

0:00

Claudia Rankine: (Audio clip): I feel like *Citizen* was just the next book that looked at the same dynamic that Toni Morrison or Frederick Douglass or James Baldwin was looking at. And so if I had published *Citizen* in 2007 or 2012 or 2015 or yesterday, it would have the same mirroring relay effect, because those events have been going on and continue to go on.

0:43 Lina Moe (LM)

That is the voice of the celebrated Black American poet Claudia Rankine, from an interview that she gave in the wake of police brutalities against Black Americans in the summer of 2020.

We often think about books as having an antagonistic relationship with time. The messy, complex timelines of drafting are hidden behind a precise date of publication; and the ultimate achievement for a book is to be able to stand outside of time altogether as a 'timeless' classic. But Rankine is not arguing for the timelessness of her book *Citizen*. Rather she is pointing out its tragically persistent timeliness. Its critique of race relations is as relevant today as it was when it was published. As I discovered in my interview in this week with guest Matthew Kirschenbaum, Rankine's *Citizen* is the an example for thinking about books not as timeless aesthetic objects but as historical archives: as living, breathing documents that don't avoid the taint of time but are instead a testament to the changing times in which they are written and read, re-written and re-read.

Hello and welcome to *Overdue*, 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. My name is Lina Moe and I'm the curator of literature at Columbia University's Rare Book and Manuscript Library.

Today we'll talk about one of the developments in publishing that allowed Rankine to update her poem to make it more responsive to unfolding events: the transformation of books into bitstream.

Authors and publishers' archives are invaluable for academic research as well as for popular books and biographies that excavate the private letters, or personal foibles revealed in the clutter that many writers leave behind. Today, we talk to an expert in digital archives who says this may be coming to an end.

Matthew Kirschenbaum is Associate Professor of English at the University of Maryland and the recipient of a Guggenheim Fellowship. His work raises crucial questions about how, given the transformations of how we write— on computers, using digital files— and how literature is published, the record of literary production is at risk.

Digital work has lots of advantages, but also so many unexpected costs. In the publishing world, like other businesses, work has been flattened into the capacious inbox— or grown to include

ephemeral text messages and slack channels. What happens when editorial correspondence with authors or their edits on a manuscript sit alongside memos from HR or requests for PTO in the same inboxes. Will those messy archives be sorted out so that the record of cultural production of the past twenty years and into the future is allowed into archives and made available for public use? Will researchers ever see what went on behind the scenes for 21st century publishing?

Digital work has also allowed the fragmentation of the book publication process in new ways—so designers may be hired on a contract basis and keep their files in personal drop boxes. Will those files be preserved? But also, the prevalence of independent contractor and gig work has implications for the security and pay of workers in publishing.

Of course, not all digital innovations in publishing are bad. Let's go back to Claudia Rankine's book of poems. One experience of digital publishing was brought home to me by a student describing how they read aloud from it in class. They started reading together from the part of the book where she lists Black Americans killed by police. The voices from the earlier editions dropped off, and the last person reading had happened to buy the most recent edition of the book. That student had the copy that documented Rankine's ongoing involvement with the text. And, because Rankin's book was digitally published, it's relatively low cost for her publisher to keep updating it. There are dozens of different editions of the book and it remains an open record, documenting the political work still to be done.

MM 5:09

Matthew Kirschenbaum, welcome to the podcast. So I want to begin with: you are concerned, I would say, with our ability to preserve cultural heritage today. Why is that? And what do you think is already or soon will be getting lost?

MK 5:26

Yeah, thank you, Lina. And thanks for having me on the podcast. I think one way into that question is just by kind of scrutinizing some of the vocabulary. So when you say "we," you're concerned with, or "our" ability, I mean that really expansively: certainly professional archivists and librarians, but also scholars, other people who have a role in the cultural sector, and some of my most recent work that would extend to commercial publishers, and even members, really—everybody—people sort of in the population at large, we all have a lot of digital stuff nowadays that were sort of concerned about, sometimes its longevity. And so in terms of what does concern me, everything that I've been writing about, really, for the last 20 years are most of what I've been reading about anyway, sort of revolves around the cultural shift to what we might think of as the "born digital." And that's, I think, as others have acknowledged, that that can be a somewhat problematic term. Nothing is really ever truly exclusively born digital. Everything always has a kind of material base and component to it. But nonetheless, and that kind of commonsensical way we write something nowadays, and as often as not where it was, it means we're opening a file on our computer to do it. And so what happens when the raw material of cultural production, all the sorts of things that archives, that manuscript archives and libraries might have collected in the past? What happens when that becomes what I've called a bit

stream? Basically, the ones and zeros we talk about all the time? How do you collect that? How do you preserve it? How do you ensure that it's accessible to posterity to future generations? How do you even sort of, kind of get your arms around what all might be kind of floating out there in the clouds and in the ether? So that's sort of the the broad backdrop for my concerns

LM

And you position this shift, really among the major shifts in publishing and print history, you know, alongside the printing press, the linotype, the laser printer, and other really huge technological and social and institutional shifts that have been studied in literary studies and material text studies. So to get into the weeds a little bit more, both of your recent books tell the story of what people do all day, who make books, both these people who write them and track changes, and those who publish them in book files. And there are many museums and special collections that showcase the history of the change of print from changing fonts and paper making over the past 200 years. Or some even have small presses on display, like the one in the reading room at Columbia special collections. But I have yet to visit a special collection that is demonstrating the path from Word files to Google Docs to InDesign files through all of the digital iterations that you track, as books get made and published today. And the broad term for that that you give is digital asset management, or DAM. So when did DAM come about? And is it unique to book publishing? Or does it govern all kinds of digital workflows?

MK

So yeah, it's that darn DAM that we could talk about digital asset management system? Let me just try to start also at the beginning of your question, because I think you're really, I mean, you're exactly sort of right to kind of see the connection between these two projects—the word processing history, and now the books dot files report on publishing books—that files actually very much grew out of track changes the word processing book, in part because I realized, having written a couple of books myself, I basically had no idea how books are made nowadays. And then I sort of looked around me kind of disciplinarily. And it seemed like a lot of people, folks who were say, early modernists could actually tell me a lot about how books were made back in the hand press era. Now, there's deep, deep reservoirs of knowledge and expertise and that, but pretty much nobody I talked to—none of my colleagues or other scholar, friends, all of whom have written books—can actually tell me how the book was made. And so I wanted to know, and that then led me to these questions of preservation that are the centerpiece of the report, to take us back to the DAM, the digital asset management system. A dam is really a database. I mean, that's what it sort of is, and kind of the most generic sense. So there's nothing particularly mysterious about it. But what I think is really significant, and where I begin the books dot files, report, and dams are used in many different sort of sectors and industries there. There's nothing particular to publishing about them. But around the year 2000—so y2k, the infamous y2k bug—right around that same time, they start to find their way into the publishing world, in part because people who work in publishing houses are having to maintain more and more data in digital form. It used to be—and this was something I learned doing the Track Changes book—that even those writers who were pretty precocious, and ahead of their times, they would send a floppy disk to their publisher back in the 80s, or 90s, thinking they had done the publisher some great favor, the first thing that publisher would do with the floppy disk they

got in the mail would be to print it out, and then turn it over to a typist, somebody who was actually called a word processor, that was their job title. The vast majority of the time, that would have been a woman. The industry was very gendered. But they would then take that and read and retype the whole thing. And so it took a long time for publishers to kind of really make the digital transition. Amazon, by the end of the 1990s was a key player not only because you could buy books through Amazon, and not even because of ebooks at the time, but rather, because publishers found themselves needing to constantly feed information about their books, about their list to Amazon for Amazon's listings. So suddenly, you had to have blurbs and first pages and metadata about the book that was available in digital form that could be pushed out to Amazon and other online vendors that started to happen by the mid and late 1990s. So publishers needed to manage more and more digital information and gradually they sort of began thinking in terms of not just storing metadata or snippets or excerpts from the book, but the book itself: the whole book in digital form. And so the book from the publisher standpoint, right up until that moment that it actually sort of gets into our hands as the familiar object that we all appreciate. The book is really a set of digital files. And those digital files live there, managed and tracked and versioned and secured in one of those digital asset management systems, which really, again, it's just a highfalutin database.

LM

And you bring out in that answer a couple of things that I want to return to later, like the issues of labor, and how also the economies of distribution drive this shift. But one thing I want to get to first is that you make this pretty compelling claim that the DAM is going to make it more difficult for publishers to give materials to libraries. Why is this?

MK

Yeah, there are a couple of different things that sort of come together, I think that sort of create the kind of problem landscape that the report sets out to address there. Certainly the technology itself, digital files and formats, files are easy to lose track of; formats go out of date; there is really no standardization across the publishing industry. So for example, I know at Columbia, you've collected at least in the past archives from Penguin Random House, you have holdings from Harper & Row. There's no sort of particular reason to think that Harper & Row, now HarperCollins and Penguin Random House have adopted the same in house sort of solutions and strategies for managing their digital information, which means that when you get their records as as archivists to process them, you'll have you'll have to reinvent the wheel not once, but twice. In addition to the predictable kinds of technological barriers and considerations. There's a whole host of what I call by way of shorthand, legalistic considerations. The publishing landscape, starting back in the 1980s, began to consolidate small independent publishing houses were bought up by larger houses, which were in turn bought up by multinational media corporations, that have their handprints in 20 different industries, and sectors publishing is just one of them. But there's the same parent corporation throughout. And so we're left with nowadays what are called the Big Five: Simon and Schuster, Hachette, Penguin Random House, HarperCollins... Who am I forgetting?

LM  
Macmillan?

MK  
Yeah. Macmillan would be the fifth. Anyway. So, the point is that, within the newly corporatized and consolidated environment of publishing, the lawyers enter into the picture, meaning that information data, the digital stuff of which books are made, it's not just cultural heritage. It's not even just intellectual property. It's also potentially a liability from the standpoint of discovery, corporate secrecy, all of the things that we sort of associate with the corporatized world nowadays. And so that means that publishers who once had this kind of very close and intimate relationship to their authors, to their readers to the sort of enterprise of literature and heritage, and all of that are now subsidiaries in a corporate structure. And they have to play by those rules. They have to listen to the corporate attorneys, and they have to follow in house protocols for information handling, or just to kind of concretize that it's commonplace at many large corporations, not just publishers to not retain email for longer than, say, 18 months, and the email is routinely purged on the kind of set schedule, if we think of email as analogous to the kind of correspondence an editor might enjoy with an author and which had traditionally been the stuff of literary archives. Nowadays, that is, it's a very vulnerable format, not only because it's email and digital, but also because it's subject to those corporate information handling protocols.

LM  
You make a great point when you talk about the flattening of documents and records into equally restrictive protocols. And this made me think of the proliferation of government records that are labeled top secret, which may or may not have any valuable information in them. And the fallout is that you create a culture that's less likely in the future to prioritize making documents public. And this seems to be happening in publishing right now.

MK  
Exactly. And there's also no commercial incentive. And so, you know, the fact that an English professor at the University of Maryland might access a publishers' archives to write, you know, some sort of amazing scholarly article about a particular author—there's no financial sort of prospect there, which means that not only for just strictly mercenary reasons, but just because everybody is so busy nowadays, we're all so pressured and our jobs our day to day, no one has time to think about preservation and posterity, certainly not if it's not directly connected to the bottom line, it's just not a priority, and it never will be.

LM  
One bottom line that we sometimes run into in New York City is a shortage of physical space. And so publishers' archives were given to libraries simply because it was too expensive to keep them in Manhattan. But with the digital servers, this too, has changed. I was really drawn to the emphasis that you place on print on demand services, and how the possibility that books won't ever go out of print, and therefore will remain viable capital assets into the future and this can change how publishers will consider or evaluate donating their files to archives?

MK

Yeah, sure. So that's another kind of crucial development in all of this. So we started to see really viable print on demand technology in the early aughts, I would say, and that was the point at which publishers realized that what's called their back list, in other words, all of the books that they've sold for many years in the past, but but which are not sort of, you know, currently, their bestsellers. The back list was actually a revenue stream that costs with print on demand technology, they didn't have to reprint 1000s or 10s of 1000s of copies of a 30 year old book, they could reprint one copy just as soon as somebody needed it and was willing to purchase it. The catch, though, was that publishers realized they didn't have the actual raw materials of the book in their own hands, meaning in the past, that would have meant access to what are called the film's photographic plates from which a book would have been printed. This would have been true for much and certainly the second half of the 20th century. Books and other printed matter were generally printed from photographic films, which would be destroyed and recycled after the book was out of print. Then with the advent of digital workflows and digital processing, publishers saw that this was the moment in which they could keep the raw physical stuff of the book in other words, the bitstream that bits and bytes, but they needed that asset management system to sort of keep it in hand. So print on demand was real, it became a kind of incentive for publishers to convert their holdings to convert their back list and also to ensure that their current titles going forward. But these were all available in digital form so that they could essentially be kept in print indefinitely.

LM

It feels like there are a lot of reprint efforts happening in the past 10 years and some of them are very beautiful, like the New York Review of Books reissues. I love having those on my shelves, and sometimes past books are rediscovered, and they are wonderfully relevant or were really relevant. You offer a Sinclair Lewis example. And another that comes to mind is Catherine Ann Porter's 1939 *Pale Horse, Pale Rider*, which I don't know if you've come across, but is a haunting story about a young woman who survives the 1918 flu epidemic. But that said, a lot of books are not like this. That is, they are not read by very many people decades later, nor do they enjoy a sudden resurgence in relevance or popularity. So if a lot of books fail to find an audience, print on demand seems to be changing the calculus for all books, when it's really only relevant for a few titles. And I'm sure that archives and publishers have always made risk calculations about the capital value of their assets and giving away things that are monetizable. But with digital assets, it really seems like this calculation has been recast. And that the balance tips in favor of holding on to these titles.

MK

I think that's right, because also, as you've pointed out, holding on to them doesn't mean taking up physical inventory space anymore, it means that it's on a server. So holding on to a book means retaining a relatively modest amount of data in digital form against the possibility that someone sometime might want to purchase a print on demand copy of it. I think there's there's another kind of aspect of this that's really interesting too, though, which is that print on demand,

but more generally, what's called digital printing, as opposed to offset printing. So offset printing would have been those plates and films I was talking about a moment ago.

Nowadays, a lot of books are actually printed on sort of industrial strength, digital inkjet printers, basically enormous industrialized versions of what we have on our own desks at home. And the quality has gotten such that offset and digital printing are increasingly indistinguishable from one another. What this means is that publishers, because of digital printing, can do smaller and smaller print runs, not quite print on demand, but almost. So instead of a print run of again, 1000s, or 10s or 1000s of copies, which is what would be cost effective for offset, you could do hundreds or even perhaps 10s of copies. And let me just give you one example of how this has played out. In a really sort of amazing, and I think important way, Claudia Rankine, who is a poet and writer, perhaps best known for a book called *Citizen* that she published in 2013. It's subtitled *An American Lyric*. And it's kind of a really intense exploration of racism in American culture. It's written partly in the first person. So it's filled with examples from her own sort of day to day of microaggressions. It's an amazingly produced book that has sold a couple of 100,000 copies, which makes it really essentially a blockbuster from the standpoint of sales of poetry. But the reason I use *Citizen* as an example is that one of the things that Claudia Rankine has done as revised a particular page of the book, which lists the names of Black Americans who have been murdered as a result of police violence. And in the earliest printings of the book, there are only a couple of names there. Then, with each subsequent printing, she has taken the opportunity to tragically, as news continues to happen to add additional names to that page, she's able to do this, because *Citizen* is being reprinted in those smaller and smaller print runs. So reprints are more frequent. And that creates the opportunity for her in this incredibly sort of powerful and monumental way for her to continuously revise that page of the book and turn it into this kind of living record of police violence.

LM

That's really helpful to point out some of the possibilities of reprinting and digital printing in that it makes literature into something that can respond in the moment, and also makes publishers accountable to the authorial goals. In that sense, I'm sure Claudia Rankine had to work with her publishers to say, "I want this to be an open-ended project." I mean, we've talked a lot about how digital files make it easier for publishers to sort of re-spin, redeploy and re-monetize these files. But they're also, I imagine, just an enormous number of costs associated with dealing with these materials. And that's certainly what we perceive on the repository, and which is, there's this proliferation of data, and what do we do with it? So is the digital asset an instrument, the DAM? Do you think that this has really facilitated this proliferation of data? Or was it a response and organizational response to the data, the avalanche of files that was already being created by writers and by publishers?

MK

Yeah, that's a great question. And I'm not sure I could sort of answer it with complete authority. It's been interesting to really sort of hear from people in the publishing world what their kind of day to day perspective on that was. But I mean, I think the answer would have to be both in the sense that this sort of conforms to a pattern that we see continually throughout technological

history, that new technology, new innovation, it's about something that we kind of are reactive and responsive to, but it's also something that is sort of always emerging out of a kind of matrix of needs, and priorities and so forth. You know, I think there are certainly costs associated with things like server space, and you know, software licensing and all that sort of thing. But the costs, I think, especially come to bear, I suspect, if you were to ask a publisher, they would tell you that the costs were in personnel, yeah, having the people sort of in-house who have the expertise that is needed to keep these systems running and to use them effectively.

LM

Well, the labor aspect of publishing, I think, is really important. And you detail some of the supply chains and outsourcing, but one of the labor issues that I found, I think that you touch on but doesn't come to the fore is thinking about publishing in the context of gig work and independent contractors, and in particular, the social costs around the dissolution of full time work that has happened. And you know, we've seen this in the last six months, the way that taxpayers have to pick up the social cost of when companies don't pay their workers. Well, when they don't pay for family or leave or unemployment insurance, sick leave. And so in your research in books, files, how do you think outsourcing has become a bigger part of the publishing business? Is this a movement out of house just for key aspects of the publishing process? Or do you think more and more parts of it are being pushed out to the fragmented labor force?

MK

Yeah, fabulous question. Yeah, there's a saying that authors don't write books, authors write manuscripts. And then publishers take the manuscript, and they turn it into the book that we hold in our hands. What we see both historically, and in terms of the research that I've done, is that the process of taking the manuscript and turning it into a book is the work of many, many hands. And as you suggest, increasingly, they are not the hands of people who are either physically sort of under the roof, if you will, of the publishing house, nor are they necessarily full time permanent employees, they are precisely freelancers, contractors, gig workers. This is true of book designers, this is true of copy editors, this is true of indexers. Publishers certainly don't keep printing presses in the basement of the building. So the actual printing of the book, that kind of magical sort of moment, where the digital file actually really becomes the book that's all contracted and outsourced, and other sort of more esoteric aspects of production as well that just have to do with really, essentially project management and kind of shepherding the book through all of these different steps. Even that can be outsourced. What publishers tend to do nowadays, are acquisitions and marketing, to generalize. Of course, publishers do many other things, too. But it seems to me at least as a kind of lay person looking in from the outside, that those are sort of the you know, that plus some editorial, those are the core functions that publishers now perform. And since books before they do become books, as actual physical commodities, are only just sets of digital files, the people who are sort of doing all of this work—the copy editing, the designing, the indexing—all of these different steps and stages, they can be anywhere in the world. And so the digital has, I think, very much facilitated that kind of casualisation and outsourcing of publishing labor. And you know, even for the actual printing of books, I visited a print publishing facility at its printing production facility as it would be called in



Indiana. But a lot of books are also now printed overseas, particularly in Asian markets. And the costs for doing that are such that it's often more cost effective to have the book printed overseas and then shipped back spending a month on the water, you know, in a shipping container, it's still more cost effective to do that than to pay a domestic printer. So publishing a book that, you know, if we go back to the many hands that are taking that manuscript and turning it into a book, those many hands belong to many people who can themselves be located literally anywhere on the planet.

LM

Well, it seems related to the archiving hurdle, which is that the outsourcing fragments where the files are coming from who's creating them how many versions there are, and this, this just increases the burden for the archivist or the fantasy archivist who might be gifted these publishing archives, which are increasingly unlikely to be gifted.

MK

Yeah. So I mean, take the example of designing a book, a book jacket. They are sort of, in some ways, the most dramatic and recognizable feature of the book, the dust jacket. That's almost entirely freelance designers who are doing that work nowadays. And historically, we do have sort of alternative versions of dust jackets and publishers, archives, that's the sort of thing that could be potentially really interesting to a scholar, interested in, you know, how a particular book was conceived and marketed and so forth. But nowadays, those alternate drafts would probably take the form of, I don't know, Adobe Illustrator files that are sitting in a freelance designer's Dropbox, and they'll be there for just as long as it takes for their Dropbox quota to overflow. At which point, you know, they'll purge projects that are more than six months old and right and so there you see the casualisation intersecting with the kind of technological underpinning the technology but technological base of the labor in ways that make prospects for archiving extremely unlikely.

LM

And because it seems difficult to draw a line around what files are important. I mean, certainly, the alternative versions of the dust jacket, I think are very interesting. But would negotiations with the bookseller be part of the archive? One would hope so, although that also seems like proprietary information, but what about the shipping forms? What about how much it costs to send things overseas? So the materials that you have the potential to bring into this archive might also change the subjects around which the repository is, you know, organize their collections. And I'm thinking here of a movement away from single author purchases? And do you think that both these challenges of the DAM might enable a kind of rethinking of those collecting impulses?

MK

I would think so, I mean, I would certainly hope so. I really kind of just like your observation about the sheer amount of perhaps very mundane materials, which because of their underlying sort of digital format and pedigree can be collected as effortlessly as anything else. So that the publishers archive in a sense, it's a literary archive, perhaps, but it's also perhaps a window into,

as you say, in the global shipping industry, and potentially of interest to a scholar of labor and logistics and so forth.

LM

I want to take a step back, because I think I didn't get to ask you about a couple of the central pieces of your book. And one of them is, you know, the multiple definitions of archive that you bring up in your discussion of digital assets. So how, in part of the computer systems, meaning of the term has kind of migrated into daily use. So when I use the word, I think of it as long term storage, I'm archiving my emails. And it's increasingly important for all kinds of companies and institutions to think about how they're going to store the huge caches of data being generated. But storage is really different than accessibility. And archive in the university context, usually points towards something that's supposed to be explored. That's ideally going to be made accessible, maybe we're going to have to overcome a lot of technical hurdles to do so setting up InDesign and reading rooms so that you can look at the files and those technical hurdles are a big piece of what you're talking about. But there's also the detachment of archives from accessibility or readability.

MK

Yeah, and I think, as you know, there's an enormous amount of confusion and contestation around the word archive, you know, as far back as the 1970s. It was a word that entered into computer systems design. And you would sort of have computer scientists suddenly talking about like the archival layer of the system. They didn't mean Hollinger boxes, they meant essentially, long term storage, often kind of both physically and operationally removed from the main user access to the system. In other words, tape rails can be stored in a dark room and painstakingly re-mounted and placed back online should some, you know, particularly troublesome person actually submit that request. But that's what archiving meant, with very little necessarily unnecessary provision for access. Certainly, you know, little regard if any regard for reconstituting the actual materiality of the object, the actual interface and look and feel of it, you know, the idea of student access scholarly access, the sorts of things we prioritize in a university setting. Now archiving just meant long term retention. And then, you know, more generally, the word has kind of had this popularization as well. So somebody who is not an archivist by training at all might speak casually about this weekend during COVID "I'm going to archive my photos on Google Drive." Right? And we hear that turn of phrase, and at least for me, my first thought is, well, you know, what does that mean, Google Drive as just a hard drive somewhere that you don't own or control? And so in what sense are you safeguarding these priceless family mementos, but the word has entered into the vocabulary? And the most troublesome aspect of that? Is it just tends to really complicate discussions with publishers. Because they'll say, "Well, you know, archiving? Well, we already have that—it's built into the DAM, right? It's on the feature list." And now, not in the sense that we mean archiving.

LM

So if you had your wish list, which of the files that are challenging for researchers—both to be able to see and study in the reading room, but also maybe get access to—what on your wishlist would publishers hand over and keep and make available?

MK

For all of my sort of technical interests, I'm a fairly traditional literary scholar. And so for me, it would be the things that have sort of the strongest analogues to physical materials and sort of artifacts from the print world. So early on, we use the example of email, which has largely replaced correspondence with literary correspondence, you know, nowadays, you know, gosh, that might include text messages or resume sessions, right? But then in addition, things like different versions and states of the manuscript as it works its way through editing and copy editing, production materials, things like those alternate versions of the dust jackets. Having said that, you know, having kind of planted that sort of more traditional flag, if you will, it's very much the case that scholars are also increasingly interested in data in aggregate big data. This I think, goes back to your question about changing, collecting the mindset of collecting, and not just making it individually focused. What could we learn if we had access to hundreds or 1000s of files that were related to the publisher's day to day operations? Would that give us any kind of insight into the publishing business as such? I think about a scholar like Donald McKenzie, a famous British scholar of the book. He reconstructed essentially the day to day workflow at Cambridge University Press back in the 18th century, by going through the presses accounting records in an almost forensic way, and sort of learned things about how books were printed in the 18th century by studying accounting records, so things that might not at all be obvious to us as having importance or bearing on the enterprise of bookmaking can turn out to be highly relevant, and certainly data mining, and visualization. And all of these kinds of big data technologies seemed to me to be highly relevant to that.

LM

Well, a lot of the things that you point out as being challenges, such as emails from HR and salary negotiations, sitting alongside authorial correspondence and emails are both huge challenges, but incredibly appealing research topics also, as we want to know more about the social institutions that are creating these books. So we didn't get to talk about *Track Changes* this time. Perhaps that will be another episode. I want to end the podcast with a question about your interaction with Special Collections. Is there an archivist or an archive that was especially important for you, or that taught you something about using Special Collections?

MK

Yeah, thanks for the question. There is lots of talk about archivists who have been really crucial collaborators and partners in my work. I think I'll always have a soft spot for the Ransom Center at the University of Texas. That was the place I went when I was researching my first book, a book called *Mechanisms*. And they had acquired the digital papers of an early hypertext novelist named Michael Joyce. And so I went there and they had just gotten the collection themselves, they were still figuring lots of things out. They were super accommodating. I actually worked with two amazing archivists. They are Katherine Stoller-Peters and Gabriella Redwine. And I, you know, they kind of invited me in and helped them sort of essentially be the guinea pig and figure out how they were going to serve this collection to users. For me, it was my first real sort of insight into what it meant to actually think about the digital long term from the

standpoint of preservation. And it led to a lot of subsequent work. So I'm grateful not only for having been well treated as a patron, but really sort of invited in as a kind of collaborator as well.

LM

That's a great place to end. Thank you so much, Matthew, for being on the podcast.

MK

Thank you, Lina.

Amanda Martin-Hardin (AMH) 44:30

Thank you for listening to Overdue Conversations. This podcast is published in partnership with Columbia University Libraries. This episode was written and researched by Lina Moe and Anirbaan Banerjee, a graduate student in English at Columbia University. It is edited by me, Amanda Martin-Hardin. Music is by Poddington Bear through Free Music Archive.

If you enjoyed this episode, be sure to like and subscribe to Overdue Conversations wherever you listen to podcasts.

Transcribed by <https://otter.ai>