

Amanda Martin-Hardin (AMH) 0:45

Welcome to *Overdue Conversations*, a podcast about the ways archives inform our discussions around history, literature, and politics. From digital publishing to reparative justice, climate change to public health, this series of overdue conversations takes archival documents out of the stacks and into the public forum to consider how collecting practices, selective reading, and erasure of past knowledge informs and distorts contemporary debates.

I'm Amanda Martin-Hardin, a history doctoral student here at Columbia, and a producer for this podcast.

In this episode, Columbia literature curator Lina Moe sits down with Trevor Owens, the head of Digital Content Management at the Library of Congress. Trevor is the first person to hold this position because it's new—in fact, digital content management is new to most institutions.

Lina and Trevor discuss the many, sometimes contradictory, challenges of dealing with digital content. How do we keep the things we want to preserve, but get rid of things that inadvertently get swept into digital archives—like personal, sensitive, or even offensive information?

Lina's conversation with Trevor is wide-ranging, covering topics including digital forensic sleuthing, recovering overwritten data on wiped hard drives, humanities versus computer science training for librarians, and the overlooked labor that keeps libraries going. Despite working at one of the largest repositories in the country, Trevor also brought up the importance of preservation at smaller community archives, like those in tribal communities. Finally, Lina and Trevor discuss the “more product less process” movement within archives, including the ethical questions raised by archival acquisitions like Stanford's 4chan collection.

It's a fascinating conversation, so let's jump in!

Lina Moe (LM) 2:52

Hi, Trevor, thanks for talking to me today on the podcast. We were originally scheduled to talk January 6, and you're based in DC. So very reasonably, you wrote to me and said, maybe another time. But we're here today. And I just saw that the House has voted to impeach a second time. There's still so many questions to be answered about what happened and what it means. So I went today to talk to you just a little bit about our immediate context, and also get to your book and your work as a digital preservationist at the Library of Congress. But I felt like we sort of had to start with the

events that contextualize our conversation. And I wanted to ask you, how would it work, say, if we wanted to create a collection around what happened on January 6, maybe more broadly, I think it would be interesting to know the sort of discussions that you have about the challenges of trying to archive social moments, as opposed to preserving the archives of an author's papers or those of a cultural institution. Do you have any thoughts about that, about how you would begin that conversation with your team? And what are the sort of big issues and parameters you would think about?

Trevor Owens (TO)

Sure. So I think that the first thing: it's good to clarify that my day job, my sort of hat at work, is really about how to work with the things that other people want to get. So we don't pick what we acquire or do those collection development kind of things. That's a thing that falls to subject matter experts. So I'm not in the best position to talk about that from an institutional perspective. But I think the main thing I would offer in that vein is that I think the questions you're asking there are always going to be tied to a particular institution, and what its goals and objectives are for building collections. And so I think starting from that will start to draw out what kinds of things someone might collect or require about really any event or situation.

LM

Well, in your book, you write about how preserving context is important. And I thought that you put that in a really interesting way, which is, on the one hand context is or can be the computer interface in which Salman Rushdie composed. So you note that he had specific writing processes that show up in the way that he used the computer. I loved that he was an avid user of sticky notes, which I feel like I am too, and it's just the worst way of organizing my thoughts! But in other broader contexts, you write--you can think about the letters and the artifacts left at the Vietnam War Memorial, for example--but you can also document how people interact with the wall, and how they behave in the vicinity. So I was wondering if you, if you're thinking about a social event there, what are the kind of maybe specifically digital documents that you would look for?

TO 5:52

Sure! I think one of the things that's really important to think about with digital information is that we've really had a pretty dramatic context collapse between a lot of different sort of lineages or traditions of materials you might collect. So the way that media sort of merges together in mixed forms, or the extent to which people are sort of doing a lot of documentation or recording things, ends up being really powerful as documentation. But at the same time, we've got a lot of really valuable documentation strategies that have developed through work in oral history and those sort of movements. And so I think when anyone's approaching thinking about what to acquire

around an event or an activity, there's now sort of a really broad set of tools available to think about documentation of sources from a moment in time. Do you want reflections? Do you want the liberation? All those kinds of things, I think, draw out a lot of different sorts of opportunities to collect. It's worth underscoring, too, that there's just such a flood of information, constant stream of information, that I think an important thing to be thinking about is that the sort of selectivity of collecting becomes all the more central. I think it's also important for everyone to be thinking about the effects of what it takes to work through and process any given collection, particularly with something that has traumatic components to it, because someone's going to have to work through and review and process material. And there's a lot of questions that come up, I think, increasingly around any number of fraught situations, in terms of who's going to do that work? Who's going to look at this material? Who [does it] serve? How and why is it being collected? Those sorts of questions.

LM 7:39

You raise a number of good points there, so I want to touch on them in turn. The first point that you draw out is that there is a flood of information, and it can't all be preserved. So people worry on the one hand about their digital footprint lasting forever; being haunted by that one Facebook post. But then on the other hand worry about losing things online: what if I have so many files, I can't access them? Or, for me, that fateful day when I switched from my Gateway computer to my new Apple laptop. I don't know if you remember that company? And then I just never switched any of the files, I just lost them and left them behind on that kind of ancient hardware. So in your job, what do you worry about getting lost? And what do you worry about sticking around that you don't think is appropriate to keep around?

TO

Oh, interesting. I think the thing, broadly speaking, that I think are of the biggest concern are for enduring access, or some of the materials that are really small institutions or organizations collect and acquire that serve very specific local communities. So that would be things like community archives, or a lot of sort of under-resourced institutions. So you think about tribal libraries and archives, there's really important perspectives and knowledge that really are anchored in those communities that are, as I was noting, under-resourced to begin with. And I think one of the challenges there is the possibilities for partnerships and collaborations. But there's also, in many cases, challenges around power dynamics and things like that, that are themselves complex for what those kinds of partnerships should look like. So I think that's the, if there's stuff that I'm I would be worried about, it would be materials in those spaces. And then if there's things that I think, in general, there's concern about sticking around, I think on that, and a lot of things that you would want to be tracking and managing and sort of records, retention

schedules, things that records management and set up detailed processes for working with in the analog world. But I think that in many cases, it's easier to keep a bunch of stuff than to parse through it in various organizations. And so there's a lot of sensitive information about people's personally identifiable information, those sorts of things. And I think it's worth underscoring just how frequent we see various major companies having data breaches, and things like that, and a lot of that data. If we all got better at getting rid of data that could cause those kinds of harms, I think we'd be in a better state as a society.

LM

That's really interesting. So a lot of personal data can just get lost in this avalanche of data, and then you don't even know that it's been being preserved, and so it's not being protected either. And to your other point about small scale institutions, I think that it's really clear in your book that you have a variety of audiences in mind. So I was thinking about the disasters that can strike small institutions. So actually, last January, there was a fire at the Museum of Chinese in America that destroyed an irreplaceable collection. It was documents that told the story of Chinese immigration to the US--from textiles and restaurant menus to tickets for passages from China to New York--destroyed all of the collection that wasn't on view. And physical disasters like this happen, and they are tragic. And I think part of your book is trying to keep digital disasters from happening, too. And now I'm thinking more about the losing side of things rather than inadvertently keeping things around. So you talk about keeping multiple digital copies of things. Can you walk me through, like, what are some of these disaster scenarios that might happen?

TO

So in this vein, I think for small organizations, the kinds of guidance that exists around just personal digital archiving, or personal digital file management is spot on. And so there's a lot of consumer level material that's just useful for that as a starting point. But I think the core sets of principles there are making multiple copies; engaging in practices to check on those copies; setting the schedule by which you're going to check them; thinking about various media and increasing those issues around thinking through how syncing happens between copies. Because so many of our approaches and systems that we use are involved through automated syncing. So you can have cascading problems between systems, but I think in a tangible or pragmatic sense, of the kinds of losses that exist. One of the things that inspired some of the work in my book, and my teaching on digital preservation is, as at a public history conference, talking with someone from a House Museum at one point, and they were really excited about this digital preservation training that we're going to do on premise, which is a preservation metadata schema. And I was talking with them a bit more about what things they were

really concerned about. And I said, “Well, the thing that, you know, I’m really worried about, is we have these amazing, unique oral history interview videos that are just on one computer at the workstation. Someone works at that computer every day.” And that’s the kind of thing where I want to sort of grab people and say, “Don’t go to that workshop, go buy a second hard drive! Make a copy of the data off that thing, take the hard drive home.” Even at the most basic level that kind of “get the boxes off the floor” mentality or approach. But I think the thing to be thinking through in any of these situations is very tangible, for what are the most pressing risks to any set of content that matters to you? And in that case, one of the most pressing risks was someone spilling a cup of coffee on that one computer, and it’d all be gone. So I think in those cases, that kind of very pragmatic risk thinking can work really well. You really sort of very quickly hit at some of the most core risks. And I think to that point, your fire example is good, too. Because if you have two copies, and there’s a hard drive in your office and the computer in your office, you know, the hard drive saves you from a cup of coffee problem, but it doesn’t save you from if there was a fire because they both burned in the same spot. But if you took that hard drive home, you’re better off. Or if you’re using a, you know, a cloud provider service, you’ve got another vector for how that works. But it also underscores that your options vary dramatically, depending on the kinds of ways you can provide access to the content. So if you were in a small organization, and you actually could make the material very openly available, you might, if the right situation worked out, it could be that you’ve put copies of things into Wikimedia Commons, and you put copies of them into a few other different places where they can actually get access to news from those points as well. So I think there’s ways to be creative about it, too.

LM

I want to step back just for a moment. Do you think that there are unique challenges to digital preservation as opposed to the preservation of paper archives? Or do you think fundamentally the same challenges apply?

TO

Yes, I think there are unquestionably things that are unique to digital information. The big challenge we face consistently with this is questions about how radically different it is, and is everything different, or to what extent is this the same, and we need to do the same thing? So I think it swings pretty dramatically between the two poles at different moments in time. So I think, on the base level, the thing that’s worth underscoring about digital information that’s pretty fundamentally different than analog information is that one of the core bases of computing is constant replication of copies. So when I open a file or load one, it’s actually like copies are made of files constantly as part of the way of experiencing an object. So that is pretty radically different. And it’s also the case that

digital objects are at the base level, fundamentally different than analog objects, and that they are sequences of bits. So there's sort of a grounded truth base to them. That's pretty different than analog material where an analog material, the way you even assert that a thing is the same as something else ends up being different in sort of a philosophical level. So you know, if I say, the physical objects, persisting forward in time is the basis of working on sort of conservation and preservation for the most part with analog material, with the exception of reformatting, in which you make a distinct new copy of something. And in contrast, all digital preservation is effectively preservation of information bitstreams. It's worth noting that digital information can't exist without being encoded in a medium. So there's always a tangible object that has a bitstream on it. But for the most part, we talked about digital preservation, we're not talking about, like, conservation of media objects themselves. We're assuming that the information is extracted off the objects and then treated in that bitstream kind of function. So that's pretty radically different. I think another thing that's worth noting in that vein is that, Lev Manovich, in [The Language of New Media](#) makes this case that one of the things that's sort of fundamentally core about new media objects is their database nature. So the claim he puts forward is that within a database, there's sort of no first row on some level, it exists for you to query it. And so in that vein, there's a linearity to analog media that ends up being pretty radically different than the base, primitive ways of facet and filter and those kinds of things that come in. And that sort of nested layering of how digital media objects work.

LM 17:08

You give some great examples of how databases organize digital content, and I was really drawn to your distinction between the database logic of Yahoo versus Google. And the way I understood how you explain this is that Yahoo, which originally started as "Jerry and David's Guide to the World Wide Web" (I'd also never heard that) worked from a logic that might be like a finding aid or an index to a book in which the important topics and agents are chosen by moderators versus Google, which created an algorithm that allowed query results to come up based on data in the system. So, links, number of keywords, page views. Is this an apt description of the comparison between the way they're organized? And how do you think that this division is playing out for libraries and librarians?

TO 18:06

Yeah, so the bit about the early days of Google and Yahoo, and the extent to which Yahoo was sort of an expert-curated way of navigating the web, and then the PageRank algorithm that, I mean, Google's drifted pretty far away from what PageRank was originally. But that idea that the graph emerges from the links between the pages was really revolutionary. But I think it is fair to say that there's emergent property of the

media versus sort of expert curation. But I think it's worth noting that that is a thing that on some level is consistent with archival practice in many ways, too, I think someone's in the kind of MPLP framework, you do some amount of description of the whole the entirety of a collection, but you try to leverage as much as possible the order and structure that the material came in as a way to persist that context forward for users to make sense of so in that sense, you've got a mixture of the inherent structure of material and how it was organized over time with expert knowledge on the front end. And I think it's also worth noting that at this point, whenever you do Google searches, one of the core things that comes up is the knowledge infoboxes, which are largely powered by structured data that comes out of Wikipedia. So the Wikipedians, I think, are themselves an interesting valence into this, as well. And that so much of the power of search for us is driven increasingly by that structured data. That is the kind of knowledge graph that comes from this crowd built encyclopedia.

LM

That's really interesting. I think that during COVID, we've started to investigate sort of where hits are coming from to our archives portal, and directly to our finding aids. And Wikipedia is a big driver of that. People come directly from Wikipedia, or people link finding needs to Wikipedia. It's interesting, because I ran into some moderators that said, or I read the comments of some moderators, who said, "you know, this isn't a giant archival repository, it shouldn't be an encyclopedia for everything." And yet, it kind of works that way. And people want to add direct links to libraries.

TO

I think the work that's gone on around GLAM wiki (sort of gallery, libraries, archives, and museums) Wikipedia over the years is really great. So there's some good progress that's been made in that space. Wikipedia is really exciting on the one hand, in that it is an amazingly successful venture that is fundamentally non-commercial, that is a tech startup sort of thing. And in that vein, I think it has a lot to offer for thinking about sustainable, non-commercial ways of technology infrastructure working. And one of the things that's particularly important about that is there really is governance to Wikipedia, like it has bureaucracy, and that's frustrating. But it's also, like, the sign of something that becomes healthier. At the same time, the last Wikipedia bit there is that it has a really, really problematic diversity of viewpoint and participant issues. And it's got its own parts of it that become toxic in terms of how fights about notability play out and all those sorts of things. So there are a lot of folks who have sort of given up on it in the struggles that come from trying to get it to be well governed and run. But I think there's still so much good there that I remain optimistic that further participation can get it to become stronger and more equitable. And there's a lot of great efforts, many of them

anchored in libraries, archives, and museums to try and help improve that knowledge base.

LM

Were you interested in libraries and collections as a kid? Did you collect anything?

TO

I mean, in terms of collecting things, I had, like, Star Wars figures. So I bought old Star Wars figures. Um, but I mean, you know, I think more or less like any kid, I loved to go to the library, checking out books. The thing that really got me interested in libraries in a big way was when I started working at the Center for History and New Media at George Mason University on Zotero. The main way that that open source reference management tool got into the hands of users was by librarians doing reference training, those sorts of things. And so I was really excited and impressed by that. And the Center for History and New Media has a long history of being anchored in the idea that historians should be involved in collecting and preserving in ways they had been a lot more in the past. So there were a lot of collecting projects going on there.

LM 22:51

So I wanted to ask you somewhat of a technical question about the sleuthing element of digital preservation, with the caveat that I've only read about this through your work. But I do want to say, Trevor, that one of the wonderful things about your book is that you equip people to ask questions about digital preservation with basically no background in it. So you describe a procedure in which using a Hex editor, you might be able to recover a bit stream of deleted files, something like the Archimedes palimpsest, which, for listeners without this information at their fingertips, is the name of a 10th century Greek copy of a previously unknown work by Archimedes that was erased and overwritten by a religious text in the 13th century. But the material traces of the older texts were still visible to scanning. And so it was rediscovered. But you say that digital works can be recovered in this way, too. So even though something is deleted, and maybe even written over, the original information can be forensically recovered. Have you ever conducted this with an archive under your care at the Library of Congress? And in what situation? Would you turn to this sort of very involved forensic work?

TO

Oh, sure. So I can give a couple examples and work my way up to sort of stack of ways that this works. And I think the main thing to underscore with this is that, it's important to understand how data is written and understood in a computing system. And that changes depending on the exact structure of it. But really the best resource to go to on understanding this is to get Matt Kirshenbaum's book on [New Media and the Forensic](#)

[Imagination](#). So Matt's book, I think I've probably read 10 or 15 different times, it's just amazingly well written and it has these great detailed dives into some of these issues. But the core things that come through for understanding that is that for the most part, when a computer deletes a file, it actually just marks it as space available to be rewritten. So what that ends up meaning is that in his book, in particular, he's got an example where he downloads a ROM for a game called mystery house. And he opens it in a hex editor, which is a way to sort of render the blocks of storage stored on the disk. And within that, he is able to browse it and find text that doesn't appear in the game, which is pretty revolutionary. And you look into it, and the text that was in that game is actually from a different game. And from that the inference becomes, what probably happened is the person who uploaded a copy of the disc, uploaded a full copy of a disc that they had actually saved different games on over time. And so you could actually see a game that was on that disk before it was overwritten with a different game. But there's also a deeper layer to that, which is when hard drives get wiped, they get overwritten a ton of times and like a lot of organizations actually have shredders for hard drives, too.

LM

Like a physical shredder?

TO

Yes, they will shred hard drives, because it's important to know if someone has put a lot of resources into it. You can in some cases, read as many as five or six previous writes backwards on a hard drive. Because every time the hard drive writes a band, it's always slightly out of alignment with the previous band, because it's worth underscoring, again, that this is all physical, tangible stuff. And so you know, there's a bid sequence that gets written onto a disk that gets read off. But if you do spend a lot of time with it, you might be able to see the bands that were there before that are partially overwritten. And then, the last example that I use in the book is interesting. It's actually a thing that is in the Library of Congress's collections, but I had nothing to do with how it was done. Before I got there, it was done by a researcher, Doug, who resigned. He's now a curator at New York Public Library working with the Performing Arts collections. And in his case, he had a situation where he opened up a text file of a musical and saw that all the text was very different than the final version. In that case, it turned out that because he rendered that document in an application, it wasn't written in, but it could still read the text, it rendered what was actually the original version of the text, because I think it may be WordPerfect, but it might be Wordstar. But in any event, that tool had a thing that was called fast saves, where when you did save, it would just append to the bottom of the file the edits. And so the way that file works, if you open it in a different piece of software, it shows you the very first version, and then lays out all of the sort of changes that occurred to

the document over time, which is really exciting for thinking about reconstructing someone's work. But that's a case where one of the metaphors I like to use with this is that, weirdly, there's a lot of nooks and crannies around how digital files get saved or moved, that ends up leaving a lot of traces that you can carry through to find unintentionally remaining material.

LM

The way you describe that last example of having both the text and the history of changes made to it is just probably a dream for a literary scholar. And it is, in a way, easier. I mean it usefully collates all of the research that one might do into a single document.

TO

Yeah, no, it definitely was, and Doug's written a few things about it. Doug, and Matt also, were both involved in writing a clear report about digital forensics for libraries and archives, which is also outstanding. So I can't recommend that enough as another source for what it really are some kind of mind blowing stories about where and how information exists and persists in different systems.

LM

I spoke with Matt about his newest report book files about the publishing industry, and how the move to digital publishing presents a number of new challenges. One, something that you've talked about is that the constant creation of new copies makes it incredibly difficult, especially if nobody is responsible or motivated for organizing those copies. The other thing that we talked about is the sort of fragmentation of the publishing work across both employees and independent contractors. So people who might not be paid well enough and not be attached firmly to a company so that they won't keep their files. And so that makes it more and more difficult to have a kind of coherent literary history of publishing. And I'd actually like to ask you some of the labor questions that his book, I think, brings to the surface. One thing that I wondered is your job, I think, sits between or includes both expertise in computing and in archiving. And I think that there is a new power to having some computer science background in the library and curatorial and archives field. And I was looking at a job listing at the Library of Congress, a digital innovation specialist, which sounded really fun. And I also saw that the minimum salary was twice that of a curator at Columbia. So I think this is somewhat of a tricky question to ask, because questions around salary are always so taboo, even though some people are working to make these more transparent in the interest of equity. But I guess my question is, do you think that the positions with computer science backgrounds, given how the market provides more lucrative alternative careers than, say, to humanities curators who aren't going to scamper off to

work for Google? Do you see pay as being a problem in libraries? And in the digital preservation field? And do you think there are ways of confronting pay disparity?

TO

Yeah, I mean, I think it's an interesting thing to get into different sorts of organizations. If you haven't checked it out, there's a relatively new book that I'm pretty sure named--Anne Helen Peterson has come out with--*Burnout*. And millennials as a generation, one of the things she gets into a lot is just the spiraling set of challenges around precarity. And work and pay for anything that someone might think in her terms, she refers to as "a cool job," like people want to have a cool job that they tell people about. And so I think librarianship falls into that or like journalism, marketing, any number of these things. And all of those are vexed with the same set of problems. In many ways, I do think it's an important thing. I think that the Digital Library Federation has a group focused on labor issues, and I think a huge project and effort they're around, particularly for grant-funded and sort of precarious positions, trying to clarify and firm up what the expectations are for when someone's getting hired as a project archivist, those sorts of things specifically with our work. There are good things and bad things about the way that the federal government aid system, the general schedule works, but I do think in general, it creates a lot more stability. And it does a lot of good things, I think, for librarians and archivists. And so that's there because there's this pay band system that's relatively well defined. There are some differences in jobs and issues there. But for the most part, when we hire librarian positions, they're not radically different than the bands that the IT units work in. But getting back to one other part of your question, our positions for digital collections specialists and senior Digital Collections specialists that are librarian jobs, have scripting and data management, manipulation--things that someone might think of as being as having been IT kinds of jobs. But we've been highly successful in getting people trained as librarians with library degrees that know how to do things in Python. They're not software engineers. That's not what we're trying to hire people to do, but they are knowledgeable about Library and Archives practice and are also able to roll up their sleeves and do the kinds of things with Regular expressions are whatever needs to be done to kind of work through an issue or a problem. And they're getting those skills from library schools, which is great, because the library schools I think are creating pretty robust course offerings to support people learning how to work with an API and do metadata transformations, those sorts of things. Many schools are teaching things about how to work with bit curator and those sorts of things, which is great, too. But then on top of that, a lot of our role ends up being needing to be able to work as product owners in various IT systems and services, sort of being able to talk with folks that really are engineers in the computing sense that know how to build and manage well designed software systems. But it is really important to develop the skill set that what someone as a product owner does in a

scrum or agile software project is being a subject matter expert, but also understanding what's reasonable or not reasonable, a lot of the constraints around IT systems. So in that vein, I find myself--I wouldn't trust myself to write computer code or scripts much at all--but I feel like I have a really good handle on what was sort of doable and not doable from that kind of product ownership side. And I think that is a space that will see more and more growth for librarians in the future, because we do need to bring that knowledge of, you know, in the language of the search of things, the business of the organization into collaboration with people who are really experts on the more infrastructure side.

LM

So I was struck by a line in your book in which he wrote that highly technical definitions of digital preservation are complicit in silencing the past. I just wanted to ask you to elaborate on this line. What were you thinking of, and can you give an example?

TO

Sure. And in some ways, the story that I gave earlier about, you know, the spilling coffee on the hard drive in the House Museum, or something like that is, I think, somewhat emblematic of this where a lot of the work on digital preservation today has really come from some big institutions: a lot of IT resources, getting into really technical sort of complex system design conversations and things like that. And so those are, that's all important and great, but it ends up often creating a space where folks at smaller organizations can be like, "well, I guess I don't have the resources to do that stuff. So I don't know if I can do anything." And so that kind of "get the boxes off the floor" mentality I think is hopefully a way to say you don't need to know everything before you can do something. And here are some things you can do in that vein. I note too, that the Digital Power Project--preserving digital objects with restricted resources--has done a lot of good in developing training and support for this. So I think that's what that is really focused on, is that the more jargon the more elitist sorts of things get built up around it, the less permeable the field is to many of the people who are the most, you know, the Lone Ranger's out there: the people at a historical society, a lot of organizations entirely run by volunteers, those are people who, I'd really love to have some basic info about how they can get involved. And I think in that vein, if you look at things like Memory Lab, which DC Public Library started, but has now been sort of franchised and spreading around beyond that, sort of setting up spaces, and public libraries that provide that kind of training to everyday folks about their family photos, and those sorts of things, you know, families or institutions, small religious organizations, all these groups that we really rely on to, to create a really broad net that documents and collects our experiences. Those groups need very different things than this sort of very complex technical jargon diagram universe.

LM

And to get it, the hardware question you talk about, how at different points in history, new media are taken up with breathless enthusiasm. And you have a great line in which you quote The National Colloquium on Oral History saying that the tape recorder is important enough to oral history to constitute a part of the definition of oral history. And so, at brief moments of time, new media, or I'm thinking here, techniques of digital preservation, become so important as to get confused with the definition of the thing that it is being used to preserve. So I wondered, on the one hand, what digital preservation tools you see are taking these kinds of breathlessly important roles today? And I think an implicit argument of your book is that this is almost always going to be wrong, or at least temporary.

TO

Yeah, I think that's a good example you have brought you up there, which I also really love. So I'm thrilled that you like to have the tape recorder, and oral history is a useful thing to turn back on. All of our media and memory and sort of mind in the broadest sense, right? So early in the book, I tried to get at a few different ways in which preservation happens. So this notion that there's a sort of folkloric function of preservation, which is about repetition and variation versus carrying forward of specific objects, versus making copies of information on media and the way that those things play out in different ways, information, artifactual folkloric and sort of frames on it. And it's worth noting that the big picture of the folkloric system is the longest running one of them. And it has really different sets of assumptions. So if you think about everything that you know, that you didn't read in a book, or read off of a physical object, that's all part of that folkloric system, which is sort of very much a part of our bodies and our the way that we socialize with each other as a part of language. Think about the way that much of the, you know, you think about something--like the Odyssey or something like that--it exists in a preservation system that predated written language, based on orality and repetition. So it's worth keeping in mind that embodiment media and the way our own faculties work in the way that we absorb technologies into our sort of extended mind, too, are all really important aspects and thinking about carrying forward information. So I think the part of it that I will come back to, that all of our media innovations have, in some sense turned inward and changed some of how we see ourselves. I think the most recent one I would point to is that change that has slowly come about of us all carrying around these, like, phones that are media capture devices and also media exchange devices. It's weird that we even call them phones, right? Because that's like the last thing you use them for at this point. But that's been transformational, and just how we live and experience the world and document it. I think

now, the fact that the interfaces increasingly to a lot of our systems are voice based is also really weird and strange.

LM

I think it'll be really interesting to see with this move towards voice based, which is basically an analog system, meaning you can listen to things at twice the speed. But at least for me, so far, familiarity with the media form, it's a lot faster still to read than it is to listen to things. So the return to the analog and the return to something that you have to listen to, from beginning to end, you can't skip around (like a Codex) I think is going to be really interesting vis-a-vis, the idea of a database or how lists are organized. So you've written about the "more product less process" movement, which prioritizes access to collections, above any item level processing or description--and maybe also above preservation--using the quote "golden minimum," which is green in misers terms. Which I think is so important, because the institutions that I've worked at have incredible backlogs. It's really one of the most serious problems that repositories face, who really have the responsibility to make their collections available to users. But I also think it's a really tricky commandment to follow. One, I get the sense that not all archivists want to work this way. In fact, some might say they got into the profession, because they love deep dives. And two, it seems to run contrary to some of the efforts to provide more context, that you highlighted also in your book, more warning and protection around sensitive archives, just a richer sense of what an archive is, so as to more fully understand it. And I think it's interesting that you give the example of 4chan, and I love that you used something this challenging that, in this case, Stanford went entirely in one direction, which is not doing hardly any processing. I imagine, because if they'd started, where would they have stopped? So I guess I just love to hear your thoughts about these two kinds of counter currents to the "more product, less process" movement, which I acknowledge is very important for access. So the counter currents are one, I guess, the joy of being in the weeds, and to any ethical duty, we might have to get into the weeds, especially with some collections.

TO

Yeah, I mean, I think some of my key takeaways from the **MPLP framework** is one, that the idea is that you're setting the defaults, and those and having decisions about when to opt for different defaults is important, too. So if you set that golden minimum default, and then the other element that comes into play, is one of the core concepts, is that what you do for description, arrangement processing, all those things. Align at the same level so that there may be things that you know, you need to do really close, and things that you're doing more bulk. But I think the other important element there is that I really love their their concept of the difference between when it's time for a shovel and a tweezers. You can still do some important work with the shovel, like, even just deciding

early on, like, what is the value of that series? And what is the risk of how much pie is in it? And just deciding, "we're just going to appraise that whole thing out at that level," can I think potentially serve both those goals at the same time. But I do think that what the right sort of level of processing work to think about doing with any set of material is in a given context for a specific purpose. So when we get to the point where we're really thinking about this is, you know, the default is this golden minimum, but then there's going to be stuff that it's important to invest more time in, or I think similarly, you know, if you think about the collection level, descriptive work and organization work, you can do a lot of detail work at that level, just having appropriate collection descriptions, and the highest level descriptions of them is still important work that is well worth doing in a nuanced way. But it's just very different than how far into the weeds you go with the individual objects associated with it in the Stanford example. I mean, that was an interesting one in that someone basically, as far as I understand, said, "I would offer you this bulk of stuff I have. Here's what I know about it." And someone at Stanford said, "that's interesting enough, let's get it and put it here, make it available," which is useful. It's also worth underscoring that I think them not having gone into do detailed processing in it is also potentially better for everybody like someone who finds their way to that and makes use of it is going to have had to it being not very easy to discover the granular parts of that body of material is probably better.

LM 44:44

Well, just to end, what do you think we missed about digital preservation, especially for maybe those users of the library who aren't that familiar with what goes on behind the scenes?

TO

The biggest thing I would say is, alongside the sort of get the boxes off the floor message, the other message I really try to hit home is that digital preservation is not primarily a technical problem, it's a social problem. It's a resourcing problem, it's a staffing problem. If you want to know how serious an organization is about digital preservation, the first people to talk to you are probably not the folks working in it or on the systems or the folks in the institutional repository. But go to the finance people and ask how they allocate their budget to support something that in many cases for a library should be a core functional spend, right? It's not a thing you project fund. It's not a thing that you do on the side. It's sort of very much at the heart of your institutional mission. And the only way to do it well is to pay people well that spend their time thinking about the problems that they're going to face, and come across who worked together to implement the best possible system to mitigate risks as they currently appear, and are doing a lot to think about what those risks are going to be in the future. So I think that that sort of preservation as a people kind of message is something particular, when this

sort of ideology around digital infrastructure and information is that systems can do things instead of people. But I think now after half a century or more of history, we have digital preservation really happening. And that people in libraries, archives and museums are doing it and making it work shows that it's really about empowering people and giving them the resources to make those decisions about risks, and not about a particular technical super solution at any given moment in time.

LM 46:32

Thank you so much, Trevor, for talking to me today. I really enjoyed the conversation.

TO 46:38

My pleasure. Take care. Yeah, it's always fun to do these, and it's really fun to do them with someone who read the book so closely. So thank you!

AMH 46:48

Thank you for listening to Overdue Conversations. This podcast is published in partnership with Columbia University Libraries. It is researched and produced by Lina Moe and Thai Jones. If you enjoyed this episode, be sure to like and subscribe to Overdue Conversations wherever you listen to podcasts.

Transcribed by <https://otter.ai>