



Subscribe, Rate and Preserve Wherever You Get Your Podcasts

Mary Kidd , Sarah Nguyen & Erica Titkemeyer

To cite this article: Mary Kidd , Sarah Nguyen & Erica Titkemeyer (2020): Subscribe, Rate and Preserve Wherever You Get Your Podcasts, Journal of Archival Organization, DOI: [10.1080/15332748.2020.1769997](https://doi.org/10.1080/15332748.2020.1769997)

To link to this article: <https://doi.org/10.1080/15332748.2020.1769997>



Published online: 17 Jul 2020.



Submit your article to this journal [↗](#)



Article views: 25



View related articles [↗](#)



View Crossmark data [↗](#)



Subscribe, Rate and Preserve Wherever You Get Your Podcasts

Mary Kidd^{a,b} , Sarah Nguyen^{a,c} , and Erica Titkemeyer^d 

^aPreserve This Podcast, New York, New York, USA; ^bPreservation and Collections Processing, New York Public Library, New York, New York, USA; ^cUniversity of Washington, Seattle, Washington, USA; ^dUniversity Libraries, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

ABSTRACT

Contemporary podcasting is used not only to extend the reach of existing radio stations and personalities but also by hundreds of independent content creators who can bypass the strictures of traditional distribution channels to deliver audio programs directly to online audiences. Preserve This Podcast (PTP) was an Andrew W. Mellon-funded outreach and education project designed to raise awareness among indie podcasters of the need to proactively preserve their content and teach basic techniques for podcast preservation. This paper discusses the preservation challenges faced by indie broadcasters and strategies the PTP team used to address these through face-to-face workshops, online outreach, and web-based tools, using the present threat of “podfade” to raise larger issues surrounding the urgency and challenges of preserving born-digital audio content.

ARTICLE HISTORY

Received 18 February 2020
Accepted 12 May 2020

KEYWORDS

Podcasting; digital preservation; rss; web archiving; born-digital audio; independent podcasting

Introduction

Podcasts are a distinct media format and preserving them poses unique challenges to both creators and archivists. Over the course of 2018–2020, the Preserve This Podcast (PTP) project devised a personal digital archiving curriculum targeted toward independent or “indie” podcasters. The curriculum served as the foundation for a suite of tools, resources, and workshops aimed at teaching independent podcast producers actionable, low-cost, and easily implementable steps toward preventing catastrophic file loss. It also explored options for how podcasters could rely less on third-party platforms to sustain their podcast’s discoverability and accessibility through open tools and better control of metadata. The lessons put forth by PTP demonstrate how indie podcasters can take personal preservation measures; however, institutions, organizations, and other funded agencies with existing digital preservation infrastructures

CONTACT Mary Kidd  kiddmary@gmail.com  Preserve This Podcast, New York, New York, USA; Preservation and Collections Processing, New York Public Library, New York, New York, USA

© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC

may also perform wide-scale preservation of podcasts. Looking at existing broadcast radio preservation initiatives could offer a guide in scaling these efforts as both radio and podcasts are known to share similar “practices, cultures, institutions and distribution systems.”¹ At the same time, the technologically complex and dispersed nature of podcasting will pose a difficult preservation challenge to those tackling born-digital audio preservation. The PTP project complements broadcast preservation efforts with personal digital archiving strategies in order to address one of the defining mass media of the last two decades and its greatest chance for preservation success.

The Preserve this Podcast Curriculum

The Preserve This Podcast project is based on a foundational personal digital archiving² curriculum introducing to podcasters how and why their digital files are at risk. The project provides practical, simple, and time-sensitive steps for podcasters to future-proof their files, protect against catastrophic file loss, and promote post-production access to some or all podcast components (i.e., raw and edited audio and project files, websites, transcripts, release forms, and artwork). In developing the curriculum, PTP engaged with and received feedback from a number of experienced practitioners working in the digital, audio and personal digital archiving and preservation fields.³ The resulting curriculum focused on four key lessons: file and folder management, storage and backups, metadata, and self-hosting. Each lesson was delivered by individual or a combination of exercises, guides, reading lists, survey datasets, and podcast episodes published by the PTP team.

Taking the personal digital archiving approach further, the PTP curriculum considers how indie podcasters work during a regular podcast production cycle, in order to understand where preservation measures could be incorporated into established workflows. In this way, podcasters could see preservation benefits to both their podcast’s longevity and to improving their day-to-day operations.

Podcast Technology 101

This section will cover a brief history of podcasting and its foundational technology. Then, a review of the ecosystem of software that podcasters, producers, and listeners use will shed light on the quickly evolving technological landscape of podcasting, which consists of layers of interdependent proprietary software and platforms.

```

<code>

<title>Episode 1: Time to Take Notice</title>

<link>https://ia803005.us.archive.org/28/items/Preserv
eThisPodcastEpisode1/PreserveThisPodcast\_Episode1.mp3</link>

<pubDate>Thu, 21 Mar 2019

09:00:00+0000</pubDate>

</code>

```

Figure 1. A snippet of PTP’s RSS feed for the podcast, which shows its RSS feed URL. The code tells podcatchers the title of the episode, the location of the episode audio file on a streaming server, and date of publication.

A Brief History of How Digital Audio Files became Podcasts

There is no prescribed definition of the podcasting genre in terms of length, style of sound, geographic origin, or other characteristics; instead, podcasts may be defined, simply, as audio files distributed via Rich Site Summary or Really Simple Syndication (RSS). There are many existing canonical resources⁴ that define and maintain RSS,⁵ but at its core RSS is a variation of extensible markup language (XML),⁶ carrying the same “encoding, transfer, and occasional internal system storage mechanism for metadata.”⁷ RSS requires defined tags, unlike XML which has more flexible attribute-value assignments.⁸ Figure 1 provides a snippet of PTP’s own RSS feed, which shows the part of the code that effectively points systems to the location of one podcast audio file hosted on the Internet Archive. As a content delivery mechanism, RSS allows systems (e.g., podcatchers) to automatically locate and download newly-updated content to the top of a single feed, obviating user website visits one-by-one to receive updates. Originally, RSS was fabricated to facilitate convenient and personalized distribution of text-based news to users of the once-popular Netscape browser.⁹ This enabled the rise of weblogs, commonly known as “blogs”, that allowed for the syndication of decentralized and independently-produced writing and image content.¹⁰ In the year 2000, RSS was enhanced to deliver media files, resulting in the creation of “audioblogs.”¹¹ In seeing that podcasting technology’s roots stem from both developments in audio technology and web content delivery, PTP took a hybrid web and digital audio preservation approach in forming its curriculum.

The Podcast “Systems Ecosystem”

PTP built a curriculum around how indie podcasters approach their day-to-day work by developing a deep understanding not only of RSS, but the underlying systems that use RSS to create the current podcast production and distribution universe. This research revealed that both podcasters and listeners are a part of and rely on a number of commercial, proprietary systems to produce and consume podcasts. These systems make up a large and multi-layered podcast “systems ecosystem.” What follows is an overview of six common systems, who uses them, and how they are interdependent on one another:

Firstly, podcast producers and audio engineers use a **digital audio workstation (DAW)** software or app (e.g., Audacity, Reaper, Final Cut Pro) to produce various audio files, including the MP3 files uploaded to podcast platforms for playback. Some DAWs can also be used to embed metadata into files; this metadata, in turn, may be automatically piped into a podcatcher directory’s RSS metadata fields. Some DAWs output proprietary project files, which may contain file path references specific to where they live within a podcaster’s workstation’s file and folder system.

The **podcast hosting platform** provides access and discovery services to podcasters, such as generating and hosting an **RSS feed**, streaming, website hosting, and data analytics (e.g., Simplecast, Audioboom, Podbean, etc.). The RSS feed, to be described in more detail in the following sections, is a file that contains structured XML metadata about podcasts. These are more commonly auto-generated by podcast hosting platforms as a service but can also be hand-coded. RSS feeds are collected by podcatchers, who in turn read, parse and deliver data to listeners. Most importantly, they contain URL references to podcast audio files.

Finally, **podcatchers**, which can take on the form of a piece of software program, a website, or smartphone application, enables listeners to subscribe, stream, and download podcast episodes (e.g., RadioPublic, iTunes, Spotify, etc.). Podcasters submit their RSS feed to podcatchers by filling out a form. The podcatcher, in turn, reads and uses the RSS to relay audio and other podcast metadata to listeners. They may also provide data analytics.

Podcasters more often than not also maintain a **website**, or choose to be a part of a network, to further their listener reach; however, these components are not required to distribute podcasts. These websites, used to host and make primary and supplementary podcast content available to listeners, can be an important source of contextual metadata, containing creator biographies, and transcriptions, and may also link to external sponsor, event, and donation sites. Websites may be hosted on a blogging or podcast platform such as Wordpress, GitHub Pages, or RadioPublic, to name a few. Many chose to use their website to host their RSS feed. **Podcast networks**

(e.g., Radiotopia, Gimlet, Pineapple Street, etc.), similar to radio networks, host, produce, or sell advertisements for podcasts. Usually, they fund the podcasts directly, as well.

With the basic podcasting ecosystem defined, the PTP curriculum focused on raising awareness about how podcasters could lessen some or all technical, financial or infrastructural dependencies on one or all podcasting systems by illustrating how and why each of these systems can or have historically failed. Failure happens, in large part, because most systems rely on the market viability of their parent companies. A perfect example of this is Soundcloud, a popular and free audio and podcast hosting platform, used especially by independent creators, that nearly went bankrupt in 2017.¹² Archivists and others focusing on podcast preservation will also need to pay close attention to shifts in podcast technology, because podcasts are shaped as much by their creators as the technological trends that back their distribution; arguably, these models shape the way cultural production is governed.¹³

Platform Dependencies

If nascent podcasting was defined by open and free web technologies like RSS, present-day podcasting seems to be more defined by a shift toward increased dependence on proprietary platforms, which will have implications for future podcast preservation efforts. Most podcasters today will subscribe to a podcast hosting platform for a recurring fee, in exchange for services that guide podcasters through all aspects of podcast distribution, hosting, promotion, monetization, and analytics. There are even platforms, such as Anchor.fm, that “[enable] users to record audio directly via a smartphone app, edit the file and add music via app, [and] upload it to Anchor’s servers”, which may mean podcast hosts could eventually supplant client-side DAWs.¹⁴ For indie podcasters, this point-and-click direct creator-to-consumer setup is ideal because it lowers the technical entry barrier into an otherwise complex world of automated distribution and play-out systems. At the same time, the conveniences afforded by platforms have disconnected podcasters from how their metadata, audio files, and RSS work under the hood. These layers of abstractions between services and end-users can contribute to lack of end-user agency, especially during instances of systems failure.¹⁵

Grant Deliverables

PTP sought to deliver the curriculum through highly accessible audio, print, and web offerings to reach and engage as many participants as

possible. The asynchronicity of the podcast and booklet gave flexibility to those who could not attend the in-person workshops conducted in select cities.

The “Preserve This Podcast” Podcast

The PTP team produced and published a five-part educational podcast miniseries featuring interviews with active podcast creators performing the lessons, weaving a personal digital archiving narrative into the language and communication medium of podcasts, in order to speak directly to their target indie podcaster audience. The first episode of the “Preserve This Podcast” podcast was launched in March 2019, with subsequent episodes released bimonthly through June 2019, and resulted in a total of 5,800 downloads¹⁶ with several positive ratings.¹⁷ The podcast was professionally produced by PTP Project Co-Lead, Molly Schwartz, in collaboration with editor Allison Behringer, sound engineer Dalton Harts, and soundtrack producer Breakmaster Cylinder. The podcast worked in tandem with an illustrated exercise booklet or “zine”, encouraging listeners to turn to specific pages; the zine, in turn, referenced specific episodes within the podcast. The printed zine, also available online through the PTP website, provided a tangible object that could be taken home by participants after workshops.¹⁸ Both the zine and podcast proved to be highly successful in teaching preservation to novices, and according to post-workshop survey feedback, 64 percent of respondents felt the zine was an effective way to learn about preservation concepts.¹⁹

The team also coordinated and conducted thirty-two in person and virtual workshops for major media and podcast centers, public and university libraries and archives, and other community spaces. During the two- to four-hour workshops, the instructors utilized the zine and podcast as teaching tools. The workshops were a critical outreach and community engagement component, allowing PTP team members to interface with podcasters. With them, PTP guided podcasters to develop a “Personal Podcast Preservation Plan”, a written commitment to short- and long-term steps toward preserving their podcast. Most of the workshops (with the exception of those held at major library, archives, and podcast conferences) admitted participants free of charge and were held outside of regular work hours, in order to be accommodating to freelancers’ schedules. With an eye toward continually improving the curriculum, the instructors distributed a post-workshop survey to participants to determine curriculum strengths and weaknesses.

PTP administered an anonymous online survey gauging preservation awareness and behaviors through listservs, social media, and other

promotional efforts. A total of 556 self-selecting participants took the survey and PTP published the results in February 2019.²⁰ A survey analysis report was coauthored by Jacob Kramer-Duffield, contracted audio strategy consultant, who summarized survey findings and found that “[t]hose who have some sort of institutional affiliation (i.e., work professionally in podcasting for an institution) are much more likely to back up all files (including raw tape and drafts), uncompressed, than independent producers (46 percent to 33 percent).”²¹ In addition, a small, yet not insignificant set of respondents (less than 7 percent) did not have a backup plan in place at all.²² The survey statistically captured the need for personal podcast preservation.

Following the lead on podcasts’ successful decentralized and open-source distribution networks, PTP similarly put measures into place to ensure all project deliverables were made openly accessible through their website and GitHub repository. Instructions for hosting a podcast preservation workshop, and descriptive instructions on how PTP went about preserving their own digital assets were published. A Creative Commons CC-BY license was applied to all PTP products, allowing anyone to repurpose or reproduce any content. Taking these measures ensured that PTP lived by its own preservation ethos and provided continued resource availability beyond the grant-funded period.

Curriculum Exercises

PTP built five interactive exercises, each of which expanded on lessons put forth by the curriculum. The first three exercises focus on standard archival and preservation tenants (e.g., file management, storage/backups, metadata) customized to address how each may be approached by a podcaster.

Get Organized!

The first PTP exercise addressed file and folder management, established by archivists working on similar personal digital archiving educational efforts as a first and foundational step. A 2012 survey conducted by archivists on the personal digital archiving behaviors of writers using word processing software discovered that “the accumulation of poorly managed, highly distributed, and unsystematically labeled files, representing works of writing in myriad versions and in various stages of completion” eclipsed other notable personal archiving problems which included format obsolescence and hard-drive failures.²³

Like writers, podcasters must similarly prioritize file management, as production workflows often result in the creation of dozens of files in various

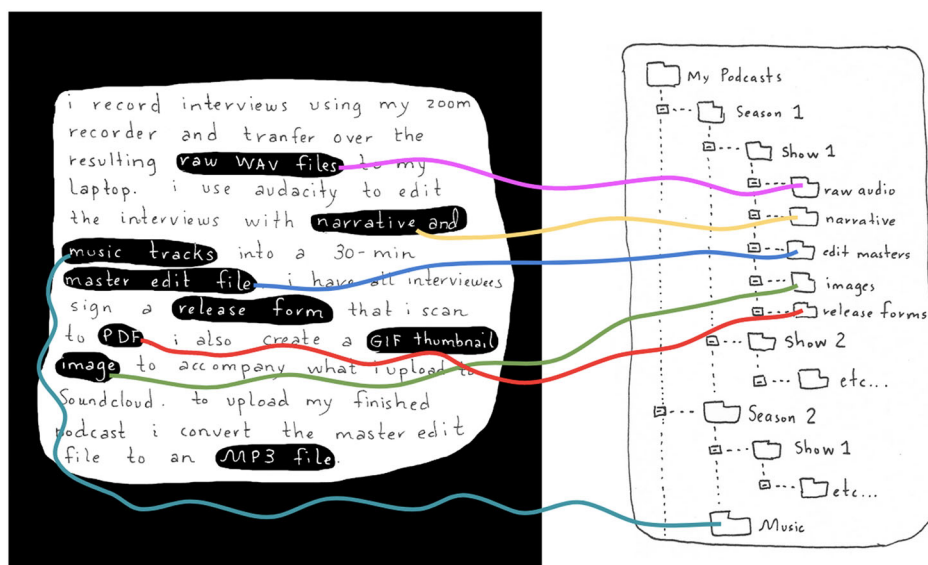


Figure 2. This illustration was presented during PTP's workshops during the "Get Organized!" exercise (page 7 of the zine) and demonstrates how podcasters can use their own production narrative to identify and inventory file types. This inventory can be used as the basis for a reusable folder template copied from project to project. Credit: Illustrated by PTP Lead Mary Kidd. This work is licensed under a Creative Commons CC-BY license.

formats, such as raw interview, tracking, project, session, soundtrack, transcript, and guest release form files. A common piece of feedback received by PTP during workshops was that indie podcasters working on teams using several different platforms were plagued by duplicate or ambiguously named files. As a way to guide podcasters toward effective organization, PTP's exercise titled "Get Organized!" guided podcasters to write and describe their entire end-to-end production process and take stock of when files are made and what systems are used to make them (Figure 2). The resulting writing was used to create a rough inventory; i.e., podcasters identified how many products were being created over the course of producing one episode by circling each file type. PTP then provided an illustrated example demonstrating how these inventories can be used as the basis for a structured folder template that podcasters may copy from project to project.

Store, Backup or Lose

In taking stock of how many digital files are made during regular production, podcasters can better estimate their needs for a file backup system, which is perhaps the most cost-prohibitive step within PTP's suggested actions. PTP promotes a "3-2-1" backup approach, which at minimum, consists of two copies of a file stored on two separate devices, and a third copy stored on a hard drive or server in a separate geographic locality.²⁴

This plan requires the implementer to not only purchase additional hardware and subscribe to a service, but put in the time and effort to maintain these systems over the long-term.

In the “Store, Backup or Lose” exercise, PTP guides participants through a web-based game, fashioned after the popular “Choose Your Own Adventure” series, illustrating common lack of redundancy and third-party platform failure scenarios based on real-life examples. One such example familiar to podcasters is the case of Myspace, which abruptly deleted fifty-three million files from its servers in 2019, resulting in artists and musicians losing unique copies of their audio files.²⁵ For those who find cloud backup services cost-prohibitive, the Internet Archive was suggested by PTP as a free option. According to post-workshop surveys, 63 percent of participants made plans to invest in a 3-2-1 backup system.²⁶

Metadata Archaeology

The third and final PTP exercise, “Metadata Archaeology”, introduces podcasters to data description as a mechanism to make podcasts discoverable and accessible past the point of “podfade.” Podfade is a podcast industry term, and describes how, of the roughly 540,000 podcasts in existence in 2017, 75 percent ceased production.²⁷ Here, participants are encouraged to visit the “2005 Podcast Core Sample Collection” hosted on the Internet Archive.²⁸ This collection of 1,155 podcasts is the result of a script, called the “podsucker” written by Jason Scott, an archivist at the Internet Archive. In 2005, the podsucker scraped RSS data and downloaded audio files off a now-obsolete podcast directory website, podcastdirectory.com. This collection provides a snapshot of some of the earliest podcasts created. The PTP exercise guides participants to pick one podcast from the collection at random, and attempt to find relevant information by performing a keyword search in Apple podcasts, within a search engine, or play back any available audio.²⁹ Exercise participants realized that many of the collection’s podcasts are unplayable, and given a lack of discernible metadata, undiscoverable beyond the stub records found in the Core Sample. Through this work performed by workshop participants and PTP team members, it was found that of a sample of 125 shows from the Core Sample, the affiliated websites of 87 percent of podcasts published in 2005 are no longer available online, and a staggering 98 percent are no longer available in Apple Podcasts.³⁰ By showing how metadata fades over time, PTP makes the case that taking the time to describe files, either by filling out format-specific fields (such as MP3 ID3 tags) or maintaining podcast websites allows for better discovery by future researchers, listeners, and archivists.

Metadata and Ownership

Podcasts are distributed and made discoverable via open RSS metadata aggregators, which makes podcast metadata especially vulnerable to third-party proprietary platform exploitation. One example of platform metadata meddling occurred in 2019, when Apple Podcasts published new metadata guidelines, including one that prevented podcasters from inputting episode numbers into the title field. Failure to conform to this rule would result in a podcaster's show "being rejected or removed" from the Apple Podcast database; Apple later rescinded this policy following podcaster community pushback.³¹

In another instance of metadata exploitation, the Luminary podcast platform, launched in early 2019 and touted as the Netflix of podcasting, populated its database with freely available podcast RSS metadata to attract listeners to its paid-tier content. Podcasters discovered that Luminary removed information from the show notes field, which proved to be "a problematic move as many publishers, particularly smaller and independent ones, rely on linking to sponsors and direct donation appeals in the show notes as integral aspects of their business models."³² These are just two examples of prominent platforms applying unilateral metadata decisions, and will certainly not be the last. It was at this point in the workshop exercises that PTP introduced participants to the PodcastRE service (podcastre.org) maintained by the Department of Communication Arts and Libraries at the University of Wisconsin, Madison. PodcastRE enables any podcaster to submit their RSS feed for free, effectively creating another backup of an important component of their podcast's metadata. Podcasters are encouraged by PTP to see metadata, along with audio files, as part of what they are creating and preserving.³³ By embedding metadata into their files (i.e., filling out ID3 tags) or contextualizing metadata alongside their files (i.e., foldering an audio file with its transcript or creating a podcast website), and backing up their files in a 3-2-1 fashion, podcasters are effectively creating an archive of evidence (i.e., provenance) for how they originally intended to describe their podcast.

In addition to re-appropriation, metadata may also be misconstrued by biased algorithms or AI systems, or deliberately misused by law enforcement to criminalize certain podcast creators and their content. Algorithms, which are touted as backing and improving end-user experiences for things like search engines, "can have implicit bias programed into them and limited knowledge that can skew results toward unexpected behaviors."³⁴ This is something podcasters should be aware of, especially in light of major search engine Google announcing in 2019 that they would be indexing the content of 2 million podcasts.³⁵ This issue is addressed in PTP's fourth episode, "Metadata Meta-awareness", featuring Kaytlin Bailey, producer of the *Oldest Profession*, a podcast about sex workers.³⁶ With search engines and

algorithms scraping searchable text such as podcast titles, descriptions, and keywords³⁷, metadata associated with podcasts like the Oldest Profession may be parsed and flagged under the problematic rubric of laws like SESTA-FOSTA as promoting sex work.³⁸ In the PTP episode, Bailey expresses these fears over her podcast content:

“The passage of SESTA-FOSTA provides the legal cover for the, like, seizure, erasure, locking out of a lot of this stuff. And so, I mean, there is a way to interpret through the law as it is broadly written that this podcast is quote unquote promoting prostitution, and that could justify a legal seizure and erasure of my work.”³⁹

For metadata reasons alone, this makes the Oldest Profession a possible vector for podcast erasure. In fact, for some podcasts, especially those that feature or discuss individuals or groups under state surveillance or who are otherwise vulnerable to police seizure, aggression or other human rights abuses, metadata misappropriation may be their biggest preservation threat.

Self-Hosting

The last section of the PTP curriculum focuses on the importance of the open ecosystem that the podcasting community relies upon—RSS—and an option for how podcasters could establish a platform-independent hosting system. This piece of the curriculum represents one of the more complicated technical lifts for indie podcasters, requiring that they take a number of steps to migrate web content and RSS hosting off proprietary platforms into open-source ones. Interestingly, while post-workshop surveys suggest podcasters felt the Self-Hosting section was hard to understand, the PTP podcast episode dedicated to RSS was the most downloaded episode of the miniseries.⁴⁰ Not all podcasters will take these steps and may see the first three PTP lessons as sufficient. However, moves by podcast platforms to place some podcasts behind “paywall gardens” threaten open RSS structures.⁴¹ By establishing a self-hosted system, podcasters can effectively eliminate their financial and technological dependence on third-party platforms, better guaranteeing their podcast’s longevity and accessibility on the web.

PTP Self-Hosting Guide

PTP formulated and published a step-by-step guide for podcasters to self-host their audio files, RSS feed, website and other common podcast components with little to no expense. These steps could be performed either during the podcast’s active production cycle or exported from the podcast platform post-production.

The guide was modeled after PTP’s own self-hosting system established out of the need to provide post-grant access to their podcast and resources.

The Guide was published and promoted with the caveat that this was not the be-all, end-all way to self-host. Additionally, PTP acknowledged that its own system, though good enough for the time being, is not entirely immune to shifts in technology, industry, and culture. The steps taken were distributed across two distinct phases: during the funded grant period (Phase I) and transitioning into the post-grant stage (Phase II).

Phase I: Grant-Funded Podcast Hosting

PTP used a combination of five primary platforms and services during the grant-funded period, of which all but two were paid. Those free services included **Google Drive** to draft, store, and collaborate over various documents (e.g., transcripts, grant narratives, notes, release forms, budget sheets, images, proposals); **GitHub Pages**⁴² as the webhost to publish [supplementary information](#) about PTP research and resources, including an online version of the zine and links to episode audio and transcripts; and lastly, like all podcasts, the RSS feed was disseminated to various free **podcatchers** for discoverability.⁴³ PTP paid for two main services: a **domain registrar**⁴⁴ that enabled PTP to register for preservethispodcast.org, and a finite subscription to **Simplecast**⁴⁵ to generate and host PTP's initial RSS feed and audio files, podcast-relevant metadata, streaming services, and listener analytics.

Phase II: Post-grant preservation

In Phase II, PTP migrated all of the project and podcast materials onto open-source and free platforms. Full documentation, code, and instructions to self-host a website based on the PTP system are available in PTP's GitHub repository as a template.⁴⁶ PTP mainly took advantage of services offered by the **Internet Archive (free)** for post-funding sustainability. These services allow all PTP podcast audio files to be hosted, described, and made streamable through RSS. PTP continues to use Google Drive, GitHub Pages, various podcatchers, and the prepaid website domain. While each PTP member has a backup copy of the Google Drive files, the preexisting Drive is made available for easy access and continued collaboration. GitHub Pages continues to host the PTP website content, but there are additional files and configurations to host PTP's new self-hosted RSS feed.⁴⁷ This RSS file updates to point to the new URLs of the audio files now hosted and streaming on the Internet Archive.⁴⁸ The new self-hosted RSS feed was re-submitted to all major podcatchers, and the PTP website domain was prepaid with existing grant funds until 2030.

In addition to sharing with podcasters its own self-hosting system, PTP let podcasters know that they can request an "HTTP 301 Move Permanently", colloquially known as a "301 redirect", to migrate all visitors,

listeners, and subscribers from their old RSS URL to their new RSS URL. This is similar to a mail forwarding service, but for web objects. Some podcasters believe that the 301 redirect still leaves subscribing via RSS in a precarious position, but for those who are already hosting their podcast and do not want to lose their current list of subscribers, initiating a 301 redirect alongside taking the steps as provided in the PTP curriculum can ensure podcasts can be found by listeners and computer systems.⁴⁹

The PTP approach to self-hosting comes with a major caveat: PTP did not put any systems into play to maintain post-grant download and listening analytics. To date, PTP has not found a viable, free, open-source analytics system. It is possible and expected that many podcasters, especially those deriving some or all of their financial well-being from their podcast, may forgo self-hosting. As mentioned in PTP Episode 5: “RSS Resuscitations”, Brad Smith, CEO of Simplecast, reported that

“[e]very time somebody download[s] something we pay for every byte of data transferred ... we are a podcast analytics and hosting service. The two things that cost us the most money are our analytics and the server farm, and all the databases and elastic search that runs that. And then just the core cost of streaming.”⁵⁰

If the core value that podcast hosts offer is analytics, self-hosting may be the most appropriate if and when the podcaster decides analytics are no longer necessary or useful to them. Although preserving podcast analytics went beyond the scope of the PTP project, it suggests further research and efforts toward self-hosted analytics solutions could deeply benefit personal podcast preservation efforts.

Outro

The sheer volume of podcasts, compounded by the ephemeral nature of web publishing, will pose a challenge to any current and future preservation initiatives, even to those performed by well-funded institutions at scale. One of the largest collecting institutions in the United States, the Library of Congress, admitted that the sheer volume of podcasts their newly announced Podcast Preservation Project to prioritize the acquisition of the audio files of currently active, non-defunct podcasts. Additionally, this project will also only ingest final edit masters, leaving behind other important artifacts such as podcast websites and unedited raw audio files.⁵¹ This collecting scope will, for example, exclude those podcasts from the 2005 Podcast Core Sample discovered by PTP workshop participants as lacking any discernible metadata traces on the world wide web.

Independent producers working in any digital or web-based medium, whether that is podcasts, music, radio, stand to benefit from taking the steps put forth by the PTP curriculum, which centers preservation action

on creators. Acting as digital mythbusters, PTP raised awareness not only of actionable preservation steps to take, but also used their platform to raise awareness over the ephemerality of the world wide web, the precariousness of commercial and proprietary software platforms, and the limitations of institutional collecting scopes. With no guarantees, creative legacies float in a sea of media, strapped to digital life vests. If they are organized, described and backed up, they stand a greater chance of reaching future listeners, scholars, and storage repositories, intact.

Notes

1. Richard Berry, "Podcasting: Considering the evolution of the medium and its association with the word 'radio'".
2. Gabriela Redwine, "Personal Digital Archiving" (DPC Technology Watch Report 15-01 December 2015, Digital Preservation Coalition, Great Britain, 2015), 2, <https://doi.org/10.7207/twr15-01> (accessed February 17, 2020).
3. Meridith Mink, "Preserve This Podcast 2018–2020," (Minkwell Research & Consulting, Madison, WI, 2020), 29, <https://static1.squarespace.com/static/5b43bbda2971143f2370ee64/t/5e46bf8bd8a89a292563aee6/1581694863413/Preserve+This+Podcast+Assessment.pdf> (accessed February 17, 2020).
4. "XML RSS," W3Schools, Refsnes Data, http://www.w3schools.com/XML/xml_rss.asp (accessed January 27, 2020).
5. Shelley Powers, *Practical RDF: Solving Problems with the Resource Description Framework* (Sebastopol, CA: O'Reilly & Associates, Inc., 2003).
6. "RSS 2.0 Specification," W3C Markup Validation Service, UserLand Software, <http://validator.w3.org/feed/docs/rss2.html> (accessed January 27, 2020).
7. Jenn, Riley, *Understanding Metadata: What Is Metadata, and What Is It For?: A Primer* (Baltimore, MD: National Information Standards Organization (NISO), 2017), 12.
8. Ibid., 8.
9. Reuven M. Lerner. "At the Forge - Syndication with RSS." *Linux Journal*, October 4, 2004, <http://www.linuxjournal.com/article/7670>.
10. Lucas Graves, "The Affordances of Blogging: A Case Study in Culture and Technological Effects," *Journal of Communication Inquiry* 31, no. 4 (2007): 343. doi: 10.1177/0196859907305446.
11. Andrew J. Bottomley, "Podcasting: A Decade in the Life of a 'New' Audio Medium: Introduction," *Journal of Radio & Audio Media* 22, no. 2 (2015): 164. doi: 10.1080/19376529.2015.1082880.
12. Daniel Sanchez. "SoundCloud Is Going Bankrupt & There's Nothing You Can Do About It." *Digital Music News*, May 2, 2017, <http://www.digitalmusicnews.com/2017/05/02/soundcloud-go-3-months/>.
13. John L. Sullivan, "The Platforms of Podcasting: Past and Present," *Social Media + Society* 5, no. 4 (2019): 8, doi: 10.1177/2056305119880002.
14. Ibid., 5.
15. S. I. Hayakawa, *Language in Thought and Action* (New York, NY: Harcourt Brace & World, 1964), 153.; Bret Victor Bret, "Up and Down the Ladder of Abstraction," Up and Down the Ladder of Abstraction: A Systematic Approach to Interactive Visualization (blog), October 2011, <http://worrydream.com/LadderOfAbstraction/> (accessed February 12, 2020); Nic Weber, "Preservation INFX 511" (Lecture slides), 2018.

16. Preserve This Podcast, "Preserve This Podcast Simplecast Analytics," mnylc/preservethispodcast/assets GitHub repository (dataset; accessed December 10, 2019), <https://doi.org/10.5281/zenodo.3673937>. This dataset shows streaming and download statistics for PTP's eleven-month subscription to Simplecast.
17. "Preserve This Podcast on Apple Podcasts," Apple Podcasts, <https://podcasts.apple.com/us/podcast/preserve-this-podcast/id1451896473> (accessed February 12, 2020).
18. Mary Kidd, "Preserve This Podcast Zine", Preserve this Podcast, <http://preservethispodcast.org/#zine> (accessed February 17, 2020).
19. Ibid., 7.
20. Jacob Kramer Duffield, "Podcast Preservation Survey Findings," Data infographic slideshow, Preserve This Podcast, https://preservethispodcast.org/assets/PodcastPreservation_SurveyFindings_Feb2019.pdf (accessed February 5, 2019).
21. Jacob Duffield Kramer, internal memo, "Preserve This Podcast Survey Summary Memo," January 10, 2019.
22. Ibid.
23. Devin Becker and Collier Nogues, "Saving-Over, Over-Saving, and the Future Mess of Writers' Digital Archives: A Survey Report on the Personal Digital Archiving Practices of Emerging Writers," *The American Archivist* 75, no. 2 (2012): 483, doi: 10.17723/aarc.75.2.t024180533382067.
24. Peter Krogh, *The DAM Book: Digital Asset Management for Photographers*, 2nd ed. (Chambersburg, PA: O'Reilly Media, 2009), 207.
25. Damon Krukowski, "History Disappeared When Myspace Lost 12 Years of Music, and It Will Happen Again," *Pitchfork*, March 20, 2019, <https://pitchfork.com/thepitch/history-disappeared-when-myspace-lost-12-years-of-music-and-it-will-happen-again/> (accessed February 17, 2020).
26. Mink, "Preserve This Podcast 2018–2020," 9.
27. "How Many of the 540,000 Podcasts Have 'Podfaded?'," Amplifi Media, <https://www.amplifimedia.com/blogstein/2018/8/22/how-many-of-the-550000-podcasts-are-actually-active> (accessed January 4, 2020).
28. "The 2005 Podcast Core Sample Collection: Free Audio: Free Download, Borrow and Streaming: Internet Archive," Internet Archive, https://archive.org/details/2005_podcastcoresample (accessed January 20, 2020).
29. John Koetsier, "Top 100 Apps Of 2019: Netflix, Uber, Spotify, Google Pay, Wish, And More," *Forbes*, last modified December 30, 2019, <https://www.forbes.com/sites/johnkoetsier/2020/12/30/top-100-apps-of-2019-netflix-uber-spotify-google-pay-wish-and-more/> (accessed February 12, 2020). PTP used Apple Podcasts for this exercise since, historically, it has set the tone for podcast publishing standards and was generally known to be the most popular podcatcher application. Spotify has since replaced Apple as the most popular podcatcher.
30. "Metadata-Archaeology_2005 Podcast Core Sample", Google Sheets, last modified January 24, 2020, https://docs.google.com/spreadsheets/d/1YxUVwxnExZY6U1HJrmhf5D5I-0xR56EBYxpKuls-r_Q/edit#gid=0. Dataset of podcasts searched for during the Metadata Archaeology exercise. Out of 136 podcast searches, 88.97% of the podcast shows could not be found on Apple Podcasts.
31. William Gallagher, "Apple's new podcast metadata rules ban episode numbers, threaten removal [u: Episode Numbers Permitted]," *AppleInsider*, <https://appleinsider.com/articles/19/02/28/apples-new-podcast-metadata-rules-ban-episode-numbers-threaten-removal> (accessed January 4, 2020).

32. Nicholas Quah, "How Luminary's Messy Debut Ended Up Roiling the Podcast Industry," *Vulture*, April 26, 2019, <https://www.vulture.com/2019/04/luminary-debut-podcast-controversy.html> (accessed February 17, 2020).
33. Jennifer Douglas, "A Call to Rethink Archival Creation: Exploring Types of Creation in Personal Archives," *Archival Science* 18 (2018): 29–49, <https://doi.org/10.1007/s10502-018-9285-8>.
34. Henry Geneva, "Research Librarians as Guides and Navigators for AI Policies at Universities," *Research Library Issues*, no. 299 (2019): 66, <https://publications.arl.org/rli299/47>.
35. Todd Spangler, "Google Indexing Content of 2 Million Podcasts, Lets Users Stream Directly From Search Pages," *Variety*, August 8, 2019, <https://variety.com/2019/digital/news/google-stream-podcasts-search-pages-1203297026/> (accessed February 17, 2020).
36. Molly Schwartz and Kaytlin Bailey, "Preserve This Podcast Episode 4 - Metadata Meta-Awareness," mp3, Preserve This Podcast, 2019, <http://preservethispodcast.org/> (accessed February 17, 2020).
37. Steve Pratt, "Google's New Way to Find Your Next Favorite Podcast," *Medium*, May 23, 2018. <https://blog.pacific-content.com/googles-new-way-to-find-your-next-favorite-podcast-452c993710bd> (accessed February 17, 2020).
38. Sarah Jeong, "Sex Trafficking Bill Is Turning into a Proxy War over Google," *The Verge*, September 14, 2017, <https://www.theverge.com/2017/9/14/16308066/sex-trafficking-bill-sesta-google-cda-230> (accessed February 17, 2020).
39. Schwartz, "Preserve This Podcast Episode 4 - Metadata Meta-Awareness".
40. Meridith Mink, "Preserve This Podcast 2018–2020," (Minkwell Research & Consulting, Madison, WI, 2020), 14, <https://static1.squarespace.com/static/5b43bbda2971143f2370ee64/t/5e46bf8bd8a89a292563aee6/1581694863413/Preserve+This+Podcast+Assessment.pdf> (accessed February 17, 2020).
41. Erik Jones, "What You Get When You Sign up for Luminary (and 8 Other Paid Podcast Platforms to Try)," *Medium*, June 5, 2019, <https://bellocollective.com/what-you-get-when-you-sign-up-for-luminary-and-8-other-paid-platforms-to-try-232af0976f2c> (accessed February 17, 2020).
42. "GitHub Pages," GitHub, <https://pages.github.com/> (accessed February 9, 2020).
43. Popular podcatcher platforms include: Apple Podcasts, Stitcher, RadioPublic, Pocket Casts, and Spotify. Wikipedia page: *List of podcatchers* has a more comprehensive "list of noteworthy podcatchers or software with podcatching capability." https://en.wikipedia.org/wiki/List_of_podcatchers
44. "Iwantmyname," Iwantmyname, <https://iwantmyname.com/> (accessed February 9, 2020).
45. "Podcast Hosting, Distribution & Analytics," Simplecast, <https://simplecast.com> (accessed February 9, 2020).
46. "mnylc/preservethispodcast," GitHub, last modified February 9, 2020, <https://github.com/mnylc/preservethispodcast> (accessed February 17, 2020).
47. <https://github.com/mnylc/preservethispodcast/blob/master/podcast.rss>
48. <https://archive.org/details/@preservethispodcast>
49. Stephen Jondrew, "Better Podcasting #065 – Self-Hosting Your Podcast RSS and Media," *GonnaGeek - Geek Podcasts, Tech, Comics, Sci-Fi, Gaming & More* (blog), February 7, 2017, <https://www.gonnageek.com/2017/02/better-podcasting-065-self-hosting-your-podcast-rss-and-media/> (accessed February 17, 2020); Allan Tépper, *Branded RSS: the definitive guide*, (TecnoTur LLC, 2019).
50. Molly Schwartz, "Episode 5: RSS Resuscitations", mp3, Preserve This Podcast, 2019, preservethispodcast.org/#episodes (accessed February 17, 2020).

51. Jennifer Waits, "Podcast #230 - The Library of Congress Launches Podcast Preservation Project," *Radio Survivor* (blog), January 29, 2020, <https://www.radiosurvivor.com/2020/01/28/podcast-230-the-library-of-congress-launches-podcast-preservation-project/> (accessed February 7, 2020).

Acknowledgements

We would like to thank the Andrew W. Mellon Foundation in supporting our work, and in particular Program Officer Patricia Hswe for her initial guidance and encouragement at the outset of grant application process; the staff at the New York Metropolitan Library Council Studio Space for providing workshop space and administrative support throughout the grant; all the indie podcasters who appeared on our podcast, Alice Y. Hom, Dan Weissman, Amanda McLoughlin, Eric Silver, and especially, Kaitlyn Bailey, Producer of *The Oldest Profession*, whose insight into metadata encouraged us to dig deeper; Jeremy Helton, Allison Behringer, Dalton Harts, Breakmaster Cylinder, Jacob Kramer-Duffield, Austin Eustice; and finally, we want to express deep gratitude to PTP co-Leads Molly Schwartz and Dana Gerber-Margie, for their tireless support for podcast preservation work, now and into the future.

Notes on contributors

Mary Kidd (@kiddarchivist) is Co-Lead of Preserve This Podcast, and also works for the New York Public Library's Preservation and Collections Processing Department as the Systems and Operations Coordinator. She was selected for a National Digital Stewardship Residency in 2015, where she worked on a digital audio preservation project at New York Public Radio. She is also an illustrator and uses drawings to teach personal preservation and archiving.

Sarah Nguyen (@snewyuen) is the Project Coordinator of Preserve This Podcast, a Research Scientist for Investigating and Archiving the Scholarly Git Experience at New York University Libraries, and a pre-doctoral candidate at the University of Washington's iSchool. Sarah's research focuses on peoples' behaviors and community building influenced by open information technologies, especially impacting individuals of marginalized identities. Offline, Sarah practices movement through dance.

Erica Titkemeyer (@etitkem) is a Project Director in the Louis Round Wilson Library at the University of North Carolina at Chapel Hill. Co-ordinating audiovisual preservation and access activities for recordings within Wilson Library, Erica collaborates across UNC University Library departments to promote and improve on library technologies that support digitized and born-digital audiovisual discoverability, use and sustainability.

Funding

This work was supported by the Andrew W. Mellon Foundation under Grant 1711-05080.

ORCID

Mary Kidd  <http://orcid.org/0000-0001-9294-0961>

Sarah Nguyen  <http://orcid.org/0000-0003-2040-390X>

Erica Titkemeyer  <http://orcid.org/0000-0002-7465-9313>