

2016

Opening Access to Fresh Air's Archives

Melody Kramer
Wikimedia Foundation

Anu Paul
WHYY

Follow this and additional works at: <http://digitalcommons.kennesaw.edu/provenance>

 Part of the [Archival Science Commons](#)

Recommended Citation

Kramer, Melody and Paul, Anu, "Opening Access to Fresh Air's Archives," *Provenance, Journal of the Society of Georgia Archivists* 34 no. 1 (2016) .

Available at: <http://digitalcommons.kennesaw.edu/provenance/vol34/iss1/6>

This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in Provenance, Journal of the Society of Georgia Archivists by an authorized administrator of DigitalCommons@Kennesaw State University. For more information, please contact digitalcommons@kennesaw.edu.

Opening Access to *Fresh Air's* Archives

Anu Paul and Melody Kramer

Introduction

For more than 35 years, *Fresh Air with Terry Gross* has consistently been one of public radio's most popular programs. The program reaches a global audience, with over 620 public radio stations broadcasting *Fresh Air*, and three million people downloading the podcast each week.

The *Fresh Air* audio archive is one of the biggest collections within American public radio. More than 8,000 interviews are contained in the archive, which was recorded and preserved in various audio formats—from analog reels to digital files—over the past 40 years.

Producing a daily public radio show while preserving past content and metadata for public access is both time-consuming and difficult, particularly for a small staff untrained in archival techniques. Over the past decade, however, *Fresh Air* has made a concerted effort to preserve its collection by digitizing and standardizing the metadata in its archives in order to ensure future public access to its material. In this paper, we detail why *Fresh Air's* archives and other audio-dominant collections deserve such urgent attention, and present a case study for how a small public radio institution successfully managed an archival project and rethought its asset management strategy. We hope the recent efforts detailed in this paper will help preserve and broaden access to *Fresh Air's* archives. We believe this work is important because of the breadth of the collection, but also time-sensitive because of how the collection was originally recorded and preserved.

Challenges of Audiovisual Preservation and Access

Fresh Air's earliest episodes were taped on analog reels and had what is called “sticky shed syndrome,” which occurs when humidity deteriorates the magnetic tape's binder, leaving a sticky residue on the tapes. Tapes that have this sticky material on them may be damaged when played back on a reel-to-reel machine. In addition to sticky shed syndrome, there has been growing concern in the audiovisual preservation community and among WHYY radio station staff about playback obsolescence. Digital audio tapes (DAT) and reel-to-reel tapes require specialized machines to be played back and access content. Many of these machines are no longer produced, which increases the difficulty of preserving and migrating content.

These are issues that the audio preservation community has long recognized. Academic literature contains many articles about archival repositories and the looming danger in storing material in non-digitized formats. In his 2015 [article about media preservation](#), Mike Casey of Indiana University's Media Digitization and Preservation Initiative stated:

Media preservation has reached a crisis point for content carried on physical audio and video formats ... [and] archival media collections could soon be considered highly endangered. ... Archives hold very large numbers of analog and physical digital recordings on obsolete audio and video formats that are actively degrading, some of which contain content with high research value. We

have a relatively short time to save these recordings.

Casey warned of a “15-20 year window of opportunity to digitally preserve legacy audio and video recordings” before material starts to degrade and equipment becomes too costly or obsolete.¹

However, there are some practical limitations in implementing a robust digital archives program at smaller and cash-strapped institutions, as Joseph A. Williams and Elizabeth M Berilla point out in their case study discussing a digital archive at the Stephen B. Luce Library at SUNY Maritime College. After mentioning Mike Kastellec’s excellent list of limitations—including “quickly antiquated technology, data redundancy, selection criteria, access issues (virtual, physical, and temporal), a discombobulated legal structure, and funding”—they write:

We are left with, therefore, numerous difficulties and no agreed-upon best practices to follow in establishing a digital institutional archives, especially at a small institution with limited funds and staff. In this regard, practicality often trumps theory, and a middle ground of digital content management must be contemplated. Because of their own idiosyncrasies, institutions must cherry pick among best practices for what works for them. In essence, every institution must develop a unique plan.²

Fresh Air has developed and evolved its plan over the past decade, and like Williams and Berilla pointed out, practicality has often played a role in determining best practices. We should note that there are several challenges faced when archiving a public radio program that is produced and broadcast on a daily basis. Significant issues include prioritizing and selecting materials for digitization; developing efficient workflows for cataloging that can be integrated with the daily production and broadcast of the show; creating best practices for creating appropriate descriptive and technical metadata; and designing a searchable asset management system and the technological infrastructure to support an Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) compliant system. In later sections, we detail how we handled these challenges.

Existing Efforts in Public Broadcasting

Before undertaking our own archival work, we looked to other public media institutions to understand how their radio and television programs were archived. We discovered that while stations thought that archiving their programs was worthwhile, only a handful of public radio initiatives—like WGBH’s Open Vault project, the American Archive of Public Broadcasting (AAPB), and NPR’s Research, Archives and Data strategy team (RAD)—had created digital asset management systems that allowed digital files to be ingested, catalogued, and disseminated to a publicly accessible site.

¹ Mike Casey, "Why Media Preservation Can't Wait: The Gathering Storm," *International Association of Sound and Audiovisual Archives* 44 (2015): 14-22, accessed May 18, 2016, http://www.avpreserve.com/wp-content/uploads/2015/04/casey_iasa_journal_44_part3.pdf.

² Joseph A. Williams and Elizabeth Berilla, “Minutes, Migration, and Migraines: Establishing a Digital Archive at a Small Institution,” *The American Archivist* 78, 1 (2015): 84.95.

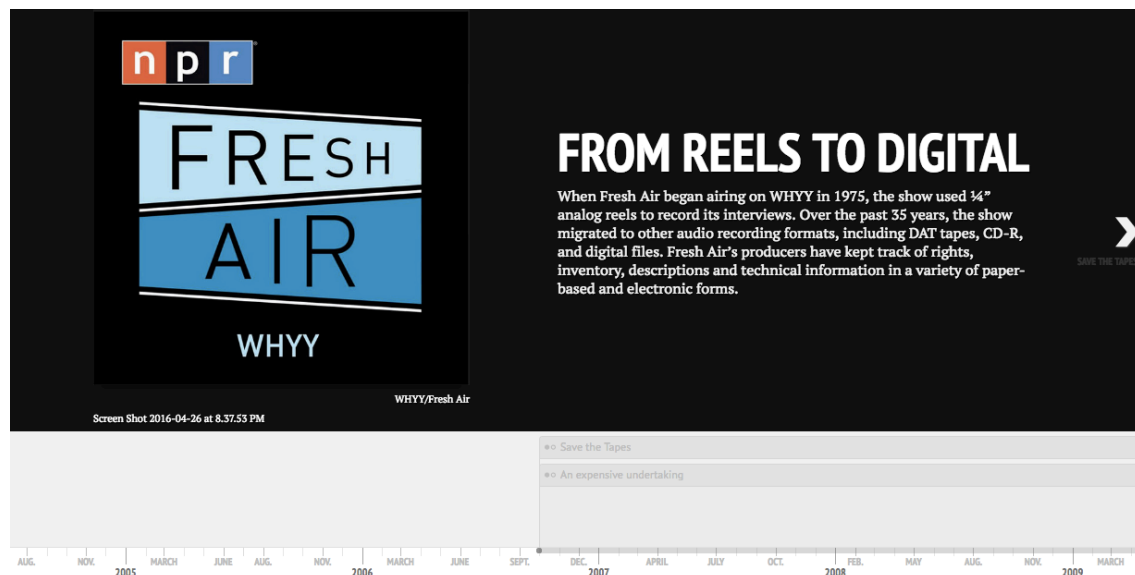
We also learned that there was no one-size-fits-all solution in public broadcasting; institutions varied in terms of size, structure, budget, and staffing, and so did their archival needs and requirements.

Talking with WBGH and NPR was very useful in the early stages of the *Fresh Air* project. These conversations helped us understand the technical challenges of building a digital asset management system from scratch; the cost of digitization of reel-to-reel tapes and CDs; sources of funding; and the staffing required to successfully implement such a large scale project. We clearly understood that the costs and effort associated with digital preservation projects would be significant, but we also knew that the rewards would be enormous. Migrating from “magnetic tape-based” formats to more powerful and easy to use “file based systems” would allow the production team greater flexibility in reusing, searching, and disseminating the *Fresh Air* program.

Our research also indicated that we needed to approach the project in two phases. Phase one would include the digitization of reel-to-reel analog tapes and CDs and the development of a simple database based on PBCore 1.1, a metadata schema designed for public broadcasting materials. Phase two would then focus on building a robust searchable database based on the PBCore 2.0 metadata schema that would allow the entire collection—containing both the files digitized from analog media as well as born-digital audio files—to be housed in a single location with a logical structure.

A Multi-Phased Project to Evolve an Asset Management Strategy

In this timeline, we detail how *Fresh Air* evolved its asset management strategy, and how the information was structured.



We will also go into more detail within this article.

Phase 1

A decade ago, WHYY engineers worked with *Fresh Air* producers to create a detailed inventory of all recorded episodes of the program. This work also identified the location of *Fresh Air*'s earliest episodes, which were recorded on both reel-to-reel tapes

and CDs. Most of the tapes were stored onsite in the WHY Y archive but some of the older episodes—from 1976 through 1980—were in an offsite storage facility.

Once staff had a detailed inventory of the tapes, it was easy for the producers and engineers to appraise the *Fresh Air* collection for its strengths and weaknesses. During the appraisal process the at-risk reel-to-reel tapes and CDs—recorded between 1976-2000—were identified and prioritized for digitization. The shows produced after 2007 were already taped in a digital format so they were not considered a priority for this phase of the project.

A *Fresh Air* program is composed of three different parts: the raw version of the interview which has not been edited; the original broadcast of the show—called an aircheck—which airs live and may contain tiny errors; and then a lightly-edited version of the original broadcast which corrects any timing or musical cue mistakes and is considered the “archival” version of the show. We had an interesting discussion internally about which of these three parts should be archived. It was decided that given our limited resources, the aircheck would be digitized and archived, and we would review the unedited raw version of the program and the archival version of the show at a later date.

The initial digitization was done by a full-time audio engineer, reporting to the media technology manager. The engineer worked for two years digitizing the reel-to-reel tapes as well as migrating audio content from CDs to the database. In the second phase of the project, a full-time project manager and two part-time metadata specialists were hired to enhance metadata and implement a robust asset management system.

Phase 2

By 2014, WHY Y was ready to implement the next phase of the project. In January 2015, WHY Y received a grant from the Council on Library and Information Resources (CLIR) to enhance the metadata of 8,000 *Fresh Air with Terry Gross* shows that aired between 1976 and 2015, and to make those shows available for research and scholarship through WorldCat.

The *Fresh Air* archives project team, in collaboration with partners at Drexel University (DUL) and TelVue Corporation, created a metadata application profile based on the PBCore 2.0 schema and designed an asset management system that would store the entire *Fresh Air* collection in a single location with a logical structure and export the database records to WorldCat for public access, with direct links to the audio files.

Metadata Enhancement

Designing the metadata involved several meetings with Drexel University consultants, *Fresh Air* producers, and developers at TelVue. These discussions allowed the archives team to understand the current and future needs and requirements of *Fresh Air* staff. They also helped to identify the metadata schema that would be most appropriate for a collection like *Fresh Air*. We decided to use PBCore 2.0 schema because it is a preferred schema for media archives and “it is designed to handle the descriptive details and hierarchical relationships needed for preservation and management of time-based media such as audio recordings and other audio visual

materials.”³ Each *Fresh Air* show has two or more segments. One segment is always an interview while the other segment can be either a review or another interview. We wanted our metadata to reflect this aspect of the show and PBCore 2.0 can handle complex hierarchical relationships like segments or episodes that are part of a larger series.

With the help of our collaborators at Drexel University, we designed a metadata application profile based on PBCore 2.0. We wanted the application profile to address three main issues. First, it had to adhere to standard vocabularies like Virtual International Authority File (VIAF) and Faceted Application of Subject Terms (FAST) for authoritative data about contributor names and subject terms. Second, it needed to enforce consistency in the application of PBCore elements, “the syntax of metadata values, and the usage of controlled vocabularies.”⁴ And finally, we wanted the metadata to be shareable, since one goal of the project was to expose the metadata for public discovery through WorldCat.⁵

Beginning in November 2015, Drexel University catalogers began establishing unique identities in the Library of Congress Name Authority File (LCNAF) for *Fresh Air* contributors. Anytime the archives team came across a contributor who was not in the LCNAF database, we recorded the name and some contextual information in a Google spreadsheet that we shared with the DUL catalogers. The DUL catalogers then created an authoritative listing for that person in LCNAF.

Technical Infrastructure

As mentioned earlier, the database we developed for the digitized files was based on PBCore 1.1. This database was later expanded so that we could take advantage of the relationships and granularity of PBCore 2.0. For instance, `<pbcorePart>` was added so that detailed metadata could be created for each segment of the show, and `<pbcoreRightsSummary>` was added to clearly indicate the rights associated with the usage of audio files.

The developers at TelVue worked on expanding the database and designed a metadata entry interface that would be simple and easy to use by professional catalogers as well as producers of the show. We wanted to make sure that the producers incorporated cataloging of the show as part of their production workflow. This was also key to ensuring that the project would be sustainable beyond the duration of the grant.

To ensure producers found it easy to use authoritative data, we integrated VIAF and FAST API into the data entry interface. This feature forced the catalogers and producers to choose controlled vocabularies and authoritative contributor names from an authorized list when they entered data in the contributor or subject fields. It also

³ Marsha Maguire, “Description of Audio Recordings,” in *ARSC Guide to Audio Preservation*, ed. Sam Brylawsky et al. (Washington, DC: CLIR, 2013): 77-105, accessed September 26, 2016, <https://www.clir.org/pubs/reports/pub164/pub164.pdf>.

⁴ Diane I. Hillmann, “Getting the word out making digital project metadata available to aggregators,” *First Monday* 11, no. 8 (August 2006), accessed September 26, 2016, <http://firstmonday.org/ojs/index.php/fm/article/view/1385/1303>.

⁵ Sarah L. Shreeves, Jenn Riley, and Liz Milewicz, “Moving towards shareable metadata,” *First Monday*, 11, no. 8, (August 2006), accessed September 26, 2016, <http://firstmonday.org/ojs/index.php/fm/article/view/1386/1304>.

enforced consistency in the way the contributor names and subject terms were applied to a record.

Access

One of the important aspects of the project was to ensure that these interviews would be available to the public as well as researchers. For this reason, one of the key components was to develop an OAI-PMH compliant system that will allow our metadata to be shared with WorldCat. Each PBCore record in the *Fresh Air* database is mapped to Dublin Core and exported as an XML file to be harvested by WorldCat. At the time of writing this article, we have approximately 1,800 records available on WorldCat. For each record, the following Dublin core fields—title, creator, contributor, date, type, rights, language, and description—are displayed on the WorldCat interface along with the streaming MP3 version of the interview. In addition, we are also creating a finding aid for the project which will be made available on the Philadelphia Area Consortium of Special Collections Libraries (PACSCL) website.

Workflow for Metadata Production

The *Fresh Air* archives team began creating the metadata for the 8,350 shows in October 2015. We began in reverse-chronological order because more recent shows had supporting documentation such as guest names, length of each segment, and transcripts in addition to the audio files. The metadata specialists listened to the audio while scanning the transcripts and then added subject terms to each segment. The workflow changed when we started archiving the older episodes which had little or no supporting information. In those episodes, Terry Gross interviewed local Philadelphia-based activists, artists, and musicians who were not nationally renowned personalities. Since we had very limited written documentation from that period, the catalogers had to listen to the entire program carefully to ensure we accurately recorded the guest name and to add relevant subject terms.

The team followed the Draft-Review-Publish format. Metadata created by the metadata specialists was reviewed by the project manager before being ingested into the database and published to WorldCat. To keep track of the metadata records, audio files, and rights and name authority work associated with the file, the *Fresh Air* team used the collaborative project management tool Trello.

With so many collaborators, we also wanted to select a communication platform that would allow us to have seamless and open communication with our partners who were located in different institutions. We identified Slack as this communication tool and have used it extensively to resolve technical and metadata issues.

Found in the Archives

Terry Gross has hosted *Fresh Air* since 1975, when it was a local show. Her interviews, and those of contributors Dave Davies and David Bianculli, have always covered a range of subjects, including literature, the performing arts, health, religion, and the issues of the day. In her own interviews with authors, artists, and entertainers, Gross's goal has always been to talk not just about their new work, but about the personal experiences that made them who they are and made their work what it is. Without shying away from difficult questions, she has created an atmosphere in which

guests feel comfortable thinking out loud, and where they find themselves drawing connections between their public personas and their inner selves.

Fresh Air has also regularly featured interviews with journalists, scholars, and experts in various fields aimed at putting the issues of our time into a larger political, cultural, and historical framework. Even in these exploratory interviews, Gross has brought the subject into sharper focus by encouraging her guests to reflect on what drew them to their work.

Fresh Air interviews often take surprising turns. What Gross assumed would be a brief chat with the late author Maurice Sendak about his new children's book, became a reflection on the recent deaths of his partner and friends, his own declining health, and his preoccupation with mortality. (20:01)⁶ The conversation left Sendak in tears, along with many listeners. Louis C.K., a comedian known for his irreverent, often lewd humor, choked up during an interview, explaining how a dark episode of his television series was inspired by the recent deaths of two self-destructive fellow comics. (45:56)⁷ An interview with the late *New York Times* columnist David Carr, about the impact of new media on journalism, evolved into a discussion about the prayer he carried in his wallet, despite his ambivalence toward religion. (37:33)⁸ These interviews testify to Gross's willingness to *listen* as well as question. There are thousands of interviews in *Fresh Air*'s archives that feel similar; the conversation takes an unexpected turn, or someone reveals something surprising and personal while talking to Terry Gross.

Archives staff have uncovered some gems while adding metadata to the database. Highlights have included:

- A 1989 conversation with *The Color Purple* author Alice Walker about her grandmother Kate's murder outside of a church when her father was eleven. "In some ways, it was the strange little secret of my family," she says. "Why was this beautiful woman killed ... and why was she blamed for her own murder?" (05:02)⁹
- A 1985 interview with civil rights leader James Farmer, on spending forty nights in a Jackson, Mississippi, jail cell during the Freedom Rides. "We were not going to bail out. We were going to stay in as long as we could stay in ... and that ended up being forty days and forty nights," he says. "They quickly found out that we were not going to bail out right away, and then what they wanted to do was make it so uncomfortable for us that we'd wish we had never come, and we'd stop others from coming." (04:27)¹⁰

⁶ Terry Gross, "Maurice Sendak: On Life, Death and Children's Lit," *Fresh Air*, NPR, September 20, 2011, accessed September 25, 2016, http://pd.npr.org/anon.npr-mp3/npr/fa/2011/09/20110920_fa_02.mp3.

⁷ Terry Gross, "Comedian Louis C.K. Laughs Through Difficulties," *Fresh Air*, NPR, July 8, 2011, accessed September 25, 2016, http://pd.npr.org/anon.npr-mp3/npr/fa/2011/12/20111213_fa_01.mp3.

⁸ Terry Gross, "David Carr: The News Diet of a Media Omnivore," *Fresh Air*, NPR, October 27, 2011, accessed September 25, 2016, http://pd.npr.org/anon.npr-mp3/npr/fa/2011/10/20111027_fa_01.mp3.

⁹ Terry Gross, "Interview with Alice Walker," *Fresh Air*, NPR, May 1, 1989, accessed October 17, 2016, <http://www.worldcat.org/oclc/958461456>.

¹⁰ Terry Gross, "Interview with James Farmer," *Fresh Air*, NPR, September 17, 1985, accessed October 17, 2016, <http://www.worldcat.org/oclc/959924338>.

- A rousing conversation with Julia Child about the rules of French cooking: “I felt that we needed fuller explanations so that if you followed one of those recipes, it should turn out exactly right. ... My feeling is that once you know everything and have digested it, it becomes part of you.” (04:00)¹¹
- The surprising candidness of political operative G. Gordon Liddy, who, on May 13, 1980, explained the extreme and unorthodox ways he strengthened his will and overcame his fear of rats. (02:23)¹²
- Steve Jobs, in 1996, on pushing for Apple to have a mouse along with their computer. “It was obvious that you needed a pointing device, and that a mouse was the best one. ... We found that people would push the wrong button [if there were three buttons] or they would look at the mouse instead of the screen. So we got it down to one button so you could never push the wrong button.” (06:23)¹³

Why Public Radio Archives are Important for the Future

The significance of public radio archives to scholars and educators examining the past cannot be understated, but it is equally important to note that public radio broadcasts play an important role in newsgathering today. As Mary Kidd, the National Digital Stewardship Resident (NDSR) at New York Public Radio, writes, “Reports or stories bolstered with material from the radio archive creates great news.” She points to a recent example of a WNYC newscast segment by reporter Andrea Bernstein examining Hillary Clinton’s evolving views on gay marriage over her political career:

Her report provides snippets found throughout her extensive collection of recordings documenting Hillary Clinton’s political career. Bernstein and NYPR’s Head Archivist Andy Lanset worked closely together to mine the archive’s PBCore-backed database, CAVAFY, as well as the actual physical archives for items ... to put together a comprehensive timeline of Clinton’s public remarks on marriage rights. This story ... demonstrates how archives are not just safekeepers of the past, but informants of the present. ... [T]he archive can continue to inspire its producers and staff and work alongside the stations’ output, rather than behind it.¹⁴

Fresh Air also uses its archives regularly for obituaries and to mine content for new interviews with repeat guests. Having access to these interviews provides producers with additional material and context, and the audience with information from sources in their own words.

¹¹ Terry Gross, “Interview with Julia Child,” *Fresh Air*, NPR, October 7, 1983, , accessed October 17, 2016, <http://www.worldcat.org/oclc/959925340>.

¹² Terry Gross, “Interview with G. Gordon Liddy,” *Fresh Air*, NPR, May 13, 1980, , accessed October 17, 2016, <http://www.worldcat.org/oclc/959925100>.

¹³ Terry Gross, “Steve Jobs: ‘Computer science is a liberal art,’” *Fresh Air*, NPR, October 6, 2011, accessed October 17, 2016, <http://www.npr.org/2011/10/06/141115121/steve-jobs-computer-science-is-a-liberal-art>.

¹⁴ Mary Kidd, “Digital Stewardship in a Radio Archive: An NDSR Project Update,” *The Signal*, January 5, 2016, <http://blogs.loc.gov/thesignal/2016/01/digital-stewardship-in-a-radio-archive-an-ndsr-project-update/>.

What's Next

The *Fresh Air* archives project lays the groundwork for creating new tools and opportunities for the public to explore and interact with the entirety of the *Fresh Air* archives. We not only envision our audience listening to the end results, but actively helping us enhance the way the archive can be discovered and shared with new audiences. We anticipate a future platform that will increase access to and visibility of the collection through an interactive public interface supporting crowdsourced metadata and tools to help reveal relationships between various data elements.

This was, in part, why we made certain decisions involving elements and attributes described in PBCore. Staff spent a lot of time deciding how we could use the structure of PBCore to allow for our future vision for the archive. We picked the schema because it could handle complex structures and hierarchical relationships contained with the *Fresh Air* program; we made sure our metadata is shareable by ensuring use of controlled vocabularies and consistency in the application of metadata values. By taking advantage of these unique features of PBCore, we hope to have built a strong foundation that will open up countless possibilities for the archive. These include:

- An app on a mobile phone that suggests *Fresh Air* episodes related to specific places between locations A and B, or an app that suggests an interview based on your location. “In West Baltimore? Listen to Terry’s conversation with Ta-Nehisi Coates.”
- The ability to link *Fresh Air* interviews about live news events to local station broadcasts about the same event. For instance, after listening to a *Fresh Air* interview on Hurricane Katrina, listeners might be directed to hear the live broadcast from WWNO on August 24, 2005.
- The ability to search the *Fresh Air* archives by emotion or activity. (Example: “I need a good laugh. Find me *Fresh Air* episodes that will put a smile on my face.” or “I’m driving in the car with my kids. I want to hear five episodes that we’ll all enjoy.”)
- The ability to find *Fresh Air* interviews with specific authors while searching local library websites for a particular book.
- The ability to link structured data from *Fresh Air* to structured data in Wikipedia so that listeners can discover that guest Maurice Sendak was born in Brooklyn and then query “I want to listen to more interviews with guests from Brooklyn,” and be directed to a conversation with Mel Brooks.
- The ability to find *Fresh Air* episodes that would work in a high school or college syllabus.

All of these are ways to deepen the connection between the public and WHYY. Enhancing the archive through metadata ensures it will be preserved, accessible, and listened to for decades to come.

Anu Paul is the Project Manager for *Fresh Air* Archive (CLIR) at WHYY. She manages the day-to-day operations of the archive including metadata creation, development of an asset management system, access, and outreach. Before joining WHYY, she worked at the Scholarly Communication Center of Rutgers University Libraries. She holds a master's degree in library and information studies from Rutgers University.

Melody Kramer formerly produced *Fresh Air's* online and social media content. She is a Peabody-award winning audio producer with bylines in *Esquire*, *National Geographic*, and on NPR, and a former Visiting Nieman Fellow at Harvard University, where she researched the future of public media membership. She currently leads audience growth and development at the Wikimedia Foundation and writes a weekly column for the Poynter Institute on the future of news.