

Webinar Series: Fair Use Code & Other Legal Tools for Software Preservation

Episode 4: Working with Sources Code & Licenses

Speakers & Facilitators: Lauren Work (University of Virginia), Daina Bouquin (Harvard | Smithsonian Center for Astrophysics), Brandon Butler (University of Virginia), Peter Jaszi (American University)

Timestamp: [0:19]

Jessica Meyerson: Welcome everyone thank you all so much for joining us today today my name is Jessica Meyerson and I'm the community advisor to the software preservation Network as well as research program officer at a Educopia Institute just a little housekeeping before we get started as always everyone but hosts and guests are asked to be muted throughout the webinar just to maximize the audio and visual quality of this recording of which you'll note to your top left that we are currently recording as we have the previous episodes in this series if you have any questions during the presentation please type them into the chat box of the zoom control panel and I will bring them up during the presentation and we'll also have time for questions at the end of every episode will be recorded transcribed and posted to the spin web site and we're currently working on this work these will be freely available for all today we're presenting episode for working with source code and licenses this is a discussion with members of the coda best practices research team and esteemed guests which include Daina Bouquin who's the head librarian at the Center for Astrophysics an organization where Harvard and Smithsonian scientists work together on space missions and research aim is responsible for all library operations and leads efforts facilitate software preservation and code attribution within the Astronomy community. She's also an advisor to SPN's EaaSI program of work the emulation as a service infrastructure, the number of archives IT advisory group as well as a member of the Center for Astrophysics scientific computation advisory. we're also joined today by Lauren Work who's the digital preservation librarian of the university of virginia library where she is responsible for the implementation of

digital preservation strategies and systems he's also the project lead for university of virginia libraries fostering communities of practice and project and serves as the UVA library node configuration coordinator for the SPN EaaSI program. Your research leads and facilitators for today's episode include Brandon Butler director of information policy at the university of virginia libraries joined by Peter Jaszi professor emeritus at American University Washington School of Law Professor Jaszi is one of the originators of the fair use best practices movement and his co-author of the software preservation code of best practices for furious along with Brandon Butler, Pat Aufderheide and Krista Cox. This is the continuation of our seven part series of webinars would you explore the fair use code as well as other legal tools for software preservation and it's co-hosted by the association of research libraries and the software preservation network and so in this fourth episode Brandon, Peter, Daina and Lauren will discuss how fair use applies when working with licensed software and so with that I'll hand it off to Brandon.

Timestamp: [3:32]

Brandon Butler: Great thanks Jess so as you see there we have a little roadmap of what we're going to do today first Peter and I are going to talk a little bit about principle 5 of the code of best practices if you've been here before for the last couple of these you know that's the document that we've been sort of marching through for the first several episodes of the webinar series and principle 5 brings us to the end of the code of the principles in the code and that principle deals with source code and so we'll talk a little bit about what's in the code of best practices about code of the source variety and then we'll also talk about something that I think is really really important for us to talk about in this community which is the relationship between fair use on the one hand and licenses or contracts on the other hand was the thing that really from the very beginning sort of was it there's the first question that happened you know as soon as we would bring this issue up whether with whether we're talking with you know lawyers or librarians there was always this question of licenses since the licenses are so prevalent in software. So once we Peter and I give a kind of overview of these legal principles then we'll have we'll hear from Daina Bouquin about her work with source code and how the kind of principles and best practices might be helpful to her and we'll hear from my colleague Lauren Work here at UVA about some of the work we're doing with licensed software and then we'll open it up we'll talk amongst ourselves and we'll also have questions from you all.

Timestamp: [5:16]

Peter Jaszi: Before you summarize principle five I wanted to make a more general observation that looks backwards and also specifically to this principle about how this code came about we've talked before and we'll be speaking a little bit again later today about the preliminary research that we did and all of the people who were so kind as to spend hours with us on the telephone making us understand about how copyright and fair use did or didn't operate in the software preservation space and we thought a little bit about what came next that is a series of

what were in effect small or focus group meetings among preservation professionals at sites across the country which we moderated and organized around topics that had come up we're currently in the research at the beginning of the project that I described a moment ago and the principles including the ones we've talked about in the last several webinars and principle five that we're discussing today come out of those small group meetings they are our best effort to sort of to write down and concretize what the groups of professionals we talked about we talked with believed were good practices that wasn't quite the end of the process because once we had done that we then filtered our good faith summaries of the group consensus through a bunch of lawyers just to make sure that the group and the sort of general perceptions of where the law stands weren't out of sync so to speak they weren't but the processes I've described it which is the same process that we've used with I don't know twelve or thirteen or fourteen different community based practice groups over the last 15 years is in itself of some interest because it's an obvious source of strength in terms of the the broad-based foundations that the final principles have but it also has been from time to time a source of I don't want to say vulnerability it's the wrong word but it's been something that that has been questioned particularly by advocates of a long strong copyright over years and the question has always taken some version of the you know lunatics and Asylum trope you know how can you leave something as important up to the law as the law up to these people who are not lawyers first and who are collectively self-interested in having as much freedom to operate within whatever domain they are in the domain of classroom teaching or the domain of documentary filmmaking or the domain of of our history writing you won't get a balanced result if you just talk to the people who always want more for less now obviously we don't believe that critique because we've gone on doing it that way for some fairly successfully for years and years and years and one of the reasons we don't believe that critique is because in practice that's not how it works out in practice when you get a group of people who work in a practice area together it turns out they they represent and they know people who represent and they feel affiliation with people who are in all sorts of different positions across the spectrum from wanting the greatest degree of access and freedom to use on the one their hand to be in quite concerned about proprietary rights on the other so one of the things that we discovered about your community which we found fascinating is that the individuals the preservationist SIF you will who were in our groups and all kinds of lines and connections to to the programming community and some of them were ex programmers some of them were current programmers some of them had been in the industry and then it moved over to the archival side some of them were still working in for-profit settings but as archivists and the result as has been true in every community that we've worked with over the 15-year history of this project is that the end product is a very balanced one it's one that doesn't just maximize the the users freedom of choice if that remains anything but also respects the legitimate interests of rights holders and we've talked about some of the instantiation of that balance over the last several weeks among them the fact that in connection with principle after principle among the limitations that the small groups felt were important to include were ones that were designed to make sure that archive will act these didn't compete directly with current active programs of of commercialization run by copyright owners or their successors that was one example this is another the fifth principle is another example of a situation in which the group of archivists that we dealt with felt very strongly that there were interests on both sides that needed to be accommodated and so in addition to being interesting in

its own right it's also a nice illustration of what I would maybe slightly slightly arrogantly but nevertheless perhaps usefully referred to as the the genius of the process not of us because all we're doing is following along taking notes but of the process itself sorry Brandon.

Timestamp: [11:56]

Brandon Butler: now that's a really really useful intervention Peter because I think we've we've jumped into the contents of the code but it's it's because it's sort of irresistible to just get into it but it's good to remember where it came from and I think this principle is as you say a really nice illustration of that. So you know principle five deals with working with source code yes it's the human readable right version of software as it's sort of written by its authors they can be reused and recompiled and broken up and and and so on and in this principle what we find is that it is fair use to to preserve and make available this kind of material but there are as always a series of limitations. And you know another refresher on the general format of the code and the principles in the code: each of these principles is accompanied by these limitations. To sort of look under the hood again what we what we do is we have these focus group discussions and we and we propose and they these things evolve as the discussions evolved but we propose a version of a principle and we poke at the principle and we shape the principle but then we also ask for the outer bounds of its application what are some things that you had better do or you had better not do if you want to stay within the realm of what this field believes to be a reasonable moderate centrist process

Timestamp: [13:38]

Peter Jaszi: usually this is where the rich discussion is the principle because it is so general is relatively easy to get people to to agree to and refine. Its the limitations that are the real test of the group process. So one good thing to remember about using this document or any of the other documents to which it is related is that the the limitations are integral with the principle the principle doesn't have any independent meaning when severed from the limitations

Timestamp: [14:26]

Brandon Butler: and so so it's you as a result of all that you move very quickly from the principle to the limitations the principle will stake out a broad fairly easily as Peter said fairly easily arrived at kind of zone but then the limitations really help you define the boundaries of that zone and so in this particular case one thing that emerged very clearly in the discussion this was that source code is very often donated by an author and the authors of source code who donate their their creations to archives and special collections are just like the authors of unpublished manuscripts and they have complex and you know important thoughts about what should be done with that kind of material and you know the law of fair use as we'll learn later on can be made subject to the limitations in a contract and the contract and a deed of gift or a donation agreement can limit your

rights just as much as anything else and so you and we enter into these agreements under those kinds of terms because it's worth it right because it's still worth it to get our hands on and to save and to make available in some way these materials. But it was really important to the groups that we spoke with as part of exercising your fair use rights to be conscientious about the limitation in donor agreements to keep them at the front of your mind and and you know fair use does not overcome those limitations you need to look to those terms to govern what you do. Relatedly and again working from analogy to unpublished manuscripts there was a lot of sensitivity to putting source code online without an express grant of permission in the first place and so there was a sense that again because this is reusable content this is code that could be broken up and deployed again in a new format and this may be code that was never released and so was maybe never intended to be used that again you should be guided in some sense by the same kinds of policy considerations that you would use with unpublished manuscripts which is not to say you would never publish it but is to say that it's a different kind of a beast than commercially available machine readable code. The next limitation asks you to limit access even by researchers to make sure that the level of access is related to the inquiry that they are engaged in either by redaction or otherwise. And again this was something that came out from the folks and collecting source code and told us you know that this kind of material is can be very sensitive and so when you're when you're making research access available again to be sensitive to how much access is appropriate to the research project which is again actually strongly rooted in fair use law itself the level of access that fair use permits always needs to be sort of calibrated to the underlying transformative purpose and then the final limitation is a very common one it's something and it's another indication of the good faith that Peter was describing before which is communities of practice consistently consistently insist on attribution of authorship and ownership in these contexts as a way of providing you know proper credit and provenance for the materials in your collections and making sure that you know the folks that are using is materials understand where they came from so those are the limitations they're few but important important to understand Peter did you want to add anything

Timestamp: [18:27]

Peter Jazsi: I just love things stand one is about the attribution one because that's so interesting as those of you who are students of copyright law as well as other things no there's nothing in the copyright law that imposes any burden of attribution on anyone under any circumstances say the narrowest range which don't apply here but what we've seen over 15 years in doing this work with different practice communities is that consistently without regard to the domain the professionals who are responsible for these best practices think that attribution is really important and so we can we can as brandon has suggested tie that to the law of fair use by characterizing it I think accurately as an aspect or a-- or a factor in determining the users good faith which does sometimes enter into legal analysis but it's also a unethical freestanding ethical imperative that i think in in in candor the different professionals we've worked with over including in this context the people who are doing software preservation your community would probably have arrived at even if there weren't a specific legal hook on which tank it's not important the other thing I wanted to

mention is the relationship between the fifth principle on the one hand and the other four principles that we've rehearsed over the last several webinar sessions on the other the first four principles are in a sense cumulative or cascading in nature they start small and they grow out from there a certain thing bunch of things you can do almost with impunity and other a bunch of things you can do on your premises with very relatively few although to the extent that they are expressed of course nevertheless specific constraints a larger another set of activities and now at principle three which you can do on your own site providing virtual access to members of your community and then fourth a set of activities that you can do online for a wider range of users in consortium and collaborative arrangements and really those principles almost demand to be read together because as I say they they accumulate or cascade I mean you can choose your you can choose your verb but one way or another they build one builds on the last. This is slightly different this is a you could say a kind of separate freestanding principle which isn't simply the inevitable outcome of the four that preceded it dealing with what we understood as we went along to be something that the community regards is a very special and sensitive case so I would urge you in thinking about the code as a whole since we now reached the end of our description the principles I would urge you to think about the the cumulative relationship between the first four principles and to appreciate the significance of the fifth principle as a kind of as a kind of freestanding one. End of thought.

Timestamp: [22:30]

Brandon Butler: Thanks Peter so then let me move to fair use of Licensing and the the starting point here I think or a good starting point is the kinds of concerns that we heard when we embarked on this project again from from you know lawyers and the librarians alike the grounding of the concern is that software everything is now almost everything now feels like it's distributed with a license I think I saw the other day that Hudson Yards this gigantic New York real estate development there's a there's a big fake staircase to nowhere that has a license on it that says if you take a picture they own it and so licensing has escaped from software and and you know infected lots of other things but software was way ahead of the curve here and has been distributed with contracts attached for a long time and the the history of you know clicking through and having to agree to things when you when you install your software has what anxiety expressed in a number of ways you know one of those is you know I can't do fair use on this software because I don't own it I'm just a license or I'm just a licensee I'm just a a mere you know I'm here entirely at the will of the copyright holder so how can I do fair use another version of the story is you know that the software license is is the beginning in the end of my my ability to use something and if the license says yes then I can and if the license says no that I can't and that's all I need to know another concern that we heard was I not only do I not own it I I'm not even a licensee of the software I just got a box of props right somebody licensed it at some point but it ain't me so how can I have any rights to use this stuff isn't it just sort of radioactive now and then finally and this was sort of ubiquitous one of my favorite things was to ask people who are concerned about licenses which term they were worried about you know sort of what is it about the license that concerns you and across the board no one has read the licenses and and [sometimes they can't,

sometimes their gone] yeah that's right the licenses are gone that is you know I'm assuming there was a license because there was always a license in this era of software but I can't find it so I don't even know what the terms are and so this was sort of where people found themselves when we started these conversations and so we actually I have the school myself and luckily get you know I had Peter involved if he could school me a little bit too about how the big picture of licensing and fair use really worked together and Peter I wonder if you could break that down for us for a little bit

Timestamp: [25:39]

Peter Jaszi: oh I'd be very happy to and and I would just emphasize that very outset of this project I think much of the skepticism and there's always skepticism that people in the community expressed about whether the exercise would be useful was around this issue of licensing it was some version or variant and Brennan has done a very good job in preceding slide of showing what all of those were of the picture of the proposition that well this period stuff is all nice and good but in the end it doesn't matter because some license somewhere whether I know what it says or not controls and that that was a sort of was something we had to talk her way and work our way through and in a way that the material that we're going over now and that is also in the in the appendix to the document itself about licensing sort of sums up what what that that working through of the material entailed and we certainly don't know that there could never be a case in which a license term could theoretically stand in the way of accomplishing some preservation project that was otherwise authorized by law under the fair use doctrine because as as you see here there are basically two kinds of authorization for anything you want to do with actually or presumptively copyrighted material one is that you've actually got permission and the other if from the rights holder and the other is that you have permission and event from the Congress and courts of the United States by law and what I think was hard for people to understand and I hope we were successful in explaining and are trying to emphasize in the appendix and in this this discussion today is that the fair use doctrine the authorization by law for a use is very potent and it's not actually easily defeated in theory or in practice by agreements licences contracts call them what you will and although it's certainly true that if under certain circumstances I were to make a deal with someone based on an exchange of value in which I promise to renounce some or all of my fair use rights with respect to a particular work it's certainly true that in the future I might be bound by that deal but that proposition which is undeniable is in itself relatively trivial in the real world of software licensing and that's for a number of reasons which are summarized here I can't remember who is going to take the first one Brandon - [that's a good question] - I'll do it. Obviously because we don't know no one knows because there are nowhere to be found we don't know the exact terms of every end-user License associated with every legacy software package commercialized within the last 50 years we do know about some and what we know is that based on the sample that we can get access to it's really really really rare if we weren't lawyers we would say unknown but since we're lawyers will say really really really rare for the agreement to include anything that looks like an express agreement on the part of the user to forego preservation activities or even to forego fair use which can involve other things than preservation of course as a

category later on nowadays as licensing has gotten more sophisticated and things are being you know distributed in in new in different ways terms like that pop up not commonly but at least occasionally but in the legacy B software period with which this project is associated they just don't seem to be there as a matter of description so that's the first reason why even if you don't know what the license associated with a particular program said you probably don't need to be tremendously worried about the possibility that it could be a restriction on your professional work over to you Brandon.

Timestamp: [31:27]

Brandon Butler: yeah and so what we what we found when we looked at the licenses that we could find was that if you pay attention to the wording of the license what the licenses do is they tell you here's what this license permits and here's what this license does not you may do this under this license you may not do that under this license right this is a license for your personal use and not for your business use this is a license for three computers and not for ten computers all of that is fine and good what it tells you is the scope of the license but as you've learned on the previous slide fair use can go beyond the scope of the license right and so we need not and should not read language that very literally and clearly says this is what's in the license has some how impliedly or you know magically excluding your legal rights which come from beyond the license and so that's really the the most common source of confusion here to me was that folks would see a license that says that kind of says this is what our license allows and this is what our license does not allow what but what what those licenses very rarely say is you know we block fair use or we block preservation you will hereby promise not to engage in fair use and I want to make sure we get over to Lauren and Daina so Peter can we move a little more quickly to the last this weekend

Timestamp: [33:07]

Peter Jaszi: there's really very little to be said about the arrest and I will do it very quickly and the the third one is really just an explanation of why the second principle operates no court would ever read a license willy-nilly to exclude fair use unless it actually said that because fair use is so important for those people in institutions who are dealing with material that they themselves didn't purchase that was donated or purchased from that they didn't purchase from the source that was donated or purchased secondhand then there's another interesting ancient but powerful principle of contract law which says contracts and that's what a license is a contract don't apply except between people who are parties to those contracts or who are in some kind of tight relationship with parties and this is the privity principle and in most cases where you're dealing with second-hand or donated material there's no privity between the archive and the original licensure so whatever the software says excuse me the license said it probably doesn't matter and then finally and this is important from a risk assessment point of view let's suppose that you were in one of the the tiny perhaps speculative range of cases in which a license might actually operate to control what would otherwise be various preservation activities and the right question is how

much trouble could you get into if you made the wrong call and the answer is not very much because the only thing you'd be liable for is the breach of the license not for copyright infringement you are after all doing fair use but for breach of the license terms and licenses when a license or any kind of contract goes to court then the court asked well how much real harm was done and if there was an even they might give compensation but almost by definition in the case of preservation good-faith preservation of a Legacy Program the actual measure of real harm in dollar which is going to be nil and

Timestamp: [35:32]

Brandon Butler: that's right so now I would love to turn it over to Daina Bouquin to talk a little bit about her work with source code - Daina?

Timestamp: [31:27]

Daina Bouquin: hi can you hear me okay? [you sound great] so I was going to talk a little bit about just like a specific case that I think is pretty illustrative of why source code and like uncompiled code non executable code more broadly has some other nuances baked into it that we're still trying to grapple with when it comes to fair use or figuring out licenses and what we can do to both preserve this content but also to document it because I'm essentially coming from a landscape where for the most part the things that our community builds code and simulations and theoretical models those are existing in a distributed capacity so there we are often not the direct stewards of these objects and instead we have to help support our community's ability to document them for potential reuse and description but also for that attribution piece that was mentioned a little bit earlier because if we want to be able to make it so the community can both reproduce this work and validate their findings we also need to make it so that they have a job to do that so doing actual Career Pathways having actual career pathways for these people depends a lot on their ability to get credit for their work and this gets a little hazy sometimes so I'm having to talk in tons of depths about the attribution piece of this but when it comes to licensing if there's not a clear way to document even the license we start to get into the realm where we're falling back a lot on attribution if that makes sense so for instance I have up here a link to something called the IDL astronomy library so IDL is a programming language a proprietary programming language that was used profusely throughout the astronomy and astrophysics community for all kinds of purposes from the 70s actually on through now so people are still using IDL but it is a proprietary language and the codes that they've written this source code but IDL actually has explicit licensing that's actually never been tested in the courts where they actually tried to prevent byte code compatibility with other environments so they are explicitly trying to tell you through their licenses that you are not allowed to kind of make this compatible with another tool however by the nature of the work that the community is doing a lot of this code is built into pipelines because it's distributed so for instance the Solar Dynamics Observatory and all of the pipeline's that are used to pull down roughly about like a terabyte a day of just like raw image files

all of that was originally written in IDL and Fortran and a lot of it is but because the community has seen how this does not scale it's not nearly as flexible as they want it to be and the licences are very expensive there's this odd window in the history of IDL where astronomers started wrapping IDL code with Python scripts and because they are kind of trying to find a way out of IDL at this phase they're falling back on attribution so they are doing their own fair use essentially of some components of IDL to do this stepping-stone essentially to more open tools and so as someone who wants to make it so people can get proper attribution for and credit for their work but also find a way to document these things so that we can capture and archive them these are the kinds of things that propose that kind of pose these new challenges because we actually so we go from this still necessary proprietary code to this what I'm calling just wild card stuff where they're purposefully mixing these things together to this open code that has now been often based on proprietary work and where we still have these documentation issues although like I can talk all I want about code meta or citation files and structured ways to document these things the licensing situation here stays a little bit strange no matter what and I kind of put a link at the bottom here because this I do not think at all is a isolated instance the programming language Julia it's a vectorized language so it kind of incorporates some of the really great functionality that are has into this much more flexible general use language is really getting picked up by the Astronomy community and one of the first astro Julia a-- packages that they've put out is an IDL package so all of these communities are trying to find ways to take these proprietary things that their work has relied on so heavily and they are trying to incorporate that into what they're doing in the open now and because we're not the direct stewards again we're not receiving this as like a deposit to us but we're having to find ways so that we can index this content so we can find it and document it and point to where it actually is we're falling a lot on attribution so I'm happy to kind of talk more about what we're doing to navigate this and the literacy challenges that this is sometimes incorporating but I will say that we're currently working with Harvard cyber law clinic on a little bit of a legal study they're doing a staged kind of study here to look at kind of what our legal risks are as an institution if we start trying to capture and distribute this sort of content because although the the the code of best practices kind of says you know for most of the time you don't want to distribute this stuff that's kind of the opposite of what the community expects so this is a very open community and they expect to be able to like find and read everything because they're all reusing each other's work because you can't redo observations and for the most part they've actually given up numeric determinacy if that means anything to you so what they're actually trying to share is a functionality of a model and the mechanisms by which it works so repeating a result isn't actually what they're trying to do they're going to get a slightly different result every time just by the nature of the thing so being able to share the source code as openly as possible even when it is this proprietary or Frankenstein type stuff is a little hazy so we're still kind of working out exactly what language and how we're going to advise our community

Timestamp: [43:30]

Brandon Butler: Wow very interesting stuff thanks Daina I really appreciate it and it's interesting to know that's an example of attribution that's sort of an all-purpose show of good faith

mechanism so okay thank you and Lauren I wonder if you could tell us a little bit about our friends at Vector Works.

Timestamp: [43:52]

Lauren Work: sure can you hear me okay yeah so kind of completely switching over now to commercial software so UVA library as Jessica said in her introduction so we're part of the cohort of what is known as the fostering the community of practice and that is under the umbrella of some of SPN's IMLS funding so it's a small cohort and our project is known as Emulation in the Archives and what that is is a fairly scoped project that focuses on a manuscript collection from a local architect here in Charlottesville and the donation from his widow included commercial software which is a CAD BIM software known as Vector Works that was part of his collection so it came in with manuals there's a couple of iterations of the software included in it and it's necessary for us to perform these preservation and access actions to the accompanying born-digital collection files that are also part of this collection so I just want to note that this project is very much still in progress so there could be updates that come with this over the next couple of months but for the scope of our project we really wanted to focus on providing access to this software and the software depending digital collection materials in an emulated environment and our reading room only so really you know with no ability to really download materials um and I wanted to stop here and say why we were anticipating the need to apply this fair use kind of principles for this developing project instead of doing something like you know buying a license for new software because this is definitely a company that's still in existence and there's a couple of reasons for this one is that the older versions of the software that are in this collection so this collection came in before my time at UVA at around I think 2010-2011 so a lot of the versions of the software predate that by at least 10 years so these are old versions of the software that correlates comes specifically to these digital archive so these versions of the software are no longer being supported or sold by this company and the other reality of the folks who are probably familiar with Kevin software and iterates extremely quickly it's the highly complex you know there's a lot of things like third-party plugins different design features libraries things that change over time so new software in 2019 would likely not support many of the features from software that was in 2005 and we need that to be able to accurately render the digital materials that are reliant on this software it's very much in our preservation mandate to make sure that we get software and software dependent signal materials off of physical mediums you know optical discs are not forever either and hard drives are not forever so this is part of our preservation mandate and we also really have a strong research and teaching use in our architectural school and a pretty strong institutional mission related to this so access to older software and files depending on that software could really be a very important teaching tool as well. And we've actually identified this as one of the user groups you know within the scope of our FCoP key projects so jumping in a little bit about our license software specifically so it's mostly Vectorworks is part of a much bigger company than Nemecheck group which is headquartered in germany and VectorWorks specifically is proprietary CAD software and the one other thing I wanted to highlight before kind of transitioning to the components of the paths that we talked to Brandon about are kind of

around this idea of authentic access to these born digital files that are part of this manuscript collection so authentic access you know in our case for the manuscript collection means that researchers and user health in an environment that under for kind of an understanding of how a work itself is rendered created develops iterated and used in the course of an architect over time so something I've been thinking about more often and in talking with the archivists working on this touching up you know ideas our own provenance what is the collection and where did it come from we have a lot of large-scale printouts that are also part of this collection are heavily marked up by architects you know you can see iterations you can see comments on these things and I've been thinking about this as somewhat analogous to what we could you know find in the 2004 files from this collection where the kind of iterative use that we'd be able to find through observing it in the version of 2004 software you know that relate to these 2004 files seeing these kind of iterative components and working documents are very much part of that so kind of final point here of the pathways of you know talking to Brandon about this couple of things that have happened given us architected the archival landscape of this collection is that updating the deed of gift which folks talked about a little bit earlier was very much part of this this came in at a time where we did not have a digital addendum as part of our practice for deeded gifts so that's something that we you know really wants to make sure has been updated for many of the reasons already highlighted here and the other big thing that came in with this collection which may be of interest to folks who may have things like this are that we had both the manuals which include the license terms as well as the license keys and this is something that you know I'll let Brandon touch on a little bit here too but this you know idea around the fact that we have this commercial software we have the keys to the commercial software environment and what that means thinking about licensing terms so I'll end there and turn it back over to Brandon for additional elaboration on that

Timestamp: [49:59]

Brandon Butler: yeah hey so you know I I think actually so that was a thank you our first of all that's a it's a great story I think it's just it fits perfectly with the kinds of use cases that we heard in the small groups as well the the notion that you know in some sense if you can't open these files in the software where they were created you're losing access to information you're not seeing the file the way it was made you're not going to see the layers you're not going to see the revision history you're not going to experience that the way the author did so the the research value is just undeniable so we're really excited to be able to work on that and you know I looked at the license I read the term then they fit in exactly with the kinds of trends that Peter and I described you know this is a license that tells you what the license tells you it tells you what the license is going to let you do and what the license is not going to let you do and we are not going to be doing things the license let us do but that's okay because section 107 of the Copyright Act lets us do the things that we want to do and the code of best practices tells us that the things we want to do are reasonable and normal and and you know justify so this might be a really good time then to see what kinds of questions folks might have had in the chat while we've still got a little time before the top of the hour yeah

Timestamp: [51:30]

Jessica Meyerson: Thank You Lauren, Daina, Brandon and Peter as Brandon said please do feel free to post any questions that you have in the chat I know it's a lot of information I think that those examples were incredible hugely illustrative and hopefully thought-provoking. um thank you again Lauren and Daina. so we'll give everyone an opportunity for the next few minutes to post any questions that you have it can also be for Brandon and Peter in regards to the contextual information they provided at the top of the hour so yeah everyone feel free. I think in the short term I'll ask a question to get things going which is Daina, back to your example, so in terms of um exactly how to work with the Cyber Law Clinic and the code best practices is informing policy and these are astrophysics in terms of how you're treating IDL the example that you provided was just questioned about similar examples or have in your professional circles at large when it comes to the research source code especially legacy proprietary research source code like like the example

Timestamp: [52:51]

Daina Bouquin: yeah so the study that we're doing with the Harvard Cyber Law Clinic what we're trying to do is have that be kind of a graduating up to more sophisticated and complicated cases for the most part the oh sorry I'm hearing a lot of echo maybe the idea that maybe my fault with somebody sorry about that so what we're doing is we're trying to move up in sophistication of the kinds of questions we're asking around proprietary source code because a lot of the community has shifted over to more open tools but so a lot of the talk about attribution or even software publishing so in our community there are software papers or code review as a peer-reviewed process and so like we have a lot of people who publish in the Journal of Open Source Software so there's tons of conversation about software and what we're capturing and who we're giving credit and how we're documenting it. But for the majority of these kind of more legacy projects that is a conversation that's going on more internally at the institutional level for the most part it's kind of almost in this web archiving discussion to some point because these are projects for instance like spec 2d there's a couple of them where they are these components of pipelines that people have incorporated into all kinds of web applications in particular and so I guess to say where we're going with that is we're starting with this one study on IDL with a couple of examples of projects they used IDL so that we can come up with kind of a vanilla language and policy to kind of advise the community on when they're sharing and building on this kind of work so they mitigate legal risks but then also graduating up from there to figure out what are the things that we want to actually try to archive as an institution what resources will we put into them and to what extent are we going to continue to maintain the licenses so some of these tools they have these bigger licensing platforms and we have an institutional license to what extent are we actually trying to kind of like emulate an application that might use like a couple of open tools but then it's also got some like potentially proprietary UNIX libraries and some oldest astrometry tools this one website might

use like four or five of those things to what extent are we going to invest in maintaining that's functionality

Timestamp: [55:39]

Jessica Meyerson: that's a really important really important institutional policy and resourcing question Daina and it definitely reflects a lot of what Lauren was talking about in terms of Digital addendum development and things like that. as far as the crossover to the curatorial threshold from from researchers and other members of your community that are going to be donating those materials yeah and actually that's an open question to all attendees here today maybe in the time that we have left if another question doesn't pop up which is is anyone willing to offer maybe some initial thoughts at least about where your organization is that in terms of thinking about software preservations impact on your existing policies especially some of the specific examples at learn and Daina highlighted in today's episode

Timestamp: [56:34]

Brandon Butler: yeah while people are thinking about that and maybe typing in the chat Daina story you gave me two I had two thoughts about that one is that there is a long and honorable tradition of using the codes of best practices as a as a tool to empower clinics that you're working with who need to help you through specific challenges that you're having so in the documentary community there's there's just a cottage industry of IP law clinics all over the country - including especially the one where Peter and I have worked together in the past where filmmakers would come to us they bring us the film and what the code does is it gives the law students a very workable framework they can apply to the film and say you know this film is within the framework of the code and then they can write an opinion letter in that opinion letter is a very powerful tool in the context of documentary film distribution and so it may well be that this code could have a similar application that folks who are faced with a tricky situation and who might have access to a law clinic like the Berkman Cyberlaw clinic at Harvard they could bring the code into that conversation as a tool that I think can provide a lot of scaffolding for the law students who might otherwise feel like they were starting from zero [that's we're hoping] yeah good yeah crazy you know Chris and Kendra are very well aware of the code so I'm sure we're in we're in the mix and then the other thing is you know the Daina's comment reminded me of my general view of fair use and licensing you know my grand unified theory of fair use some licensing on a time scale is that you know going forward if you can start fresh you want to use open licenses and that will make everyone's life easier and you know there's a lot of discussion about licensing for good in the context of software and communities who use open licenses and that's great but the problem is we've got a half a century plus of people who didn't have open licenses or weren't using them consistently and then they display and so fair use to me is just the the solvent that can really help us make progress there so I'm very hopeful they know that that it'll be helpful for you

Timestamp: [59:01]

Daina Bouquin: well thank you both I mean I think that that's kind of just the land that we're in right now is this big transitional period and even just on the literacy standpoint teaching people about explicit licensing and all of that is still an issue

Timestamp: [59:16]

Brandon Butler: for sure all right well we're about at the top Jessica was there anything else

Timestamp: [59:21]

Jessica Meyerson: no I think that's it I would encourage everyone I know it's always a lot of information to process in a single episode so if you think of questions as you start to parse the examples that were provided by Lauren and Daina today and you think about them in relation to your own practice in your own organizational context please don't hesitate to forward those questions on to myself the contact information for our speakers will be provided or the rest of the research team for the code of best practices for fair use in software preservation and we'll try our best to field responses to those questions and bring them out into the open because I'm sure if you have them other people will have them as well. and with that I'll just say huge thanks again the entire code of best practices research team to your facilitators brandon and peter and one thanks to our esteemed guests today Dana Bourquin and Lauren Work and also to all of our attendees for joining us in discussion today I'm joined us next week same time same place episode 5 where we're going to be looking at understanding the circumvention rules and the preservation exemptions around software preservation this will feature Kendra Albert of the Harvard Law School cyber law clinic which was mentioned several times in this episode as well as Jonathan Band whose counsel to the library copyright Alliance and Lindsay molds of rhizome the next week's episode will be facilitated by Brandon Butler oh pardon by Krista Cox of the association of research libraries as well as Peter Jaszi of the Washington School of Law at American University and thanks again everyone for joining us today we'll see you next time bye