



## **A More Complete Record: The Case for Archival Partnerships** The Radcliffe Workshop on Technology & Archival Processing

**Friday, April 13**  
**Knafel Center, Radcliffe Institute | *Public***

### **Transcript**

#### **9:00-9:15 Opening Remarks**

- **Jane Kamensky**, Pforzheimer Foundation Director of the Schlesinger Library on the History of Women in America, Radcliffe Institute for Advanced Study and Professor of History, Harvard University

#### **9:15-10:30 The Past and Present in the Future | *Keynote Conversation***

- **Brewster Kahle**, Founder of the Internet Archive
- **Jill Lepore**, David Woods Kemper '41 Professor of American History, Harvard University

- Good morning, everyone. Welcome or welcome back to the second day of this Radcliffe workshop on technology and archival processing. I'm Jane Kamensky. I'm the Pforzheimer Foundation director of the Schlesinger Library and a professor of history at Harvard, and I'm delighted to welcome you here on behalf of all my colleagues at Schlesinger and Radcliffe and Harvard.

Because some of you weren't here yesterday, I'm going to ask those who were to keep their headphones in for the next 90 seconds or so as I repeat a little bit of the framing with which our conversations began and recap some of yesterday's highlights-- a very quick loop back, I promise. This workshop series began in 2011 with the goal of leveraging technology to improve access to archival information. And since then, in mostly annual iterations, it has explored various stuck places in the world of special collections where technology might be used as a kind of crowbar to pry open heaps of treasure that would otherwise remain more or less locked away.

This year is the Schlesinger Library's 75th anniversary-- an occasion to take stock looking backwards and looking forwards, and this is part of that project. Working with an advisory group of historians, archivists, public intellectuals, and other history makers, we've dedicated this anniversary Radcliffe workshop to furthering one of our highest strategic goals, which is expanding the diversity of special collections to help researchers create a more complete and nuanced historical record.

And what we've wanted to do here is to explore how we might best do that not only through acquisitions-- not only through building the collections of individual repositories-- but through collaborations through archival partnerships. So the idea, as you heard in many different ways yesterday, was to bring together the different kinds of resources that different kinds of institutions have-- depths of staffing, in some, contexts, in others, relationships, accumulated wisdom, community. And the idea is that we can be stronger of greater service to the present and the future together.

We want to think about access alongside ownership, especially as we wrestle with the proliferation of born digital and digitized collections, which break down physical boundaries in so many ways. We want to talk about growing pipelines of staff expertise as well as collections and so on. And we hope what we'll come away with after this day are robust and exportable models of collaboration, especially in this instance between predominantly white institutions often rich in endowment balances and other measures of accumulated wealth and historically black colleges and universities which are so rich in collections and in the dedication and deep knowledge of their staff and contributor communities.

Yesterday's keynote conversation between professors Craig Wilder and Beverly Guy-Sheftall entitled "The Past is Present" focused on how we got here-- on the entwined and often painful histories of predominantly white institutions like this one and historically black colleges and universities, which, of course, were not-- our secretary of education to the contrary-- the nation's first exemplars of school choice.

Wilder toured the post-2001 scouring of elite institutional archives, first in Yale and Brown and then Harvard and Princeton and Georgetown and others-- efforts that were often student-led until very recently to surface these universities' connections to slavery. And though he didn't quite say this in his remarks, I think there was a whiff in what he said of the ways that predominantly white institutions have taken pleasure in their pain-- in other words, that some of these efforts at exploring the painful past have become part of the brand of institutions like Harvard, Yale, Princeton, and Brown.

Beverly Guy-Sheftall, in reviewing the ways in which she had built the archives of Spelman College under the auspices of the Women's Research and Resource Center, told a quite different tale of institutional anxiety about what archives held as well as of cross-institutional distrust about past partnerships or collaborations that had been vexed in various ways-- so this legacy of scouring of painful pasts on the one hand and different kinds of efforts to salvage pasts that are shot through with anxiety and complication on the other.

Then there was a panel of scholars who had done research in Africana collections, both in majority and in HBCU contexts, who talked about the joys of discovery and the ease of discovery, which sometimes go together and sometimes don't. And the ease of discovery is, of course, resource and staff-dependent in so many ways. They talked about the vexed decisions of contributors of collections, including themselves, about where to place their papers and about the sometimes competing demands of community on the one hand and resource and access on the others. And I guess it's breaking down that boundary that, in some ways, today's workshop is all about.

There was a hint in that discussion about a digital future in which the question of community, location, and access might recede as physical place became less germane in a post-paper world. And we didn't quite dig into this question of technology as a potential solution space for the anxieties and distrust and historical inequities that the scholars and the keynoters have surfaced, and that is very much the topic of today's keynote and then workshop panels.

Today's keynote conversation, "The Past and Present and the Future," looks forward to the kinds of new collaborative solutions that may be found in technology. It features two very different presenter interlocutors-- the historian, journalist, and all-around American treasure who is my colleague Jill Lepore, and the technologist and visionary who is Brewster Kahle.

The conversation in some ways continues the work of a spectacular annals of technology essay that Lepore published about Kahle and his projects in *The New Yorker* in January 2015 called, "The cob web-- can the internet be archived?" The central irony of that essay and maybe of the internet itself is that everything you think is gone isn't and everything you think will be lasting won't unless we think about it very, very boldly, as Kahle has, through his babies, the Internet Archive and its remarkable Wayback Machine which has become-- even more since Lepore published that profile-- a truly vital part of the fabric of researchers' everyday lives.

Today, Lepore and Kahle reopen that conversation to help us solve the related problem of making the highest, best uses of and providing the highest, best access to the kinds of vital manuscript and rare print materials that our workshop participants talked about yesterday afternoon. As I did yesterday, I will introduce each speaker in very compressed fashion, and then they'll each offer short remarks and have a conversation as a duo and then a conversation with all of us here. After the keynote address, there will be a 15-minute break, and then the Invitational portion of the workshop will begin at 10:30.

Jill Lepore is the David Woods Kemper '41 professor of American history here at Harvard. She is also a staff writer at *The New Yorker* where her beat includes the future of the past in many guises with a special focus on topics including gender and sexuality, privacy and secrecy, law and literature, and the history and destiny of various forms of media. A prize-winning teacher as well as a decorated author, she teaches classes in evidence, historical methods, humanistic inquiry, and the broad sweep of American history.

Much of her scholarship explores absences and asymmetries in the historical record with a particular emphasis on the histories and technologies of evidence. As a wide-ranging and prolific essayist, Lepore writes about American history, law, literature, and politics. Her 12th book, *These Truths-- a History of the United States* will be published by Norton in September.

Brewster Kahle, digital librarian and founder of the Internet Archive, has been working to provide universal access to all knowledge for more than 25 years. He earned a bachelor of science in 1982 from the Massachusetts Institute of Technology, where his studies focused on artificial intelligence. In 1983, when the rest of us were beginning to use cordless phones the size of loaves of bread, Kahle helped to start Thinking Machines, a parallel supercomputer maker, serving there as lead engineer for six years.

Since the mid-1980s, he has focused on developing technologies for information discovery and digital libraries. In 1989, he invented the internet's first publishing system, the Wide Area Information Server, founding WAIS Inc., a pioneering electronic publishing company, which he sold to America Online in 1995 when AOL was the hottest new thing.

The following year, Kahle founded Internet Archive, which may be the world's largest digital library. Its stark Greek portico logo should by this point be tattooed on most researchers' foreheads. Kahle also founded Alexa Internet, which helps catalog the web and which he sold to Amazon in 1999. And whether that Alexa is the grandmother of our present home invader, he can perhaps tell us during the Q&A.

Today, in addition to Internet Archive, he serves on the boards of the Electronic Frontier Foundation of Public Knowledge of the European Archive and of the Television Archive. It's an honor and a pleasure to welcome you both here today. Thank you.

[APPLAUSE]

- Oh, I'm over here. Right. OK. Thanks so much, Jane, for that lovely introduction. It's really a delight. It's really a delight. Thank you, Brewster. See, we needed Brewster here. I, in any case, needed Brewster here for the most elementary technological support.

It's a real pleasure to be here with you all today, and I want to thank the Schlesinger Library at Radcliffe for hosting this really vitally important and urgent conversation. It's also just an incredible honor to share a stage with Brewster Kahle, one of my great intellectual heroes. So I want to talk today a little bit about how I see the problem. I'm fully confident that Brewster will get up here and provide the solution.

So what I want to do is offer a parable about the problem of the asymmetry of the historical record that as it is manifest in archives in the United States and where technology plays a role in actually worsening that asymmetry. And I'll tell you the story of Joe Gould and Augusta Savage, pictured here. I had an assignment to write a short essay about Joe Gould a couple of years ago, and I ended up writing a short book called *Joe Gould's Teeth*.

I was working very quickly through a mystery that I really wanted to solve. Gould was a homeless bohemian New Yorker. He lived largely on the streets for most of his short life, even though he had been an undergraduate at Harvard before having his first mental breakdown and moving to Greenwich Village right before the First World War. Gould was himself obsessed with the asymmetry of the historical record, and he decided to write what he called *An Oral History of Our Time*.

What he wanted to do was wander through New York and all over the world and write down what the inarticulate said. He would record the lives, the daily sayings, just the happenstance of the lives of immigrants, of African Americans, and of the poor and of the homeless like himself, and he would create the most important historical record ever produced. He considered himself to be like a human tape recording machine.

Gould's *Oral History* was the subject of great fascination during his lifetime. He wandered down the streets of New York carrying these little black and white composition notebooks under his great cloak. And he was written about a great deal, but nowhere more powerfully than in an essay written by The New Yorker writer and famous profiler Joseph Mitchell, who wrote this sort of adoring portrait of the crazy bohemian, clown-like Joe Gould in 1942 called "Professor Sea Gull."

And in this essay, Mitchell-- it was a very affectionate essay. And Mitchell explained that Gould had arranged-- he couldn't carry all of these hundreds and hundreds of volumes of the *Oral History of Our Time* around with him, but he stored them at friends' houses. And he had arranged to give them to Harvard and the Smithsonian the end of his life so that this great archival account of the lives of ordinary Americans could be published posthumously.

Later in his life though, Joe Mitchell, The New Yorker writer who's pictured here, recanted his story of Joe Gould and said that, in fact, soon after he met Gould, he realized that the *Oral History of Our Time* did not exist-- that Gould was insane and had made up this historical record. The reason I ended up writing about this-- it really puzzled me, because I was not convinced that Gould had made up writing the *Oral History of Our Time*. I thought Mitchell might have been wrong about that.

And so I started looking into the life of Joe Gould. And it turns out Gould is the kind of guy who is extremely easy to trace in the archives. Gould had what was known as "hypergraphia." He could not stop writing. He wrote more or less constantly.

And if you look for Joe Gould in Art of the Archive or NUCMC or any other database or record of unpublished papers, Gould comes up all over the place. He wrote pretty much to everyone who lived in his time and era. He kept an elaborate diary, and he did indeed write *An Oral History of Our Time*.

Even going into the Harvard University Archives, his undergraduate records fill many boxes, which is unusual. Most undergraduate records are a single piece of paper. It's like a transcript.

But Gould left this incredible trail behind. And I became convinced that the Oral History of Our Time did exist and that it probably was somewhere.

And as a historian, if you read Gould's description of that document-- what he was trying to compile-- it's very much in the spirit even of this very conference-- the way that he wanted to archive the lives of ordinary people and make sure that their lives were recorded and not left behind. He used the technology that was available to him, which was scribbling in these notebooks day after day after day everything he heard on the subway, in Harlem, in Greenwich Village, in the New York Public Library, in the Schomburg library. Everywhere he went, he would write down what people said.

So it's fascinating to find it, and I came across a lot more. Everywhere I looked I found writings by Joe Gould. His extraordinary diaries are held by New York University in its archive. Joe Mitchell's own drafts of his essays about Joe Gould are held at the New York Public Library, and they offer up great detail.

And indeed, I found in the New York Public Library that after Mitchell's first essay appeared, readers wrote to him congratulating him on the work and suggesting that they had, in fact, read parts of Joe Gould's Oral History of Our Time. So when Mitchell wrote in 1964 that the account did not exist, people started writing to him saying, oh, actually, you're wrong. Joe Gould's History of our Time does exist. I've read some of it.

And people began offering to send parts of it to him. This one woman, Florence Lowe from upstate New York, said she had been very close with Joe Gould in the 1920s. And he had given her as a gift some of the volumes of the Oral History of Our Time. She offered to send them to Mitchell.

And she did send them to Mitchell, who didn't write back and didn't publish them and didn't recant his recantation. This is one of the volumes of Joe Gould's Oral History of Our Time, which is also called *Meo Tempore-- My Time--* as you can see here.

But in doing all this research and the abundant and abundantly discoverable Joe Gould papers in libraries across the country, I've encountered something really troubling. A lot of intimations that much of Joe Gould's diary and Joe Gould's Oral History of Our Time concerned a single person-- an artist in Harlem named Augusta Savage, who was the most important sculptor of the Harlem Renaissance. And there was a lot about how the scraps that I found about Savage suggested that there was something quite untoward about Joe Gould's fascination with Augusta Savage and about chronicling her every writing and doing.

Savage was born in Florida and with the great migration had moved to Harlem, had become an extraordinarily influential and generative artist. She opened a school for young artists. She trained up a lot of artists while working for most of her life also as a laundress. She's an extraordinary, extraordinary person.

But I kept encountering these hints of sexual violence in scraps of very hard to find scraps about Savage in the archives. And there was a good deal of evidence among the artistic community that Gould was considered a monster. Alice Neel painted him here with three penises. He famously went around naked.

So I got really concerned that there was a much darker story behind why Joe Gould's Oral History of Our Time no longer existed and that it involved Augusta Savage. So then I turned my research project upside down and tried to do what I had done for Joe Gould for Augusta Savage-- that is, to find everything that she ever wrote and everything that she left behind to try to chronicle her life.

And you'd think given how prominent she was as an artist that this would have been a fairly straightforward piece of work. You might think that to be the case. Augusta Savage does have papers. They're discoverable through the ArchiveGrid. They're housed at the Schomburg Center.

They are two tiny boxes containing hardly anything except for a photocopy of Augusta Savage's FBI files. She was, of course, surveilled by the FBI, as were essentially all Harlem writers and artists, and her FBI file looks like this. This is a pretty good representation of how Augusta Savage's life survives in the archives insofar as it is discoverable.

You can look for Augusta Savage in the files of the Rosenwald Fund-- the Julius Rosenwald Fund, which funded a lot of her work-- and she does not come up. She is not listed with that collection which is at Fisk University, because the collection is not listed in an ArchiveGrid by subject name. Even if you look through the record there, you won't find Savage's name.

The finding guide of the Rosenwald Fund is not available in any searchable form online. Very recently it's been put available online as a scanned PDF. And you can look here. If you search for Savage, you get zero hits in that finding guide. It's not a searchable finding guide.

In fact, there are many, many, many Savage papers in the Rosenwald Fund archives at Fisk University where she wrote an entire history of her life to apply for that funding-- where she reported on her art, where she wrote in great detail about her experience traveling when she finally got to go to Europe to undertake her work, the influence of other artists on her work, her relationships with other writers-- Dorothy West, Zora Neale Hurston, Claude McKay. There's an extraordinary wealth of information about August Savage, in fact, in the Fisk University archives that are almost impossible to find using any of the research tools that we now take for granted as offering us the ability to find what we're looking for.

In fact, this is not solely about Augusta Savage's archives. It's also about her own work. Savage was apparently-- as far as I've been able to piece together-- so tormented by Joe Gould. There's considerable evidence that he repeatedly raped her. She was unable to go to the police. He was arrested many times for attacking other women.

But New York writers-- famous New Yorker writers in particular-- bailed him out of jail and served as character witnesses, and he never served time. He was eventually confined to an insane asylum where I believe he was lobotomized. But to escape his constant harassment, Savage left New York, left all of her friends, and moved to Saugerties, New York where she had an art studio.

And this is what it looks like now. It hasn't been restored. It has largely been abandoned. Savage also made an effort to destroy her own work. I think she was very much wounded by the experience that she had with a person who purported to be archiving her very life. Thanks a lot.

[APPLAUSE]

- All right. I'm going to have to fiddle around just a moment.

We have an opportunity-- if we think boldly and differently-- to leverage technologies to do something quite different than what our institutions have done so far with archival records. This is our day. We have an opportunity if we're willing to be bold.

I'm trying to figure out how-- as the organization I run-- how to frame it. And the only way that I've been able to come up with-- a top down idea-- is to try to go bottom up. So the Internet Archive is a nonprofit library located in San Francisco-- this imposing white building. We have lots and lots of digital stuff.

We've been really organized to try to get the published record of humankind to a generation that turns to their screens to answer questions. That's the problem that if it's not online, it's as if it doesn't exist. And right now, the 20th century doesn't exist to those that turn to screens. It's locked up in our libraries, and we're not making them available. And it is an enormous problem.

We actually have better access to things from the 19th century than the 20th century. And the 21st century is doing fairly well, but it's highly biased towards commercial interests. So we've been basically digitizing like nuts to go and make things available and boldly putting things up and finding out if anybody is upset. And we find that people generally aren't upset, which is kind of great.

But the idea of gathering as many lawyers as you can find, putting them in a room, and finding out why all the things you can't do something is going to lead you into paralysis and not doing anything at all. But going and finding out if you make things available and then, do people get upset, tends to be a much better way about this new digital transition-- at least that we've found. And we've got a lot of materials up there.

But when it comes to archival materials, this all comes from people that are obsessed about something in particular. At least for our organization, we are a community of communities.



People come to us and say, this has to be done. And we say, OK, how can we help? And that is an open offer.

I thought I would just give a couple of examples. Timothy Leary was very important to a bunch of people in our communities, and so he was a very well-documented guy. He knew he was going to be famous. He went and did clippings of all of his sightings from the time he was in Harvard until the time he died in LA. So there was a lot of record, yet his heirs basically needed to make some money out of-- how do you deal with all of this?

And so they found the highest bidder, which is New York Public Library, that went and wanted to archive his stuff. So they basically took it. And you know that last scene of Raiders of the Lost Ark where there's sort of the things-- you're going down into the third-level basement, and you're never going to see it again? That-- that's what happened to his archival records.

Except New York Public Library didn't want it all. They just wanted the original stuff that could go in those nice boxes or whatever, and they dumpstered a lot of rest of the stuff. So our buddies just went to those dumpsters and said, let's get it, and let's go and do what we can. And so they basically took the clipping files.

And we said, let's just photograph them, and let's just put them up. Let's take the videos. Let's take his LP collection with the little hand notes and stuff like that-- all these sorts of things around Timothy Leary-- and let's put them up. And it's still a work in action as it's going along really by volunteer obsessed people's efforts.

Marion Stokes-- Marion Stokes is a civil rights activist from the Philadelphia area, and she had a short-lived community television program. But she got obsessed that the record on television was not being recorded terribly well. So she said, I'm going to do it. And she just had a large number of VHS tapes and Beta tapes and just recorded news 24 hours a day. Well, it became 24 hours when news went 24 hours. But she just recorded the news in the Philadelphia area, and she's kind of a hoarder.

When she died, her son said, what do we do with all of this? No libraries wanted it except for us, so we said, we'll take it. We're not quite sure what we're going to do with it, but a shipping container full of videotapes came our way.

But she also had all of the rest of the records of her life. And her son said, let's just do this. So we hired a recent graduate of library school, and she's basically just digitizing all the things-- all the pamphlets, the workers, the communist rags from Philadelphia at that time, the posters, all of these materials. And there's some real gems in here.

So her community television had this Haverford professor on. And it turns out that decades later, we find out he was one of the guys that broke in to the FBI offices that stayed secret for decades and basically expose COINTELPRO, which was this awesome FBI program. We have recordings because it was done as a happenstance. So thank you to basically Marion Stokes's

obsessions and her son's interest in her legacy and our ability to just kind of move through the copyright problems.

Ted Nelson-- Ted Nelson invented hypertext, which became the world wide web-- awesome man, also vigilant about recording himself. But he doesn't want to put this stuff up yet, because there's all sorts of privacy issues and all these issues. Oh, there's always issues.

So we gave him a shipping container. He's sort of decked it out as his. And we hope that he keeps it there as opposed to selling it to some organization that will bury it in the third level basement.

But we're trying to digitize it, and our Rogue Archivist Jason Scott-- really, that's his title, rogue archivist. Good title, by the way. He said, well, what can we do? And he said, how about his junk mail cartons? So he'd been collecting junk mail all the way through the '60s and '70s.

And Jason Scott said, well, let's just digitize these. Let's do it as a for-pay project with a Kickstarter. And there's some guys up in upstate New York that are going and digitizing his junk mail. It's awesome. It's really popular actually out there in the net. So this is sort of-- I'm just painting some portraits of alternative ways of, how do you get around all of the rights problems that are generally around?

Then there's Artinian tangos. So the woman that's in the center there-- that's her grandmother to her left Artinian tango dancing in the '20s and '30s. They had this awesome record collection. She had this. She didn't know what to do with it. She donated it to the Internet Archive. We digitized it. We put it all up. All is good.

And it's really helped her family stay together where it sort of rotated around these media type. And there's this collection now that talks about her family, and people are using it for all sorts of things. It's a different kind of archive, but I suggest we can do this now in interesting ways.

A couple more-- born digital. Websites-- how do you go and collect websites? We try to collect all websites, but we have now 1,000 librarians that are going and building subject-based collections to make sure they're really done well and they're done at the right frequency. And we've been able to do this at scale.

So these are actually students. There's a program IMLS funded to have lots of school children go and say what it is they're interested in. And this is the 007 spies collection from the Rocky Hill school. But we've also worked with other people in different areas to go and make sure that the record is kept very well from these born digital eras which are also pretty tricky.

Just thought I'd throw this in. There are other nontraditional things you can do with these archives. We have GeoCities, which was a 1990s home-- it had a lot of those animated GIFs like butterflies or pizzas that spin-- those. So our engineers went and pulled all of those out-- there are four million of them-- and then made a search engine out of them. And it's just a

nontraditional use of archives that were collected for one reason, but we could do something quite else.

Television news is starting to be used in different ways, for instance, to go and find political ads. So this is an archive built for one reason. And then we basically were able to look for all of the political ads to find out how Citizens United is corrupting our country. And by the way, it is, because you can see the amounts of money going into these different battleground states to go and change behavior.

And so I think there's three major things. We've got to get it to be cheap and easy and figure out the right issues to be able to get through. There's digitizing your basement. It's the boxes o' stuff, and they come in lots of different kinds of things.

So I actually had the pleasure of being involved in a patent lawsuit. It wasn't about me. Samsung and Apple were going at it, and we could blow up their patents, which was always a delight for me. So they basically digitized my basement to blow up those patents. And so they paid for digitizing this stuff and making it available online, because that's, of course, what I would do.

Another is, how do you handle the hard drives and the cell phones full o' stuff? It's like, I don't know, right? When somebody passes, you get these things. What do I do with this?

And it's full of all these materials that weren't organized in their day, so what are we going to do with them after the fact? Should we even preserve them, and how do you preserve them inexpensively? Otherwise you end up with forensics units that just take forever and cost too much, and so you don't do very much. So this is a real challenge-- don't have a good answer for it.

And then there's the-- uh oh-- there's the cloud. What do we do with people that live on the cloud, and how do we go and archive those so that they're relevant and actually preserved? Because every one of these companies is going to be dead. I don't know when, but they're going to be dead. And so we have to go and work on these materials.

I'd say it is our day. This can be something that we can do differently if we act boldly, and the Internet Archive is up for acting with those that want to act boldly. Thank you very much.

[APPLAUSE]

- That was fascinating. Thank you. So we'll chat for a little bit up here, and then we'll take questions from the room. I'll start off if that's all right.

I'm so fascinated by your brazenness in just grabbing stuff and putting it up, and it seems to mainly work to do that. Libraries that are housed in universities have a different kind of legal

accountability and a different administrative structure that they're answerable to. Are there ways out of that?

- Yeah. It's all personality-driven. Really, it's personalities. There's more reasons for no than you can count. And it's always, well, you can do this because of blank, which is made up. And I can't because of blank, which is made up. And the answer really is, I'm just not going to do it.

Nothing that we're doing I'd suggest is cavalier. We're very careful to try to not tread on commerce. That's the landmine that causes people to throw lawyers at you.

And we also try not to break any confidences. So when we're dealing with archival materials, there's real privacy issues. If you're dealing with things that are-- John Perry Barlow just passed, and I remember talking to him a couple of years ago about his email. And he said, yeah. He's kept all his email.

I said, I don't keep my email. Because we have a document retention policy in our organization, which means destroying documents because lawyers might come. So he has all of this, but it has all sorts of weird things in it. So he's a very public figure. So how do you make these things available?

And what we've done as a compromise is we're trying to go and keep the raw logs, but we've made up a collection site so that other people can upload to the Internet Archive all of their memories, photographs, things, email recipients that they've gotten from him and adding those into the Internet Archive's collection. And now we've done this with Aaron Swartz. We've done it with a bunch of others that have been sort of like, let's collect the materials from the community right now. What are the stories? What did this guy do to impact?

And I think the answer is just go forward-- that you can do this. You can do it by doing steps at a time. Try something. Don't lead with "no." Let's try.

So I don't think it's because we're any particularly different from anybody in this room. We have a different circumstance, but everybody's got a different circumstance. It's those of us that are compelled to make something happen. And the opportunity to getting things to a large number of people-- people that you didn't think you were going to get to-- by putting things online is so great. It's wonderful.

And I can tell you the reason why I'm sitting here is because I was bold. If you're bold, you get to sit here next year. The idea of doing something and putting it out there and finding out what works, what doesn't-- that's the only way we move forward.

- And so in terms of what you collect, most of what you describe seem to be things that came to you. Is that correct, or are there things that you go out and seek?

- We've been actively grabbing everything we possibly can off of the world wide web, television, radio, books digitizing, but we don't have enough money to do much of it. So exactly what gets done-- we built the technologies, and then we work with people that have collections. So often there's a collection, there might be money, and there's us, and then there's these triangles.

But we try to make it so cheap that you often don't even need the person with money. Otherwise you end up in this grant cycle, which just never ends. My god. It's like, [GROAN], we're going to write another grant proposal. But you can just start doing it, and let's start with what's easy. And the web is easy. We can collect 1 billion pages a week.

- So let me just jump in there, because I would assert that what's easy is actually the powerful and the wealthy-- so just to press you a little bit on that, right? I mean, my point about collecting is that historically, what has been collected is what is easy, what comes to you, what comes to you from people. Take [INAUDIBLE], a person with very great self-regard, a very accomplished, important person whose life should absolutely be chronicled and preserved, no two doubts about it.

But the real interesting work, say for instance now on undocumented immigrants-- the calling of these people undocumented immigrants a big problem for people who are undocumented that don't have papers. And whatever papers they have, they tend to be-- migrant populations might be storing their papers in some distant friends, right? Looking forward, it'll be very difficult to chronicle the experience as historians during this time of real political debate about immigration and border control.

We're not doing a very good job, say, documenting that population. So I would say that's a hard problem, not the easy problem. And if your approach is, let's go after what's easy, what's left out? Do you ever go after something because it's hard?

- We go out after things that people are obsessed about getting done. And that tends to be everything from somebody's mother in the case of Marion Stokes to baseball cards or old things. The Internet Archive doesn't have enough money to have specialists in all of these fields. So what we try to be is tools to make it really easy for those that want to share to make it really easy to share. It shouldn't cost you to give something away.

But in this country, if you give something to the public, you get a pat on the back and a tax donation-- except on the internet. You put something up that's popular, it can go and sink you for ISP bills. Or you have to go and put on YouTube, which knows what happens and then somebody else's ads. So we try to make it so that those that want to give things away can do it easily.

Then there's, how do you get things online-- digitizing it? And we've built these digitization machines. We've tried different approaches towards having other people do it. We found that in general, it works much better if we employ somebody at \$0.10 to \$0.12 a page to digitize,

say, books than to try to even give away these technologies, because people don't use them even when we've given them away.

We've tried selling them, and we thought that would make it so people would use them-- no. We tried giving them away. People swore that they were going to use them-- no. And they would just sit around, right? They'd go into that closet-- that digitization thing, the grant-funded thing. The grant ran out.

So we're trying to find mechanisms to go and get things out there, and cell phones are helping a lot. A lot of people are just photographing things. Also, there are these underground collections that exist.

But I think you're fundamentally right. We're spotty. The Internet Archive-- if you look down the collections and just go through, it's spotty as to what's in there, except for things like the web collection which I think is pretty broad-based. The television collection is quite broad-based. My appeal out of this is, let's work together to go and make it. Maybe it'll still be spotty, but let's fill in some of the spots that you care about.

- Could you maybe then just walk us through this? I was serious in saying, look. I presented a problem. I knew you'd have a solution. Imagine I come to you. No one's ever written a biography of Augusta Savage, because it's very difficult to do, right?

But surely there are materials out there. It's a question of finding them, which is not what you do. What you do is make them available, right? But walk me through a collaboration.

Let's say the Schaumburg Library-- a bunch people got together and said, we have a lot of energy. We really want to have this incredible Augusta Savage collection. We want to have all of her work in it from every museum with every scrap of paper. We're going to decode the FBI file. We're going to do oral histories.

How does that work? Walk me through that collaboration. What are the resources a team of people interested in that would need to provide versus what you provide?

- So the Internet Archive provides free hosting. So there's an Upload button on the home page. Hit it. Try it. Just hit Upload. Upload something. Just try it, and we will try to preserve it. We'll OCR it. We'll make it available. If it's a video file, then we'll put it into a video viewer and the like.

- And there's no fee for any of it?

- There's no fee for anything. If you start hitting it a lot, we'll notice. And we'll go and say, oh. Let's make a collection. Let's name the collection, and what other tools do you need?

We have a little collections group that watches for people that succeed. There are lots of people that say they're going to succeed and don't, so we just look for those that actually hit the button a lot. So that tends to be a way in.

We have tools for digitizing videos on all different types really, really inexpensively. Because if you go through the classic vendors, it's extraordinarily expensive. You just end up in tens of thousands of dollars and grant hell forever.

But if you just say, all right, let's go. What can we do that's easy? And we basically can try to get a lot of that done. If there's bound materials so you don't want to go and scissor the books-- please don't scissor the books-- then we can offer that as well.

I'd say the web collecting is kind of a real opportunity, because it's relatively easy to do if you're on it. So it's going and collecting. Your Savage case isn't a good one, but maybe your current people that are heroes in your field-- let's go, and let's make sure that their materials are well-captured from the net and that their born digital materials are there. And then we communicate with people while they're still alive, because it's a lot easier to have them, I think, put it together.

I hope that Ted Nelson example turns into something. Basically, it's really cheap to not throw stuff away. It's actually really expensive to throw things away. So if he gives somebody, say, a shipping container that's in a warehouse and say, while you're alive, build your mausoleum here.

What's important to you? What would you want people to see when they walk into your little space after you're gone? And basically we'll clunk it closed, digitize what we can. But mostly it'll be closed in those archival sense. Going and keeping that kind of material alive I think would be a very different thing than what archivists do after the fact when their accessioning in somebody's collection. I think that there's something there that's in the physical areas.

So free hosting, digitisation help, trying to surface some of these materials, some mechanisms of how do you do the therapy sessions with heirs-- we think a lot of our job is therapy. We're talking people through the costs and the rights issues. And are they going to get in trouble?

Fundamentally, people are trying to figure out if they're better off or worse off for being online. And if they think they're going to be better off, they're going to do it. If they're going to feel like they're going to be worse off, they're going to find some reason to keep it from happening.

And start small. Start with somebody. Start with something. Just get off the beach and into the water, and see if you get bit. We've been around for 21 years. It's been working.

- So do you ever have objections-- I'm going to go to questions in a minute, I think-- but do you ever have objections from people? I run a library at a university. I have this amazing collection. I don't have the resources to make it available digitally.

I think there's some web crawling that could be done that could amplify the collection and add to the collection. There's a bunch of other things I'd like to do. I have some old VCR tapes that are a part of it. Was the objection like, then I've given it to you, and my library doesn't get credit for it? It doesn't have a home in my institution. We are giving away our content in a way.

- Right. Yes. There's this sort of, where does it live? Who owns it?

- Who gets credit for it in terms of-- does our university? Does our library get it?

- This used to be a much bigger problem 10 years ago than it seems to be now. A large number of people feel like it's OK to have things up on the cloud. The Internet Archive is like a cloud provider. So yeah, it's not on your home servers in your own institutional repository structure, but it gets a lot more visibility. And it is blended in with other collections, which makes it so that they're a lot more seen.

So I think there's trade offs there, but it's definitely part of the equation. People are trying to figure out, should we do this? Are we giving up the gem? Are we selling out? Are we doing right by that figure by doing this? Those are the questions that people are trying to figure out, and we are as well-- and trying to figure out, how do you craft these?

So for a lot of collections on the Internet Archive, there is a separate web page that somebody has that is that collection or whatever it is. And they just link into the Internet Archive as free hosting. Yes, you can get to it by coming to the Internet Archive. But actually, if you do a Google search, you'll probably end up on that home page.

And so you can organize it and contextualize it as you want to, because our tools aren't very good at that. We can do something, but we can't do a lot. So having your own home page and structure is a good thing. Then it's going and being accessed in multiple different ways, and that's a way of trying to balance out these main problems that you see in this area.

I think the current term is "unification." If we can unify collections across institutions-- maybe it is that there are these different Savage materials in different particular people's archives. Can they go and host them in such a way they can be interlinked? And the web was the promise of interlinked. But if you've tried these federated services, they really don't work very well unfortunately, just technically.

So another potential is host them on the Internet Archive. They'll have their different sides to it. But from an end user's perspective, they can aggregate and cross search and cross find. So we're trying. And by going and saying, well, is it now ours as well, at least if there are no rights issues such that the people posting it say that it can be downloaded then these things can be downloaded in bulk and even rehosted and researched in bulk, which I highly recommend.

But if there are some rights issues from privacy or copyright, we do have lending structures and mechanisms of having things be stream only. But if you go and turn those knobs up-- those



restriction knobs-- the use will plummet, and you can see it on the download counts. So err towards openness in this generation, and it will be found much more often in that way.

- Well then, I'm fascinated by this question of the federation model or the unification model. Because is that something that we should anticipate being perfected in some short term of years?

Just to give an example, for instance, a lot of archives happen to have an individual paper collection's passport applications that go back. The first passports were like 1782. There's just a long history of passport applications. Ancestry.com has a lot of passport applications, because that's one of the things that it collects and digitizes.

But would there be a time that you could search through all archives for their digitized passport collections that would be in any set of papers? Could they be federated by type of-- I'm just trying to imagine what would be the utility of such a thing. Yes, you can write, I want to look for everything in any collection that involves Augusta Savage.

Theoretically, we're supposed to be able to do that now. It's just that the asymmetry of funding and resources means you can only do that for basically wealthy people whose papers are at very well-endowed institutions. So I'm trying to think if there's something about-- the federated thing's going to require a awful lot of tagging that would likely be extremely asymmetrical. So again, we're sort of left with this problem.

- It is a huge problem. So let's just take the Ancestry thing. Genealogy and a lot of what we're talking about overlaps in ways that sometimes may make the professionals uncomfortable, but they overlap. We got good digitizing microfilm. And so we just went out to ask any libraries, hey, just give us your microfilm. We'll just burrow through it.

And basically, everyone said no, except for Allen County Public Library that had all the census records-- the census and the American Indian ship records. And so we just burrowed through all of that, and we put it online. And Ancestry just took it and put better indexing on it and put it up-- and also FamilySearch, which is the Mormon equivalent-- more open, more free access-type folks.

And so what we're trying to do is build a bed of materials that other people can build services on. Some of them are going to be for pay, and some of them are going to be free. And it's a little heart-wringing when somebody makes a lot of money off your stuff, but you get over it. And sometimes they give you gifts, and sometimes they just rip you off. But at least it's available in these new and different services.

And what we'd like to do is then get other people's collections to start to match in, but we find that most organizations really aren't oriented towards bulk sharing. I don't know. It's sort of a crime of our world. We think of ourselves in the library and archives world as up for sharing,

but it's often based on the onesie-twosie thing. Maybe it's because we feel like we're going to lose something if we share in bulk.

But in bulk is often the way that new services get built. And it's based on aggregated understandings of the past that you can do statistically now are really exciting, but it has to be done based on bulk access. So at least let's get the public domain publicly accessible in bulk and let different services come up.

And that's a little bit different approach than the let's federate all of these things that are in separate places and try to weave them together into one user interface that makes sense. We've tried that latter approach, and it's been a bear. It often ends up with really patchworky things that don't last very long.

People are still shuffling around with their institutional repositories, I think largely not succeeding. They're very expensive to maintain. And often, they end up fragmenting the record into all these little bits and pieces that all need metadata, so it doesn't get done.

So I don't know. I think there are some different approaches we can do if we think at scale. It's a little different, I think, than exactly what you're looking for and what you've been able to leverage by being here.

- Well, I'd love to hear questions, comments, thoughts from audience. If people could come up and use the mic so everyone can hear you, that would be terrific. Do you know if we're supposed to go till 10:30 or 10:15?

- I have no idea.

- Can we ask?

- Well, this is good.

- So there's a lot on the table. I wanted to ask a question. I'm really struck by the technologist and the historian who started the conversation with symmetry, and I really feel like there's tremendous asymmetry in where the conversation ended. And so I wanted to ask Bruce more specifically.

You seemed not to fully appreciate the implications that Jill proposed and the concern about the erasure, elisions, the hidden stories that grow precisely out of the messiness of human beings-- the story of Joe Gould's sexual violence. And in fact, I think for most people in the room, they understand precisely that the nature of scholarship depends in part on that kind of discovery-- that the things that are the everyday, although important down the road, tend to be the things less problematic in terms of scholarship.

And so I wanted you-- I don't know if this is a technology personality in general, because I hear this in 1,000 other conversations about shiny technology that's going to solve our problems. But it seemed like you guys were like this. And so maybe an obvious question would be, isn't the Internet Archive going to go away one day by the same terms that you describe that these cloud-based services-- that our lives-- we need get them up online? So that's sort of an obvious question. How do you archive yourself with this shiny technology?

And then I think for me, the other question is the question of taking. It's sort of a question about the law of unintended consequences, which is to say that isn't it possible that if your model is actually successful that there will be even greater self-editing of people who want their material online to become the obvious record of the past when, in fact, part of what we get today in archives is the kind of accidental archiving of the messiness of our lives?

Some are brave and know that their story is going to be told in ways that their lived experience was not fully understood. But others-- there's a lot of accidental discovery. And it seems to me that the unintended consequence of this technology would be that the public has even greater understanding that their future record, their future selves are going to be subject to this federated, broad-based, everything's there approach.

And I'm not being a Luddite in this conversation. I'm just sort of wanting you to think about that if you're not thinking about privacy in the way that helps scholarship really advance knowledge about how messy we are, and it's all shiny and perfect and it's just going to get there and we're to be bold, then we may find that the archives themselves really don't tell these stories in the future any longer in the way that they have.

[APPLAUSE]

- Before your first question, I think there was kind of a question. Maybe it was kind of a framing or an accusation. Maybe that's too strong, but I you're right.

So from a technology perspective-- and I'm a technologist. I come from that angle. I've looked for the new capabilities that is able to be done with the technologies that make it so that we don't have to face every past problem plus new ones. So what is now afforded to us based on those technologies that were difficult before?

So for instance, we can do contemporaneous archiving at scale now, where in the old days, we pretty much had to wait for people to die. And then we got their stuff. Not only does that not work in the digital world because it's already gone, the average life of a web page is 100 days before it's either changed or deleted. So we actually have to do things contemporaneous.

But it's not your private drafts of the great novel that are in Microsoft Word on your hard drive that are making it out there. So we have a very different view on what people thought of that as maybe edited or how people can go and craft what it is that's going to be saved about them

more than how our traditional archives are done. That's absolutely true if we take the things that are publicly out there.

And most of the internet archives work has been on things that have been publiced in some way or another-- not necessarily published, but made public. So we don't collect people's private Facebook groups or email collections in general unless they're deposited. And if they're deposited, often they're deposited with some sorts of restrictions on them in such a way that they are not available-- so either to particular communities that are available or not the general public-- because of the privacy issue. So we're trying not to be cavalier about the privacy aspect.

Damn, now he had two other questions. I forgot them.

- [INAUDIBLE]. How do you archive yourself? That was it.

- Oh, yeah. We think that access drives preservation. Things that are preserved are things that are cherished. Things that are cherished are things that are used, and so our approach in general has been to try to be used.

I think the concept of dark archives is a terrible idea, especially in the digital domain. Because trying to actually read things on floppies or things that you stashed away 10, 15 years ago is really, really hard, and people don't know that they want to unless it's kept public. So in general, we're trying to be used then have multiple copies.

The Library of Alexandria version one is probably best known for not being here anymore. And if they had made other copies in India or China, we'd have the other works of Aristotle, the other plays of Euripides, but we don't. So we tried putting copies in a new library of Alexandria-- kind of cool. It's a great place. Please visit-- and then also in Amsterdam. So we've got the Middle East, a flood zone in Amsterdam, and we've got earthquakes in California. What could go wrong?

So I think what we want basically more copies of these materials, but making copies in bulk come up with rights and privacy issues as well. And so how do you go and structure an international library system that will withstand regime changes and walls going up and down? What happens to libraries as they're burned? And they're burned by governments, and it will happen to us.

- All right. Let's go to another question. If I could just ask people to introduce themselves-- not everybody's been here yesterday as well.

- Hi. I'm Melissa Goodman. I'm the codirector of the science program here, and I'm also an astronomer professor here. So my question has to do with astronomy, photographs, and Jill's books and going back to your possible future collaboration.

So before I say anything-- so photographs, right? When you search for photographs now, it's kind of amazing. You can type somebody's name. You can type, I want a picture of a cat. You can type, I want a picture of zebra. And contextually, because these algorithms have been trained, they can find these photographs. And that's amazing.

So that's place one, and that's gotten unbelievably better-- I hope you all know that-- in the last 10 years. Astronomy is notorious-- Brewster probably knows this-- for sharing data openly and federating archives in exactly the way that you're hoping could happen in history. Of course, astronomy data are born digital mostly now. I have interesting projects about historical data.

But anyway, they have coordinates. They have numbers. They are much easier to search than history. So now we get to Jill's books. So Jill, I'm going to flatter you and say that the thing I find the most amazing about your books is how you could possibly collect that much information and synthesize it into a really good story in your head but also figure out what's connected to what.

So here's my question. Let's say that some beneficent world of artificial intelligence-- we'll leave the other part till next week-- figures out how to connect the facts that Jill wants to connect from across these archives-- not write the book for her, but give her a starting point of, here's this information I found according to the kinds of natural language searches that you did. And I kind of tried to organize it for you in a way that you seem to like-- to find zebras, for example. You call that thing a zebra. Is that useful to you, Jill, and is that possible in our lifetimes, Brewster?

- That's a great question. And there are actually some tools that attempt to do that-- not so much for scholars yet, but for journalists. There's one kind of beta program. I think it's called FactFinder. And what it purports to do-- if you're a journalist, let's say you want to write a story about this upcoming Supreme Court case, and it's on a proprietary measure. Well, you could put a query into FactFinder, and it will do what exactly what you just said.

That is to say, it will do a kind of crawl and search and present you with an organized, a connected-- there's a bit of AI to it-- a connected set of statements of fact about this issue that you want to report on. It would almost generate the story. It doesn't really work right now, but let's assume it could one day work. It's not implausible, right?

So no. For me, that sounds like having sex with a robot. There's no way I would ever do that. The whole joy of doing research is actually reading the stuff and making the connections in your head. That's the intellectual labor of scholarship. I would not outsource.

Really, it's like giving birth in a test tube. There's nothing about that that appeals to me as a thinker or writer. I think it would appeal greatly to a college freshman who doesn't want to take my class. So in fact, I think it's quite dangerous.

It's not that it's sinister. I see where there are uses for such a thing. But in the same way that people will say, the thing about Google as a search engine-- if you wanted to design a machine that decontextualized information and made it essentially meaningless, you would have designed Google search engine. That is what it does.

And that is why we get student papers that look the way they look, because people have actually-- they've absorbed information. They've taken it as output from a machine that took it out of context. So with all due respect to Brewster, it is amazing work. I am a Luddite at some level about certain of these things.

So I think it could be built. I think it would be quite problematic for people to use it, because the work of finding, the work of discovery is an important elemental human activity. There's a whole really interesting research about how how we investigate comes from tracking animals in this very primal way. That's what hunting is. We want and need to look at those tracks and see that branch and follow that, and then we get to eat. That is who we are.

[INTERPOSING VOICES]

- No, no, no. See, I want to hunt at the supermarket instead of hunting in the wild. That's what I'm saying. I think it's a partnership-- a human-computer partnership.

And so the same way I'm pretty sure you don't want to look through your phone for every picture of a cat when you want to show people just pictures of cats-- maybe you do-- I think that I never would have done that 10 years ago, because the software was so horrible. But I will do that now. So I'm not saying that we're going to replace you ever, don't worry. I'll go away, but I'm curious what you think.

- It's happening, and it's happening at scale. And you point to Google. A lot of what's underneath Google is a knowledge graph that is a set of-- it's not just keyword search anymore. More and more, journal literature at scale is being fact-finded and distilled.

So the exciting thing-- problematic in many ways-- of what we're getting into is we have people. We have a network that's connecting these people. We have basically digitized and content materials, but we've got this machinery that is going and making it so that you can ask a couple-word query to Google and come up with pretty awesome answers.

I'm glad that you're not spending all of your time and only your time using those, but it's happening at scale. This computer mashing of what's going on is important. And if you want to have an influence on how people are going to perceive their world-- maybe it's mostly the college freshman, but there are a lot more college freshman than Jill Lepores-- is put your stuff in a place and a way that it can be learned from.

Because these AIs are stupid. They can only learn from what it is they find, and so let them find what we think is the breadth of our culture. Otherwise, we will have the monoculture that we're fearing.

- I'm going to take three questions in a row. I want to make sure that everybody in line gets a chance to speak, and then we'll answer them in a kind of group if that's OK.

- Good morning. I just wanted to thank you again for being here this morning. My name is Dominique Lester. I'm the Teenie Harris archivist at the Carnegie Museum of Art.

And I had a question related to just maybe your thoughts on how-- if you're web archiving and web monitoring, what can that look like beyond personal and organizational papers into local, state, and federal government municipalities? And how can web monitoring or web archiving be used as democratic defense for individuals looking at policy changes that might not always be publicly reported? Or how can black and brown individuals in certain communities use the internet and internet archiving and website monitoring as a sign of defense for themselves, their rights, and their liberties?

- Great, thank you.

- Hi. Patrina Jackson, Iowa State University Head of Special Collections University Archives. My question is, many marginalized communities are very hesitant or have a lot of mistrust of large entities because of issues of erasure. This seems like more for access instead of a permanent preservation, because I'm not sure exactly what happens with the analog materials that are given to the Internet Archives.

And also, beyond that, who's doing the outreach for the Internet Archives? Who do they look like? What demographic do they represent, and how would people even know to put their materials on the Internet Archives? Because it almost seems like it's duplicating the same as it ever was.

If the people who are doing outreach are mostly white people, and if the people who know about it are mostly white people, basically it just seems like it's duplicating same as it ever was. So how are you addressing that in your outreach for this access archive? That's what it seems like to me.

- Yeah, thanks. And then one more, and I promise we'll get to these.

- Thank you both. I'm Cecily Marcus. I research African American history, and I think that my work starts somewhere around where you're working, Dr. Lepore. I'll talk about [INAUDIBLE] Search later this afternoon, but basically it's an aggregator that brings together primary source material about African American history from archives all over the country.

So I'm sure you know already that there's letters between Augusta Savage and WEB Du Bois thanks to the great work of Aaron Rubinstein and UMass Amherst. But it can only present what has been collected, what has been digitized, and what has been made freely available, which is not very much material, especially around African American history.

There is also an example with Internet Archive of incredible work by [INAUDIBLE] then at Payne Theological Seminary who didn't have the infrastructure to put her material from Payne Theological, a historically black seminary in Ohio, up on their own digital collection. They didn't want to build their own digital collection. So for their centenary, they digitized a bunch of material, and there's a collection on Internet Archive that we can aggregate for research with an API.

It seems that there is a model for not duplicating digital collections all over the country in every single small college and institution and working with community archives to develop these collections that otherwise wouldn't be available. So I wonder if you could talk or think a little bit about how to make those materials that wouldn't otherwise be available through Internet Archive as sort of an activist repository for community archives.

- These are for you-- all three.

- OK. So our website collecting being done at municipal and state levels and national levels-- and the answer is yes, not as much as they necessarily should. Almost every state in the United States subscribes to our thing called Archive-It. It's a subscription-based service that those guys can go and say, these are the sites they want archived at this frequency.

And it's also done also at the country level. A lot of countries now collect their web pages, because we kind of embarrass them into it-- to go and make sure that .nz or .it are done completely. Often, they'll contract with us. We'll collect it and then hand it back to them.

It has been happening, but I'd say there are not as many libraries involved as we'd like. And we'd love to work with your libraries much more. So please let me know. Let's set up an Archive-It account. Let's try something. Just get off the beach.

- Could you just speak to the defense issue? The specific question--

- [INAUDIBLE]

- Yeah, the specific question whether with a regard that-- the state and municipal stuff is being collected, but are there ways in which the Internet Archive has or has imagined collaborating in ways that those collections could advance civil rights or defenses against government intrusion? Is there a really interesting Black Lives Matter campaign going on with student surveillance or webcam collections or body cam, dash cam-- are there projects like that that you could tell us about?



- Are there collections that focus on those particular issues and trying to make sure are they well archived?

- There have been communities that use Internet Archiving or web monitoring as a sense of democratic establishment of defense.

- Good. So Archive-It is one way, but that's an institutional system. So there are black Lives Matters, for instance, collections that spring up immediately. They tend to be volunteer collections based on lots of people sending in URLs and things that are important to be saved and around pretty much all-- mostly in United States but now starting to be more international. Events happen, these collections spring up and are available.

There's a more grassroots way of doing it, which is Save Page Now. Save Page Now-- people have gone and made it so that they can archive web pages as they happen by just prepending [web.archive.org/save](http://web.archive.org/save), and it saves it and gives you a reliable long-term URL. That's now being hit-- that button is being hit 80 times a second.

It was completely surprising to us, and it's basically because people are putting bots on it. And people are starting to use it to go and archive any link that they've ever tweeted just to make sure that they're there, because it's a disappearing record. Right now, the government is just going in and just taking things away.

And so a lot of the reason why we know of these things is the Internet Archive. We've become actually much more prominent in journalism now, because so much is being wiped away, whether it's the climate change stuff at the beginning of the administration or whitewashing people's resumes to try to make it so that they can look like government officials and not be laughed at. Or there's tweets that are just actively being-- they just live for a very short period of time. So the Internet Archive is being used in a more active way towards building a public record that people can hopefully trust, and it's based on provenance of the organization and some of the technologies that we've done.

- I would just say I have seen instances where, say, a city council is making some kind of housing ruling or something, and they put it up on their website. And then someone points it out, and then they take it down. And it's disappeared, and you can never prove that they even said that at the city council meeting.

What Internet Archive does is they let you say, please crawl that website every half hour or every hour. And there are all kinds of instances where people do capture stuff that was said and was in a transcript. But you can't prove it, because it's disappeared.

So it's kind of perfect for that use, and journalists use it that way all the time and so do activists. I don't know in what community that happens, but people do use it. It's very easy to do. It's a click thing. You just ask for that, and then that exists. It's a way to hold people accountable for things they've said or posted. It's one way.

- Second question was about distrust of others taking care of the archival record, what happens to the analog materials, and what's the bias on outreach. Some of those are easier for me to answer than the others. We thought we would just try to have one digital copy of everything ever published. We thought that would be hard enough, and we found that libraries were throwing things away.

We didn't like this very much, so we've tried to figure out how to keep one copy of everything that we could get a hold of as well. And so we bought a warehouse, became very packed up, and it's a very preservation-oriented off-site repository. And so we filled up that warehouse, so we bought another one. So we're getting better at long-term archival storage.

And so I recommend don't deaccession. Just don't. And if you are, deaccession to us, and we'll try to hold on to materials within our broad framework we would like to keep hold of.

In terms of outreach, we do almost no outreach. The Internet Archive is really an internet entity, so there's almost no outreach. I'd say that the biggest bias we have is towards the young, because they know of us. Until Jill Lepore's paper called "Cobweb," there was nothing-- literally nothing-- in the intellectual press about the Wayback Machine for the 20 years that it existed.

The only thing was in 1996 when it came up, Slate went and said, [GROAN]. Who needs it? It's garbage. Don't save it. Crap. That was the only-- and then there was Jill's article.

But it's become majorly known within people that use the web, and so it's interesting. I'd say if there's a bias in our outreach, it's really towards the young. And we don't have a marketing department. I'm here. But basically, we just stick to our knitting, because we don't have the funding to be able to do much. And that's--

- I'm sorry. We do have to close up very briefly. So if I could ask the last two people to put very quickly the last two questions on the table, and then we'll try to bring it to an end.

- OK. My name is DeLisa Minor Harris. I'm a Fisk University special collections librarian. This is for Ms. Lepore. You spoke about the gaps in materials. Specifically, you focused on the finding aids and lack of a certain HBCU.

And I just wanted to know, from you, what personally are you currently doing to address this maybe common issue? And how are you using your voice, your name, and what you've done-- because you have a nice resume-- too address the issue that you discussed? And also, I'd love to talk to you more about Augusta Savage.

- Great, thank you.

- So my name is Monika Rhue. I'm director of the James B. Duke Memorial Library at Johnson C. Smith University and also the board chair of the HBCU Library Alliance.

And so my question is, how is the Internet Archive being preserved for long-term? Because you talk about funding. You talk about things in a cloud could be very expensive. So talk a bit about the long-term preservation of the things that you're collecting as well.

- Thanks. I'm going to just briefly thank you for that fantastic question. I published an article in The New Yorker about the Joe Gould Augusta Savage story. I ended up publishing a book, which I really fought to have an extensive notes section. And I heard from a lot of people who said, oh, I have an Augusta Savage letter, or I have this, or we have an Augusta Savage sculpture.

And I think if I were a more entrepreneurial person, I would have contacted Brewster and made a website and tried to do something and gather that material together. That's not what I do. I try to cast light on it.

Writing the profile of the Internet Archive was a way to say, people don't know about this, and more people should be using it. And more different kinds of people should be using this. That is what I try to do. I think the storage question you actually spoke about quite a bit with regard to the duplicate copies around the world.

- Copies, copies, copies, copies. Lots of copies keep stuff safe.

- All right. Well, I know this conversation will continue as we move into the next panel. So Jane, do you want to take us out?

- Thank you both so much. It's been like drinking through a fire hose to hear both of you, and I think you've given us lots of questions and problems that we'll cover.