Anne Wootton & Bailey Smith
co-founders, Pop Up Archive

Michael O'Connell, producer, It's All Journalism
My name is Michael O'Connell. I'm here today with Anne Wootton and Bailey Smith, the co-founders of Pop Up Archive. Good morning. Or good day, actually. It's good morning for you. It's good day for me.

Anne Wootton, co-founder, Pop Up Archive
Hi.

Michael O'Connell
You're on the West Coast, right?

Anne Wootton
We are. We're based in Oakland.

Michael O'Connell
OK. And you just got back from South by South West [SXSW]. I assume that you were there to talk about Pop Up Archive. Tell me a little bit about what Pop Up Archive is and what you're trying to accomplish with it.

Anne Wootton
Sure. So, Pop Up Archive is a Web application that makes audio searchable, leveraging technology for automated transcription and entity extraction and, as a result, making that audio accessible to many more people than ever before. It started as a master's thesis work, when Bailey and I met at the School of Information at [the University of California] Berkeley and we've been in this area ever since.

Michael O'Connell
So who are you doing this for? Who do you see as your audience?

Bailey Smith, co-founder, Pop Up Archive
Definitely journalists and generally media producers, but we're also looking at small archives that have interesting audio collections. We think that this is a valuable proposition to bring all these people to one table and enable them to share media with one another.

Michael O'Connell
Just so I get an understanding of it, how does your product work? What does it do? Is it just transcription or does it doe something else?

Anne Wootton

Anyone using Pop Up Archive has the option of adding one or multiple audio files and organizing those files in sort of buckets or collections through which they can track certain projects. But whether or not they choose to organize, there is metadata that's already automatically created and that's what really makes the audio searchable. And that does include the transcriptions. It also includes, what I mentioned, entity extraction technology. That means being able to identify keywords, people's names, places, popular themes or subjects.

And then, there are options to kind of go beyond that in terms of relational data, keeping track of people interviewed or interviewers, hosts, producers, you know. The sky's the limit when it comes to just how much you want to track all your audio, but the problem is that most people don't have the time to put that type of effort in themselves. And so, automating the process makes audio, which otherwise is just kind of as opaque as the filenames and however many nested folders you may have on your desktop. We're trying to break that mold.

Michael O'Connell
So is it then something a journalist would use behind the scenes to sort of organize and to search the data they've collected through audio? Or is it something else?

Anne Wootton
Absolutely. We've been talking to producers for a while now about their workflows and kind of the often cumbersome steps that they go through to backup certain media but not other media and share that media with teams they may be working with or just re-access it for their own purposes and the trouble that they face in doing that effectively and quickly.

But then absolutely, we see this as being kind of an API for all unfinished content that anyone can search. There were always be privacy concerns, but there's a lot of public material out there that, you know, sometimes the best way of accessing audio today is doing a YouTube search and hoping that you find something close to what you're looking for and there's just that material could be accessible in much more effective ways.

Michael O'Connell
So, is there a search aspect to it? Is that what you're sort of saying?

Anne Wootton
Absolutely yes.

Michael O'Connell
How does that work? Are you pulling on the metadata for that?

Anne Wootton
We are. We have a few different ways of indexing the material and storing metadata based on how we've kind of identified its datatype and then, yeah, searching automated

transcriptions and keywords and kind of weighting the metadata depending on how much there is.

Michael O'Connell
What inspired you to come up with this?

Bailey Smith
I guess that would be our work with The Kitchen Sisters. We were masters students at the University of California at Berkeley in the School of Information, and so, we were working on a final thesis. And Anne met up with The Kitchen Sisters at, I think, some event, a party maybe? And got to talking with them and she said, "Well, I'm interested in digital archiving." And they said, "Oh, that's wonderful. We have all this material, and we're losing it and we'd love it if you can come on board and try to help us build something out of" what they call their accidental archive, which is a really charming way to describe like the material they've gathered over a 30-year relationship of creating audio for radio.

And, Anne started off trying to help them get a grant for digitization, which helped us realize that there's this difficulty for smaller producers. There are archival systems out there, but they are cumbersome, I think for a small shop to get into place and start using. We wanted to help them find a way to make that process easier, and sort of came to the realization that there wasn't anything out there that was really built for them and that we could step in and make that happen.

Michael O'Connell
OK. So, what did you end up doing for them?

Anne Wootton
For The Kitchen Sisters we actually augmented open source software called Omega to build plugins that support broadcast media in particular and that would let them back their material up at the Internet archive that's kind of a trusted digital repository from the preservation and access standpoint, the software and granting agencies are coming from as well as published material to SoundCloud to engage with their growing audience there. We sort of shifted away from that model now, although we're really happy to release that software, be a part of that community. We found that even that has kind of the simplest WordPress-esque approach to archival software was still a little bit too complicated for your average producer, your average consumer. So, that's really the need that Pop Up Archive is looking to fill.

Michael O'Connell
What were you doing at SXSW? Were you just sort of rolling this out and saying, "Here, this is what I've got. Or, this is what we developed." What kind of feedback were you getting from people?

Bailey Smith

SXSW was also nice. First of all, we really love Austin. It was really exciting to be there and it was so warm and there's so many interesting people. We were there to give a digital archiving workshop. So, we sat down for four hours with a group of people that were all interested in archiving. Yes, we demoed our site for the first time. Right now, it's in beta, so if you'd like to use it, you need to send us a request and we'll get back to you as it becomes available.

Bailey Smith

But we got to work with a few like really great producers and watched them in action for the first time and get some of the testing done and see how they interact with the product, which was really exciting and I think we definitely learned a lot in Austin that day. We'll be incorporating that work as we go forward.

Michael O'Connell

So what type of feedback were people giving you?

Anne Wootton

You know, the feedback that a lot of Web-based products get during their first round of beta testing. You know, just, "This page isn't intuitive" or "You need more explanation here" or "The user flow doesn't feel quite right." And you can't really see that happening until you're watching over someone's shoulder as they're interact with it. So, it's invaluable to have that kind of feedback early on.

Michael O'Connell

It's difficult to sort of anticipate all of the ways people are going to use things. There are plenty of examples, you know, in the online world of products being rolled out and them being used in a very different way than their creators had imagined them, for both good and bad, I guess.

Anne Wootton

Exactly. And we're still very much in the process of testing a lot of hypotheses around and thinking about, you know, we got into this working The Kitchen Sisters on content the value of which was abundantly clear to us and we love all the different content-creators we've encountered over the course of this process. But I think the question is then how do you most effectively make audio searchable and what are the most appropriate approaches for doing so.

Michael O'Connell

Now, you had a Knight Foundation backing on this?

Anne Wootton

Yeah, we were winners of the Knight News Challenge Data Round in 2012.

Michael O'Connell

And what has that been able to allow you to do?

Anne Wootton

It's given us a generously long ramp to kind of continue attacking this problem and considering, like I just said, the different approaches for providing the type of service that we want to provide and hopefully getting it right. So we're working, there's the two of us, and then we're working with the Public Radio Exchange to build out the platform, the repository of content, the Web application, the search functionality.

Michael O'Connell

So, what is your thinking at this point? Is this something that you're creating that people can access? Is this something that people are going to be able to purchase and use in their own business with their own material?

Anne Wootton

Yeah, so we have several different potential markets that we are fond of for different reasons and continue to sort of bat around for those reasons. But, I think fundamentally we're excited to be creating an API for all of the audio that's kind of falling by the wayside right now. And there are some really exciting efforts happening in public media and in general, whether the DPLA, the Digital Public Library of America, is building an API for all sorts of incredible archival and library content from across the United States.

Somewhat similarly, the Public Media Platform is the result of five major public media organizations coming together and saying, "We need a way to systematically distribute public media content as it's produced." And that's great. That's great for the produced work, but there is definitely this long tail of all this other content and I think that we're focusing on building that API and building a search functionality that works well with the content we've attracted so far, and we'll continue to attract and then consider from there who to focus on when it comes to providing services because there are, right now, a lot of groups of people that we would like to please and we realize that you can never please everyone, so we'll do the best we can.

Michael O'Connell

I can see how this would be really useful. Actually, I was at the Library of Congress yesterday coincidentally doing some research on a story. I had to access some audio files from the Vietnam Veterans Project. I'm writing about a Vietnam vet for a freelance story. And, you know, it just struck me that they had all of these audio interviews that they've done over the years and so much of that is, I'll use the word that you used, which is "opaque." You have the name of what's on the disc, but you don't necessarily have the information short of just accessing the file and going through it all. So, just from a search and a functionality aspect of it, I can see the immediate use for something like this.

Anne Wootton

Yeah, your story sounds like similar stories that we've heard from a lot of other producers and journalists and content creators. There's —

Bailey Smith
The obit scenario?

Anne Wootton
Yeah, there's a radio station in New Hampshire that we talked to, but it could really be anywhere cause it's a classic scenario, the obituary scenario. And they knew that they had the audio of a particular senator sitting in the station somewhere, but because of the staff turnover that they've had, they lacked that institutional memory. They lacked that path back to that audio and had no way of searching for it.

Michael O'Connell
Right.

Anne Wootton
Yeah, and just the fact that the technology really is just now, I think audio is kind of a latecomer to the Web in a lot of ways, even behind video, hence the YouTube scenario where people are searching on video sites to find audio content. For any number of reasons, but it's been hard, if nothing else, to store this stuff and serve it up consistently without a significant cost, and so, the fact that, increasingly, there's no reason not to save everything is appealing, but I think it's only worth it if you can search that material. So, search is really a big piece of this problem.

Michael O'Connell
And I actually work at a radio station, a radio news station, Federal News Radio here in Washington, D.C., and our focus is covering the federal government. And we produce tons and tons of interviews every week that the challenge for us is how to translate that audio content to the online environment in a way that people can use it. And we'll have five, seven-minute audio files available for people of interviews. But, do we pull out snippets? Do we pull out things that are very interesting? Short of us going through and listening to the whole thing to pull out bits, sometimes it can be really time consuming. But knowing that there's also a lot of value in the content that's there, just finding a way to make this easier and more available to your readers, that's incredibly valuable and incredibly useful to getting your website, you know, increasing the traffic. These people want to get this material.

Bailey Smith
And that's nice to hear. We definitely feel you on creating information about your files requires a lot of manual labor. Anything that we can do to cut time out of that process is really what we're interested in. If you went to the site today, for instance, and you extract a file into the site. This transcription and keyword extraction will happen in the background. But you'd also be able to customize the metadata about your file and we have a simple form that makes that possible. I think that's one of the barriers. You start organiz-

ing something and over time, things changes. The way you do it changes. Like, if we can make that happen in a standardized, easy way, every time you want to add the name of an interviewer, you do it in this one field and that's fine. Every time you add the name "interviewee" it's in one field. It's all Woottonized in a way that I think can make this process a lot less painful.

Michael O'Connell
So, where do you want to go next with this? What's on the horizon for you?

Anne Wootton
Well, I think the SXSW workshop was the very beginning of a lot of testing and many iterations to come in the next few months as we hammer out the rest of this technology. We're pretty excited about expanding. We'll include video content, and seeing just how robust of an API we can develop and getting content to kind of test that case and prove that point. We have a lot of testing to do and I think we'll probably be reassessing later this year, as I mentioned earlier, thinking about which communities are best to focus on in this domain and kind of grow the service from there.

Michael O'Connell
Do you have a timetable of when you think you'll be able to roll this product out?

Anne Wootton
People can request invites to beta test the service at our website at PopUpArchive.org. We'll start sharing the service with people towards the end of this month in beta-test mode. I think that we will be further publicizing the service and offering a more stable release this spring. And I think Pop Up Archive will be up and fully functioning this year.

Michael O'Connell
Oh, that's great.

Anne Wootton
Yeah, definitely. I think there's an endless list of features that we want to continue to roll out and the timeline on that is a little bit longer, depending on the feature. But in terms of a baseline product, yeah.

Michael O'Connell
Were there any things beyond your experience with Pop Up Archive at SXSW? Any other things that you saw there that kind of impressed you that you took away from that?

Anne Wootton
We went on an awesome bike tour of the east side of Austin, as part of, there's a project funded by the Corporation for Public Broadcasting called "Local Lore" with The Association of Independents in Radio and public media radio station partners. Maybe there's

some television stations as well. Public media station partners across the country that are working with producers, one of which is at KUT in Austin, called "Austin Music Maps." So they had a panel at South By and took a few folks on a bike tour to some of Austin's greatest and not necessarily best-known music spots and secret spots. And that was really cool.

Bailey Smith
And it ended with two-step lessons. That was a good day.

Michael O'Connell
Cool

Anne Wootton
Love two-stepping.

Michael O'Connell
It's always nice when you go out on these business trips and find something else that's just fun to do.

Anne Wootton
Oh yeah. We make a point of it.

Michael O'Connell
Is there anything else you want to talk about as far as audio, Pop Up archive that you want people to know?

Anne Wootton
We love seeing kind of the content that trickles up or down to us or comes through the cracks as people hear about what we're trying to do with this service.

Michael O'Connell
Having worked on this project, what are your observations on what's going on, certainly with audio in particular, in trends in journalism now?

Bailey Smith
One of the things that we think is really going to be useful for this service that because, anyway, there's a tendency to be fewer full-time reporters in a newsroom and more free-lancers and those freelancers have several jobs. So, they maybe have audio that they're creating for three different programs and for them, organizing that is more prob-lematic because they may have something that's on a shared drive at one station and then other material that's elsewhere. And it's helpful to be able to unite that for them like in one personal collection. So, I think that could be very helpful and that's an interesting trend, I think, that there's movement between stations. And like Anne said earlier, there's some loss of institutional memory because there's been a lot of turnover at newsrooms.

Michael O'Connell

You said before you were masters students at, was it California?

Anne Wootton

Yeah, Berkeley

Michael O'Connell

What experience do you have in journalism beyond developing this? Do you see your-selves more as developer people? What is your approach to journalism at this point?

Anne Wootton

We both came to the School of Information looking to kind of bone up on our technical skills, because we both come from humanities backgrounds. I did a bit of newspaper and magazine journalism when I was in college and beyond. And Bailey worked on a radio documentary actually. And we were interested in coming at media from a different perspective and being kind of frustrated by the current prospectus for media as we knew it. And I think being at the I School provided a really useful and necessary change of perspective and set of skills that we wouldn't otherwise have in terms of being able to, I mean, we had to take programming classes as part of our curriculum. So, just under-standing the design cycle and understanding engineers and being able to talk about technology in ways beyond what we could do when we came into this program.

Bailey Smith

Yeah.

Michael O'Connell

What would you say to somebody who wanted to pursue a more technical career in journalism? What should they be looking for in a school? In the types of opportunities that they should pursue?

Bailey Smith

You know, at Berkeley, my husband actually went through the journalism program and graduated the year before I did from the masters program at the J School at Berkeley, and I managed to get over there and take a few classes and really enjoyed it. And now they have a digital media track, which is great.

I don't know. While you're in school, if you're in school, take a programming class. There's intro level classes and take whatever language is available. I think Python is a good place to start or learn some Javascript for front end programming. And there's all sorts of ways to do that. There's sites online, like [Codecademy,](#) where you can go and like just learn some basics. CCS and HTML2 are really good to know.

Anne Wootton

Yeah. Even if you're not in journalism school, I think J-schools are increasingly offering really great programs in digital media and those types of skill sets. But there are a lot of resources out there. There are free classes. There are meet ups and popup classrooms, certainly in cities across the United States. So, that's a good place to start.

Michael O'Connell

Thank you very much for your time. I've been talking to Anne Wootton and Bailey Smith, the co-founders of Pop Up Archive. Again, where can people find out information about this?

Anne Wootton

You can check out our website, PopUpArchive.org. You can request and invitation to try out the Pop Up Archive service and transcription stuff there. And you can follow us on Twitter @PopUpArchive.