White Paper Report

Report ID: 109390 Application Number: HD-51791-13 Project Director: Nikki Silva (kitchen@kitchensisters.org) Institution: Kitchen Sisters Productions Reporting Period: 5/1/2013-4/30/2014 Report Due: 7/31/2014 Date Submitted: 7/29/2014

1. Cover page

Interim performance report Grant #HD-51791-13 Pop Up Archive: Standardized Preservation and Distribution of Culturally Significant Audio Project Director: Nikki Silva The Kitchen Sisters July 31, 2014

2. Narrative Description

PROJECT ACTIVITIES

We are pleased to report that we have successfully completed our objectives for the NEH Digital Humanities Start Up Grant within the grant period, and are moving forward with continued expansion, development, and sustaining of the Pop Up Archive tools for preserving and creating access to digital sound.

Partnerships with the Internet Archive and the Public Radio Exchange (PRX) — whether preexisting or formalized over the course of the NEH funding — were critical to achieving these goals, and provided us with significant engineering acumen and efficiency, trust within our constituent communities, access to large potential user bases, and unlimited public digital preservation of audio files and related metadata.

Major project activities included:

- Building a centralized web community and repository for public media records. Over the course of 2013, with NEH funding, additional support from the John S. and James L. Knight Foundation, and partners at PRX, we co-developed a centralized repository (platform) for archival sound, www.popuparchive.org, that enables search and preservation for historically significant digital audio, including public media.
- Extensive user testing with stakeholders in public media and audio archiving. While we benefited from the early input of many different potential users, we quickly realized that our initial user personas encompassed too many disparate communities. After building our Omeka plug-ins for small to mid-sized audio archives, we have now expanded our focus to adequately serve radio producers, media stations and larger university and institutional archives.
- Assessment and R&D of speech-to-text solutions for archival audio recordings. We worked closely with researchers at the BBC to identify the best automatic speech recognition technology to help automate the creation of metadata for — and ultimately searchability of — recorded sound collections. We identified the best course of action for implementing such tools and are taking steps to proceed with the development and integration of this type of software (which expands beyond the scope of our Start Up grant).
- Active use of and contribution to the PBCore metadata schema. In addition to developing Pop Up Archive's metadata schema based on best practices from the PBCore standard, we created the Pop Up Archive repository to ingest PBCore records directly from partners such as Illinois Public Media (WILL). We are also members of the Association of Moving Image Archivists sub-committee on the ongoing development of the PBCore standard and integration with the Public Media Platform API. Finally, leveraging The Kitchen Sisters' archive of early cassette and non-digital recordings, we are contributing participants in the PBCore-compliant <u>California Audio Visual</u> Preservation Project.
- **Trainings and community engagement** (see "Project Accomplishments" and Appendix for more specifics). We have conducted numerous trainings, webinars, and beta

testing/pilot programs with the Pop Up Archive tools and for the audio archiving/public media communities at large. There are currently over 8,000 public items comprising thousands of audio files accessible through Pop Up Archive. Over 800 Pop Up Archive public collections are stored at the Internet Archive and have over 20,000 downloads to date.

PROJECT ACCOMPLISHMENTS

The Digital Humanities Start Up Grant phase of Pop Up Archive entailed an academic survey of methods for preserving and accessing culturally significant public media and oral history content, the release of software plug-ins for Omeka, and the development of a centralized database of thousands of archival audio records, files, and transcripts. We successfully accomplished all of the objectives proposed in our NEH Start Up grant application, with several valuable lessons and critical insights along the way:

Completion of Omeka plug-ins for audio preservation and access.

Our first Start Up Grant deliverable was to finalize, perfect, and test software plug-ins that contributed to the ongoing development of Omeka, the free web publishing platform supported by George Mason University and widely adopted across humanities collections. These plug-ins enable standardization of oral history and public media metadata, safeguarding of material at the Internet Archive, and sharing through web services like SoundCloud: the Internet Archive plug-in sends files and metadata to the Internet Archive for permanent preservation, and the SoundCloud plug-in sends files to the user's SoundCloud account, with an option to publicize the content or keep it private.

A centralized repository (platform) and tools for archival sound.

Pop Up Archive (www.popuparchive.org):

- Enables anyone to **search**, filter, and access sound whether their own collection or material from national media organizations and archives
- Enables anyone to **preserve** heir sound in a cloud-based, accessible digital location.

We debuted a prototype of Pop Up Archive at SXSW Interactive in March 2013 and conducted beta testing from April through October 2013. Pop Up Archive launched formally on November 21, 2013 and has over 1000 users. Institutional partners and customers of the service include the Pacifica Radio Archives, WFMT Chicago's Complete Studs Terkel Digital Archive, Stanford's Hoover Institution, and NPR.

Through our Kitchen Sisters archive prototype with Omeka, and from surveying the public media/small archive ecosystem, we established early on that many potential users do not want to maintain their own servers. Thus, Pop Up Archive is a web application and presents little to no technological barriers to entry for interested users, beyond an Internet connection. For larger organizations (media stations and institutional archives), we provide enterprise services to integrate with existing content management systems, APIs, RSS feeds, and third parties such as SoundCloud for distribution, the Internet Archive for preservation, and Amara for crowdsourced transcript refinement and translation. In designing this system took great care to identify rights management concerns and enable both public and private access options.

Automated speech-to-text software. One exciting, unexpected development in the project has been our ability to leverage automated transcription (speech-to-text and keyword extraction software) to help make large quantities of sound searchable. Automatic transcription, searchability, and finding aids are among the primary tools our platform can offer. We feel that this shift toward audio searchability is an important shift that reflects the needs and abilities of

independent radio producers, audio archives and other stewards of culturally significant audio.

Standardized vocabulary for describing audiovisual assets. We finalized our standardized vocabulary (the Pop Up Archive Element Set) for describing audiovisual assets in early 2013. The vocabulary will be continually updated to reflect industry needs and mappings to PBCore (pbcore.org), the formal standard for public broadcast metadata. The vocabulary maps to and from other common standards that include MediaRSS, iTunes, and SoundCloud feeds as well as the NPR API.

Trainings and community engagement. Using The Kitchen Sisters archive as a prototype, we led an archiving workshop at SXSW Interactive and contributed to a panel presentation on audio archiving at the Integrated Media Association conference in March 2013. We had over 80 registrants for a webinar for members of the Association of Independents in Radio in April 2013. We presented at the Digital Preservation 2013 conference in July 2013 sponsored by the Library of Congress and participated in a "Story-First Technology" panel at the Public Radio Programming Directors Conference in September 2013. We presented to multiple departments at NPR in October 2013 and gave the keynote address at TechRaking 4: Hearing the News in November 2013 (see Appendices).

AUDIENCES

Pop Up Archive is working with hundreds of small archives, radio stations, and individual content creators. We're working with retired journalists to surface decades-old source material, and we've helped institutions like the Chicago History Museum liberate the recently-digitized Studs Terkel audio archive. We are also members of the <u>National Digital Stewardship Alliance's</u> <u>Innovation Working Group</u>. However, we are by no means finished with the work of integrating as seamlessly as possible into existing (often complex and convoluted) workflows at larger organizations.

User testing in 2013 helped us define interest for Pop Up Archive among journalists, archivists, and the institutions supporting them (stations, newsrooms, universities and other institutional archives). We saw over 100 participants in our online archival trainings and webinars, not including audiences at conference presentations. Over 500 people requested invitation-only access to our beta. Approximately 50% of those requests created beta accounts and tested the system.

Pop Up Archive has over 1100 individual users, with over 14,014 unique visitors and 74,174 page views from launch on November 21, 2013 through July 16th, 2014. For the month from June 15, 2014 through July 15, 2014, visitors to <u>popuparchive.org</u> spent an average of over 4.25 minutes on the site.

Pop Up Archive's primary audiences are radio producers, journalists, archivists, oral historians, academic researchers, and the institutions supporting each of these stakeholder constituencies. Most visitors to the Pop Up Archive website are based in the United States (75.73%), Canada (3.66%), and the UK (3.62%). Within the United States, most visitors to the Pop Up Archive site come from the Bay Area (Oakland and San Francisco, 13.6%), Chicago (6.24%), New York (5.8%), and Los Angeles (2.56%).

Participating institutions include: the Public Radio Exchange, the Chicago History Museum, WFMT Chicago, NPR, KQED, KALW, KCRW, WILL, WTJU, Storycorps, Princeton Theological Seminary, Presbyterian Historical Society, Snap Judgment, Center for Investigative Reporting, Pacifica Radio Archives, Circus World Museum, Wisconsin Historical Society, Columbia College Chicago, Detroit Sound Conservancy, and FIXT POINT.

EVALUATION

Pop Up Archive has been evaluated through ongoing and unforgiving real-world tests: will users pay for this service? Will funders fund it? The answer to both questions seems to be "yes."

While Pop Up Archive was still in private beta in 2013, 35 of 42 respondents to the site's exit survey (over 80%) said they would be "disappointed" if they could never use Pop Up Archive again. Of those 35 respondents, over 60% responded that they would sign up for a paid (as opposed to free) plan.

We are currently working toward a sustainable, replicable, scalable business model for the service. Our intention has always been for institutional/enterprise contracts to subsidize the resources we provide to less well-funded organizations and individuals. We continue to believe that this goal is achievable, but we know that we will need to continue to build out features on top of our speech recognition and tagging capabilities that help customers ascribe a monetary value to the service we provide.

As was noted at the Preserving Digital News panel at the Digital Preservation 2014 meeting in July 2014, most newsrooms — and archives — ascribe some value to their archival content, but that value is vague and unquantified. For that reason, we continue to explore technical solutions that would help these stakeholders pin down the exact value of their archives. For example, we want to help archives and stations programmatically embed textual information about audiovisual assets into their web pages, so those pages will be automatically indexed by search engines and results in more web traffic, page views, and visibility for the institutions. We also want to help our constituents do a better job of sharing and reusing their audiovisual content, through better embed tools and capabilities to share segments or specific moments of audiovisual content, as opposed to the entire source file.

Pop Up Archive has preserved and indexed over 10,000 hours of sound so far, and popuparchive.org currently provides access to over 8,040 public audio items comprising many more audio files. We anticipate that number will continue to grow quickly as we onboard more partner organizations. Over 800 Pop Up Archive public collections are stored at the Internet Archive (over 5,400 items) and have over 20,000 downloads to date.

As Lisa Hollenbach, a poetry scholar working toward her PhD at USC, explains, "I'm particularly interested in the generation of searchable transcripts and better tagging that would help me better navigate which recordings might be of interest to me and save me time laboring on my own transcription... Recovering contextual data with sound is often, I find, very challenging."

The public loves Pop Up Archive! Here's a sampling of our press coverage:

Jun 19, 2014 | PBCORE.ORG

PBCore data exchange with the Pop Up Archive

You may have heard about the Pop Up Archive, which if you haven't, you should. From my vantage point, they are solving a key problem of radio producers (and anyone creating digital audio) by providing: - An embeddable audio player for each asset...

Dec 21, 2013 | MAKEUSEOF.COM

15 Awesome Websites And Online Services That Were Born In 2013

While developers are definitely preferring to make apps for mobile platforms like iOS and Android, the Web isn't going anywhere for a while, so you will still have some cool Internet services and apps being made. And while you might think 2013...

Dec 16, 2013 | OPENCULTURE.COM

Studs Terkel Interviews Bob Dylan, Shel Silverstein, Maya Angelou & More in New Audio Trove

The recently-launched digitization project PopUp Archive hosts a great selection of Studs Terkel audio. The site's archive of interviews comes from Terkel's 1952-1997 radio show, "The Studs Terkel Program," on the Chicago station WFMT. While...

Dec 13, 2013 | IJNET.ORG

Online sound library Pop Up Archive revamps audio for the digital age

Updated Dec. 13, 2013, at 7:00 a.m. Eastern. Journalists who work with recorded sound know just how painstaking it can be to sift through hours upon hours of audio to find the right clip for a story. Then comes the time-consuming task of...

Dec 3, 2013 | MAKEUSEOF.COM

Pop Up Archive Transcribes and Tags Sounds, Searches Historical Broadcasts Pop Up Archive is a new tool to help journalists, media, archivists, historians and others easily find and reuse sound.

Nov 29, 2013 | THENEXTWEB.COM

Pop Up Archive serves up thousands of hours of 'lost' radio broadcasts

Launched last week in cahoots with the Public Radio Exchange, Pop Up Archive serves up thousands of hours of lost radio broadcasts, including...

CONTINUATION OF THE PROJECT

Pop Up Archive will definitely continue after the Digital Humanities Start Up Grant period. We've only scratched the surface of what's possible. We're more dedicated than ever before to our goal of uniting audio recordings and the historic voices from disparate places and eras therein, and making it possible for anyone to graze the surface of or dive deep into our nation's rich and varied oral history.

Following the research and tests we conducted during the Digital Humanities Start Up Grant, we applied — and were recently thrilled to receive — a Digital Humanities Implementation Grant. With these funds (in addition to growing revenue and additional outside investment — see "Long Term Impact for more details), we will implement our own automated transcription, keyword extraction, and speaker differentiation toolkit, inspired by similar work at the BBC World Service Archive, and advised by/in collaboration with the BBC's Research and Development team. We will also hire additional staff in order to keep up with demand for the Pop Up Archive service as well as growing engineering and web development needs.

Pop Up Archive's innovation lies in the suite of tools we have developed to automatically make large quantities of sound searchable with timestamps and keywords. Automatic transcription, searchability, and finding aids are among the primary tools our platform can offer. We accomplish this through a combination of speech-to-text and entity extraction software, which

we initially developed using Google's free speech-to-text API, Open Calais, and the Yahoo Analysis API.

Continual improvements in speech-to-text technology make it possible for the first time to apply automated processes to widely varying audio collections of disparate voices, accents, and dialects. We are currently implementing a robust speech-to-text toolkit architected in part by the BBC R&D team. We are building on the BBC's work to automatically identify different speakers within audio and augment automated processes through crowdsourced contributions to transcripts (Amara), speaker identification (Wikimedia), and cataloging with controlled vocabularies and DPBedia for linked data.

We intend to process over 30,000 hours of digital sound from sources that include the American Archive of Public Broadcasting, the official Studs Terkel archive, oral history archives such as the Pacifica Radio Archives, and PRX's independent producer and public radio station communities.

We will continue educating public media and oral history communities about best practices for preserving, sharing, and reusing digital sound through a shared web space, blog posts, webinars, conference presentations, and our free individual archiving service for publicly preserving sound and making it searchable.

Below is a sampling of institutions/audiences that are either Pop Up Archive's current partners or who have expressed a need to benefit from better audio management tools in letters of support for our follow-on NEH Digital Humanities Implementation Grant:

- The American Archive of Public Broadcasting: Started in 2007 by the Corporation for Public Broadcasting, the AAPB is continued today by WGBH and the Library of Congress, with 40,000+ hours of digitized archival broadcasts spanning over 50 years from U.S. radio stations.
- Independent radio producer holdings: Media creators, institutions, and scholars frequently request access to archival material from independent producers like The Kitchen Sisters as they amass interviews and source material for new radio works, books, documentaries, and scholarly output.
- **Studs Terkel Archive:** The Library of Congress is digitizing 7,500 recordings from 40 years of interviews conducted by oral historian Studs Terkel.
- **"Orphaned works" from retired journalists:** Veteran reporters maintain significant audio collections amassed over the course of decades in the field, but are unclear about their right to digitize and share these materials with a new generation of journalists.
- Illinois Public Media: Stations like WILL produce constant streams of audio content, adherent to the PBCore metadata standard, but without established methods for access, preservation, or reuse.
- **KALW:** Like many small public radio stations, KALW's website has changed multiple times in the past decade, leaving entire years of audio and contextual content stranded on standalone servers without finding aids.
- Pacifica Radio Archives: Digitization of the Pacifica holdings (the largest radio archive

in the United States) is funded by numerous federal agencies and private foundations, but access for scholars, producers, and the public is restricted to a single web portal and the Internet Archive.

- **Center for Investigative Reporting**: The 2nd-largest investigative reporting nonprofit in the world (after the BBC) relies on a massive archive of primary source transcripts to cover stories that span years and require significant access to and reuse of archival material.
- **PRX's Radiotopia podcast network**: Podcasts such as 99% Invisible draw large audiences, but are not preserved or indexed.

LONG TERM IMPACT

Pop Up Archive has generated substantial public and industry interest in its relatively short (21 month) lifetime, establishing a reputation as the vanguard in audio access — i.e. searchability — and preservation.

In addition to partnerships with the Public Radio Exchange and the Internet Archive, Pop Up Archive's long-term impact is being realized through additional outside funding as well as community projects and programs across the country. Anecdotal evidence suggests that a large number of stakeholders in digital archiving, A/V preservation, academic oral history, and public radio have heard of Pop Up Archive. The Pop Up Archive website sees a continually growing stream of visitors.

Public media newsrooms and shows across the country (Center for Investigative Reporting, CUNY Television, KALW, KCRW, The Kitchen Sisters, KQED, NPR, Snap Judgment, WBUR, WILL) are integrating the Pop Up Archive tools into both their production workflows as well as new efforts to revitalize archival their content — efforts that might not ever have been undertaken without Pop Up Archive's extremely user-friendly, low-barrier-to-entry tools.

Pop Up Archive is used for a variety of educational purposes:

- Professors and students at Columbia College Chicago use Pop Up Archive for oral history coursework.
- As part of the Studs Terkel Complete Digital Archive, over a dozen library science students from Dominican University in River Forest, IL use Pop Up Archive to transcribe and catalog thousands of Terkel interviews.
- We intend to keep expanding the use of Pop Up Archive in classrooms of all types, whether graduate-level oral history programs (i.e. Brown University, Williams College) or elementary school classrooms (Chicago Public Schools).

We are excited to help make audio and video more accessible and resusable for students around the world, specifically through the development of easy-to-use web-based tools for exploring, sharing, and remixing audiovisual content.

Having received a \$300,000 award from the John S. and James L. Knight Foundation in 2012, Pop Up Archive secured an additional \$100,000 in Knight funding in 2014. Pop Up Archive was also thrilled to learn that our proposal with PRX for an NEH Implementation Grant (\$325,000)

was approved in July 2014.

GRANT PRODUCTS

The main product from this Digital Humanities Start Up Grant is Pop Up Archive (<u>www.popuparchive.org</u>): a live web platform and publicly accessible front-facing website. The site is a Ruby on Rails application deployed with Heroku; media is stored at Amazon Web Services and/or the Internet Archive. AngularJS powers Pop Up Archive's front end; search is powered by ElasticSearch. The site will be maintained indefinitely. The Pop Up Archive codebase from this phase of development is open source and available at <u>https://github.com/PRX/pop-up-archive.</u>

3. Appendices

We have provided links to relevant web resources below:

Blog posts:

- Pop Up Archive featured in the Library of Congress Signal digital preservation blog
- Pop Up Archive contributed to the Signal blog in a post featuring Emily Gore, Director of Content at the Digital Public Library of America
- Pop Up Archive has made numerous contributions to the PBS Idea Lab blog:
 - Pop Up Archive Breathes New Life into Dusty Media
 - Pop Up Archive and PRX Launch Audio Archiving Service
 - Pop Up Archive Makes Audio Searchable, Findable, Reusable
 - Pop Up Archive: Giving a New Voice to Sound on the Web
 - Pop Up Archive on Creative Commons and digital audio copyright
 - Pop Up Archive's Tumblr frequently addresses digital audio preservation/access issues:
 - Fragility of digital audio and role of Internet Archive
 - New speech-to-text opportunities for public media and archives
 - Recent advances in speech-to-text technology

More details on our educational outreach:

- Initial UC-Berkeley Master's thesis on public radio preservation, May 2012
- SXSW Interactive workshop: Build an Archive and Make it Count, March 2013
- Audio archiving panel at the Integrated Media Association conference, March 2013
- Association of Independents in Radio webinar (screencast), April 2013
- Digital Preservation 2013 conference (video), July 2013
- "Story-First Technology" panel, Public Radio Programming Directors Conference, September 2013
- Keynote address at TechRaking 4: Hearing the News (audio), November 2013
- News App Archiving Designation (collaboration with Knight/Mozilla OpenNews and the Newseum), March 2014

Screenshots (see this and following pages):

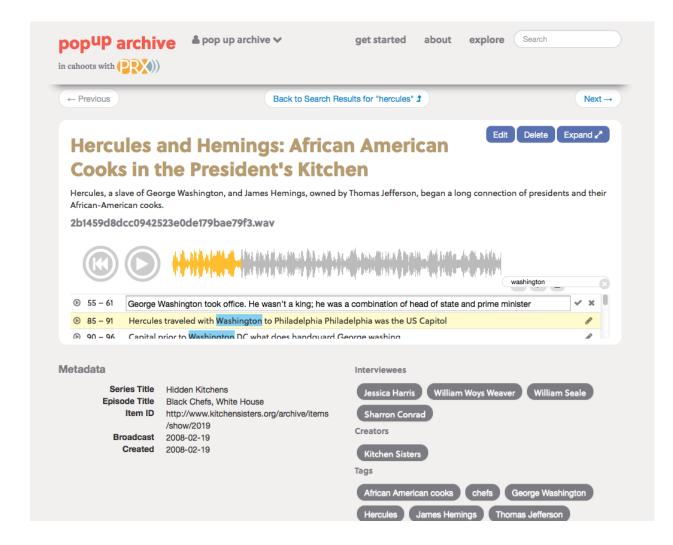
Uploading to Pop Up Archive collections:

popup archive a pop up archive Image: Constraint of the second seco	home a	bout explore	news Searc	ch
Unsorted Files				Batch Upload
This is a temporary place for your audio. Select audio below to move it to	your collections.	. Then: listen, see tr	anscripts, and a	add tags!
Upload () Upload				
Collections				Batch Edit
Pop Up Archive collections can be private or shared publicly. Learn more	about privacy.			
Create a Collection VOA test feed You're	US	Hidden Kitchens		TAL test files

Time-stamped search results with metadata facets:

in cahoots with (PRX))	
← Previous Search	*george washington"
interviewee	1 to 10 of 13 items
Zephyr Wright, Cook William Woys Weave William Seale, Histor Sharron Conrad, Culi Robert Creeley Jessica Harris, Culina Chef Joe Randall producer	 Hercules and Hemings: African American Cooks in the President's Kitchen The story of Hercules and Hemings, the African American cooks who fed the nation's Founding Fathers. Hercules, one of George Washington's slaves, was chef of Washington's kitchen at Mt. Vernon and later in the nation's early capital in Philadelphia. James Hemings, who, along with his sister Sally Hemings, was a slave of Thomas Jefferson, became Jefferson's chef de cuisine in Paris and also at Monticello. Also featured is an archival oral history interview with Zephyr Wright who cooked for Lyndon and Lady Bird Johnson in the White House. 0.55 George Washington took office, he wasn't the king. He was a combination of head of state and 1.35 George Washington enjoyed good cuisine. My name is Joe Randall, I founded the 1:40 African American Chefs Hall of Fame. Wherever was George Washington was, at the
The Kitchen Sisters	(1) African American cooks chefs George Washington Hercules
WILL-AM Geislers, Louise WILL Radio	 Andromeda 21: News Media, Education, and Black Americans with Dick Gregory and Hank Parker, March 11, 1971 Dick Gregory discusses racism in American culture, starting with bias in the media. Mentions a fear of an American dictatorship aided by the FBI and the press. Talks of the difference between "emancipation" and "liberation" and problems and fallacies of the American education system.
host Christopher Lydon	(1) (14:10 can we talk about liberty in America Greg Bell so we should have known someone wrong me (1) (14:35 the S Pen for black folks that all my life man I've been celebrating George Washington (14:40 George Washington birthday and I do not appreciate celebrating birthday lif master of the mix.
tag	African American Civil Rights Gregory, Dick Racism Media and Journalism
Racism African American Civ	(3) (3) Andromeda 21: Afro-American Culture, with Dick Gregory and Hank Parker, 1971

Pop Up Archive item view:



Preservation through the Internet Archive:

 Archive
 Web Video Texts Audio Software About Account TVNews CopenLibrary
 DoppUP archive

 Home
 Audio Books & Poetry | Community Audio | Computers & Technology | Grateful Dead | Live Music Archive | Music & Arts | Netlabels | News & Public Affairs | Non-English Audio | Podcasts | Radio Programs | Spirituality & Religion

 Search:
 Community Audio ÷ Q Advanced Search
 Anonymous User (oph or join us)
 Upload

Audio Archive > Community Audio > Pop Up Archive

Spotlight Item	Welcome to Pop Up Archive 636 item	
idato bung tomo pertama 945.mp3 here is no description for this em	Pop Up Archive helps you organize your media to make it searchable, reusable, and accessible, without requiring technical expertise or substantial resources. We're building an archive of sound from journalists, oral history collections, and accidental archivists from around the world.	Last Week more 1. pidato bung tomo pertama 1945.mp3 25 downloads
bout the Archive ackground equently Asked Questions	Browse by Subject / Keywords All items (most recently added first)	Interview with Shel
	Recently Reviewed Items (<u>more</u>)	Silverstein 6 downloads 4. <u>Test Public Archive</u> 5 downloads
	This Just In (<u>more</u>)	5. Cow Palace
	The California Report 2 hours ago	Interviews 5 downloads
	The California Report 2 hours ago	

Integration with Amara for crowdsourced transcription and translation:

<mark>o</mark> amara	Q	🖵 Subti	tle Videos			Pro Services your Languages	Help	edison@po	
	cifica Rad iving History		hives	l					
Pacifica Radio	Dashboard	Videos	Members	Activity	Settings				
Considered by historians a the oldest and most import world, the Pacifica Radio A political, cultural and artist	tant audio collections in the Archives chronicles the			Hel	ping out is	easy!			
Show all ↓		Just find a video that interests you and click the button below it to get started!							
1 Video	2 Members	Learn more about subtitling in Amara.							
		Videos that n	eed your hel	р					
Member Tools		Tell us what la	nguages you s	peak and we'll pe	ersonalize this li	ist for you			
Your role: Owner									
Leave team									