

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)

Jess Meyerson:

Hey, everybody. Welcome, welcome. Happy Monday.

Lauren Work:

Hey.

Brandon Butler:

Hey, Jess.

Jess Meyerson:

Slides look good.

All right, getting started.

Welcome everyone, thank you all so much for joining us today. My name is Jessica Meyerson and I'm the Community Advisor to the Software Preservation Network as well as Research Program Officer at 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. Speaking 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 do type them into the chat box of the Zoom control panel and I will bring them up during the presentations. We'll also have time for questions at the end. Every episode will be recorded, transcribed, and posted to the SPN website. We're currently working on this workflow. These will be freely available for all.

Today we're presenting episode four, Working with Source Code and Licenses. This is a discussion with members of the Code of Best Practices Research Team and esteemed guests, which include Daina Bouquin, who is the Head Librarian at the Center for Astrophysics, an organization where Harvard and Smithsonian scientists work together on space missions and research. Daina's responsible for all library operations and leads efforts to facilitate software preservation and code attribution within the astronomy community. She's also an advisor to SPN's Easy Program of Work, the emulation as a service infrastructure, and a member of Archive's IT

advisory group, as well as a member of the Center for Astrophysics Scientific Computation Advisory Board.

We're also joined today by Lauren Work, who is the Digital Preservation Librarian of the University of Virginia library where she is responsible for the implementation of digital preservation strategies and systems. She's also the Project Lead for University of Virginia Library's Fostering Communities of Practice and Project and serves as the UVA Library Node Configuration Coordinator for the SPN Easy Program Report.

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 of Washington School of Law. Professor Jaszi is one of the Originators of the Fair Use Best Practices movement and is Co-Author of the Software Preservation Code of Best Practices for Fair Use, along with Brandon Butler, Pat Aufderheide, and Krista Cox.

This is the continuation of our seven part series of webinars which explore the Fair Use Code as well as other legal tools for software preservation and its 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 source code. With that, I'll hand it off to Brandon.

Brandon Butler:

Great. Thanks, Jess.

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 and the code and that principle deals with source code. We'll talk a little bit about what's in the Code of Best Practices about code, of the source variety.

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. It was the thing that really, from the very beginning, sort of is the first question that happened as soon as we would bring this issue up, whether we're talking with lawyers or librarians, there was always this question of licenses, licenses are so prevalent in software.

Once we, Peter and I give a kind of overview of these legal principles, then we'll hear from Daina Bouquin about her work with source code and how the kinds of principles and the 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. Then we'll open it up, we'll talk amongst ourselves and we'll also have questions from you all.

Peter Jaszi:

Brandon.

Brandon Butler:

Yes, Peter?

Peter Jaszi:

Before you summarize Principle 5, 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've talked a little bit about what came next. That is a series of what were in effect small or focused group meetings among preservation professionals at sites across the country, which we moderated and organized around topics that had come up were currently in the research at the beginning of the project that I described a moment ago.

The principles, including the ones we've talked about in the last several webinars and Principle 5 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 who 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 have described, which is the same process that we've used with ... I don't know ... 12 or 13 or 14 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 broad based foundations that the final principles have but it also 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 has been questioned, particularly by advocates of long strong copyright over years.

The question has always taken some version of the lunatics and asylum trope. 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 film making or the domain of art 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 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 hand to being quite concerned about proprietary rights on the other.

One of the things that we discovered about your community, which we found fascinating, is that the individuals, the preservationists, if you will, who were in our groups had all kinds of lines and connections 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 had 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 user's freedom of choice, if that term means anything, but also respects the legitimate interests of rights holders.

We've talked about some of the instantiations 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 one that were designed to make sure that archival activities didn't compete directly with current active programs of commercialization run by copyright owners or their successors. That was one example. This is another.

The Fifth Principle is another example of the 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 arrogantly but

nevertheless perhaps usefully refer to as 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.

Brandon Butler:

No, that's a really useful intervention, Peter, because I think we've jumped into the content of the code because it's sort of irresistible to just get into it, but it's good to remember where it came from. I think this principle is, as you say, a really nice illustration of that.

Principle 5 deals with working with Source Code. It's the human readable, right, version of software as it's sort of written by its authors that can be reused and recompiled and broken up and so on. In this principle, what we find is that it is fair use to preserve and make available this kind of material but there are, as always, a series of limitations.

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 do is we have these focus group discussions and we propose and these things evolve as the discussions evolve, but we propose a version of a principle. 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 practice?

Peter Jaszi:

And usually, this is where the rich discussion is.

Brandon Butler:

Yeah.

Peter Jaszi:

The principle, because it is so general is relatively easy to get people to agree to and refine. It's 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 limitations are integral with the principle. The principle doesn't have any independent meaning when severed from the limitations.

Brandon Butler:

As a result of all that, you move very quickly from the principle to the limitations. The principle will stake out a broad fairly ... As Peter said, fairly easily arrived at kind of zone, but then the limitations really help you define the boundaries of that zone. In this particular case, one thing that emerged very clearly in the discussions was that Source Code is very often donated by an author and the authors of Source Code who donate their creations to archives and special collection are just like the authors of unpublished manuscripts and they have complex and important thoughts about what should be done with that kind of material.

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 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 limitations in donor agreements, to keep them at the front of your mind and 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's a lot of sensitivity to putting Source Code online without an expressed 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. Again, this was something that came out from the focus groups and collecting Source Code and told us that this kind of material can be very sensitive and so when you're making research access available, again, to be

sensitive to how much access is appropriate to the research project, which is again, is actually 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.

Then the final limitation is a very common one 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 proper credit and provenance for the materials in your collections and making sure that the folks that are using those materials understand where they came from.

Those are the limitations. They're few but important. Important to understand.

Peter, did you want to add anything to those thoughts?

Peter Jaszi:

Yeah, just a couple of things to add. 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 know there's nothing in the copyright law that imposes any burden of attribution on anyone under any circumstances, save 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.

So 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 factor in determining the user's good faith, which does sometimes enter into legal analysis but it's also a free standing ethical imperative that I think in candor, the different professionals we've worked with over years, including in this context the people who are doing software preservation, your community, would probably override that even if there weren't a specific legal hook on which to hang it. It's that 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 [inaudible 00:20:12] 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, another bunch of things that you can do on your premises with very relatively few, although to the extent that they're expressed, of course, nevertheless, specific constraints, another set of activities, and now at Principle 3 which you can do on your own site, providing virtual access to members of your community.

Then fourth, a set of activities that you can do online for a wider range of users in consortial and collaborative arrangements, and really those principles almost demand to be read together because as I say, they cumulate or cascade, I mean 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 free standing 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 as a very special and sensitive case.

I would urge you in thinking about the code as a whole, since we've now reached the end of our description of the principles, I would urge you to think about the cumulative relationship between the first four principles and to appreciate the significance of the Fifth Principle as a kind of free standing one. End of thought.

Brandon Butler:

Thanks, Peter, so then let me move to Fair Use and Licensing.

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 lawyers and 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 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 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 history of clicking through and having to agree to

things when you install your software has led to anxiety expressed in a number of ways.

One of those is “I can’t do Fair Use on this software because I don’t own it. I’m just a licensor or I’m just a licensee. I’m just a mere ... I’m here entirely at the will of the copyright holder, so how can I do fair use?” Another version of the story is that “The software license is the beginning and the end of my ability to use something and if the license says yes, then I can and if the license says no, then I can’t and that’s all I need to know.”

Another concern that we heard was “Not only do I not own it, I’m not even a licensee of the software. I just got a box of [inaudible 00:24:30],” 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?”

Then finally, and this was sort of ubiquitous, one of my favorite things was to ask people who were 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 often-

Peter Jaszi:

Sometimes they can’t. Sometimes they’re gone.

Brandon Butler:

They can't. Yeah, that's right, the licenses are gone. That is, "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."

This was sort of where people found themselves when we started these conversations and so I had to school myself and, luckily, I had Peter involved and he could school me a little bit, too, about how the big picture of licensing in Fair Use really worked together.

Peter, I wonder if you could break that down for us for a little bit?

Peter Jaszi:

Oh, I'd be very happy to and I would just emphasize that at the very outset of this project, I think, much of the skepticism ... 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 Brandon has done a very good job in the preceding slide of showing what all of those were of the proposition that, well, this Fair Use 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.

That was something that we had to talk our way and work our way through when in a way the material that we're going over now and that is also in the appendix to the document itself about licensing sort of sums up what that working through of the material entailed. We certainly don't know that there could never be a case in which a licensed 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 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 from the rights holder and the other is that you have permission in effect 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 discussion today, is that the Fair Use doctrine, the authorization by law for use is very potent and it's not actually easily defeated in theory or in practice by agreements, licenses, contracts, call them what you will.

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 was going to take the first one, Brandon?

Brandon Butler:

That's a good question, we can just go back and-

Peter Jaszi:

Oh, I'll do it.

Brandon Butler:

Okay.

Peter Jaszi:

Obviously, because we don't know ... No one knows, because they're 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, we'll 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 distributed in new and different ways, terms like that pop up, not commonly, but at least occasionally, but in the legacy software period with which this project is associated, they just don't seem to be there, as a matter of description. 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.

Brandon Butler:

Yeah, and so 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 10 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," as somehow impliedly or magically excluding your legal rights, which come from beyond the license. That's really the most common source of confusion here, to me, was that folks would see a license that says, "This is what our license allows and this is what our license does not allow."

But what those licenses very rarely say is, "We block Fair Use or we block preservation." "You will hereby promise not to engage in Fair Use." I want to make

sure we get over to Lauren and Daina, so Peter, can we move a little more quickly through the last-

Peter Jaszi:

Of course we can. There's really very little to be said about the rest and I will do it very quickly.

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 and institutions who are dealing with material that they themselves didn't purchase, that was donated or that they didn't purchase from the source, that was donated or purchased second hand, 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.

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 licensor. So whatever 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 tiny, perhaps, speculative range of cases in which a license might actually operate to control what otherwise be Fair Use preservation activities, then the right question is how much trouble could you get into if you made the wrong call?

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 reach of the license terms and when a license or any kind of contract goes to court, then the court asks, "How much real harm was done?" And if there was any, 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 dollars is going to be nil.

End.

Brandon Butler:

That's right.

Now I would love to turn it over to Daina Bouquin to talk a little bit about her work with Source Code.

Daina.

Daina Bouquin:

Hi, can you hear me okay?

Brandon Butler:

Yeah, you sound great.

Daina Bouquin:

Okay, 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.

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 for 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.

Doing actual career pathways, having actual career pathways for these people depends a lot on their ability to get credit for their work. This gets a little hazy sometimes, so I'm happy to talk in tons of depth 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.

For instance, I have up here a link to something called the IDL Astronomy Library, so IDL is 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. People are still using IDL but it is a proprietary language and the codes that they've written is source code but IDL actually has explicit licensing that's actually never been tested in the courts, where they actually tried to prevent bite code compatibility with other environments.

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.

For instance, the Solar Dynamics Observatory and all of the pipelines they're used to pull down roughly like about 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 still 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 licenses are very expensive, there's this odd window in the history of IDL where astronomers started wrapping IDL code with Python scripts.

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 steppingstone essentially to more open tools. As someone who's 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 kind of pose these new challenges, because we actually ... so we go from this still necessary proprietary code to this, what I'm

calling, Wildcard 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 R 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 packages that they've put out is an IDL package.

All of these communities are trying to find ways to take these proprietary things that they're 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 the Harvard, the 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 code of best practices kind of says, "For most of the time you don't want to distribute this stuff," that's kind of the opposite of what the community expects. This is a very open community and they expect to be able to find and read everything, because they're all reusing each others' work, because you can't redo observations.

For the most part, they've actually given up numeric determinacy, if that means anything to you. What they're actually trying to share is a functionality of a model and the mechanisms by which it works. 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.

Brandon Butler:

Wow, very interesting stuff.

Thanks, Daina, really appreciate it.

Daina Bouquin:

Yeah.

Brandon Butler:

It's interesting. It's an example of attribution, a sort of an all purpose show of good faith mechanism.

Okay, thank you, and Lauren, I wonder if you could tell us a little bit about our friends at [inaudible 00:43:50]?

Lauren Work:

Sure, can you all hear me okay?

Brandon Butler:

Yes.

Lauren Work:

Okay.

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.

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 CADD-BIM software known as Vectorworks 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 actions to the accompanying Born Digital Collection Files that are also part of this collection.

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 these software dependent digital collection materials in an emulated environment and are reading room only, so really with no ability to really download materials.

I wanted to stop here and say why we're anticipating the need to apply this Fair Use kind of principles for this developing project instead of doing something like buying a license for new software, because this is definitely a company that's still in existence.

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. A lot of the versions of the software predate that by at least 10 years. These are old versions of the software that correlates kind of specifically to these digital ... and so these versions of the software are no longer being supported or sold by this company.

The other reality with folks who are probably familiar with CADD-BIM software, it iterates extremely quickly. It's highly complex, you know, there's a lot of things like

third party plug-ins, different design features, libraries, things that change over time, so newest offer 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 ensure that we get software and software dependent digital materials off the physical mediums. You know, optical disks are not forever either and hard drives are not forever, so this is part of our preservation mandate. 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 dependent 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 this scope of our SEO key projects.

Jumping in a little bit about our licensed software specifically, so it's mostly Vectorworks is part of a much bigger company, the NEMETSCHKE Group which is headquartered in Germany and Vectorworks specifically is proprietary CADD software. 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 this Born Digital Files that are part of this manuscript collection.

Authentic access in our case, from the manuscript collection, means that researchers and user ... [inaudible 00:47:55] in an environment that allows for kind of an understanding as how work itself is rendered, created, developed, iterated and used in the course of an architect over time. Something I've been thinking

about more often in talking with the archivist working on this, touching up ideas around 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 that are heavily marked up by architects. 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 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 a part of that.

Kind of final point here of the pathways of talking to Brandon about this, a couple of things that have happened, given 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 deed of gifts, so that's something that we really want to make sure has been updated for many of the reasons already highlighted here.

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 includes the license terms, as well as the license keys. This is something that I'll let Brandon touch on a little bit here, too, but this 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.

Brandon Butler:

Yeah, hey, so you know, I think actually, so thank you, Lauren, first of all. It's a great story. I think it fits perfectly with the kinds of use cases that we heard in the small groups as well. The notion that 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 it the way the author did, so the research value is just undeniable, so we're really excited to be able to work on that.

I looked at the license, I read the terms, and they fit in exactly with the kinds of trends that Peter and I described. This is a license that tells you what the license tells you. It tells you what the license was going to let you do and what the license is not going to let you do. 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 justified.

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-

Jess Meyerson:

Yeah, 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.

Thank you again, Lauren and Daina.

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.

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 exactly how the work with the Cyber Law Clinic and the Code of Best Practices is informing policy [inaudible 00:52:25] for Astrophysics, in terms of how you're treating IDL, the example that you provided, how much ... question about similar examples are happening in your professional circles at large when it comes to research source code, especially legacy proprietary research source code, like the example that you provided.

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 ...

Sorry, I'm hearing a lot of echo.

Peter Jaszi:

That maybe my phone. Would someone ...

Brandon Butler:

There you go, Peter.

Daina Bouquin:

Oh, better.

Brandon Butler:

Sorry.

Daina Bouquin:

Sorry about that.

Daina Bouquin:

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. 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.

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 that 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. 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 [inaudible 00:55:28] libraries and some old 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 functionality?

Jess Meyerson:

Yeah, that's a 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 researchers and other members of your community that are going to be donating those materials.

Daina Bouquin:

Yeah.

Jess Meyerson:

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 at in terms of thinking about software preservation's impact on your existing policies, especially some of the specific examples that Lauren and Daina highlighted in today's episode?

Brandon Butler:

Yeah, while people are thinking about that, and maybe typing in the chat, Daina's story gave me ... 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 tool to empower clinics that you're working with, who need to help you through specific challenges that you're having.

In the documentary film community, there's just a cottage industry of IP law clinics all over the country who including ... Especially the one where Peter and I have worked together in the past where filmmakers would come to us. They bring us a film and what the code does is it gives the law students a very workable framework they can apply to the film and say, "This film is within the framework of the code," and then they can write an opinion letter and that opinion letter is a very powerful tool in the context of documentary film distribution.

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 Burkman 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.

Jess Meyerson:

That's what we're hoping.

Brandon Butler:

Yeah, good.

Yeah, you know, Chris and Kendra are very well aware of the code, so I'm sure we're in the mix with them, and then the other thing is Daina's comment reminded me of my general view of Fair Use and Licensing. You know my grand unified theory of Fair Use and Licensing on time scale is that going forward if you can start fresh, you want to use open licenses and that will make everyone's life easier.

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.

Jess Meyerson:

Exactly.

Brandon Butler:

And then they disappear ... and so Fair Use to me is just the solvent that can really help us make progress there, so I'm very hopeful, Daina, that it'll be helpful for you.

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 a literacy standpoint, teaching people about explicit licensing and all of that is still an issue.

Brandon Butler:

For sure. All right, well, we're about at the top, Jessica, was there anything else?

Jess 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 and 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 and 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.

With that, I'll just say huge thanks, again, to the entire Code of Best Practices Research Team, to your facilitators, Brandon and Peter, and a warm thanks to our esteemed guests today, Daina Bouquin and Lauren Work, and also to all of our attendees for joining us in discussion today.

Join us next week. Same time, same place, Episode Five, 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, who's Counsel to the Library Copyright Alliance and Lindsey Molds of RISEM.

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.

Brandon Butler:

Bye, all.