)

MEDIA:

Type: Moving image/Original footage (3)

Physical format: Betacam SP (3)

Total running time: 00:52:43

PEOPLE:

Producer: Hoyt, Austin

Interviewee: McFarlane, Robert C.

Interviewer: Hoyt, Austin

Camerasperson: Hopkins, Terry

Sound: Jennings, Ed

Figure 3. Record page

The record page is the most significant page in our prototype. For the majority of users, who arrive via a recommended link or a public search engine, it will be the first page they see. Moreover, for scholars, this is where they will spend time interacting directly with the content.

The first goal of this page is to provide the researcher with access to the content. This is fairly straightforward in the “*Interview with Robert McFarlane*” example above, as we have digitized the video and provided a transcription of its audio. A secondary goal of this page is to facilitate the researchers’ *use* of the content. To that end, taking into consideration our scholars’ requests, we have developed several features to make our archival media content more easy to incorporate into a scholar’s workflow.

Metadata

New search: 

 See this record in relation to others.

RELATED:

 Interview with Robert McFarlane, 1996 (Part 2)

 Interview with Robert McFarlane, 1996 (Part 3)

SOURCE:

Element: Interview with Robert McFarlane, 1996

Series: American Experience

Program: Reagan: An American Crusade

(c) 1996 WGBH Educational Foundation

DESCRIPTION:

Element Description: Robert McFarlane was the National Security Advisor from October, 1983 until November, 1985. He discusses the legacy of President Ronald Reagan, including Reagan's thoughts and actions surrounding the Strategic Defense Initiative, the Soviet Union, the Iran-Contra Affair, and American diplomacy in general.

Type: Interview

Topics: Soviet Union, National Security Advisor (United States), Marxism, Nuclear weapons, Antinuclear movement, Nuclear freeze, MGM-31 Pershing, SS-20 Missile, Strategic Defense Initiative, Korean Air Lines Incident, 1983, Armageddon, Communist party, Sound of music (Motion picture), Williamsburg Economic Summit (1983), Iran-Contra Affair, 1985-1990, Geopolitics, Iran Hostage Crisis, 1979-1981, Air Force One (Presidential aircraft)

MEDIA:

Type: Moving image/Original footage (3)

Physical format: Betacam SP (3)

Total running time: 00:52:43

PEOPLE:

Producer: Hoyt, Austin

Interviewee: McFarlane, Robert C.

Interviewer: Hoyt, Austin

Cameraperson: Hopkins, Terry

Sound: Jennings, Ed

Covered: McFarlane, Robert C., Hoyt, Austin, Reagan, Ronald, Nixon, Richard M. (Richard Milhous), 1913-1994, Kissinger, Henry, 1923- , Ford, Gerald R., 1913-2006, Carter, Jimmy, 1924- , Gorbachev, Mikhail Sargeevich, 1931- , Thatcher, Margaret, Shultz, George Pratt, 1920- , Deputy National Security Advisor, Habib, Philip Charles, 1920- , Baker, James Addison, 1930- , Deaver

The extensive scholar-requested metadata can be found in the right column of the record page. This serves much of the function of traditional archival finding aids or library catalogs. Here a researcher will find the contextual information necessary to determine the authenticity and relevancy of an item. They can quickly learn the item's larger context by looking to the source information for the record's collection, series, and program memberships. They can also read an overview of the content in its description and determine the type of material cataloged. Basic information about physical media instances, dates, locations, and participants are also included here, providing the researcher with clues to the environment in which the archived content was created.

Included with the record are the topics, places, and people pulled from the transcript. Each of these are hot-linked, allowing the researcher to perform a new search with a click using our taxonomy.

Also present in the right column are links to related records (in this case, additional parts of this interview) and to similar records. "Related records" share root relationships such as a parent conceptual record (*Interview with Robert McFarland*). "Similar" records are calculated by the open source Apache Solr search engine using several catalog fields as variables.

Michael K., Regan, Donald T., White House Chief of Staff, Khomeini, Ruhollah, Weinberger, Caspar W., Casey, William J., Buckley, William Francis, Saffire, William, 1929- , Secretary of State, United States Secretary of Defense, North Atlantic Treaty Organization, KGB, Group of Seven countries, United States. Central Intelligence Agency, National Security Council (U.S.)

DATE:

Covered: 1980s

Created: Jun 06, 1996

Issued: Jun 25, 1997

Digitized: Sep 24, 2008

PLACE:

Covered: Soviet Union, United States, Middle East, Washington (inhabited place), Soviet Union (former nation/state /empire), Geneva (inhabited place), The Pentagon, United States (nation), White House, London (inhabited place), Germany (nation), France (nation), Kremlin (Moscow, Russia), China (nation), Oval Office, Iran (nation), Beirut (inhabited place), Santa Barbara (inhabited place)

SIMILAR:

 Nixon introduces Kissinger

 Interview with Stansfield Turner, 1991

 Air Strikes

 Interview with Elliot Abrams, 1992

 Troops and gunboats, 1969.

Figure 4. Right column metadata (edited to fit on page)

Transcript Synchronization and Search

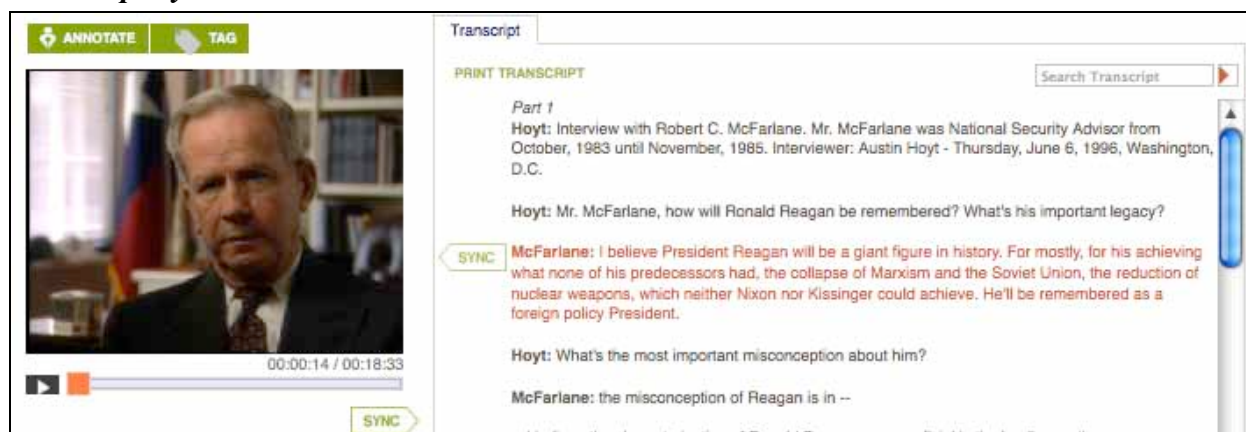


Figure 5. Transcript synchronization

When playing the video or audio on the left, the relevant section of the transcript on the right is automatically highlighted in red so the viewer can follow along while orienting themselves to the media as a whole.

In order to identify content of interest, researchers can use the search function to locate specific language in the transcript. The search results highlight paragraphs that include the search term, as shown below in a search for “Marxism”.

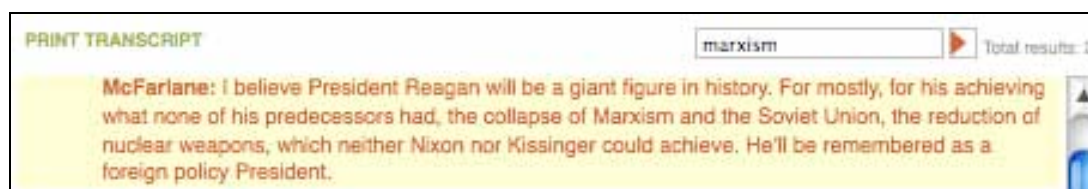


Figure 6. Transcript search

The “sync” function allows a researcher to use the transcript to jump to a specific point in the media or vice versa. For example, after searching and finding the mention of “Marxism” in the transcript, the researcher can immediately play that associated point in the media by rolling over the paragraph with their mouse and clicking on the “sync” button that appears. Likewise, if a researcher is viewing or listening to the media and would like to view the speaker’s words in the transcript, clicking the “sync” button underneath the media will allow them to quickly read the associated section of the transcript.

The “search within transcript” and “synchronize” features answer our scholars’ requests to dive quickly into relevant parts of the content without having to watch or fast-forward through entire videos or audio recordings.

Print Transcript

Our scholars were adamant that they wanted to be able to print the transcripts of the media materials, so we included a “print transcript” button within the transcript area. The printed version includes the record title, the transcript text, and citation information.

Share



Figure 7. Share button

In order to allow scholars to share records with their peers, we also provided a “share” link on every record page. Clicking this link brings up a “ShareThis” interface customized to match our design scheme.



Figure 8. Share options

Through this free service, users can share the record page’s URL via email, Facebook, Twitter, and other social networking tools. (While users can also cut and paste the address from the browser bar, we have found that some mistrust the permanence of URLs and prefer to use an application instead.) This feature responds to our scholars’ desire to send relevant records to colleagues and collaborators in their field or even, as one scholar candidly put it, “email it to myself for later!”

Save



Figure 9. Save to folder

In addition to the “share” function, we also created a space – “MyPage” – where users can save and organize records for themselves within a structure that allows the user to create their own folders or save to a central folder. (“MyPage” is discussed in more detail below.)

Of course a user can bookmark the record page as well, but not all of our scholars trust the bookmarks to return them to the same content. The “save” function raised their confidence that records would be available to them when they returned, an issue of concern when it comes to web resources.

Cite



Figure 10. Citation tool

One significant barrier to the use of archival media materials in scholarship is academic researchers’ unfamiliarity with the media content format and lack of precedent in written scholarship. While finished films and archival manuscript collections have well-established citation protocols, scholars face a conundrum when they consider how to cite resources such as raw archival film materials provided through online technology. For this reason, we decided to help by providing citations for each record in three standard styles: MLA, APA, and Chicago. Having the “Cite” function also encouraged scholars to think about the content as something worth crediting in their work. This feature was simple to implement and garnered rave reviews from our advisors and user testing subjects.

Our prototype site also offers machine-readable data points that can be exploited by tools like Zotero, a citation management tool created by the George Mason University’s Center for History and New Media. While none of our three scholars made use of Zotero, some of the newer faculty and graduate students we later consulted in our user testing either used it or were familiar with it. By integrating with these tools, we hope to make it easier for researchers to cite and organize their bibliographic records for archival media online, thereby removing doubts about this “new” format.

Embed

While this function is not shown on the record above, we also designed a feature to enable scholars to embed media materials into their own work, be it PowerPoint, HTML, or text. We did not fully implement it however because the materials contained in the prototype had rights restrictions and, to a lesser extent, because our scholars did not request it for their work. Once we have a collection that is cleared for embedding and an audience ready to use this feature, we have it on-hand and ready to implement.

As noted above, we have had some experience with embedding through our work with Columbia University on *Vietnam: A Television History*, an online collection that will feature broadcast and background content from the series as well as curricular materials. In that case, Columbia developed their own environment for embedding our archival materials. Like the “save” feature, their interface allows a scholar to cut and paste the URL of an item into their work in order to link or revisit the record page from an external application. We are planning to implement the embed feature fully with the Spring 2010 release of Open Vault 2.0, a new version of the site that will include increased functionality in coordination with the launch of the *Vietnam* collection.

Annotate



Figure 11. Annotation tool



Figure 12. Annotations

One complaint we heard from our scholars was that note taking for archival media sources requires them to balance their note taking equipment (pen and paper, laptop computer, etc.) alongside the “player” equipment (DVD player, CD player, tape player, etc.). By using web media, we were able to eliminate one of these factors as the video or audio can be controlled on the same screen as one’s notes. Taking this a step further, we incorporated an annotation tool into the media interface so that our scholars would not even have to switch between windows on their computer screen.

In our annotation interface, a scholar can select an “in” and “out” point – or a segment of the media – and take notes that apply only to that segment. They can also take notes that apply to the entire video. Our open-source time-code capture annotation tool allows scholars to easily and immediately revisit the materials referenced by their annotations. Once a scholar has created their annotation, they then have the choice to store their annotation in a folder of their creation. In addition, they can choose to make their annotation private or public, sharing their work or

storing it for their use only. Upon return to the record, their personal annotations will then appear below the video and the scholar can click on them to read them and to synchronize the video player to the relevant segment. Public annotations will be available to anyone who accesses the record.

Tag



Figure 13. Tagging tool

Our approach to tagging evolved in discussion with our scholars. In the beginning, we envisioned researchers enhancing our metadata by adding descriptive tags to the media. For example, if an interview discussed a certain topic but we failed to catalog it under that topic, a scholar could add it here. While our scholars saw the utility of enhancing the metadata, however, they greatly mistrusted user-generated metadata that was not their own. In other words, they could imagine using the tags for themselves and contributing to our catalog, but they did not want to rely on, or even necessarily encounter, tags created by other users. Instead, our scholars envisioned using tags primarily for organizational purposes. For example, a professor might tag an asset “Thursday class” in order to revisit it in preparation for that lecture. While “Thursday class” would have no meaning for other users of the asset, it is a perfectly valid use of the tagging feature.

While we hope that user tagging will prove a rich source of metadata, we are aware of the challenges in encouraging participation. For example, the Library of Congress’ Flickr Commons project has struggled with this issue, finding only a small percentage of user-contributed tags helpful for enhancing metadata.¹ To make the activity of tagging a bit more familiar to our

¹ Michelle Springer, Beth Dulabahn, Phil Michel, Barbara Natanson, David Reser, David Woodward, and Helena Zinkham. October 30, 2008. For the Common Good: The Library of Congress Flickr Pilot Project. Library of Congress Prints and Photographs Division. www.loc.gov/rr/print/flickr_report_final.pdf

scholars and to encourage their contribution of valuable metadata, we developed a tagging feature that implements a type-ahead function.



Figure 14. Type-ahead LCSH tagging

As the user begins to enter their tag, a drop-down menu appears with suggested terms from the Library of Congress Subject Headings. The user can select one of these terms or override them and continue creating their own identifier.

SEARCH

The screenshot shows the OpenVault Research website interface. At the top, there's a navigation bar with 'HOME', 'openvault RESEARCH', and 'WGBH MEDIA LIBRARY AND ARCHIVES'. User information for 'Courtney Michael' is visible, including a 'Profile' link, a 'logout' button, and a 'My Page' section showing '27 Items'. On the left, there's a 'BROWSE:' section with a 'choose category' dropdown and a 'SEARCH:' section with a search box containing 'Iran' and an 'Advanced Search' link. Below the search box is a 'NARROW:' section showing 'Iran (39 of 492 records)' and filters for 'MEDIA' (Video: 14, Transcript: 12), 'TYPE' (Interview: 24, Moving image/Original footage: 5), 'TOPICS' (Iran Hostage Crisis, 1979-1981: 21, etc.), and 'PEOPLE' (Carter, Jimmy, 1924-: 11, etc.). The main content area displays 'Results 1-10 of 39' with options to 'SAVE THIS SEARCH', 'SAVE RECORDS', and 'SHARE'. It includes a table of search results with columns for Title, Media, Date Created, Place, and Series Title. The results list various historical events and interviews related to Iran, such as 'Carter Breaks Diplomatic Relations with Iran' and 'Shah of Iran visits Jimmy Carter at the White House'.

Title	Media	Date Created	Place	Series Title
<input type="checkbox"/> Carter Breaks Diplomatic Relations with Iran President Carter's press conference breaking relations with and imposing sanctions against Iran in response to the hostage crisis.		Apr 07, 1980	Iran United States	American Experience
<input type="checkbox"/> Shah of Iran visits Jimmy Carter at the White House Footage of the Shah's arrival and speeches by Carter and the Shah.		Nov 15, 1977	Washington	American Experience
<input type="checkbox"/> Interview with Stansfield Turner, 1990 Stansfield Turner discusses the Carter administration's response to the Iran Hostage Crisis.		Dec 29, 1990	Middle East Iran	FRONTLINE
<input type="checkbox"/> Interview with Zbigniew Brzezinski, 2001 Former National Security Advisor Brzezinski discusses his service under President Carter.		Oct 24, 2001	Middle East Soviet Union Iran Afghanistan United States	American Experience
<input type="checkbox"/> Interview with Robert McFarlane, 1996 Robert McFarlane was the National Security Advisor from October, 1983 until November, 1985. He discusses the legacy of President Ronald Reagan.		Jun 06, 1996	Soviet Union United States Middle East	American Experience
<input type="checkbox"/> Interview with Nguyen Khanh, April 1981				Vietnam: A Television History

Figure 15. “Iran” Search Results page

From both the record page and the homepage, a user can begin a new search from the search box. The image above illustrates the results of a search on the term “Iran.”

Many of the “Search Results” page features are intended to give a quick idea of the scope of the archive. This is key to convincing scholars that the archive is large enough and covers a broad enough selection of relevant topics to deserve their time and attention. On the search results page, the search navigation appears on the left and the results on the right. Similar to the record page, a user can use the “save” function to save their search and resulting individual records, and they can also share their search results by clicking “Share” at the top of the page.

Advanced Search



Figure 16. Advanced Search

In addition to the basic search feature, we have also provided an advanced search feature where users can execute Boolean searches through the interface above, adding or subtracting terms from their search query.

Faceted Search



Figure 17. Faceted Search narrowed through three facets

We have also employed a faceted search system which allows further sophistication in identifying and classifying search results. As a user performs and layers searches, metadata is pulled from the results and summarized in the left column. Users can then further refine their results by selecting one of the metadata terms (facets) and applying it as a “filter” to narrow their options. Once applied, users can also broaden their found set by removing individual filters.

Perusing and applying facets helps orient the user within the collection and illuminate connections to their initial search term. The numbers to the right of each facet indicate the number of assets within their results relevant to that term. In addition, the bar at the bottom of the filtering area provides a visual cue as to how many resources a user has in their found set when compared against pre-filtered results. Each color band correlates to the application of a different filter. (While we also experimented with showing faceted relationships relative to the complete collection, we found the visualization could not easily and simply accommodate the mass of materials.)

Gallery View

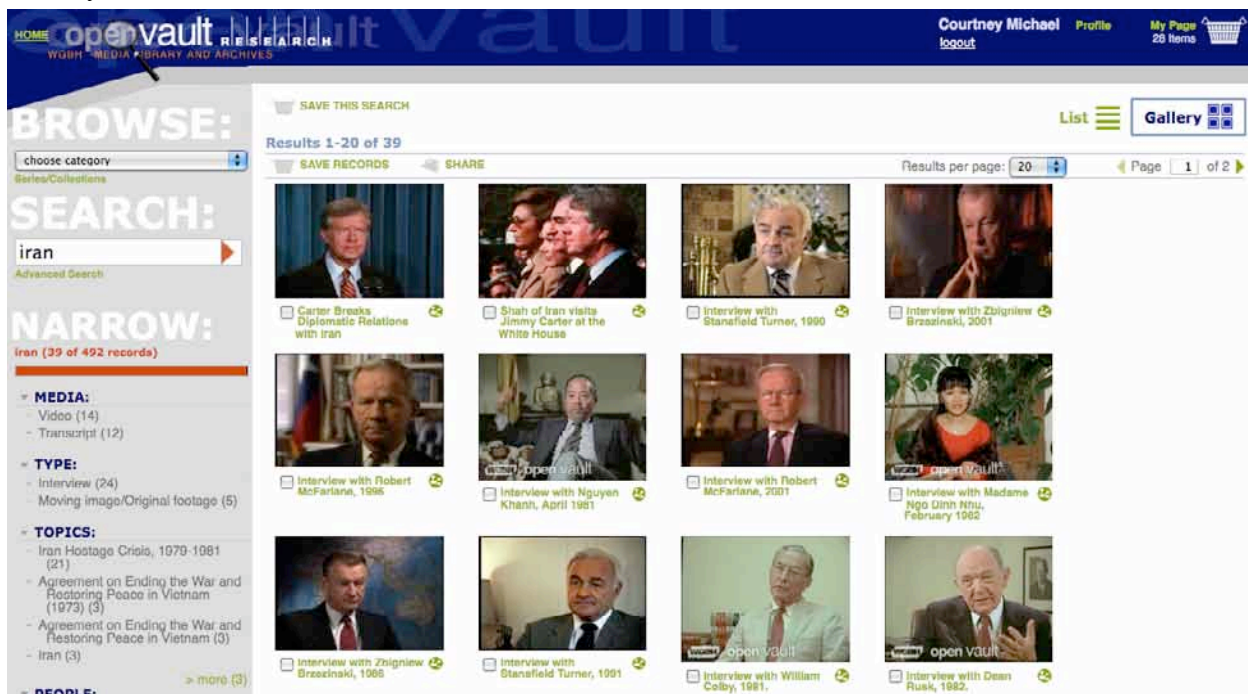


Figure 18. Gallery View

At the strong urging of Dr. Dower (who promotes “image driven scholarship”), we also added a “Gallery View” option in addition to the standard “List View” in order to allow scholars to scan representative thumbnail images quickly. When a user places the cursor over an image, a small box with summary metadata appears. This view is especially useful for viewing a collection of still images or quickly perusing a larger version of the thumbnails from the list view of the search results.

MY PAGE

The screenshot shows the 'My Page' interface of the Open Vault Research system. The top navigation bar includes 'HOME', 'open vault RESEARCH', 'WGBH MEDIA LIBRARY AND ARCHIVES', 'Courtney Michael', 'Profile', 'My Page', and a shopping cart icon. The left sidebar, titled 'MY PAGE:', contains a 'Folder' section with 'All resources', 'Iran Project', 'Interesting stuff', and 'McFarlane Interviews'. Below this is a 'Show' section with 'Any', 'Saved records', and 'Annotations'. The 'Tags' section lists various terms like 'appointee', 'brady', 'Democrat', 'desaparecidos', 'Hostages', 'hostages', 'Iran', 'Hostage Crisis, 1979-1981', 'jfk', 'marxism', 'marcia', 'brady', 'NSA', 'political appointee', 'press conferences', 'Rosalynn Carter', 'Ruhollah Khomeini', 'SDI', 'Tag', 'Vietnam', and 'Vietnam'. The main content area, titled 'Results 1-10', shows a list of records with columns for 'Title', 'Date', and 'Folders'. The records are: 1. 'Carter Breaks Diplomatic Relations with Iran' (Apr 28, 2009), 2. 'Carter's Last Hours in Office, tape 2' (Apr 28, 2009), 3. 'Chinese school in the Japanese era' (Apr 29, 2009), 4. 'Chinese school in the Japanese era' (Apr 29, 2009), 5. 'Chinese school in the Japanese era' (May 04, 2009), 6. 'Footage of the Madres de Plaza de Mayo marching' (Apr 28, 2009), 7. 'Interview with Haydee Lemos, 1992' (Apr 28, 2009), and 8. 'Interview with Haydee Lemos, 1992' (Apr 28, 2009). Each record has a checkbox and a '> edit' link in the 'Folders' column.

Figure 19. My Page

When a user saves an annotation, an asset, or a search query, the data can be found on “My Page.” This is a summary of the user’s activity in the archive. Scholars can come to “My Page” to find their notes, saved records, tags, and saved searches. They can organize items into folders and filter them by type of activity. In addition, in our next iteration of this page we will enable users to export records and annotations as PDFs or text files for use in their own applications.

Our scholars appreciated having a place within the site to save their work but were a bit intimidated by the management functions offered. They felt more comfortable cutting and pasting text into their own word processing software. In the future, we would like to keep up with the latest research management tools employed by digitally oriented scholars while also facilitating more traditional methods.

THE RELATIONSHIP MAP

As WGBH’s materials are multi-leveled and multi-faceted, our search interfaces work well because they allow the user to narrow and expand their results within an increasingly familiar framework. What the faceted navigation lacks, however, is a multi-dimensional view that can illustrate the relationships between materials and provide a visual representation of the media archive.

One of the first discoveries we made working with scholars was that they required an intense level of contextual information about a particular archival resource. As described above, we translated this need into a high level of cataloging at the item and even sub-item (or shot-log) level. While they want this density of information made available to them, however, the scholars

also want to be able to quickly pinpoint the exact content that will enhance their work. This paradox led us to experiment with alternative methods of communicating our metadata.

As a starting point, we decided to explore abstract graph-based visualizations to help answer the questions “What is the scope of your collection?” and “If I am interested in this item, what else in your collection might be worth my time?”

In response, we designed a visualization illustrating the depth of content in each asset and the explicit and implicit relationships between assets. While we did not have the resources to fully explore and implement this tool, the prototype functionality identified related items and displayed these relationships dynamically, which proved exciting to our users. We feel that this design innovation has brought us many steps closer to communicating our collection’s multi-dimensional aspect to our users.

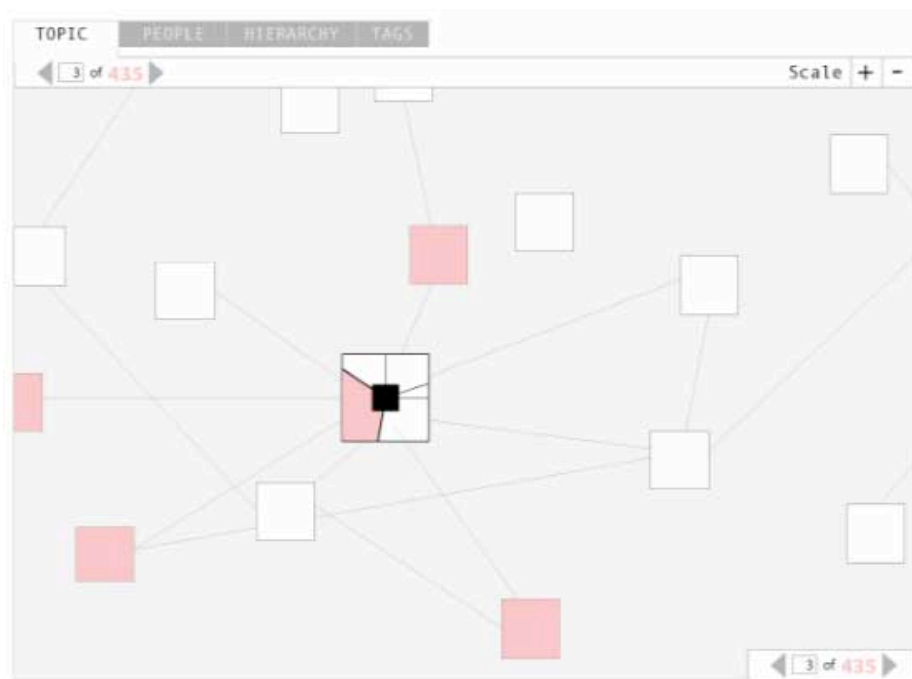


Figure 20. Relationship Map design concept

In Figure 20, the highlighted central thumbnail image indicates the central record, i.e. the record of interest to the user. The map connects content within this pertinent record to other records. Our objective here was to visually dissect the record into the different topics (or people, or places) it covers. The segments of the square depicting the central record represent the percentage of that record which addresses a specific topic. In this way, a researcher can understand visually the breakdown of topics within a record and make an informed decision about how much time to invest in exploring it.

Because this design requires additional labor-intensive cataloging processes wherein transcripts are segmented or chapterized and assigned topics and timecodes in order to create these percentage breakdowns, this functionality was not implemented. Based on this original design

concept, however, we built a dynamic radial graph that illustrates the relationships among and between records through the lenses of “people,” “topic” and “hierarchy” (collection context).²

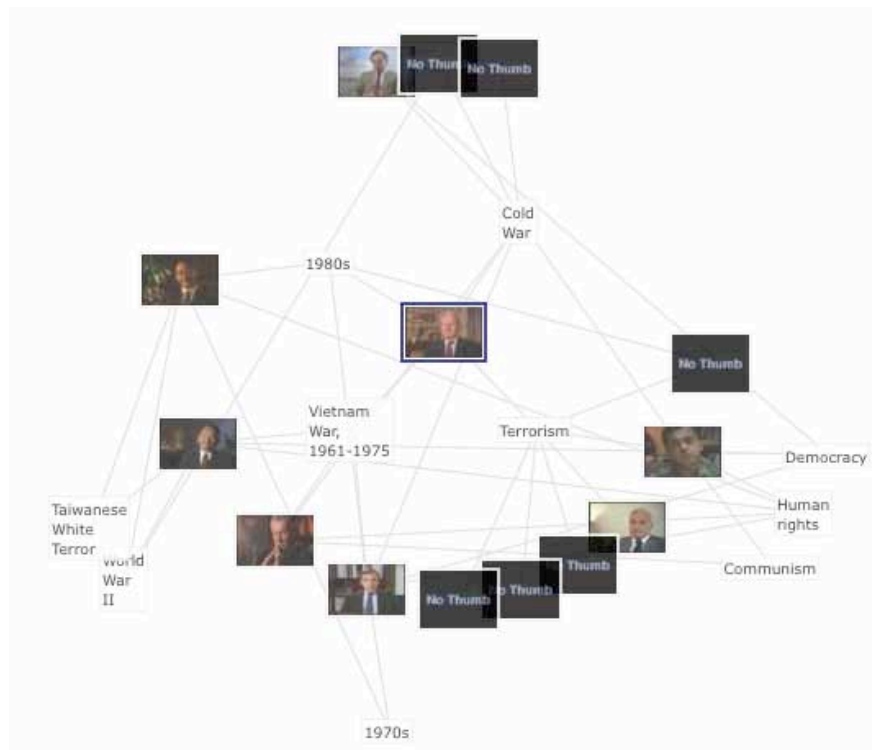


Figure 21. Radial Graph

In this example, the central square contains an interview with former National Security Advisor Robert McFarlane. Users can see that he discusses terrorism, Vietnam, the Cold War, and the 1980s in general. They can go one step further and see what other records in the collection touch on these same topics by following the lines connecting the central record to topic squares and related records. They can then click on each record (thumbnail image) or on a topic (word) to re-center the map dynamically, exposing new relationships to explore. Finally, they can also zoom in and out to discover records located farther afield. Dr. Winn was especially excited about this capability as he imagined using it to answer his research question – how did other filmmakers address the same topic in different productions? Using this tool, he can start with an *American Experience* interview with Robert McFarlane and jump to an interview filmed for *Americas*. From this data map, he can begin to tease out how public television producers in the 1980s and 1990s explored the idea of democracy.

A map like this requires intensive cataloging and thus presents challenges. For example, records are cataloged to varying degrees of completeness, skewing their relevancy and therefore their proximity values on the map. For example, a record with little data may not appear on the map at all, while a record that has been fully cataloged may connect to dozens of other records, making for a crowded and hard-to-use map display. In the latter case, we experimented with

² This graph was built using the JavaScript Information Visualization Toolkit: <http://thejit.org/>. We experimented with force-directed graphs and radial graphs. While the force-directed graph was able communicate relevance through proximity, we found that the radial graph, given our data, was more user-friendly.

automatically pruning certain ill-defined relationships in order to avoid overcrowding and overlapping data, but this too presents an issue as a researcher may mistrust the pruning mechanism. Further user testing of this feature in the future will help improve its usability and clarify its limitations.

Another challenge we have pinpointed regarding the relationship map is a lack of contextual information. The user often gets lost within the map. We need to better communicate the relationships conveyed by the connecting lines and the relevancy (expressed through the proximity) of one record to another. In some ways our challenges stem from too much data, but we hope they can be solved with continued experimentation and refinement.

We originally conceptualized the relationship map as both a browse tool and a homepage feature. However, we found that it tended to intimidate users as the first introduction to the site. The last thing we wanted to do was turn users away from the collection because they did not understand a map on the homepage, so we incorporated the map into the prototype in less prominent ways. Users encounter a link to it on the record page as well as on the search results page.

While this feature requires further implementation and testing to determine its utility, we believe that sophisticated data visualization techniques such as this will ideally save our users hours of time reading through pages of text. We found, however, that senior scholars, accustomed to spending many hours mining an archive, are intrigued yet skeptical of some visualization techniques. Nevertheless, we are excited about the opportunity this offers to expose our collection in a new way, and we anticipate that the next generation of “digital natives” will embrace these tools as part of their research, analysis, and scholarly production workflows.³

PROTOTYPE TECHNOLOGY

One of the goals of the prototype was to review and revise our technology infrastructure in an effort to make it more robust and scaleable.

In order to implement the interface functionalities described above, we embarked on a path of discovery into the repository world. Our original Open Vault website, based on CWIS, or “Collection Workflow Integration System,” has limitations we wanted to overcome in the course of this project. Therefore, as planned, we re-evaluated our CWIS repository software to determine its suitability for the prototype and future development of a scholarly archive. Our considerations were guided by our interviews with the scholars, who emphasized the importance of a strong and consistent metadata model that captured the complex relationships inherent in a media archive. While we expected the scholars to be interested in contextual metadata, they were also interested in the production of the media and its relationship to television episodes and series. This need for a rich content model reinforced our inclination to use Fedora as our datastore, but we also looked at Greenstone, KORA and DSpace (which was subsequently merged with Fedora in the DuraSpace project). We also considered building upon a content

³ Our work on this feature is further described in two articles: “Visualizing Media Archives: A Case Study” by Chris Beer, Courtney Michael, and Mayo Todorovic, *Code4Lib Journal*, Issue 6, March 3, 2009. <http://journal.code4lib.org/articles/1119> and “Visualizing Television Archives” by Courtney Michael, Mayo Todorovic and Chris Beer, *ASIS&T Bulletin* (Special Issue on Data Visualization), June 2009.

management system such as Django or Drupal. At the end of our deliberations, Fedora (fedora-commons.org) stood out because of its flexibility, XML-based storage, and track record in institutional repository and digital library applications.

As a next step, we contracted with Thornton Staples of the Fedora Commons organization and spent two days developing an initial content model that would allow us to describe the types of materials and structural relationships between them. Our content model was inspired by the atomistic design of the Functional Requirements of Bibliographic Records (FRBR), which separates basic descriptive and administrative metadata from the technical instantiation data. This allowed us to create a consistent search and discovery layer across all of our content, irrespective of the underlying content types. The process to develop this model was iterative and involved collaboration and discussion outside of this project, forcing us to think more broadly about the elements of media production, workflows, and how to align existing metadata with the needs of our audience. As a consequence, we expect this content model for media will prove to have lasting value beyond the prototype and beyond WGBH and the public broadcasting sphere. To this end, we have already shared the approach in several papers and presentations, as described more fully in the Dissemination section below.⁴

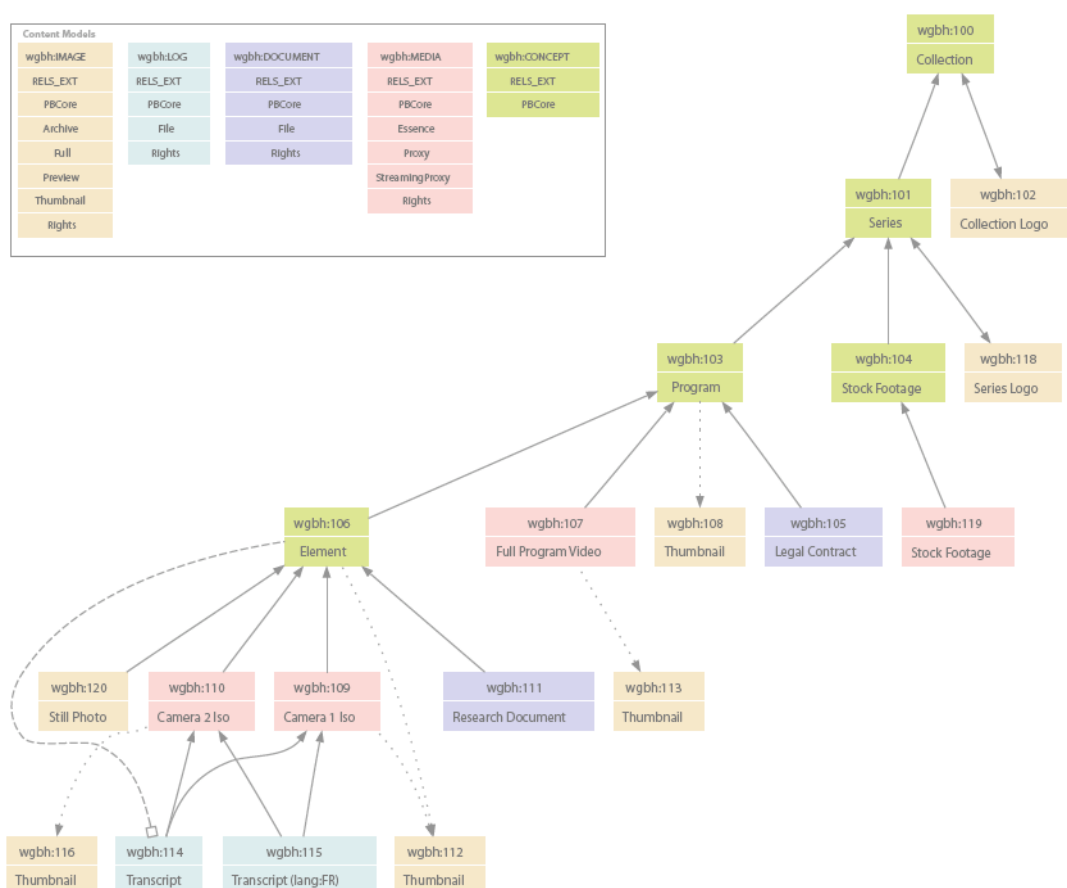


Figure 22. WGBH Digital Library Content Model

⁴ See our article on this process: "Developing a Flexible Content Model for Media Repositories: A Case Study" by Christopher Beer, Peter Pinch and Karen Cariani in the Proceedings of the Joint Conference on Digital Libraries, June 2009

Inspired by best practices (such as FRBR), we leveraged existing standards in public broadcasting and added standards from other fields as necessary. As a result, our prototype repository stores both PBCore and Dublin Core descriptive metadata.

The scholars' focus on transcripts led us to use the robust TEI (Text Encoding Initiative) XML mark-up standard for encoding them. Not only did this allow us to enhance our transcripts with standardized metadata (such as place names, geography, etc.), but we also combined TEI with elements of SMIL (synchronized multimedia integration language) to allow for real-time synchronization of the transcript with the video.⁵ This mark-up has been used to support one of our key features: the ability for users to search across transcripts and play back video segments synchronized to search results.

We knew from the start that search was a key feature, and we felt that a faceted search interface would best meet our users' simultaneous needs to search for relevant content and understand the scope of the collection. Consequently, we chose to use the Lucene-based Solr enterprise search application (<http://lucene.apache.org/solr>) because of its demonstrated integration with Fedora and a rich feature set that included full-text search, relevance ranking, and similarity metrics. We were very pleased with the results, finding that this open source software was both powerful and relatively easy to configure. (Like Fedora, Solr offers web-service interfaces, which makes integration with other tools much easier. Solr is now being adopted by many web projects at WGBH.)

When it came to the user interface components, our rapid development cycle and lack of a clear best option demanded a custom approach. At the time (2008), many of the established Fedora-based products were developed for the Fedora 2.x release ([Muradora](#), [FEZ](#) and the [Fascinator](#), among others). While a number of nascent Fedora 3.x interfaces were emerging (including the Drupal-based [Islandora](#) project developed by the University of Prince Edward Island for institutional repository projects and the Ruby on Rails-based [ActiveFedora](#) project developed by MediaShelf for digital library projects), none were at a suitable phase of development (and documentation) to employ. Instead, we chose to develop our own prototype interface, bVault. It is a PHP application based on the [Zend Framework](#) that accesses the web-services interfaces of the underlying software. Some key user interface features, including video segmenting and annotation, were implemented using the [jQuery](#) and jQuery UI javascript libraries but tightly coupled to the original development.

Code for bVault is available under an MIT-style Open Source license at <http://launchpad.net/bvault>. A list of open-source resources that guided the development of this prototype is included in Appendix B.

⁵ We explored several workflows for synchronizing the video with the transcripts and experimented with a fee-for-service model with Ramp.com. (Formerly "Everyzing", this services is used by NPR for automated, close-to-real-time radio news synchronization). We concluded that adding in paragraph-level time code by hand using SMIL was more cost effective for this project. Developer Chris Beer wrote tools to help us quickly synchronize the media to the transcript for this project, a tool we continue to use in our *Vietnam* project. This tool is available at <http://github.com/cbeer/ave-sync>. In the future, however, automated speech/text recognition may facilitate this process and would further media search technologies tremendously. WGBH is in the market for a free and open source answer to this challenge and would willingly provide media for experimentation towards a solution.

TECHNOLOGY EVALUATION

After the initial development, evaluation, and dissemination of the prototype and its discoveries, the team conducted an additional landscape review in late 2009 in preparation for this grant's final report and our ongoing work in support of the *Vietnam* project. Fedora continues to be a strong choice, having proven it provides both sufficient and flexible architecture while backed by a sustainable and growing organization. Solr released Solr 1.4, which increases its performance and feature sets and is also a sound choice. While other similar projects have advanced, matured, or are being developed, the existing prototype infrastructure is adequate, appropriate, and maintainable.

Our review of appropriate user interface applications a year later showed new development and maturity in the digital library communities compared to when we started in 2008. Our user interface, as mentioned, was a custom development effort. While it has seen deployment at Columbia University and has attracted interest by other organizations, it lacks a solid community for sustainable development in the future. However, developing it was key to our ability to work with our scholar advisors in demonstrating the functional requirements of our media archive. In the interests of creating the best possible conditions for sustainability, we have also begun evaluating new projects that have emerged from the digital humanities and library technology fields. The remaining new applications can be divided into two sections, those designed for digital humanities/digital libraries and those designed as next-generation library catalogs. A complete list and discussion is provided in Appendix C. We are particularly excited about Project Blacklight, which we plan to incorporate into Open Vault 2.0 on top of Fedora in Spring 2010.

FEEDBACK AND REVISIONS

We launched our prototype site in November 2008 and invited our scholars to use it and formally report back to us in Spring 2009. We then enlisted the help of the WGBH Research department to conduct user testing sessions with a group of less seasoned yet more computer savvy scholars. We were also fortunate to be able to visit one of Dr. Dower's "Visualizing Cultures" classes at MIT where we presented our prototype and gathered the students' feedback. From these three sources – our scholars, user testing, and the MIT students – we collected reactions and critiques of the prototype design and functionality in order to implement revisions and improve our product.

SCHOLAR FEEDBACK

All of our scholars agreed on the importance of incorporating media content into the work of historical research, and all came to realize the wealth of material in the WGBH archive. Although they may have suspected this in the past, actually working with the materials made the notion clearer and more urgent. In addition, they clearly saw the need and value in developing a prototype that could allow easier access to this material. They all enjoyed the process of helping us develop this prototype and felt the WGBH team listened to their needs and responded throughout the development.

While they enthusiastically supported the project's goals, specific criticisms led us to make revisions to the project's design. An initial concern was our early choice to use the relationship map as the homepage. While the scholars saw the ultimate value in the map, all three found it confusing as an introduction to the collection. In addition, the scholars stressed the importance of being able to easily and reliably search the collection through different metadata, search functions, and browse capabilities, and this feedback guided additional changes to the design.

Finally, the scholars' expectations regarding content changed over the course of their testing. At first, the raw materials seemed to satisfy their needs, but by the end they also wanted access to the final programs as a short cut for searching out what else might be potentially useful in the rest of the collection. As a consequence, we are considering the rights and cost impact of adding these final programs to the materials available in future iterations of the project. In the meantime, we added functionality to direct users to their local library's copy of the finished films.

For further detail on the scholars' responses, please see Appendix D.

USER TESTING

The WGBH Research Department helped us to conduct user testing sessions with seven local scholars at the beginning of their careers. A mixture of graduate students and new faculty ranging in age from 28 to 37 spent 45-50 minutes each discussing the site and providing invaluable feedback on its navigation and layout. We again heard that the relationship map on the homepage was not helpful. We also heard that our faceted navigation was most useful when the facets were located on the left side of the page (rather than across the top as we had previously designed.) Finally, this group of users gave specific feedback about the types of metadata displayed on the search results pages, helping us to better prioritize our information hierarchy on that page.

Overall, their response was enthusiastic. They were excited to have both the full transcript and the video side by side – one exclaimed “This is the meat AND the potatoes!” Another participant valued the relationship map as it allowed him to navigate from one found record to other relevant records. He testified that his visits to archives, physical or online, are made all the more useful when he finds what he was originally looking for but is also surprised by new discoveries.

A new history faculty member appreciated the achievement of our prototype in presenting volumes of information in an organized manner. He summarized the unknown wealth of archives in the following way: “A good archive is kind of like the iceberg. What comes to us in memory is this little bit, and there is all of this other stuff you can see in the archive – in all of its glory and all of its terror. [...] You guys have beat a lot of other things in terms of being the scholar and the student's friend in terms of the information. [...] Consider me a groupie!”

For further detail on this focus group's response, see Appendix E.

MIT “VISUALIZING CULTURES” CLASS

Drs. John Dower and Shigeru Miyagawa invited us to attend their “Visualizing Cultures” class to share the process of creating a digital library of visual materials with their students. Composed of graduate and undergraduate students from all majors, this class of “digital native” generation

students were able to immediately understand, use, and critique our design. Our WGBH team was both thrilled and surprised at their level of intelligent feedback. The experience helped confirm our suspicion that future generations of scholars will take for granted this type of technology and this level of access to archival moving image materials.

After we presented the prototype, the students completed an assignment where they spent time on their own exploring the site. Echoing many of the improvements suggested by both our scholars and through the user testing, these students added new ideas for future implementation. For example, they appreciated the relationship map but wanted to curate and create their own maps between assets. They commented that their professors do not generally use video in their classes but that as media becomes further integrated into scholarly workflows, scholarship will become increasingly rich and able to convey complex topics in new and different ways.

The revisions suggested from all three sources are implemented in the current prototype website. They are also informing the design of Open Vault 2.0, and the Fedora repository architecture developed for this project will be refined and will undergird Open Vault 2.0. The enthusiasm for our project displayed by all three user groups was encouraging for our project team and has helped us build a small but dedicated group of stakeholders to whom we will turn for future projects in the scholarly communications space.

SUSTAINABILITY

In order to bring this work to the next stage, WGBH worked with a business consultant to help develop a sustainable business plan. After considering a number of companies, we chose to work with Ithaka Consulting due to their strong background in working with non-profits and the academic community (our target audience), their enthusiasm about the project, and their immediate understanding of our needs and potential constraints.

Ithaka was tasked with providing guidance on how best to sustain the digitization and delivery of WGBH's valuable archival content. After several months of research, they delivered a memo outlining the landscape for digital video in research and teaching based both on a review of the literature and a survey of currently available content. In this report, they focused on presenting a detailed description of scholars' needs relating to this content, outlining the benefits and challenges of different revenue-generating models and exploring different organizational models that can support this work. Their conclusions are as follows:

Scholars' Needs: There is some demand for high-quality digital archival footage for both research and teaching purposes, and the digitization and indexing of this content could go a long way to lowering current barriers to use-related to challenges in access and discovery. There may also be other ways to expand the scholarly value of the collection—for example, by posting and promoting a detailed finding aid, or by establishing closer ties between unedited footage and final broadcast programming. However, it is important to acknowledge that this kind of content is not yet seen as a “must have” by the academy, and that no single archive, particularly a curated digital collection, can meet the full range of highly-specific research needs.

Revenue Generation: The high cost of digitizing and indexing moving image content means that it will be critical to generate outside sources of revenue to continue this work. This will be a challenging prospect. Though WGBH's archival footage could prove valuable to a wide range of scholars, it seems clear that demand for research-related uses of a single archive will probably be insufficient to drive sustaining revenue. Digital video is already widely used in humanities and social science classrooms, however. The appeal of this material for use in teaching could help drive adoption by academic libraries, providing the revenue needed to fuel new content creation while also enhancing the visibility of the collection's applications for research. Analysis suggests that shaping WGBH's archival footage into a product targeted to the academic library market will provide the best chance to generate meaningful ongoing income; however, this option would also create significant new costs and high expectations around content and functionality. Furthermore, the current challenging economic climate has left academic libraries with less available money to invest in new products. Still, Ithaka recommends that a focus on this market provides the best chance for success.

Organizational Models: Taking on the full responsibility of shaping WGBH's archival footage into a product for the library market would effectively require starting a small business as a digital resource publisher. It would necessitate the addition of new skills and expertise not currently possessed by the WGBH team, and it would create high expectations among the user community around quality, service, and functionality. As a consequence, WGBH may want to consider ways to partner with other organizations to outsource certain activities related to the packaging of WGBH's archival footage while also planning what the future might look like if no suitable partner can be found.

Given the above conclusions, Ithaka outlined three options for consideration by the MLA.

Option 1: Partner with a Publisher of Digital Resources.

One option would be to work with a publisher of digital resources to outsource many of the functions associated with creating and delivering digital video to a scholarly and academic audience. The content would benefit from the partner's established expertise in developing products that are attractive to libraries and useful to scholars and students, helping to achieve the central mission of making the MLA collection more useful for research and teaching. The raw footage would be published alongside other relevant and related materials, enhancing both the value and discoverability of the content. In addition, because (in most arrangements) the publisher would take on much of the expense of creating and updating the content and the platform as well as marketing and sales, this option would eliminate most of the direct costs that would otherwise be incurred by WGBH—and might, in fact, provide some modest amount of royalty revenue that could be channeled to other initiatives.

Option 2: Offer Subscriptions to a Self-Published Product.

Another option involves offering a self-published, packaged version of MLA content through institutional subscriptions. In this model, we would need in-house capacity to perform all the essential functions related to project management, content development, technical development, and revenue generation. Meeting the library community's high expectations around content updates, platform enhancements, and user support would be expensive, and we would need to invest significant resources to add capacity and expertise in sales and marketing. If we made

these investments, it is still not clear that there is sufficient market demand to support a product built entirely with MLA content. Although preliminary research indicated that offering a subscription-based product may provide a reliable and predictable revenue stream for streaming digital resources, further analysis of both the costs of supporting such an effort and the current market opportunity for WGBH's particular content suggests that this would be a risky approach to adopt.

Option 3: Support an Open Access Resource

A third option would be to offer an open access version of Open Vault Research based on what we feel we could sustain through grants, sponsorships, and institutional support. Without the need to meet the higher demands of paying customers, the content creation and technology development would slow to a much less expensive pace and eliminate many costs around sales, marketing, and user support. This model has several other benefits—in particular, it fits with organizational inclinations toward open access and full control over the intellectual assets of the archive and the platform. However, the success of this model would depend on the ability to control costs and attract support from WGBH and PBS administrators, grantors, and sponsors in alignment with the project's mission and goals; if support in these areas were to wane, sustaining online access to the resource might be difficult.

Ithaka's full recommendations and results are contained in their final report included as Appendix F.

OUR CONCLUSIONS REGARDING SUSTAINABILITY

The WGBH MLA's Deliberations

Of Ithaka's suggestions, a combination of Options 1 and 3 seem the best solution for our circumstance. We do not feel Option 2 is viable for us because the costs of ramping up our capability to market the service along with the costs of building a broad enough offering are too high and too hard to offset.

	Year 1	Year 2	Year 3	Year 4	Year 5	Totals
Cost	\$283,000	\$373,000	\$477,000	\$529,000	\$532,000	\$2,194,000
Revenue	\$0	\$109,000	\$218,000	\$381,000	\$545,000	\$1,253,000

Based upon the chart above, the project would reach sustainable levels (annual cost < annual revenue) at year five when costs start to level out and revenue is increasing. Please note that the estimated revenue (based on discussions with vendors) shows no revenue for first year (which will be devoted to set up and marketing). However, the break-even point of overall cost = overall revenue, is much further in the future which would require too long-term an institutional investment for an uncertain return in a changing technology environment. In addition, because the WGBH collection is limited in scope, it is unclear whether it alone could generate enough interest to build sufficient revenue, and unfortunately this question cannot be answered definitively without the investment of the start-up costs. (To view the full cost model, please see Appendix G.)

The appeal of Option 1 is that it allows us to partner with an organization that already has the

infrastructure, clients, and market expertise for content delivery, thus placing most of the burden of running the business on the vendor. However, a remaining challenge for us is how to deliver that material to the vendor. This would take significant work on our part due to digitization and rights clearances, activities that to date have been grant-funded. Since a condition of much of our grant-funded work requires that the material be made available for free, we could not use past work and would have to directly invest institutional resources in this production. Further, the requirement of free access aligns strongly with the mission of WGBH as well as public media in general.

Option 3 reflects our current practice, which unfortunately correlates to very slow growth and uncertainty for sustainability far in the future. However, this option fits best with our public media mission and would allow us more flexibility in making materials available for free. In addition, as WGBH is very committed to the American Archive initiative (see below for more detail), we do not want to commit our content to an exclusive vendor.

For Option 3, the costs to sustain Open Vault activity would be about \$200,000 annually. This would include digitizing content, continuing to put new materials online, and some new technological updates. This does not include marketing, promotion, or curating special collections targeted for revenue generation. One possible opportunity however may be to create an endowment that can cover costs of sustaining Open Vault along with grants, sponsorships, and institutional support. An endowment of \$10,000,000 would enable us to add materials to the collection on an annual basis while we continue to seek supplemental grant funding. The endowment funds could also help with grant matches. While activity and additions to the site could wax and wane with our success in pursuing grant opportunities, the endowment would nevertheless allow a sustained rate of progress. The costs associated with building an endowment would mostly be absorbed by WGBH, although supporting funds for marketing, outreach, and promotions would be helpful in raising the profile of our resources.

The WGBH MLA's Conclusions

After various discussions within WGBH, our inclination is to explore a hybrid of Options 1 and 3. Through this hybrid approach, we will provide access to content for free yet still reap the benefit of a revenue stream by partnering with a commercial vendor to create a subscription-based limited collection. This dual approach reflects that we are committed both by mission and by some granting agencies to keep materials available at no cost, yet the ability to generate sustainable revenues is difficult. Consequently, our next step will be to build a pilot project that experiments with this dual-path approach. Such a pilot will provide a crucial opportunity to build trust with a vendor and test the notion that allowing free access will not impede the potential for subscriptions.

A model already exists for selling subscription-based collections containing media to academic libraries. We are currently in conversation with the two commercial vendors in that space, and while they are wary of representing the materials non-exclusively, they both have indicated a willingness to explore new possibilities. Consequently, we are working to see if we can build a model that satisfies both partners' needs – free availability and revenue growth.

The American Archive

As we move ahead with this process, other developments may influence the path we take as well.

Since we started this project, the American Archive Initiative at the Corporation for Public Broadcasting (CPB) has gained momentum. This initiative promises to create a CPB-funded national archive for public television and radio materials. It plans to focus first on the preservation needs of public media stations and later on access for educational and public use. A pilot project has just been completed with 25 stations to test the level of effort and cost needed for full-blown implementation. In addition, CPB has recently hired an Executive Director, and it is actively posting RFPs to move the initiative along.

As the premiere producer in the public media system and the steward of public broadcasting's largest, most active archive, WGBH will pursue opportunities to be an involved partner and leader in the American Archive. As a consequence, we are hesitant to pursue commercial revenue options that may inhibit our ability to fully contribute content to the initiative. However, we also do not want to delay our own plans to move forward with the Open Vault Research digital library, particularly when our efforts in this area have served as a model for the American Archive. Since the Archive's initial focus will be on preservation rather than access, we can further contribute to solving the challenges that they eventually will face while preserving a scholar-defined focus that their collection, targeted to a broader audience, will lack. In addition, the future success of the American Archive will provide a pool of new resources that can populate Open Vault Research.

DISSEMINATION

Disseminating our work has possibly been the most rewarding activity of this project. In addition to receiving excellent constructive feedback from various academic communities, our papers and presentations have opened new doors for the Media Library & Archives, allowing us to build new ties with external institutions and experts and broadening the audience for MLA activities.

Publications and Presentations

Over the course of this project, MLA staff have published four papers and given eleven conference presentations. A complete list of papers and presentations is included in Appendix H.

Particular highlights include the February 2009 Code4Lib conference, which focused on technology solutions for libraries and cultural institutions. Project staff Chris Beer and Courtney Michael presented the relationship map from the project prototype, illustrating its function and potential by creating an association map of attendee profiles. This presentation was later revised into an article published in the Code4Lib journal.⁶

We also presented our work at the 2009 Annual International Conference on Open Repositories.⁷ In addition to speaking on our efforts to expose content to scholars, we presented a poster on the

⁶ "Visualizing Media Archives: A Case Study" by Chris Beer, Courtney Michael, and Mayo Todorovic, *Code4Lib Journal*, Issue 6, March 3, 2009. <http://journal.code4lib.org/articles/1119>

⁷ "Disseminating Broadcast Archives: Exposing WGBH Materials for Scholarly Use" Chris Beer and Courtney Michael, <http://hdl.handle.net/1853/28411> and "Opening the Vault: Providing Digital Access to Archival Moving Images", <http://smartech.gatech.edu/dspace/handle/1853/28410>

project that was honored with the Best Poster prize. This conference had the added benefit of connecting us with new contacts at Glasgow Caledonian University, where they have a Fedora repository that manages spoken word recordings and BBC materials. This relationship could result in a transatlantic project experimenting with international repositories and shared data.

A presentation at the Joint Conference on Digital Libraries titled “Developing a Flexible Content Model for Media Repositories: A Case Study” was met with great interest and published in the conference proceedings. In addition, a presentation at AMIA sparked an enthusiastic response to our discussion of open-source solutions for online access to media collections. As a result of this and WGBH’s leadership involvement (see below under Meetings Convened), the AMIA now has an Open Source committee.

Meetings Convened

As a result of our active participation in conferences and other activities, we have become more involved in the community and have spearheaded discussions to further our shared work and explore opportunities for collaboration.

In January 2009 WGBH hosted a meeting for moving image archives interested in open source technology solutions. Karen Cariani convened the meeting, which included an overview presentation by Phase II project developer Chris Beer, “Repositories: What Are They And What Are They Good For?” We also heard from Seth Kauffmann about CollectiveAccess and from Dave McCarn about PBCore. This group met again at the AMIA conference in November 2009 to discuss how AMIA members could interact and inform the open source community about content management solutions needed for moving image archivists. The meeting had over 30 attendees from a broad range of archives and has now become an official committee of AMIA. In order to broaden the discussion, Karen Cariani is currently trying to bring the media archive community in Duraspace together with this newly formed committee of AMIA because they have a common purpose and mission.

The WGBH MLA also hosted a lunch in March 2009 for local Fedora users. Attendees included Ari Davidow and Andrea Medina-Smith of the Jewish Women’s Archive and Rob Chavez and Ted Blizzard of the New England Journal of Medicine. Each organization shared information on their Fedora repositories and other open source technologies they have incorporated into their projects. Primarily a networking and information sharing gathering, this group met again in October 2009 with the addition of Anoop Kumar of Tufts University.

Conversations towards Future Collaborations

One of the first conversations we had was with Dr. Dower’s Visualizing Cultures (VC) Team at MIT. At our meeting with Dr. Dower, co-director Dr. Shigeru Miyagawa, and Program Director Scott Shunk, we discussed ways we could continue build upon our Phase II work and provide access to our content for the VC project. Drs Dower and Miyagawa have primarily focused on the visual history of Japan and China and would very much like to begin to incorporate WGBH materials from Vietnam and Taiwan. Their site offers not only a curated aggregation of historical materials, but also narrative, “synthesis”, and historical analyses of the collections. For the VC team, this project has offered a glimpse into a repository of moving image materials to mine for further intellectual work on their site and it has offered workflow tools for their process of

gathering and culling materials. For WGBH, the use of our materials by MIT scholars (and others) would provide us with the opportunity to continue analyzing how best to serve our academic audiences. In addition, further work could be done through providing access to our content for K-12 educators through VC's teacher connections and WGBH's *Teachers' Domain* project. The next step in this collaboration is to cost out the possibility of making our China and Taiwan materials accessible online and to provide Dr. Dower with our *Vietnam* materials when they launch in 2010.

In a conversation with Linda Frueh, Regional Director for the Internet Archive, we explored opportunities to collaborate around grant funding, storage, and digitization. In addition, we discussed the possibility of cross-linking with materials on the Internet Archive and having the IA conduct "curated crawls" of WGBH websites to archive them and create special collections. Most pertinent to the Phase II project was discussion of an extension to the Zotero research management software which would create a scholarly information commons where scholars could drag links into Zotero and have them automatically archived by the Internet Archive. This kind of work would help address scholars' concerns that web materials are impermanent. While the Internet Archive could be one solution for this challenge, educating scholars about permanent URLs, DOIs and trusted digital repository policies could also help alleviate anxiety and promote creating special collections through venues such as YouTube.

Conversations for Information Sharing

We also spoke with Dr. Mark Williams and Anthony Helm from Dartmouth College about Project Bamboo. Similar to the Visualizing Cultures project, their interest is in creating scholar-curated collections, something our current prototype does not do (except for personal use). They introduced the plan and objectives of their project, and we were able to demonstrate our prototype for them. We found that we share the common goal of allowing scholars to use, annotate, cite, and aggregate multimedia source materials online, and we face the common challenges of authentication, video quality, and user metrics for scholarly access. In addition, the idea of aggregating digital objects from external repositories and creating an "action" layer (ability to annotate, clip, etc.) on top of these collected resources poses an additional interesting challenge. Since then, we have had follow up conversations and are continuing to explore opportunities to share knowledge. In addition, our relationship with Dr. Williams has been especially productive as he will serve as an academic advisor for our planned IMLS-funded Boston Local News Archive.

Additionally, we spoke with Nancy Silver and Andy Maltz from the Academy of Motion Picture Arts & Sciences Film Archive, who are experimenting with Fedora-Commons as an interim preservation repository for their feature film archive. Through this conversation, we learned that while our project deals with using Fedora as the back-end for a website that provides access copies, their project focuses on preservation of high-resolution files. As a consequence, we encouraged them to speak with NYU, one of WGBH's partners in an NDIIPP project to preserve digital television that has faced similar challenges (though it uses a DSpace repository system rather than Fedora). We also encouraged them to explore ODRL as a metadata standard for their rights information, but they have not yet tackled that issue.⁸

⁸ See *The Digital Dilemma: Strategic Issues in Archiving and Accessing Digital Motion Picture Materials*. The Science and Technology Council of the Academy of Motion Picture Arts and Sciences. Beverly Hills, Calif. :

We also met with Matt Zumwalt of MediaShelf, who offered some tips as an experienced Fedora developer and well-connected community member. We discussed our challenges with large file sizes and Matt informed us that Fedora is looking to separate out the file storage layer of the architecture. He suggested we speak with Sun Microsystems (who previously developed Fedora-compliant hardware) and Dave Tarrant at the University of Southampton who works on JISC funded repository preservation projects. We also discussed the BagIt standard in development at the Library of Congress (for packaging materials and metadata and delivering them as a “bag” to repositories) and our need for authentication, security and transcription solutions.

Overall, these opportunities to exchange knowledge and ideas with others working in the field have been invaluable to our understanding both of the contributions we can make and the benefits we can share through collaboration.

CONCLUSION

As a result of this project, WGBH is well positioned to move ahead launching a full-scale version of Open Vault Research. Through Phase II’s development of an accessible archival prototype, we addressed all of the identified challenges in making media materials available online for scholars, including exposure of archival collections, access to materials, and technical challenges. We also successfully achieved the project goals through the following outcomes.

Goals	Outcomes
WGBH will have a model system created to meet the needs of scholars and educators who seek online access to the MLA’s media materials.	The Open Vault Research prototype meets these needs, and elements of the prototype will be further integrated into Open Vault 2.0, scheduled to launch this spring.
WGBH will develop a content delivery system built in open-source software that can potentially be utilized for other projects.	WGBH’s Fedora- and Solr-based solutions to this project’s technical challenges will also guide other WGBH projects including <i>Vietnam</i> and <i>Teachers’ Domain</i> , and they have been explored by projects beyond WGBH as well.
WGBH will draft a preliminary business plan that will enable it to move forward with the Phase Three implementation of the Digital Library.	A business plan developed in partnership with Ithaka includes recommendations for moving ahead and a complete landscape review (Appendix F).
WGBH will map cost projections for expanding the system and adding additional content.	Cost projections show that the most viable solution may combine a revenue-generating publishing partnership with the pursuit of institutional, foundation, and sponsor support for broader access.

Academy of Motion Picture Arts and Sciences, 2007. Also, please see a presentation by Nancy Silver, “Building a Digital Preservation System for Digital Motion Picture Materials,” at the 2009 NDIIPP Partners Meeting, Session #4 Digital Preservation of Dynamic Content.

WGBH will have knowledge of the types of program elements, content, and media from the archive that are useful for scholars who specialize in subject areas covered by its history programming.	Information collected from our partner scholars has identified resources of interest as well as the types of content and level of metadata necessary for them to work effectively in the collection.
---	--

LESSONS LEARNED

WGBH will continue to disseminate key lessons learned and participate in communities working towards the same objectives with the goal of increasing institutional knowledge and seeking out collaborative efforts. Our innovations in this space have contributed a new model for accessing media archives online.

Some crucial points that we can pass on include:

- **Video takes time.** To increase the use of media in scholarly work we need to help scholars reduce their time invested and increase their rate of discovery.
- **Metadata is expensive and essential.** However, if full metadata is not an option, it is still beneficial to publish what is available. Context is also essential for interpretation of these visual and audio assets, and this can be expressed both visually and in metadata.
- **User-generated content is not simple.** It is important to differentiate and segregate visitor contributions from authoritative metadata, yet still make it useful and an opportunity for sharing knowledge.
- **Finding aids are not the only way.** There are other useful methods of access for an archive. For example, facets can provide the same or similar context, but they require deep cataloging.
- **Visualization** is both exciting and daunting for humanities scholars. Linked data makes it possible, but further research is needed.

From our work with scholars, we have learned that the use of digital media in scholarly research is still evolving and in its infancy. The availability of such materials is still scarce and therefore practice is just emerging. Working closely with a few researchers highlighted the essential need for good cataloging and rich metadata as a precursor to this scholarly use. From our concurrent *Vietnam* project with Columbia, however, we have also learned that teachers advocate the use of digital media much more passionately than researchers. By the time those who are being taught become the next generation of researchers, we hope the many issues we face will be resolved, and we also hope that our work will contribute to resolving them.

IMPACT

The community of digital library builders has acknowledged WGBH as media experts among universities and repositories and all have an interest in the expertise we have developed, from rights issues to digital media file formats.

As a result of our Phase II work, we have been able to further develop and share our expertise in the following areas:

- WGBH staff have learned more about the scholarly and pedagogical needs in higher education, and how the MLA's resources can best be utilized there.
- WGBH has raised awareness among scholars and educators about the potential of moving image and audio media as primary resources for research and teaching.
- WGBH has crafted a prototype and collected valuable information on how to further develop Open Vault for use by scholars and educators.
- WGBH has shared both its software and conceptual models for the project with other institutions, and we will bring this knowledge to inform our involvement in the American Archive project. (As noted above, see Appendix B for access to the project's source code and other resources.)

This project has also allowed us to hire a developer to focus on the Digital Library's specific technical needs for the prototype while exploring and developing open source opportunities. WGBH now has that capacity and we will work hard to keep it. On a technological front, WGBH successfully modeled a repository using Fedora and gained experience working with other open source solutions. We feel we can easily utilize Fedora for other projects at WGBH, so this experience has opened new opportunities organizationally. In addition, we are one of the few Fedora projects to build a content model for media objects. As a consequence, we have shared this model with the community through bVault (<https://launchpad.net/bvault>).

As noted above, the WGBH Open Vault Research prototype is available at openvaultresearch.wgbh.org, and we will continue to share and answer questions about the project. Many of the technical solutions and features developed for the prototype are being incorporated into Open Vault 2.0, which will feature the *Vietnam: A Television History* collection. This can be accessed at openvault.wgbh.org after the launch in Spring 2010.

We are thrilled that our work from this project has informed and inspired other projects, both within and without WGBH. *Teachers Domain*, WGBH's K-12 digital media library, is contemplating utilizing a Fedora repository for their resources. And, in addition to employing it for the *Vietnam* project, Columbia University is considering using this model for their overall digital library efforts to handle video in their collection.

Further, we anticipate the impact of this work will continue to spread – several other media archive projects have been looking at our content model and Fedora solution. In addition, we plan to fully develop and expand our own prototype in the next phase of the Digital Library. The tools we built in response to the scholars' needs will also be used in Open Vault, and most of the media digitized for the pilot will remain in Open Vault for public access as well. Moreover, we are integral to the further development of PBCore through funding from CPB, and we hope to be further involved in the American Archive initiative, thus utilizing our experience to help other archival media projects.

Our Digital Library prototype has also fueled new relationships with the open source community and the digital library community, as we have learned from their approaches to shared challenges

and contributed our own solutions. Inspired by open source's possibilities for archivists, we have been active in developing an open source committee within AMIA, and we also are working to bring the Duraspace community together with AMIA to share ideas and solutions.

FUTURE PLANS AND OPPORTUNITIES

Through Phase II, we have learned that there is increasing awareness of what is in the WGBH collection and how it can be used, but we have also learned that users need to know more – more of them need to know that we exist, and they need to know what we have easily. This is a great moment to respond to that demand, and we hope to capitalize on the investment that WGBH and the Mellon Foundation has made in identifying opportunities for our collection.

Working with Ithaka and drawing on our experience, we have been able to cost out specifically what it takes to support a scholarly collection on Open Vault and consider different models for sustainability. Ithaka's cost model will be extremely useful for us for future planning, whether it is for revenue-generating options or for grant budgets. However we are at a crossroad in determining if we can stay true to our mission of free access and still fuel the activity. We feel there is a definite conflict and tension between the revenue generating publishing model and the American Archive mission-driven access model, and we hope that we find success in creating a middle ground.

Our work on this project has guided our vision regarding the landscape of what is needed in the world of digital libraries and archives and what WGBH can help develop. As a result of our work, we have identified additional opportunities for progress in the field that would benefit moving image archives and the scholarly community as a whole.

- Development of an open source media digital asset system (or solution using the cloud) that is a trusted preservation repository TDR Open Source DAM. Work remains to be done in creating an integrated sustainable solution in order to reduce DAM costs for small archives.
- Development of an open source media rights tracking system that fits into the system above in order to support access and reduce costs.
- Collaboration with universities to build tools that allow teachers/scholars to curate media from different collections, thus promoting the further use of digital media archives in research and education.
- Testing user-generated metadata as a low-cost solution to the cataloging problem. Inviting user-tagging to supplement a bare-bones catalog could offer a valuable approach to the labor-intensive cataloging of moving image collections.

Funding permitting, our own priority for continuing work in this field is to leverage our progress by implementing the prototype in Phase III of the Digital Library. While we have referred to the end product of Phase II as a prototype, it is already a fully functioning and accessible scholarly resource – all that remains is to populate the framework with content. We are aware of the challenges and expenses that accompany this, but we are heartened by the quality of the tools we have designed and the enthusiastic responses from our scholar/testers. Consequently, we will aggressively pursue opportunities to build out the site in our efforts to create critical masses of resources that can sustain scholarly interest.

OUR THANKS TO THE ANDREW W. MELLON FOUNDATION

On behalf of WGBH and the Media Library and Archives, we would like to extend our thanks to the Andrew W. Mellon Foundation for its support of the Digital Library's Phase II project,

Developing a Public Television Content Delivery System for Academic Institutions. We are grateful for the opportunity to undertake this work and we hope you share our pride in our progress towards creating a digital public media archive for scholarly use. We look forward to pursuing opportunities to continue our efforts, and we welcome future occasions to collaborate with the Foundation and/or its grantees in creating new structures for opening archival resources.

BIBLIOGRAPHY

WORKS CITED

“Building a Digital Preservation System for Digital Motion Picture Materials” presentation by Nancy Silver, at the 2009 NDIIPP Partners Meeting, Session #4 Digital Preservation of Dynamic Content.

Beer, Chris, Courtney Michael, and Mayo Todorovic. “Visualizing Media Archives: A Case Study” *Code4Lib Journal*, Issue 6, March 3, 2009. <http://journal.code4lib.org/articles/1119>

Beer, Chris, Courtney Michael, and Mayo Todorovic. “Visualizing Television Archives” *ASIS&T Bulletin* (Special Issue on Data Visualization), June 2009.

Cariani, Karen, Christopher Beer and Peter Pinch. "Developing a Flexible Content Model for Media Repositories: A Case Study" in the *Proceedings of the Joint Conference on Digital Libraries*, June 2009

Dulabahn, Beth, Phil Michel, Barbara Natanson, David Reser, Michelle Springer, David Woodward, and Helena Zinkham. October 30, 2008. “For the Common Good: The Library of Congress Flickr Pilot Project” Library of Congress Prints and Photographs Division. www.loc.gov/rr/print/flickr_report_final.pdf

Hoffman, Gretchen L. 2009. “Applying the User-Centered Paradigm to Cataloging Standards in Theory and Practice: Problems and Prospects.” In Jacob, Elin K. and Kwasnik, Barbara, Eds. *Proceedings 2009 North American Symposium on Knowledge Organization* Vol 2, pages pp. 27-34, Syracuse, New York. <http://dlist.sir.arizona.edu/2629/>.

Schaffner, Jennifer. 2009. “The Metadata is the Interface: Better Description for Better Discovery of Archives and Special Collections, Synthesized from User Studies.” Report Produced by OCLC Research. <http://www.oclc.org/programs/publications/reports/2009-06.pdf>

The Science and Technology Council of the Academy of Motion Picture Arts and Sciences. *The digital dilemma: strategic issues in archiving and accessing digital motion picture materials*. Beverly Hills, CA: Academy of Motion Picture Arts and Sciences, 2007.

WORKS CONSULTED

Beard, I. and Jantz, R. 2007. *Proposed Digital Architecture for Video/Moving Image Object*. http://RUcore.libraries.rutgers.edu/collab/ref/ctm_sawg_video_content_v2.pdf

Daniel, R., Lagoze, C., and Payette, S. D. “A Metadata Architecture for Digital Libraries” in *Proceedings of the Advances in Digital Libraries Conference* (April 22 - 24, 1998). ADL. IEEE Computer Society, Washington, DC, 276.

Dushay, N. and Lagoze, C. 2000. "Modeling Decisions for Digital Content" Technical Report. UMI Order Number: TR2000-1807, Cornell University.

Garreau, Joel. "Smithsonian Click-n-Drag Itsself Forward; Cyber Thinkers, Curators Discuss Digital Challenges" *Washington Post*, January 6, 2009. <http://www.washingtonpost.com/wp-dyn/content/article/2009/01/25/AR2009012502179.html>

Guthrie, Kevin, Rebecca Griffiths and Nancy Maron. May 2008. "Sustainability and Revenue Models for Online Academic Resources: An Ithaka Report prepared for the Strategic Content Alliance." JISC. jisc.ac.uk/contentalliance.

Ide, M., and Weisse, L. 2006. "Appraisal and Selection Guidelines for Public Television" in *Preserving Digital Public Television*. <http://www.ptvdigitalarchive.org/work.htm>

Kaufman, Peter B. 2009. "On Building a New Market for Culture: Virtue and Necessity in a screen-based economy." Report prepared for the Strategic Content Alliance, JISC. jisc.ac.uk/contentalliance.

(Kaufman, Peter B.) June 2009. "Video Use and Higher Education: Options for the Future." Intelligent Television, New York University and the Copyright Clearance Center. http://library.nyu.edu/about/Video_Use_in_Higher_Education.pdf

Leah, Andrea. "Context! Context! Context! Describing Moving Images at the Collection Level" in *The Moving Image*, 6.1, Spring 2006.

McGrath, Kelley. "Facet-Based Search and Navigation With LCSH: Problems and Opportunities." *Code4Lib Journal*, Issue 1, 2007-12-17. <http://journal.code4lib.org/articles/23>

"Revealing Design Treasures from the Amazon" presentation by Jared Spool at Information Architecture Summit, Saturday, March 21, 2009. Find an overview of his talk at <http://iasummit.org/2009/program/presentations/revealing-design-treasures-from-the-amazon/>

Sizemore, C. and Oliver, S. 2008. *Wikipedia as a controlled vocabulary*. <http://www.slideshare.net/guest2c797e/wikipedia-as-controlled-vocabulary>

Toms, Elaine G., "Serendipitous Information Retrieval" *Proceedings of the First DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries*, http://www.ercim.org/publication/ws-proceedings/DelNoe01/3_Toms.pdf

UVa Library Digital Initiatives. 2008. "UVa Digital Library Implementation Plan." http://www.lib.virginia.edu/digital/resndev/fedora_imp/

Viégas, F. and Wattenberg, M. "Shakespeare, god, and lonely hearts: transforming data access with many eyes" in *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries* (Pittsburgh PA, PA, USA, June 16 - 20, 2008). JCDL 2008. ACM, New York, NY, 145-146.

DOI= <http://doi.acm.org/10.1145/1378889.1378914>

Yee, K. P., D. Fisher, R. Dhamija, and M. A. Hearst (2001). "Animated exploration of dynamic graphs with radial layout." <http://citeseer.lcs.mit.edu/yee01animated.html>

APPENDICES

- A. TEI Transcript Sample
- B. Open Source Code Resources
- C. Digital Library Applications Review
- D. Scholar Reports
- E. User Testing Report
- F. Ithaka Reports
- G. Ithaka Cost Model for Sustainability Plan
- H. Dissemination