Pop Up Radio Archive: A Solution for Independent Producers

Anne Wootton, Bailey Smith, and Christen Penny
UC Berkeley School of Information
May 4, 2012
Abstract
Independent producers of broadcast content face major challenges: a lack of archival and technical training, a lack of resources, and a lack of understanding about how to best preserve and create access to cultural audio artifacts. Without understandable, easy-to-implement systems, producers and listeners are hindered in their ability to find and reuse content — and there is a constant risk that irreplaceable broadcast artifacts, and the numerous voices documented in those artifacts, will be lost forever. Existing digital asset management* solutions do not adequately serve independent radio producers, who “have all these hard drives at home” and “dream of a system where [retrieval is easier]” (Silva). Pop Up Radio Archive is a free, open-source, scalable archive system for broadcast content, built with an initial 33 broadcast episodes and hundreds of related media files. The solution gives producers a system for storing content with metadata that is semantically meaningful and that conforms to established industry standards. Using web services, sound files are seamlessly uploaded to the Internet Archive for permanent preservation, and producers are given the option of sharing their content through SoundCloud for public consumption. The Pop Up Radio Archive project is supported by an educational component, including producer trainings and a user manual.

*Definitions for terms in bold are available in Appendix 1, Glossary.
Executive Summary

Our report consists of the following sections:

• Introduction ..................................................................................................................4
  o Why archive?
  o Pop Up Radio Archive: a solution for independent producers
  o The future of Pop Up Radio Archive
  o Team biographies

• Project Timeline: September 2011 - June 2013 ................................................................7

• Ecosystem Review .......................................................................................................9
  o Competitive analysis
  o Connecting producers to storage solutions
  o Interactive media development
  o Content management
  o Roadblocks and solutions
  o Literature Review

• User Research Summary ................................................................................................14

• Implementation ...........................................................................................................15
  o Choosing Omeka
  o PBCore plug-in
  o Internet Archive and SoundCloud plug-ins
  o Web presentation using HTML5 and WordPress
  o Community partnerships
  o Generating awareness about the system

• Literature Review .......................................................................................................24
  o Archival systems
  o Evolution of radio
  o User engagement

• Appendix 1: Glossary ...................................................................................................27

• Appendix 2: Technological issues facing broadcast radio ...........................................29
  o The changing face of web archiving
  o Complex, inconsistent metadata standards and quality
  o Insufficient systems for semantic description and reuse/distribution of content
  o Goals/best practices for broadcast audio archiving

• Appendix 3: User Research Documentation ..................................................................31
  o Methodology
  o Interviews with producers
  o Interviews with listeners
  o Survey
  o Recommendations

• Works Consulted ..........................................................................................................37

• Acknowledgements .....................................................................................................40
Introduction
Why archive?
Independent radio producers compile many hours of sound in the course of the creation of a broadcast piece — but much of this information is lost in the process. In the 1990s, the Library of Congress began collecting anecdotal evidence on the survival and preservation of the past 130 years of American sound recordings. The evidence came from “cultural institutions, professional archivists, record companies, broadcasters, collectors, researchers, and interested members of the general public,” as well as the Library’s own efforts to maintain a national collection of recorded sound. Working with the National Recording Preservation Board, the Library identified the obstacles faced by those dedicated to preserving historic sound recordings as “a serious national problem” (Bamberger vi).

When Davia Nelson and Nikki Silva, known to the public radio community as The Kitchen Sisters, set out to create a six-minute story for NPR, they devote weeks to conducting interviews, researching, note-taking and revisiting sources. They record hours of audio in the field and collect print and digital resources that they keep in folders and on various hard drives. Later, in the editing room, they make painstaking decisions over which audio clips to include in the final broadcast piece. But what happens when Davia or Nikki needs to access that raw material months later? The files on the hard drive are stored in multiple fixed locations — homes, offices, storage units, and sometimes on equipment that is no longer viable. Moreover, these files are indexed by nothing more than their filenames, making them hard to find. The diagram below represents the creation of a single audio file; many of these files must come together to produce a single broadcast piece.

Figure 1: The Life of a File
Today, audio content is created with greater ease and speed than ever before by an increasing number of people. Independent radio producers find themselves in possession of culturally significant collections that lack archival systems and structured metadata. While widely varied in their intentions and audiences, such producers are similar in that they lack resources, technical expertise, and archival training — but face constant pressure to meet demand for their content using new web technologies and platforms. At its core, Pop Up Radio Archive addresses the challenge of enabling any independent radio producer to share digital audio content in ways that are meaningful and useful to the public, without the need to employ an archivist.

“Keeping digital content requires a significant information technology infrastructure, meaning both systems and people,” noted Carl Fleischhauer, Project Coordinator for the Library of Congress, in a speech before the Preservation Conference at the National Archives and Records Administration in 2003. “That may be fine for larger organizations but what about smaller or independent libraries and archives? Small sound archives are clearly not in a position to mount this level of IT infrastructure. What are they to do?”

**Pop Up Radio Archive: a solution for independent producers**

Pop Up Radio Archive acts as a protector, organizer, exhibitor and advocate for radio producers’ content. The Pop Up Radio Archive system integrates WordPress and Omeka, a web publishing platform, as a scalable archival system for broadcast audio and related material. The system gives producers an easy access point for uploading content. Sound files are seamlessly uploaded to the Internet Archive for permanent preservation with the option of social sharing through SoundCloud, all at no cost to producers. Our system will be an easy-to-implement, well-documented, and widely adopted method for storing and accessing independent producers’ content. Without such a system, the public risks losing critical and irreplaceable cultural objects, the voices of underserved populations stored in these artifacts, and an invaluable opportunity to diversify access in an era of media consolidation.

The Pop Up Radio Archive project makes use of the accidental archive compiled by The Kitchen Sisters, San Francisco-based independent producers and oral historians who have worked in public radio for over 30 years. Their collection of audio and related media includes memorable and sound-rich stories about the lives, histories, art, and rituals of people from all walks of life — the well known and unsung — who have shaped America’s diverse cultural heritage. The Kitchen Sisters’ work is recognized for its cinematic quality, its layering of sound and archival audio with interviews and music, and its feeling of place. For any given broadcast piece, hours of recordings are condensed into a segment of only a few minutes. Aside from syndicated broadcasts, The Kitchen Sisters collection has never been shared with the public, though the public has played a significant role in the creation of their work. As oral historians, The Kitchen Sisters feel an obligation to share this unpublished content — but their collection poses challenges common to many independent radio producers: the lack of an official archivist and the resources of larger productions that have succeeded in adding semantic meaning to audio content.
The future of Pop-Up Radio Archive

Our project began with an undertaking to better understand the broadcast radio ecosystem through relevant literature reviews, stakeholder interviews, and usability testing. As veterans in the public radio community, The Kitchen Sisters have been instrumental in securing interviews with relevant public media stakeholders. Analysis of existing standards and interviews with radio content consumers and producers have helped us identify the metadata necessary to adhere to industry conventions and to provide finding aids for researchers and potential listeners — for example, details about interviewees, topics, and geographic locations.

Our final product entails both an archival system (including open-source software and plug-ins) and an educational component (including a user manual for other independent producers).

Our research has also addressed the growing desire in public media to engage with audiences through online crowdsourcing and user-generated content. Our system will aim to reinforce existing audience participation habits and encourage new forms of participation, while recognizing that not all consumers of public radio are interested in, for example, generating and sharing their own recorded audio commentary. Allowing listeners to help organize content enables producers to understand how audiences interpret their work. However, we are also mindful of the importance of the authority upon which archival descriptions are historically based. To that end, The Kitchen Sisters’ files are catalogued with authoritative metadata by Kitchen Sisters staff. User-generated tags can provide additional means of interpreting and accessing content.

Team biographies

The primary team members are Masters 2012 candidates at the UC-Berkeley School of Information, working with two engineers, two I School faculty advisers, the three-person Kitchen Sisters staff, and relevant industry stakeholders.

Bailey Smith, new product development and interaction design: Before the I School, Bailey worked as an editor, journalist, web master, and information architect. At the I School, she has worked as a user experience researcher and designer for Code for America, a government 2.0 non-profit, and Inkling, a tablet publisher. She has also engaged intimately with radio production as the producer of the radio documentary, *Local Hire*, an exploration of the rise and fall of film production in North Carolina. [http://bailey-smith.com/](http://bailey-smith.com/)

Anne Wootton, project management and standards: Before arriving in California, Anne planned and executed a newspaper digitization project at Brown University ([http://dl.lib.brown.edu/dbdh/](http://dl.lib.brown.edu/dbdh/)). With a background in journalism and interest in digital archiving and new media, Anne spent summer 2011 working with The Kitchen Sisters to identify preservation and access opportunities for independent radio.

Christen Penny, user research: With a background in database technical support and quality assurance, Christen has worked with software products used for political and corporate communication strategies. At the I School, Christen co-authored a usability and feasibility study with UC Berkeley’s disabled students program. [http://christenpenny.com/](http://christenpenny.com/)

Daniel Vizzini, lead engineer and web developer: Having majored in math at UNC Chapel Hill, Danny followed the path of least resistance and became an actuary. After realizing how boring this job was, he matriculated at Berkeley to study Civil Systems Engineering. Having earned his Master’s, he is currently employed as a Research Engineer in the College of Engineering at UC Berkeley, developing a server and mobile apps. He is fascinated with the impending transformation of radio onto the web.
Project Timeline: September 2011 - June 2013

September-December 2011:
- Interviews with relevant stakeholders, literature review of public broadcast ecosystem.
- Identification of Kitchen Sisters content for inclusion in archive.
- Review of existing data standards for broadcast media (PBCore — see Implementation section).
- Backup of 1000 Hidden Kitchens audio files at Internet Archive.

January 2012:
- Final evaluation of digital archive systems and existing broadcast audio archival efforts.
- Installation of Omeka to The Kitchen Sisters’ server. Initial draft of user manual for independent producers.
- Interviews with potential users, community stakeholders.

February 2012:
- Initial engineering phase (started Feb. 16): development of Omeka plug-ins that push content to Internet Archive and SoundCloud.
- Creation of Hidden Kitchens sample metadata in adherence with PBCore standard and producer-driven needs.

March 2012:
- Cataloging of 33 Hidden Kitchens broadcasts according to established metadata schema.
- Ongoing development of Internet Archive/SoundCloud plug-ins.
- Begin engineering of Omeka plug-in with PBCore-adherent element set and output format.

April 2012:
- Debugging of existing plug-ins.
- Addition of 1000 Hidden Kitchens audio files (broadcast pieces and raw interviews) with metadata to Omeka system.
- Quality control and expansion of existing metadata.
- User testing and training with The Kitchen Sisters staff, interns, and end-users.
- Integration of Omeka archive theme with existing Kitchen Sisters WordPress theme.
- Addition of ancillary media (photos, recipes, phone messages) to Omeka system.

May 2012:
- Presentation of archival system to Berkeley I School community.
- Concentrated outreach to relevant industry stakeholders (Public Radio Exchange, Association of Independents in Radio, Association of Recorded Sound Collections, Audiovisual Archive Network).
• Collaboration with ongoing public media grant initiatives involving The Kitchen Sisters, KQED, Corporation for Public Broadcasting, and Zeega.

June 2012-May 2013:
• Further software development to make user interface more robust and functional. Addition of web presentation capabilities using time-stamped audio files, SoundCloud, HTML5, WordPress and Omeka API (scheduled for release in summer 2012).
• Seeking of additional funds through public media and journalism incentives.
• Continued outreach to relevant industry stakeholders through conference presentations and one-on-one trainings with public media producers.
• Publication of centralized website (in collaboration with public radio organizations) including archival resources for independent producers of public media (including Pop Up Radio Archive user manual, annotated links, Omeka plug-in software package).
Ecosystem Review

Competitive analysis
The Pop Up Radio Archive team identified three main types of comparable services for independent radio producers: digital asset management systems, archival systems, and radio distribution services. Each of these services addresses only certain segments of the comprehensive environment that Pop Up Radio Archive aims to create for producers.

Connecting producers to storage solutions
Many producers’ digital audio is stored on a collection of hard drives in multiple locations, making content impossible to access at a moment’s notice (Silva). Moreover, storage devices are prone to malfunction and require frequent updating to prevent them from becoming obsolete and incompatible with new devices and software. For example, according to Nathan Dalton, The Kitchen Sisters’ project manager, the duo’s sound engineer keeps old versions of machines just so he can open old files (3 Feb. 2012). To address these storage issues, existing services must provide the following:

1. **Integration with a guaranteed, long-term digital storage provider.** Content management systems push data to a reliable location for long-term storage.

2. **Storage consolidation** - Media from various hard drives can be extracted and stored in a single, secure location. Metadata from multiple locations is consolidated in one database to eliminate problems such as broken filepaths.

3. **Unique URIs for content** - Content is accessible any time, any place through a Unique Resource Identifier (URI). A URI is a web address; content is accessed from the cloud by typing the URI into a browser’s address bar or by calling the URI using web services.

Interactive media development
Radio producers want to create interactive audio experiences using emerging Web 2.0 technologies such as HTML5, CSS3, and Popcorn.js. However, a great deal of foundational work is required in order for their content to be manipulated.

1. **Web services functionality** - Data is accessible in widely adopted web output formats such as JSON or XML. Using these formats with an application programming interface (API) enables web application developers to connect directly to producers’ content and manipulate it to create interactive web applications.

2. **Avenue for presentation and narrative exhibition** - Producers can seamlessly integrate their existing website with their audio archive, manipulating content with an easy-to-use graphical user interface (GUIs).

3. **One-click media distribution** - Producers can easily share the content they archive with relevant communities.

Content management
Content management entails tracking information about the audio files stored in a collection.

1. **Standards-based metadata** - Through standards-based metadata, content can be easily accessed and understood by others who use the same standard. In radio, the dominant metadata standard is called PBCore. PBCore is used by major radio
institutions such as WNYC and the Corporation for Public Broadcasting.

2. *Rich metadata search*- Consistent metadata records allow producers and users to search archived content using fields based on PBCore (Episode, Series, and Contributor) as well as other finding aids (location and tags).

3. *Media reuse and repurposing*- Storage consolidation and rich metadata search make content easier to access. This improved functionality allows producers to rebroadcast or repurpose raw interviews and other content.

4. *Automatic creation of multiple file formats (transcoding)*- Without multiple file formats (AIF, MP3, WAV, OGG), audio cannot be heard across multiple browsers.

5. *Integration with producer workflows*- Archival solutions must take into account the workflows of independent radio producers (interactions with current storage solutions, web publishing, hard drives, etc.), and model systems to meet their needs and reduce workload.

6. *Copyright protection and rights management*- Producers need to control whether their content is publicly accessible for appropriation and reuse. Producers also need a means of tracking rights to third-party content within their own work — for example, through Contributor or Rights metadata fields.

The chart below provides a competitive analysis of each of the relevant services and products that Pop Up Radio Archive encompasses. Digital Asset Management solutions are represented by companies like Dalet Digital Media Systems, a leader in asset management software and consultation for media. Dalet’s clients include Sirius XM Radio and the BBC. Archival Systems include DSpace and Fedora; these open-source solutions are ideal for organizations with complex metadata needs, such as current users MIT and The National Science Digital Library. Radio Distribution is represented by Public Radio Exchange (PRX), which is primarily a *web-based platform for broadcast radio distribution*. PRX organizes content for distribution and puts producers’ “data house[s] in order” for mobile application development, with clients that include This American Life and RadioLab (“PRX: Our Approach”).
Figure 2: Competitive analysis

Roadblocks and solutions
The Pop Up Radio Archive team identified the following potential obstacles to the implementation and adoption of our archive system:

*Filing is not always easier than piling.* When organizing information, it’s easier to pile than file — at least in the short term. However, failing to organize up front carries a cost in the time required for information retrieval. It is crucial that producers realize that a small investment in time and resources to implement the Pop Up Radio Archive system can be recouped in time saved searching for information and in improved audience access and engagement.

*The old crowds out the new.* Migration from legacy systems can prove challenging because producers need to extract data from their current system and upload it to the new system. If producers are currently using proprietary systems, the money they will save on an open-source solution may be a compelling reason to make the switch. The ability to make content available using web services may also be a draw of open-source systems. The Pop Up Radio Archive project will attempt to ease this transition by creating an easy-to-understand user manual.
Independent radio is not known for its deep pockets. Devoting time to a project means devoting money, and small producers have limited funding. Funding institutions such as the Knight Foundation have a growing interest in technology and journalism, and grants from these institutions can support archival efforts, but they are not a universal solution.

Connecting with audiences is hard. As new alternatives for audio content distribution (such as SoundCloud, Spotify, and Facebook) are developed, users’ expectations and producers’ needs change. The ability of the user (whether a content consumer, producer, or both) to access content is paramount to independent radio, which relies on the engagement of niche audiences for success. Many radio listeners still catch their favorite programs while driving in the car — but they increasingly expect to access “broadcast” content through websites, podcasts, and mobile applications such as iTunes, Audible, and PRX (Public Radio Tech Survey PRTS3: 2010). "I used to get my favorite radio shows through iTunes, but now I’m getting them through [a mobile application]," said Ike Sriskandarajah, a public radio listener and producer based at Public Radio International (PRI) in Boston.

The creation of interactive user experiences requires the organization and markup of digital files that can be passed to distributors or used for other purposes within the organization. While user engagement is a high priority for many producers, new technologies for crowdsourcing content and supporting user interaction have yet to be mastered by the broadcast radio community — and listener habits don’t form overnight. Employing existing (and self-selecting) public radio listener communities, crowdsourcing could be leveraged to transcribe interviews (enabling full-text search) and to tag audio files (creating a folksonomy, a collaborative method of organization).

Just because you build it doesn’t mean they’ll come. Public media institutions around the world have had varying levels of success engaging users through message boards, online commenting, and other text-focused mediums. In summer 2010, WNYC found success with their “Map Your Moves” (Avigan) call for crowdsourced data and subsequent visualization. By contrast, our literature review and stakeholder interviews have shown that little success has been achieved for user-generated audio artifacts. For example, in summer 2011, NPR Music attempted to generate user-contributed audio through Soundcloud (Hilton), however, relatively few audience members participated. Crowdsourced transcription services (Causer) enable multimedia content to be indexed and linked, thus enabling easier, previously impossible, retrieval capabilities. Crowdsourced proof-reading (Distributed Proofreaders) and verification (NYPL Map Warper) services rely on public knowledge to correct mistakes, thus producing more accurate data. However, the most successful crowdsourcing efforts related to cultural heritage often attract a specific demographic, requiring substantial buy-in from a community. Few shared or re-usable components exist for the projects mentioned above, and such projects are usually backed either by ample government funding or big corporations. The Pop Up Radio Archive system can be coupled with pre-existing software to enable crowdsourced tagging and social media sharing.

Within the challenge of enabling independent producers to share digital content in ways that are meaningful and useful, we anticipate several technology gaps: a lack of standard metadata, a lack of technical or archival training
on the part of independent producers, and inconsistencies in workflow that hinder producers’ interactions with their audiences. We anticipate barriers beyond these technology gaps as well: independent producers lack the time and resources to build their own archival solutions from the ground up. We have committed our own time, through extensive coursework at UC Berkeley and hundreds of volunteer hours, to the creation of a scalable system and user manual that can be easily implemented by any independent producer. Because producers often lack the time even to educate themselves about the potential return on investment for adopting a particular archival system, our project includes a sample implementation of the archive using Kitchen Sisters content. We are working closely with The Kitchen Sisters on iterative design and implementation of the archive, recognizing that their adoption of our service is paramount to its diffusion among independent producers in the broadcast community.
User Research Summary

The Pop-Up Radio Archive project synthesizes disparate services and products, including digital asset management, archival systems, and distribution to meet the needs of independent radio producers. Our research is informed by a survey of public radio listeners and interviews with consumers and producers of public radio content, including tech friendly listeners, producers, and online engagement specialists. We also spoke to developers and content distributors at PRX, annotation and archival system developers that work on the Open Annotation Collaboration (OAC), and digital content specialists at American Public Media. Through interviews we were able to better understand stakeholders’ needs as they search, browse, and retrieve content online. Producers and listeners alike expressed an interest in easy-to-access content that is as intuitive for producers to organize as it is for listeners to locate. Because every radio producer is unique, the solution is not one-size-fits-all.

The Kitchen Sisters serve as informants for this project. They are also the first producers to implement (and benefit from) the Pop Up Radio Archive system. While buy-in from producers of digital content is important, listeners are a critical component of this study. Producers benefit from increased engagement when listeners can personalize their online experiences. For example, a Kitchen Sisters listener in Houston could search their website for shows about the city, benefiting from additional context in the form of audio, video or images of a particular topic or location. These valuable cultural assets give voice to members of communities who have, in some cases, left indelible marks on their neighborhoods. The Pop Up Radio Archive project takes oral history one step further by making relevant content easily accessible to invested listeners.
Implementation

The Pop Up Radio Archive project makes use of 33 broadcast episodes and hundreds of related media files created between 2004 and 2009 for the Hidden Kitchens series. In addition to the audio broadcasts, background interviews, and phone messages from which the Hidden Kitchen series was created, The Kitchen Sisters’ collection includes hundreds of related digital photographs, press clippings, and recipes — but this ancillary media has not been catalogued and thus is not searchable. Though some of the raw interviews have been minimally cataloged, they lack robust metadata. Metadata can track relationships between people, geographic places, and interviews or pieces themselves. Pop Up Radio Archive enables producers to add metadata to their content so it can be easily discovered.

Choosing Omeka

After close examination of many open-source archival tools (see Ecosystem section), we identified our main constraints to be cost, installation/configuration requirements, and ease of use. We learned from other organizations’ experiences with a range of archival tools. Our final choice was between Collective Access, Archivists’ Toolkit, ResourceSpace, and Omeka. The following chart summarizes our selection criteria:

<table>
<thead>
<tr>
<th></th>
<th>Collective Access</th>
<th>Archivist’s Toolkit</th>
<th>Resource Space</th>
<th>Omeka</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of organizations using the software</td>
<td>40+</td>
<td>100+</td>
<td>Unknown</td>
<td>150+</td>
</tr>
<tr>
<td>Web-based</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Active forums</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Focus on web presentation, theming</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Easy installation</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Adherence to standards</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Flexibility with multimedia</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>WordPress integration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open-source</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

![Figure 3: Archiving Software Selection Criteria](image)

We chose to implement Omeka because it is open source, has light server requirements, and has a simple web-based interface. Powering over 150 websites, with an active developer email group of over 350 members, Omeka is an open-source framework designed for organizations to archive their digital content (“Omeka Development”).

In addition to its ease of deployment, Omeka has several other advantages. It has a paid staff out of George Mason University that develops and maintains its core. Like any good open-source project, it also has a large and active community of volunteer developers. Content organization in Omeka is based on the Dublin Core standard and can be easily extended (see PBCore plug-in section). For their directory of digital collections at New York City area libraries, the Metropolitan New York
Library Council (METRO) chose Omeka over WordPress and the proprietary contentDM system for its “ease of installation” and “strong and flexible approach to metadata representation” (Kucsma).

Omeka also offers several options for integration with WordPress, a popular blogging software used by many independent producers. Omeka provides “a collections-focused web publishing platform that offers both rigorous adherence to standards and interoperability with the collections professional’s toolkit and the design flexibility, interpretive opportunities, and ease of use of popular web authoring tools,” writes Tom Scheinfeldt, Managing Director at the Center for History and New Media at George Mason University (Scheinfeldt).

Omeka also provides the architecture to enable developers (including Pop-up Radio Archive) to expand Omeka’s functionality through plug-ins. Pop Up Radio Archive desired a basic platform or system to which we could add simple software plug-ins unique to public media content. Certain systems offer more robust options than Omeka for handling relational data — for example, tracking an “authority” such as a certain interviewee and all of the ways that authority is connected to different assets (audio, video, images, text) or geographic locations. But Omeka’s simplicity, flexibility, and extensive plug-in library make it possible to find other ways of tracking such relationships — for example, through controlled metadata and hierarchical collections. We ultimately decided that Omeka’s low cost of adoption was a benefit that outweighed the potential limitations of a less robust data model. “Meeting researchers’ needs for access to materials trumps archiving perfection in archival description and arrangement,” writes Lisa Spiro in the 2009 report Archival Management Software: A Report for the Council on Library and Information Resources. Omeka could be described as the WordPress of web-based archival software. By offering a basic set of digital archival exhibition tools, it fulfills a critical need in lowering the bar for deployment. If even Silicon Valley start-ups have trouble finding capable systems engineers to manage their servers, not-for-profits certainly do not have the resources to attract or pay for such talent. For this reason, Omeka suits their needs. Rather than force producers to implement a system that requires substantial configuration, we chose to extend Omeka’s existing functionality to mimic some of the better data models that exist in other open-source software. Through software plug-ins, Pop Up Radio Archive spares producers the complexity of configuring and maintaining a relational database.

**PBCore plug-in**
Omeka’s organizational system is based on Dublin Core, a metadata standard developed by the Online Computer Library Center beginning in 1995. Dublin Core includes a set of vocabulary terms (such as Title, Creator, and Description) to help make content retrieval easier. Omeka allows for the addition of geolocation, free tags, and “Item Type” metadata to Dublin Core records. Item Types include vocabulary terms that are particular to certain types of media (such as documents, still images, and oral history). Building on the Dublin Core schema, the Corporation for Public Broadcasting helped fund development of the PBCore standard, which was first released in 2005.
We set out to build a plug-in for Omeka that integrates the existing Dublin Core and Oral History Item Type elements with additional elements unique to public broadcast content, in keeping with the PBCore standard. The plug-in enables producers to catalog and organize their content while adhering to the standards of the public broadcast community, without making painstakingly conscious efforts to do so. The plug-in generates PBCore XML output for any item in a producer’s collection, taking geolocation and free tags into account, and thus enabling producers to create robust, standards-compliant records with the click of a button.

The following list contains our PBCore Element Set field titles, definitions based on the PBCore schema, and the associated PBCore schema elements:

**Date Created:** The original date that the media item was created. In the case of raw interviews, the date on which the interview was conducted. PBCore recommends date format ISO8601 (YYYY-MM-DD). `pbcoreAssetDate dateType="Created"`

**Identifier:** The catalog number assigned to the media item by the producers, if they have a pre-existing cataloging system. Best practice is to identify the media item (whether analog or digital) by means of an unambiguous string or number corresponding to an established or formal identification system if one exists. `pbcoreIdentifier source="[your organization here]"`

**Item Title:** The name given to the media item you are cataloging. It is the unique name everyone should use to refer to or search for a particular media item. If no proper or given title is available, e.g., photographs or segments harvested from a longer work or program, a "supplied title" must be invented and used to name the media item. Be considerate of and sensitive to the end user who is attempting to search for your media item. `pbcoreTitle`
**Episode Title**: The episode or piece to which a media item contributed. `pbcoreTitle titleType="Episode"`

**Series Title**: If applicable, the larger series to which the media item contributed. `pbcoreTitle titleType="Series"`

**Description**: Uses free-form text to report abstracts, or summaries about the intellectual content of a media item you are cataloging. The information may be in the form of a paragraph giving an individual program description or brief content reviews. `pbcoreDescription`

**Creator**: Identifies a person or organization primarily responsible for creating a media item. The creator may be considered an author and could be one or more people, a business, organization, group, project or service. (For personal names use "LastName, FirstName MiddleName, Suffix") `pbcoreCreator`

**Rights**: Information about rights to the media item. Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights. `pbcoreRightsSummary`

**Physical Format***: The format of a particular version or rendition of a media item as it exists in an actual physical form that occupies physical space (e.g., a tape on a shelf), rather than as a digital file residing on a server or hard drive. `pbcoreInstantiation → instantiationPhysical`

**Physical Location**: An address for a physical media item. For an organization or producer acting as caretaker of a media resource, this may contain information about a specific shelf location for an item, including an organization's name, department name, shelf ID, and contact information. `pbcoreInstantiation → instantiationLocation`

**Digital Format***: The format of a particular version or rendition of a media item as it exists in its digital form, i.e., as a digital file on a server or hard drive. `pbcoreInstantiation → instantiationDigital`

**Digital Location**: An address for a digital media item. This employs an unambiguous reference or identifier for a digital rendition/instantiation of a media item and may include domain, path, filename or html page. If used with the Internet Archive plug-in, this field will automatically populate with URIs for each digital file format created by the Internet Archive. `pbcoreInstantiation → instantiationIdentifier`

**Transcription**: The textual transcription of the media item. `pbcoreAnnotation annotationType="Transcription"`

**Date Peg**: A way to track the relevance of a given media item to a holiday or other date. `pbcoreAnnotation annotationType="Date Peg"`

**Notes**: Any other notes or information about the media item, including legacy metadata. `pbcoreAnnotation annotationType="Notes"`

**Music/Sound Used**: Details on music or other sound clips that contributed to the media item. May include title, artist, album, timestamp, producer, and record label information. `pbcoreAnnotation annotationType="MusicUsed"`

**Oral History Item Type**: `pbcoreAssetDate dateType="Broadcast"

**Broadcast Date**: The date on which the audio was originally broadcast. `pbcoreAssetDate dateType="Broadcast"

**Interviewee**: The person(s) being interviewed. For personal names use "LastName, FirstName MiddleName, Suffix") `pbcoreContributor → contributor where contributorRole is "Interviewee.""
Interviewer: The person(s) conducting the interview. For personal names use "LastName, FirstName MiddleName, Suffix")

pbcoreContributor → contributor where contributorRole is “Interviewer.”

Host: The person hosting the broadcast piece. (For personal names use "LastName, FirstName MiddleName, Suffix")

pbcoreContributor → contributor where contributorRole is “Host.”

Duration: Provides a timestamp for the overall length or duration of the audio. Represents the playback time. pbcoreInstantiation → instantiationDuration

Internet Archive and SoundCloud plug-ins

After identifying an optimal workflow and data storage plan for broadcast content, we contracted engineer Daniel Vizzini to build two plug-ins for Omeka. Our plug-ins improve Omeka in two ways. From an archival standpoint, the server that hosts Omeka and its data is typically rented space in the cloud. A change in credit card number can delete this data. From an outreach standpoint, building the front-end to an Omeka-powered website involves a significant investment in time, so...
Omeka sites are often lightly trafficked. Pop Up Radio Archive addresses these issues by pushing content from Omeka to other locations on the web. For posterity, data is sent to the Internet Archive, whose job is to store digital media in perpetuity. For accessibility, the data is sent to SoundCloud, an audio repository with growing popularity and a more agreeable interface for interacting with audio content.

Omeka acts as a middleman between producers and these services. Omeka is written in PHP and uses MySQL relational database management and the Apache web server. Because its installation does not require root access, Omeka can be run on server space provided by popular vendors such as DreamHost, without the considerable expertise needed for server deployment. Omeka’s core functionality includes the ability to get data from a local machine (e.g., a radio producer’s laptop) to the user’s database. Our plug-ins provide the cURL code to get data from the database to more accessible, better maintained third parties. Passing data to these third parties is easier said than done. Below is a synopsis of the technical components of the plug-ins.

**cURL**: Both plug-ins are cURL scripts at their core. cURL is an open-source command-line utility that runs HTTP calls which, in this case, are implemented through a PHP object. The Internet Archive plug-in sends HTTP PUT requests to an API that mimics the popular S3 service of Amazon Web Services. The Internet Archive’s S3-like API must create a bucket after the first file is received before subsequent files are sent. Two or more files are sent to the Internet Archive: the media item(s) and one file of PBCore XML output for the item(s). The XML file is sent first, thus creating a bucket. In the meantime, a while loop runs to determine when the bucket has been created, after which subsequent files are sent in a multithreaded process. This currently presents UI problems, as users must wait for up to two minutes before the process completes. This issue will be addressed in future versions of the plug-in. The SoundCloud plug-in has no such problems with bucket creation, and its scripts can simply be sent in a multithreaded process.

**Authentication**: The Internet Archive uses public-key authentication. On installation, the Internet Archive plug-in prompts users to enter an Internet Archive-provided public key, and saves this key in Omeka’s persistent memory. This public key is then passed to the Internet Archive in the headers of the HTTP PUT request. In the case of SoundCloud, OAuth authentication is used. On installation, users are prompted to log into SoundCloud, and a token is returned to the plug-in. This token is saved in Omeka’s persistent memory and sent in the header of HTTP POST requests to authenticate tracks that posted to SoundCloud.

**Omeka Admin Integration**: All of these processes run in Omeka’s after_save_item hook, which is implemented after an Omeka item is saved. Plug-in users can specify whether they would like to upload content to the Internet Archive and/or SoundCloud. Using Omeka’s Model-View-Controller framework, future versions of the plug-in will provide a means for users to monitor files uploaded at the Internet Archive and SoundCloud from their Omeka admin dashboard.
When producers upload new content and/or metadata to Omeka, files are pushed to the Internet Archive staging server and assigned permanent URIs in multiple formats. Producers are also given the option of uploading files to SoundCloud with public/private controls. Omeka pulls the URIs of transcoded formats from the Internet Archive for as many formats as applicable. These URIs are stored in the producer’s database and reflected in PBCore XML output. This output is also stored at the Internet Archive. When producers edit existing content and/or metadata (including the creation of new metadata fields) in Omeka, changes are pushed to the Internet Archive. Updated metadata (including new URIs, if applicable) is also pushed to the Internet Archive.

Web presentation using HTML5 and WordPress

The next phase of Pop Up Radio Archive will include development of front-end web presentation capabilities for content stored within our archival system. This development will include enhanced integration with WordPress and HTML5 technology. We have conducted an exploratory phase of engineering on a test implementation using HTML5, Popcorn.js, and SoundCloud on a WordPress-powered site.

Current WordPress-Omeka integration includes the ability to override WordPress menus and redirect to Omeka sites, and an exhibition plug-in for WordPress that enables users to add Omeka slideshows directly to their WordPress blogs. Future iterations of Pop Up Radio Archive will integrate WordPress with the Omeka API scheduled for release in summer 2012 as well as services such as Embed.ly, which facilitate the creation of rich web presentations using time-stamped multimedia files.

While The Kitchen Sisters’ current public-facing WordPress site includes basic functionality for playing audio files using plug-ins, there is plenty of room for improvement. Few WordPress plug-

[Diagram: Process Flow for Sending a File to the Internet Archive and SoundCloud]
ins effectively integrate images, video, and text while linking content with the listening experience.

One such existing WordPress audio plug-in embeds a SoundCloud player within a WordPress page. Since Pop Up Radio users will have a built-in option to push content to SoundCloud, we sought to create a plug-in with an HTML5-based player that visualizes ancillary media through timestamped comments. Our test implementation uses client-side Javascript to call the SoundCloud API for audio files and comments associated with a given user’s SoundCloud account. To visualize timestamped annotations pulled within the comments, we combined HTML5 with Popcorn.js. The Popcorn.js library enables multiple methods for pulling annotation content from sources across the web, whether images or text from web sites, Wikipedia entries, Google Maps, or Twitter feeds. We linked user-generated comments from the SoundCloud API to our Popcorn.js object, and displayed these comments according to the type of media referenced within each one. We devised a simple format, according to which the Popcorn.js player checks a comment for one of three critical terms ("I" = image, "F" = footnote, "W" = Wikipedia) as well as a delivered end time. The annotation’s start time is read from the comment’s timestamp. Our implementation required substantial client-side scripting, which helped to achieve a more dynamic WordPress page presentation. The final implementation will make most sense as a formalized plug-in for WordPress, written in PHP. This will push some of the functionality further from the client side, while preserving the ability for users to create dynamic audiovisual presentations for their WordPress pages.

**Community partnerships**

In September 2011, we began the research phase of our project by reaching out to the 14-person staff at the Public Radio Exchange (PRX), which was formed in 2003 by the Corporation for Public Broadcasting, the National Endowment of the Arts, and the Ford Foundation. Today, PRX stands as the primary web-based distribution platform for broadcast radio, where stations shop content and producers push content to the public. PRX has 137,500 members — of whom about 7,500 are independent radio producers (“PRX Advanced Search”). We interviewed PRX staff to ensure that our metadata schema fits most producer needs, and we were also informed by their experience developing mobile applications for *This American Life* and *RadioLab*. The Pop Up Radio Archive team has been collaborating with the Internet Archive since the summer of 2011. We sought out the Internet Archive as a trusted digital repository. The Internet Archive has also been influential in guiding the development of our Internet Archive plug-in.

Following the October 2011 opening of SoundCloud’s first U.S. office in San Francisco, we met with relevant company representatives and were encouraged by their offer of an unlimited premium membership for The Kitchen Sisters. SoundCloud’s work on timestamped annotations, their robust API, and their active listener and developer community (over five million registered users, adding 20,000 registered users/day, with over 100 mobile apps) (Ehrlich 2011) are valuable as we consider web presentation with HTML5 and listener engagement with online audio content. In turn, SoundCloud benefits from Pop Up Radio Archive as we enable The Kitchen Sisters to automatically publish their content to SoundCloud, thus integrating The Kitchen Sisters and SoundCloud communities.

**Generating awareness about the system**

The Pop Up Radio Archive team has worked closely with open-source communities at the Internet Archive and Omeka to develop plug-ins specific to the broadcast radio community and generate awareness about our work. Plans for the archive system include further integration with WordPress. Our plug-ins will become part of existing Omeka and WordPress code bases for further use and development by others. The Kitchen Sisters’ adoption of our system is a key step toward creating awareness among other independent producers and persuading relevant community stakeholders of the
system’s value. In addition to on-site educational trainings with The Kitchen Sisters, we have produced a free user manual for independent radio producers, which will be made available online in summer 2012.

Through The Kitchen Sisters’ broad network as well as our own contacts, we anticipate a phase of publicizing the Pop Up Radio Archive system among PRX’s 7,500 independent producer members, aiming to reach at least 10 percent of this group in the first year of the system’s existence. Other venues for raising awareness include national conferences of recorded sound professionals and ongoing collaborations among The Kitchen Sisters, KQED, and the Corporation for Public Broadcasting. We also anticipate potential use of the system by producers and collectors of content other than broadcast radio material: for example, film and other audiovisual content.

Initial impact will be measured by The Kitchen Sisters’ adoption of the archive system and use of its web exhibition components. Pop Up Radio Archive will increase the interactive functionality of The Kitchen Sisters’ site and will not only retain current audience members, but also increase traffic, average time spent on the site, and return visits to the site. Impact on web traffic will be assessed by monitoring The Kitchen Sisters’ unique visitors through Google Analytics, as well as through instances of file downloads and listens tracked by the Internet Archive and SoundCloud. Stats for “on air” listeners, website page views, comments on the site, and shares (via Facebook, email, etc.) will also be tracked. Reliable statistics about online listenership are exceedingly difficult to obtain, given the proliferation of channels, amounts of data, and lack of standards — but as the technology changes and improves, listener habits are also changing, and new listener groups are being generated. While public broadcast radio listenership is concentrated in older demographics, younger age groups are responsible for much of the activity around online content — and they are only getting older.

Pop Up Radio Archive does not end with The Kitchen Sisters; we expect that other producers will adopt the system as well. Therefore, the project includes a phase of spreading awareness about the system and encouraging adoption. Beginning in summer 2012, the Pop Up Radio Archive team will work with other producers to aid them with installation of the system, including travel to nationwide industry meetings and conferences. In addition to feedback gathered during user training, qualitative data and testimonials will be gathered through evaluation surveys, which will be used to refine and further market our solution.
Literature Review

Our search for literature related to multimedia archiving and trends in broadcast media resulted in a well-rounded body of work. While many of the following works focus on a particular aspect of an archival system or industry problem, none provide a holistic view of independent production workflows or offer recommendations from installation to implementation. Literature about archives of various forms, such as digital photographs of missing museum art in Iraq and digitized television shows, informed our basic archival requirements.

The Evolution of Radio

Cohen argues that digital journalism and digital humanities should collaborate. Areas of overlap include: the presentation of complex data (in particular, mining and representing data from extensive archives), public engagement, connecting narratives to source material, and a growing need to engage with developers and technologists.

Malesky outlines dramatic changes at NPR and other public media institutions in terms of organization, archiving, and access. NPR uses an audio and transcript archive database based on the Collective Access software, and their archive is integrated with their news and web production workflows. To improve workflows, they plan to excise physical formats and begin the transition to a “born-digital” collection. Born digital formats will ease the creation of archival metadata and integrated workflows across systems that are currently segregated.

This paper gives a background to digital sound through 2007 and considers how podcasting relates to the future of radio. Menduni argues that radio is in a state of flux. He underscores this state of transition by outlining four phases in radio’s development: the birth of the popular use of digital music, web radio, music for free, and the iPod and podcasting (9). While Menduni’s phases end with podcasting, which began around 2004, Menduni predicts that podcasting is merely a transitory phase between the iPod and the mobile phone. Digitized radio files opened channels of communication that were previously reserved for the elite. With digitization, “almost everybody could broadcast” and “almost every existing radio station could ‘webcast,’ thus bypassing many “forms of social control and censorship” (Menduni). With smartphones you no longer need to download, and streaming allows for spur of the moment program consumption that fits an immediate mood.

Archival Systems

This report is a result a Library of Congress study on the preservation of recorded sound. The report finds that much of America’s heritage of recorded sound has already been destroyed and that what remains is often inaccessible to the public. “Public institutions, libraries, and archives hold an estimated 46 million recordings, but few institutions know the full extent of their holdings or their physical location” (Bamberger 12). The Library of Congress’ ongoing investigations into the state of recorded sound resulted in The National Recording Preservation Act of 2000 (Public Law 106-474). The
Library of Congress National Recording Preservation Board and the National Recording Registry were also created in response to the dire need for improved preservation and cataloging practices.


The transfer of analog to digital is a main component of this study. While digitizing analog content is out of scope for Pop Up Radio Archive, similar issues arise when dealing with multiple digital formats stored on various devices. Fleischhauer notes the inherent problems of media formats (CDs, DVDs, etc.) that are “obsolescent” (2). His proposed solution, in addition to the digitization of analog media, is the careful creation of a virtual package that includes audio, images, and metadata. However, Fleischhauer notes that the creation of these resources is technically demanding, and he foresees that creating and maintaining these resources will be problematic for small archives. While he does not claim any definitive solution, he proposes that small archives may require third-party institutions to serve as “keepers of the bits” (9).


Manus asserts that Omeka’s biggest challenges in creating an open-source community are addressing the needs of audiences with disparate skill levels and resources and creating documentation for end users and developers. This work is necessary to create a community with a long-term commitment to Omeka-powered content and Omeka itself.


Serb addresses potential for revenue generation through digital archives. Monetary benefits to stations and producers are achieved through reduction of operational costs as access and workflows are improved and through possible income from media reuse and licensing of content (95, 97). Serb contends, “Media which is not managed and cannot be retrieved, or contains inaccurate data is not an asset, it is a wasted resource.” More effective digital asset management results in an increased demand for content (99).


Spiro’s report addresses the problem of revealing hidden collections. Spiro defines hidden collections as, “materials held in special collections, archives, and other restricted or relatively inaccessible settings.” She concludes that in order to catalog and expose these hidden collections, content producers must adopt technological solutions. However, to achieve successful system integration, users must consider existing workflows and desired features. The paper offers guidelines for evaluating potential archival solutions. Spiro recommends that to inform their decisions, organizations ask if they need:

- A system for accessing information online?
- Particular formats for importing and exporting data?
- Support for functions like accessioning and report generation?
- Commercial or open-source software?


Metadata standards and accuracy are critical components of digital archives. Wactlar and Christel address metadata as it relates to video content, which differs from audio content in terms of file sizes.
and access needs. Given the relatively large size of video files, metadata is especially important, since it can help users determine quickly whether or not desired content resides within a given file. Users rarely have the patience to search through hours of unindexed material (2-3). Metadata enables large digital assets to be “simultaneously protected and accessed,” Wactlar writes. “Without metadata, a thousand-hour digital video archive is reduced to a terabyte or greater jumble of bits; with metadata, those thousand hours can become a valuable information resource” (2).


While “Recommended Metadata Guidelines” focuses specifically on digital television assets, lessons from the television industry can be applied to digital archives of broadcast audio content. Television archives must relocate content for resale and reuse in new productions, make content available for scholarly research, and provide metadata to distributors. Weisse et al. also profile the PBCore metadata standard as a necessary component of digital asset management in order to foster interoperability among producers.

User Engagement


“The Exchange” is an early example of efforts to better unite listeners with radio using web-based approaches. The paper discusses radio as art, including a digital media production context that producers can use as an ‘open studio’ and the public can use to interact with published content. Both producers and users benefit from the freedom offered by a system that is independent of government institutions and commercial facilities (130).


The MAIA’s archive illustrates the successful use of Omeka to create a viable online archive. In war-torn Iraq, many valuable pieces of art have disappeared from the country’s museums. The MAIA is collecting photographs of physical artifacts as they appeared in museums by soliciting crowdsourced images. Pop Up Radio Archive has considered crowdsourcing as a way to gather metadata for The Kitchen Sisters’ audio files, such as user-generated tags and transcripts.


Klein explains how new interactive web technologies influence listener habits, through a study of BBC message boards. She asserts that compared to “other traditional forms of media,” radio has been at the forefront of measures to encourage “listener feedback and participation” (11). The transition to Internet-based activity, such as message board participation, is another facet to an already open line of communication (12). Unlike its predecessors, Internet-mediated communication has given listeners the means to engage in ongoing, asynchronous interaction not only with producers, but with one another as well (14). According to Klein’s survey results, some listeners use online tools to elicit responses from producers and staff, while others prefer to comment and discuss ideas with other listeners. An organization should consider whether its goal is to create a space for user-to-organization or user-to-user communication.
Appendix 1: Glossary

**API**: An application programming interface (API) is a tool for building applications on the web. The API does this by allowing a program to use predefined functions to interact with the API’s source, so that the programmer does not have to write them from scratch.

**crowdsourcing**: As oral historians, independent radio producers have always drawn from the “crowd” for their material. The Kitchen Sisters’ phone line solicitations for listener-contributed audio, dating back to the Lost and Found Sound series in the 1990s, could be considered one of the earliest successful attempts at crowdsourcing.

**digital asset management**: Digital asset management is the system of information organization and retrieval for digital files.

**HTML5**: HTML5 is a markup language for the structure and presentation of web content. It improves on earlier versions of the language by supporting multimedia and remaining human and machine-readable.

**Internet Archive**: The Internet Archive (www.archive.org) is a 501(c)(3) non-profit that was founded to build an Internet library. Its purposes include offering permanent access for researchers, historians, scholars, people with disabilities, and the general public to historical collections that exist in digital format.

**legacy system**: A legacy system is an old technology, system, or program that continues to be used. These systems still function and meet at least some of the users' needs. However, new technology may be more efficient. When new technology is adopted, data from legacy systems must be migrated to the new system.

**metadata mapping**: Metadata mapping is the process copying metadata from one metadata field to another. In order to achieve a smooth metadata mapping you must analyze the metadata fields of your present metadata system and “map” them to equivalent fields in the new metadata system. (e.g., DublinCore to PBCore)

**metadata**: Metadata describes other data. For instance, metadata for an audio file might include fields for “Interviewee” or “Location.”

**OAC**: The Open Annotation Collaboration (www.openannotation.org) is a group dedicated to creating and spreading the use of easily accessible, interoperable annotations. The OAC contributes to a “resource-driven” interchange format that allows annotations to refer to different types of media.

**Omeka**: Omeka is an open-source web-publishing platform for the display of collections and exhibitions.

**PBCore**: PBCore is a metadata standard for media. It is widely adopted among the radio community. PBCore includes rich descriptors for radio and television programs.
**persona**: A persona is a fictional character that represents groups of users with similar behavioral patterns.

**plug-in**: A software plug-in is a program that adds functionality to another program. For instance, a Firefox plug-in might make it possible to block pop-up ads.

**Popcorn.js**: Popcorn.js is a client-side JavaScript library that interacts with HTML5 to display desired content at timestamps within a given audio or video file.

**relational data**: A relational databases uses a group of tables that have defined relationships to one another. In each table, relationship columns define how rows of data should be related to data in other tables. When a query is entered, results can include a complex response that includes data from many different tables.

**schema**: A schema describes the structure of metadata. Metadata schemas usually include a limited number of elements (such as title, format, etc.). By adhering to a schema, a collection achieves uniformity.

**semantic**: Semantically meaningful metadata gives contextual and relational information about the content it describes. For instance, computers don’t process audio the same way human ears do, so we must supplement audio files with semantic descriptions to add meaning to files that humans can understand.

**SoundCloud**: SoundCloud.com is an online audio distribution platform. SoundCloud incorporates a social component by enabling users to record and upload sounds, then share them privately or publicly to blogs, websites and social networks. The SoundCloud service is easily accessed through mobile devices, and the SoundCloud platform has been used for the creation of hundreds of apps.

**transcoding**: The process of converting a media file from one format to another.

**URI**: A uniform resource identifier (URI) is an address that points to resources on the Internet.

**user engagement**: Online user engagement is when users of a site or service are actively involved in the site as evidenced by repeated or sustained site visits or contribution to the site.

**Web 2.0**: Web 2.0 is a set of principles that describes the web as a platform for information sharing, interoperability, and collaboration. A Web 2.0 site allows users to actively create user-generated content as opposed to passively viewing content that was created for them.

**web services**: Web services are a way of integrating web applications by allowing them to share data with one another regardless of what operating system or programming language each party uses.

**Zeega**: Zeega.org is an open-source HTML5 platform for creating interactive documentaries and inventing new forms of storytelling, making it easy to collaboratively produce, curate and publish participatory multimedia projects online, on mobile devices and in physical spaces.
Appendix 2: Technological Issues facing Broadcast Radio

The changing face of web archiving

Few radio producers have formally trained archivists on staff, or are archivists themselves. The growing role of information systems in the management of archival content has posed challenges when moving from physical archives to digital archives, in that lack of familiarity with standards and the technology itself can pose unforeseen new challenges. If content was not organized in the first place, making it available in an organized and discoverable way online is no small feat. Countless “orphaned” collections have also emerged as a result of digital production, and the continuous creation of new content necessitates new approaches to the storage and organization of material. Many consumer-oriented tools like SoundCloud are easy enough for a novice to use, but these solutions are often short-term and do not apply to all content (for example, ancillary media such as photos and video), so additional expertise and knowledge is required. As Nikki Silva noted in an interview, “I have all these hard drives at home... I dream of a system where retrieval is easier.”

Complex, inconsistent metadata standards and quality

Metadata is important for archival projects, but it is used in widely varying ways across different broadcast archives. A key challenge is that many existing standards are oriented more toward traditional (physical) archival description, and may not take advantage of newer technology or the needs and capabilities of emerging technology. PBCore, the metadata standard for broadcast content, attempts to account for these challenges. While not yet in wide use, the standard is gaining traction among producers, stations, and distributors. Acknowledging the value of standardized metadata, Matt MacDonald of PRX said that “at some point [PRX] needs to become PBCore compliant.”

Metadata is often created to serve the unique needs of an organization. External requirements that may foster greater interoperability among stations and distributors are sometimes elusive, and standards alone are often not enough. Connecting to external service providers or even looking at broadcast audio materials in relation to larger collections is a challenge, and the tools currently used to distribute content ask a limited set of questions with no standardization. As a result, metadata quality can be lacking, and these inconsistencies limit the ways that information can be reused in the future. As archivists at WGBH wrote in their recommended metadata guidelines for digital television:

*With the proliferation of digital production and distribution workflows which rely on a wide variety of equipment by different manufacturers that must ‘speak’ to each other, it is no longer viable for stations to utilize such individualistic systems... it is increasingly important that stations begin to adopt common standards for cataloging and tracking so that program files can be successfully changed. (Weisse et al. 3)*

Insufficient systems for semantic description and reuse/distribution of content

Stemming from the challenges of creating quality metadata, existing archival audio systems do not adequately support structured organization and presentation of content. A great deal of information is created during the broadcast audio production and editing process, but much of this information is lost when pieces are compiled. Losing the contextual information about where clips come from, or timestamped information for where certain content starts and ends, limits the potential of new technologies (such as the array of new HTML5 media players that could take advantage of this
metadata). Information about geographic locations, interviewees, and other relevant semantic description does not exist in the archives of many independent producers. As Nikki Silva noted, “I throw [raw interview audio] into ProTools and I just begin listening, making edit marks, naming each segment, and it goes into an audio bin and it has everything labeled... That notation exists in the session but is lost outside of ProTools.”

The semantic descriptions we create today could become less relevant in the future — or relevant at different levels of granularity. For example, even if location data is added to audio records, will it be sufficient in the future to only have a place name associated with those records? Is it the producer’s responsibility to provide exact latitude and longitude coordinates, making it easier to pin their content on a map and thus become more relevant to anyone with a smartphone? Or will technology anticipate these needs and reliably extract geolocation coordinates from a generic place name?

Goals/best practices for broadcast audio archiving

Metadata fields
The following types of additional metadata for broadcast audio collections are an important step toward creating robust, interoperable archives:

- **Geospatial metadata** for location-based retrieval of resources. As user expectations for place-based mobile applications increase, geospatial metadata not only describes resources themselves, but also places referenced within a work.
- **Timestamped metadata** for use with annotation and commenting. Producers should make use of supplementary media (images, text, links) and curate that media temporally in relation to a given audio file.
- **Relational data about producers/creators/content managers** to better track the various authors of any given content. Providing a structured way for this information to be stored will make it possible to link individuals and their roles to resources.
- **Auto-suggest for controlled vocabulary** to reduce inconsistencies of master data.

Standards and Quality
To address variance in metadata usage and creation as well as rapidly-changing technologies across the public media landscape, we recommend the following improvements to metadata standards and quality:

- **Adherence to PBCore**, including import/export standards-compliant functionality.
- **Transcoding and data compression** to make media accessible across platforms.
Appendix 3: User Research Documentation

Methodology

This study encompasses two data gathering methodologies: interviews and a survey. Initially, we met with The Kitchen Sisters to learn about how they track digital artifacts (images, raw interviews, finished broadcast pieces). After understanding their current methods, we asked them to describe their ideal archival system. Since the goal of this project is to create a scalable system, the Pop Up Radio Archive team also interviewed other public radio producers to assess the overall needs of the community.

Public radio listeners were interviewed about how they discover and interact with public radio shows. Concurrent with interviews, the team administered a survey in order to substantiate our findings. The listener perspective is crucial in understanding how producers can improve access to content.

- **Interviews:** The Kitchen Sisters informed much of the preliminary needs assessment. In the summer of 2011, Anne Wootton, a member of the Pop Up Radio Archive team, worked with The Kitchen Sisters to identify opportunities for archiving independent producers’ radio content. Wootton’s experience served to document much tacit knowledge around workflows in the public radio industry as well as the nuances of archiving digital versus physical, or analog, content. Interviews with other members of the public radio community provided context on trends in the realm of digital media and adoption (or lack thereof) of metadata standards. From these interviews, we gleaned information on the current state of digital archive solutions in the public radio community. This included an assessment of existing open-source software solutions. The storage of digital files must also conform to the needs of distributors, such as PRX. The Pop Up Radio Archive team interviewed staff at PRX to better understand distributor workflows and to ensure that our archival system complements producers’ current distribution methods.

  Interviews were digitally recorded and transcribed. Interviews began informally in the summer of 2011, with formal interviews taking place from September 2011 through May 2012. In all, The Pop Up Radio Archive team conducted about fifteen interviews with relevant players in the public radio arena. Additionally, three public radio listeners were interviewed in February and March 2012. These interviews provided valuable input on how listeners interact with public radio in addition to informing a related listener survey.

- **Survey:** The survey was administered both online and in person. The intent of the survey was to gather metrics on ways that listeners interact with public radio programs. In total, 116 participants answered the survey with a 2.5-minute average completion time. The first 27 respondents completed a hard copy version of the survey during a project showcase at UC Berkeley’s School of Information. The remaining 89 participants were recruited through the researchers’ online networks. The findings do not represent a random sample.

Interviews with producers

The first formal, exploratory meeting with The Kitchen Sisters occurred in October 2011. Our goal was to understand The Kitchen Sisters’ ecosphere and identify use cases that represent the producers’ end-to-end process as well as "nice to have" features. We then refined these use cases through interviews with two other independent radio producers as well as archivists and staff at national public media organizations.
**Metadata Standards:** While each producer may require domain-specific fields (for example, The Kitchen Sisters need a "recipes" field), a core set of industry-wide, machine-readable metadata is important for interoperability between distributors and producers. MacDonald notes that PRX encourages producers to use as much metadata as possible because it “helps producers get discovered.” When producers upload content to PRX for distribution to stations, certain metadata (such as breaks, timing cues, and content advisories) is mandatory. The inclusion of other information, such as ancillary media and date pegs, is optional and at the discretion of each producer. However, if producers desire to communicate with other entities, standards such as PBCore can play an important role in making description consistent from one archive to the next. Courtney Michael, Project Manager at the WGBH Media Library & Archives, notes that WGBH’s Open Vault project adheres to PBCore, even though the standard is not well supported. Our interviewees generally agree that, in public radio, PBCore reigns as the universally accepted standard — but it is an imperfect system and compliance is not a priority for many producers.

**Existing Tools and Systems:** Many larger entities, such as NPR, invest substantial resources to build in-house systems that tailor to their exact needs. Generally, these systems are not scalable or even available to others. Many smaller producers lack the funds to create or purchase comparable systems, and instead employ ad hoc methods for storage and organization of their archival material. For The Kitchen Sisters and other producers, open-source tools are a viable option. Even though open-source products are free, “they are not ready out of the box,” notes Kara Van Malssen, Senior Consultant at AudioVisual Preservation Solutions. “There’s no ‘one click install’ kind of thing.” This is complicated by the fact that many independent producers also lack the funds for dedicated technologists to install and customize open-source systems. The two open-source systems that our interviewees have the most experience with are Collective Access and Omeka. Our interviewees stressed several significant differences between Collective Access and Omeka: Collective Access requires a dedicated server, while Omeka does not, and Omeka offers WordPress integration, which is currently lacking in Collective Access. Also, because Omeka was designed primarily for web exposition, “there’s not much flexibility” when it comes to major collections management software configurations (Van Malssen). (See Ecosystem Review and Implementation sections.)

**Reuse:** In order to reuse content, producers need to be able to quickly locate relevant material. Access to reusable audio enables producers to save money and time. Additionally, access to existing content is crucial when stories get repackaged (Sriskandarajah). The Kitchen Sisters encountered one such scenario when asked to contribute to a September 11th memorial show. The nation’s mood had shifted over the course of the decade following the event, and The Kitchen Sisters wanted to revisit and repackage content to fit the current tone (Nikki). This required a time consumptive and difficult search for raw footage from previously recorded interviews stored in disparate locations.

**Rights:** Rights management is a major concern for most of our interviewees. Producers need a way to track rights to third-party content within their own work. For example, The Kitchen Sisters layer music throughout most of their finished pieces without any formal mechanism for tracking the provenance of audio clips. Regardless of the rights that NPR obtains for sound clips used in syndicated pieces, the lack of rights metadata becomes problematic down the road when producers revisit content to create compilations or specials — they “have to do rights stuff” (Silva). Concerns over proper attribution hinder
producers in their ability to reuse content. PRX deals with rights issues from a distributor standpoint, communicating with producers to retroactively give rights to particular pieces (MacDonald). If audio content has not been adequately organized, rights issues can be difficult to resolve without listening to pieces in their entirety — and even then require substantial background knowledge or research.

**Geolocation and Location Data:** While some of our interviewees express enthusiasm about the potential for incorporating geolocation data with their broadcast content, others raise questions about the ultimate value of such information. For The Kitchen Sisters, geolocation metadata could be an important component of future web-based features — for example, providing a given listener who accesses a radio story about barbecue in Austin, Texas with a map of related sites of interest. Mobile applications provide increasing uses for location data.

**Audio File Formats:** Audio file formats are especially important for consistent cross-browser experiences and in multiple bitrates for mobile consumption. For mobile, access to “low quality” bitrates is important (MacDonald). A significant number of listeners access audio content via mobile downloads and streaming (see Survey section), so multiple file formats and bitrates must be considered by producers who wish to reach as wide an audience as possible.

**Interviews with listeners**

The Pop Up Radio Archive Team recruited three public radio listeners through the researchers’ social networks. Interviewees were asked to describe their listening habits and how they search for desired content. Their responses were compiled to inform survey questions which were then administered to a larger sample. For a fuller, more complex picture of listener habits, users would need to be actively recruited outside of the researchers’ social networks and the survey sample size would need to increase.

Our listener interviews focused on online streaming and podcast downloading habits. One interviewee expressed frustration about many independent radio producers’ inconsistent use of iTunes fields such as album, series, and track/episode number, as well as his appreciation for thumbnail photos, which he said improve the search and listening experience. All three interviewees have discovered shows through iTunes recommendations. While one user relies solely on iTunes for radio content (“it manages subscriptions well”), the other two interviewees supplement their consumption through other mobile applications, such as Stitcher. “I used to get my favorite radio shows through iTunes, but now I’m getting them through Stitcher,” said Ike Skriskandarajah of PRI. “The iTunes store doesn’t have a good interface for streaming. The one drawback to the mobile app is that it told everyone on Facebook that I listened to like 39 shows last week. I didn’t mean to tell people I listen to all of those things.” All of our interviewees prefer to stream rather than download audio files. Only one interviewee interacts with or contributes to his favorite radio shows, by rating episodes on iTunes (“it’s like a compliment”), emailing with producers to offer kudos, and sharing favorite episodes through social media.

**Survey**

The survey is composed of sixteen questions. Beyond basic demographics (gender, age, employment status), respondents were asked about commute times and discovery of public radio shows. Our assessment of survey results sought to better understand how listeners interact with public media,
whether discovering new shows, searching for content, or exploring particular topics of interest. Participants were not prescreened for whether or not they listen to public radio. Therefore, the survey includes the question, “Do you listen to public radio?” Of the 166 people who answered the survey, only the 96 respondents (n=96) who listen to public radio are included in this analysis. Notable results of the initial survey, along with further analysis, are discussed below.

**Demographics:** The majority of the respondents who listen to public radio are female (72%), and most respondents were between 26 and 38 years of age.

**Overall Access:** Of the 96 respondents (n=96) who listen to public radio: 46 listen via podcasts, 20 listen via smartphone applications, 44 listen via online streaming, and 73 listen to live broadcasts. The possible responses are not mutually exclusive — respondents were asked to choose all options that apply. Even with many new ways to access audio content, broadcast radio still dominates as the chosen method for accessing radio content.

<table>
<thead>
<tr>
<th>Age group:</th>
<th>Listen to shows via smartphone applications</th>
<th>Listen to shows via online streaming</th>
<th>Listen to shows via podcast downloads</th>
<th>Listen to shows via radio (i.e. live broadcasts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-23</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>24-26</td>
<td>4%</td>
<td>5%</td>
<td>6%</td>
<td>7%</td>
</tr>
<tr>
<td>27-29</td>
<td>2%</td>
<td>13%</td>
<td>8%</td>
<td>10%</td>
</tr>
<tr>
<td>30-32</td>
<td>4%</td>
<td>11%</td>
<td>8%</td>
<td>13%</td>
</tr>
<tr>
<td>33-35</td>
<td>6%</td>
<td>8%</td>
<td>9%</td>
<td>9%</td>
</tr>
<tr>
<td>36-38</td>
<td>2%</td>
<td>3%</td>
<td>6%</td>
<td>13%</td>
</tr>
<tr>
<td>39-41</td>
<td>1%</td>
<td>3%</td>
<td>2%</td>
<td>5%</td>
</tr>
<tr>
<td>42-45</td>
<td>0%</td>
<td>1%</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>46-48</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>4%</td>
</tr>
<tr>
<td>49 and older</td>
<td>1%</td>
<td>1%</td>
<td>3%</td>
<td>11%</td>
</tr>
<tr>
<td><strong>All age groups</strong></td>
<td><strong>21%</strong></td>
<td><strong>46%</strong></td>
<td><strong>48%</strong></td>
<td><strong>76%</strong></td>
</tr>
</tbody>
</table>

*96 respondents listen to public radio. When added together, percentages exceed 100%, because respondents were allowed to select multiple responses.

**Supplemental Modes of Access:** Newer technologies allow listeners to access content when and where they want. The ability to timeshift content and access digital media via mobile devices enables users to customize their listening experiences. 73 of the 96 respondents (76%) who listen to public radio listen via terrestrial broadcast waves. 51 of those 73 (70%) supplement terrestrial access with alternative listening methods (podcasts, smartphone applications, and online streaming). The chart below focuses on how these public radio listeners supplement their terrestrial access to content.
### Age group:

<table>
<thead>
<tr>
<th>Age group:</th>
<th>Listens to shows via smartphone applications:</th>
<th>Listens to shows via online streaming:</th>
<th>Listens to shows via podcast downloads:</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-23</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
</tr>
<tr>
<td>24-26</td>
<td>5%</td>
<td>7%</td>
<td>8%</td>
</tr>
<tr>
<td>27-29</td>
<td>3%</td>
<td>16%</td>
<td>11%</td>
</tr>
<tr>
<td>30-32</td>
<td>5%</td>
<td>15%</td>
<td>11%</td>
</tr>
<tr>
<td>33-35</td>
<td>8%</td>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td>36-38</td>
<td>3%</td>
<td>4%</td>
<td>8%</td>
</tr>
<tr>
<td>39-41</td>
<td>1%</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>42-45</td>
<td>0%</td>
<td>1%</td>
<td>3%</td>
</tr>
<tr>
<td>46-48</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>49 and older</td>
<td>1%</td>
<td>1%</td>
<td>4%</td>
</tr>
<tr>
<td><strong>All age groups</strong></td>
<td><strong>27%</strong></td>
<td><strong>60%</strong></td>
<td><strong>63%</strong></td>
</tr>
</tbody>
</table>

*96 respondents listen to public radio. Of those 96, 73 listen live via terrestrial radio. This table focuses on the supplemental listening habits of those 73 (n=73). When added together, these percentages exceed 100%, because respondents were allowed to select multiple responses.

**Search:** 63 of the 96 public radio listeners (66%) actively search for new radio content. Of the 63 who search for content, 42 searched for a particular episode, 29 searched for a particular subject, and 14 searched for similar programs.

**Discovery of New Shows:** 74 of the 96 public radio listeners (77%) answered “yes” to the question “Do you have a favorite radio show?” Few respondents discovered their favorite shows through social media (3 of 74), iTunes (4 of 74), or NPR’s website (5 of 74). 27 of the 74 public radio listeners who have a favorite radio show learned of their favorite shows via “word of mouth,” illustrating that personal networks play a significant role in introducing listeners to new content. However, 53 of the 74 of the respondents discovered their favorite shows when they “heard the program[s] on the radio.”

**User Engagement:** 22 of 96 public radio listeners have contributed to their favorite public radio shows. An average of 5 of those 22 contributed via either call ins, emails, or by contributing pieces. The most prevalent way that respondents contribute to shows is by “liking” the program on social media (14 of 22). It is possible that of the noted options (call ins, email, contributing pieces, and “liking” the program on social media), “liking” requires the least initiative — it is a relatively passive acknowledgement.

**Recommendations**

**Using Omeka, create a “one click” archival solution that adheres to PBCore**

The cost and timesaving features of Omeka make it a more practical choice than Collective Access, as well as scalable for independent producers. As the current standard for public broadcast material, PBCore should be a component of any broadcast archive system. A “one click” solution reduces the risk of user error and allows producers (and their interns) to archive digital content with minimal customization. This solution includes uploading files from hard drives to web-based repositories. The suggested repositories are the Internet Archive and SoundCloud.
Enable backup/transcoding of files at the Internet Archive with social sharing through SoundCloud

Founded with the mission of building an Internet library, the Internet Archive offers free, permanent storage of digital collections. Regardless of the device, most anyone with an Internet connection and a web browser can access files stored at the Internet Archive. Uploading archival content and associated metadata to the Internet Archive enables the Pop Up Radio Archive system to survive beyond the lifespan of a producer’s server. Additionally, the Internet Archive creates and stores files in multiple formats (.aif, .mp3, and .ogg), which assures that files can be streamed across browsers. 74% of surveyed public radio listeners access content in ways that rely either on mobile data plans or available bandwidth. Therefore, future iterations of Pop Up Radio Archive will transcode audio files into not only multiple formats, but also multiple bitrates, particularly the low bitrates required for mobile streaming. SoundCloud provides storage of audio files that can be accessed using the SoundCloud application for iPhone and Android.

Create a User Manual

Independent producers often lack dedicated technologists, and they rely on volunteers to organize data. A user manual would help make the Pop Up Radio Archive system meaningful to as wide an audience as possible. At a minimum, the user manual should cover installation, management, and use of the Pop Up Radio Archive system.

Continue to prioritize broadcast distribution and online access

According to industry statistics collected by Arbitron, public radio listenership peaks in the 55 to 64 age range. Terrestrial broadcast radio is still the most common way of accessing content (73% of our survey participants listen to content as it airs). According to a 2010 study conducted by Jacobs Media, 35% of public radio listeners own a smartphone, and 90% use the Internet. As millennials and other younger generations age, their Internet habits will inform how they access public radio content. Additionally, the younger sample profiled in our survey demonstrates that listeners are using new technology to access radio content. Online and mobile access to content should be a high priority as producers plan for distribution. Sixty-five percent of public radio listeners consume shows while commuting to and from school or work, via either podcasts or live broadcast. Over three quarters of that 65% listen to live broadcasts. The most common way that respondents discover their favorite shows is by listening to live broadcasts. Those who commute in automobiles are captive audience members who, if they like a program, will tune in every day at the same time.

Administer a follow-up survey

After the initial installation for The Kitchen Sisters is complete, and their archive has been populated with the Hidden Kitchens series, the Pop Up Radio Archive team will administer another survey with wider reach, in part with the intention of surmising whether respondents’ overall technological proficiency correlates to the way they access radio content.
Works Consulted


Bell, Erin Personal Interview. 9 Dec. 2011. Phone.


MacDonald, Matt, and Rekha Murthy. PRX. 4 Nov. 2011. Phone.


“Representing People’s Names in Dublin Core.” Web. 6 Apr. 2012.

Vann Malssen, Kara. Personal Interview. 2 Feb. 2012
Acknowledgements

Our work would not be possible without guidance and contributions from the following:

Erin Bell, for providing the boilerplate for our PBCore Element Set plug-in.
Nathan Dalton, for always making everything happen.
Patty Fung, for her keen archivist’s instinct helping us prepare Hidden Kitchens content.
Mario Furloni, for inspiring us and the Berkeley Big Ideas voters with the moving images he captured on Kearny Street.
Bob Glushko, for holding us accountable while being the most hospitable adviser we could imagine.
Ian Hill, for talking with us about what user engagement really means.
Internet Archive, for saving all the bits.
Berkeley’s School of Information, for supporting this project with grants, facilities, and goodwill.
Members of the fall 2011 INFO228 course at the I School, for contributing to our review of the broadcast ecosystem.
Jake Johnson, for being our guide to the Internet Archive’s mysterious staging server.
The Kitchen Sisters, for without these accidental archivists we would never be archivists ourselves.
Ray Larson, for his patience with our endless questions (and sometimes exasperation).
Dave Lester, for stepping in to help us whip our Omeka plug-ins into shape.
Lori Lindberg, for planting a seed and lending her archival ear from the earliest stages of this project.
Kris Carpenter Negulescu, for invaluable advice and perspective on how to make a system that works.
Jacob Portnoff, for basically writing a WordPress plug-in overnight.
Dave Rice, for always answering our PBCore questions with responses ten times as long — and informative.
Nancy Roberts, for being the best Big Ideas mentor a team could ask for.
Sam Robinson, for his thoughtful comments and questions as he helped catalog Hidden Kitchens content.
SoundCloud, for making our lives easy with a premium membership and lots of support.
Francesco Spagnolo, for reminding us of the magic of physical objects.
Marcos Suerio Bal and the WNYC Archive, for providing sample PBCore records and general metadata advice.
Evan Tennenbaum, for when you walked into the Kitchen Sisters office, the rest was history.
Pamela Vadakan, for always being willing to chat about the future of audiovisual preservation. And how are you doing?
Daniel Vizzini, for we extend our utmost gratitude to this freelance engineer-turned-integral member of our team.
Jeff Zych, for his fine teaching skills and patient ear.

Pop Up Radio Archive is a community effort that spans institutions, airwaves and time zones. Deep gratitude to all of our interviewees, and the countless people whose names do not appear on this list.