

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

Episode 2: Beginning the Preservation Workflow

Featuring Special Guests: Leslie Johnston (National Archives and Records Administration; Henry Lowood (Curation for Science and Technology, Stanford University Libraries)

Episode 2 Transcript

Transcript

Jess Meyerson:

So welcome, everyone. Thank you so much for joining us for today's webinar. My name is Jessica Meyerson. I'm the Community Advisor to the Software Preservation Network, along with my colleague and Director, Katherine Skinner and the rest of the Spin staff, maybe a couple of whom are on today, and members as well. I'm also the Research Program Officer at Educopia Institute.

So just a little bit of housekeeping before we get started and dive into our program. All but the hosts and the guests will be muted throughout the webinar, just to maximize the audio and visual quality of this recording. So if you have any

questions during the presentation, we ask that you type them into the chat box, the Zoom chat box, which we will be monitoring closely throughout. So that if we do have a backlog of questions throughout the course of the presentation or the discussion between Brandon, Leslie and Henry, we will cue those up and they'll be ready for Q&A. We will go in the order that they were presented.

So every episode will be recorded. We're recording right now. It will be transcribed and posted to the Spin website, freely available for all. As a reminder, today, we're presenting episode two, Beginning the Preservation Workflow. And this is a discussion with members of the Code of Best Practices research team and two of our esteemed guests.

The first, Leslie Johnston, Director of Digital Preservation for the National Archives and Records Administration. With the responsibility to develop and execute a digital preservation strategy for the agency, Ms. Johnston has over 35 years of experience in the cultural higher education and federal communities, including the Getty, Stanford and Harvard University libraries, and the Library of Congress, where she worked with digitized and born digital collections, setting and applying standards and overseeing the development of the digital content management and delivery systems and services. Her expertise includes digital collection management systems and infrastructure design, digital preservation systems and standards for digital collections.

We're also joined today by Henry Lowood, curator for history of science and technology Collections and film and media studies collections at Stanford Libraries. After being trained in the history of science and technology and receiving his PhD from UC Berkeley, over a period of 35 years, Henry has combined interests in

history, technological innovation and the history of digital games and simulations to head several long-term projects at Stanford, including How They Got Game, the history and culture of interactive simulations in video games in the Stanford Humanities Lab and Stanford Libraries. The Silicon Valley archives in the Stanford Libraries, and Archiving Virtual Worlds collections, hosted by the Internet Archive, just to name a few.

He's led Stanford's work on game and virtual world preservation and the Preserving Virtual Worlds Project funded by the US Library of Congress and the Institute for Museum and Library Services and the Game Citation Project, also funded by the IMLS. Henry is also the author of numerous articles and essays on the history of Silicon Valley, and the development of digital game technology and culture.

And your research leads and facilitators for today's episode are Brandon Butler, director of information policy at the University of Virginia Libraries, who's joined by Peter Jaszi, professor emeritus at American University Washington College 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 in Software Preservation, along with Brandon, Pat Aufderheide and Krista Cox.

So this is the continuation of our seven part series of webinars that explore the fair use code and other legal tools for software preservation, co-hosted by the Association of Research Libraries and the Software Preservation Network. And with that, I will hand it off to Brandon.

Brandon Butler:

Great, thanks, Jess. So we're really excited to join you in this second episode in our seven webinar series. By the time this thing is over, it's going to be warm. So I'm pretty excited about that.

So we're going to talk today about the first two principles in the code of best practices and fair use for software... whoa, software preservation. And let me give you a little overview, roadmap of the talk. So I'm going to start off by talking a little bit about principle one. And then Peter Jaszi will talk a little bit about principal two, and then we're going to hand it over to our guests, Henry and Leslie, to talk about their experiences in the field.

So first, Henry is going to talk a little bit about the Cabrinety Collection and some of his adventures in the world of agreements and contracts, in particular, among other things. And then Leslie will talk about the joys of collecting and preserving the digital government record.

Then, we'll have time for discussion. And I think we want to make sure that folks are able to ask whatever questions are on your mind. And if you were on the last webinar, and there were things you didn't get to bring up last time, there's no bar on talking about whatever is on your mind related to the code. But do keep in mind, we're only going to have Henry and Leslie today. So if you've got questions for them, do be sure to get those in while we've got them on the line.

So the first thing I want to talk about is the very first principle of the code. And I think we probably mentioned last week, that the code overall is structured in a kind of progression. So that each principle follows roughly, chronologically a workflow of

software preservation, starting with a box of media on your desk, and ending with a cyber utopia where everybody can read everything, no matter how old it is.

The first principle is getting from box of stuff on your desk, to software in your collection, or software that's a part of your tool set for managing your collection, as Leslie will talk about later. And so, principle one covers accessioning, stabilizing, evaluating and describing those digital objects, software objects.

And that includes the kinds of things, and we didn't talk a lot about the process of developing the codes, but the codes are developed by speaking with practitioners, like Henry and Leslie, in confidence and in small groups as well as individually about what they do. And so this process tries to mirror of what people told us they do when they're trying to preserve software.

And so that includes things like making multiple disc images of original media, documenting what the original media packaging might have looked like, other materials that were associated with the software, making notes about that. But also running the software, so that you can tell what it is and what it does and how it works.

All of that stuff that happens sort of at your workbench. When you're processing the collection, you're figuring out what's in it, you're getting it from unstable, unmanaged, to stable, managed, described, it's ready to be a thing that you're caring for. And so that's why I chose our little forensic computer pal over there as the image. I know my colleague at work has one of those things, and it costs more

than my first car. Fair use better lets you use that thing, because it's way too expensive to just sit there on your desk.

Each principle has a set of limitations that are associated with that principle. And this is the way that all of these codes work. In the focus group discussions that we led, in the process of developing the codes, we probed the consensus to say, well, at what point would you feel uncomfortable? Or, what are the things that you feel are important to be considered as part of working your way through a situation like this one?

And so, one of the first ones, which seems sort of maybe goes without saying, but it came up a lot and with some vigor was, you really need to have preservation activities that relate to your mission. This was something, literally the first thing that anyone, the first thing that any group, within 10 minutes of our first focus group discussion, this emerged, we put out a hypothetical question. And the group immediately said, well, why are we doing this, right? You need to tell me that this is a part of my mission, or I'm not going to do it. So you need to have a relationship between the preservation activity and the mission of your institution.

For donated material, donor agreements are just so important. We heard over and over again, as you're conducting this activity, you're protected by fair use. But donor agreements intervene at the same time, and you can't sort of run headlong forward without remembering the other sources of obligation that might come into play.

Reasonable care, again, at this stage, to identify content that's sensitive for non-copyright reasons. This is part of being a good actor. So it's sometimes referred to as kind of a fifth factor and lawyers and law professors debate about whether

this is good or not. But I don't think there's much debating about the fact that judges actually care whether you show that you're a good actor.

And so the community, actually, without us telling them that a judge is going to make you do it anyway said, this is what we do. We take care to take account of the things that we're processing, that might need to be flagged for non-copyright reasons. And that's part of being a good professional.

Descriptions should be created, expressing and shared to facilitate discovery within and where possible beyond the institution. And this was another one where we heard over and over again, we're not going to do this if we can't actually make these things findable. Part of what it is to preserve something is to describe it and to make it a part of a collection that someone can find and use, whether they're inside or outside of whatever circle of users we're thinking about. We want to describe it in a way that it's findable to people. So that was an important part of stage one, when you're getting started, consider your user.

And then, finally, at this stage, and this is important, thinking chronologically, right? For the purposes of principle one, the people who are handling software for this purpose should be personnel, including staff, volunteers, contractors. We don't need to see your badge, but there needs to be an affiliation, whether at the home institution, or at a partner institution, or a vendor, who are directly engaged in this kind of activity. You need to be doing preservation at the preservation stage, containing access for that part of the work.

Access comes later. And we'll talk about the terms on which access is provided later. But in this phase, access is limited to the people who are doing the preservation work.

So those were the limitations. They're I think fairly like robust and interesting, but not limiting in a way that I think is going to constrain anyone from doing what they see as an important part of their mission. And I think that's a good place to end up.

So now, I will turn it over to Peter Jaszi to talk a little bit about principle two.

Peter Jaszi:

Thank you, Brandon. And I want to start by emphasizing something that you have just said, and that I think we said already. And we were discussing the project last week, and that is that we learned very early, when we started to talk to the professionals in the field, who were kind enough to work with us, including very early on, Henry and Leslie, that the mission of software preservation was not simply a preservation mission.

And of course, this is true of almost all if not all archival activities. That is to say, preservation doesn't make any long-term sense, certainly not as a way of expending significant resources, if it isn't for purposes of encouraging and facilitating access.

So the next three principles in our code are really all about different varieties of access to preserve material. And one-

Jess Meyerson:

Peter?

Peter Jaszi:

Yes, indeed.

Jess Meyerson:

Just a brief request, is there any way that you could speak slightly closer to the mic?

Peter Jaszi:

I'm sure, I can.

Jess Meyerson:

Okay, perfect. Thank you so much.

Peter Jaszi:

Not at all. What I had said a moment ago is that one of the first things we learned is that the preservation mission and the access mission are all tied up together. One can't really be separated from the other. So the next three principles, including the

one I'm about to talk about, are focused on the access side of the preservation activity.

And one of the first things we learned when we started to talk with the generous experts in the field, is that it's often important for collecting institutions to create and make available visual and audio documentation of legacy software in operation. That might include screenshots and videos of software in operation, or software in operation being controlled by an expert user. There are lots and lots of dimensions of software, which are difficult or impossible to capture fully in a textual description or even in an emulation experience.

And all of that is good and straightforward and intimately mission related and in some sense, profoundly non-controversial. But copyright is an issue, because the protection that copyright provides, actually goes beyond code itself, and extends as well, at least theoretically, to various kinds of software generated displays.

So you have at least to think when you create products or versions of software, to document its operation that are going to be shared and seen by various publics, you have to at least think about whether or not you can do that in compliance with copyright law. And the happy answer is, yes, this turns out to be really a very straightforward, fair use question.

Last time, we talked a little bit about the ruling concept in contemporary US fair use jurisprudence, which is this idea of using something for a transformative purpose and this is a classic example. Obviously, when I present documentation of the operation of a software program, I'm doing something very, very different from

what that program, whatever functionality that program was originally designed to accomplish.

So this was easy. It didn't take the small groups that deliberated with us about the appropriate metes and bounds of these best practices, very long to decide that in the broadest sense, this was a clear, obvious and important example of fair use.

And then we have the limitations and those are pretty straightforward as well. The more the better, where context is concerned, is the first limitation. Fair use transformative purposes are always easier to demonstrate when you are showing and telling more about the context of the thing that is being demonstrated. It won't always be possible to provide rich context, but when it is, it should be done.

The second limiting proposition here is really just a restatement of a general fair use concept that I explained a little bit last time. Everyone cares, including the courts, that when you use something without having express permission to do so, the extent of your use should be commensurate, appropriate, proportional, pick your word, they all mean more or less the same thing to the purpose. So the more extensive your roles, the more documentation is justified.

And then finally, there's the last limitation, which came up and a number of our groups thought it was important to include. I have to say that this one is, for the moment, I fervently hope that will change, more of an aspirational than a real limitation. That is to say, were the copyright owners of legacy software themselves to go into the business, so to speak, of providing extensive online documentation and their legacy products in action, especially if they were to figure out a way to

monetize that activity, then perhaps collecting institutions would want to step back and leave that market to them.

But so far at least, there have been a few if any indications that that is or is likely to be taking place. So remember it, but for the moment, I think, don't feel particularly constrained by it. And that's really all there is to it. This is about as close to a carte blanche as we're going to see with respect to any of the fair use propositions that are contained in the code itself.

Brandon Butler:

Thanks. There we go. Thanks, Peter. And so we've said a little bit about those two scenarios in overview, but I think it's going to be really interesting and useful for you all to hear a little bit from Henry and Leslie, and especially I think, and they will correct me if I'm wrong and their illustrations will bear me out or not. But I think they have a nice complimentary aspect to their two use cases, because Henry is really someone who's collecting software for software's sake and Leslie is collecting digital documents and she needs that software to make sense of the documents that she'll tell you more about later.

And those were really, in a way, the two mega overarching use cases that we've heard about over and over again. So starting with Henry, I'm really excited to hear from you all about what it's really like to do this stuff.

Henry Lowood:

Okay. Thank you. Thank you for inviting me and I'm really happy to talk to all of you out there. Well, I'll be talking mostly about a collection at Stanford, the Cabrinety Collection, which is a collection on the history of micro computing software and features a collection of about 15,000 to 20,000 pieces of software. We don't know an exact number, never really have, because there's a lot of magazines with software and all sorts of things, everything you could imagine that could complicate an exact count.

I'm going to focus I think mostly on situation one about accessioning, stabilizing, evaluating and describing digital objects, specifically around a project to create disc images from original media. And in doing that, I'm going to focus on the second limitation, limitation B, just to remind you what that one says is where materials have been donated, their preservation should be undertaken in light of the terms of donor agreements, which may limit reuse and access.

They may limit reuse and access, but of course, donor agreements can also argue, help you with access, they can augment access in some ways. Use case again will be the Cabrinety Collection and in particular one of the projects we carried out with the Cabrinety Collection with the National Institute for Standards and Technology. Specifically what's called the NSR, the National Software Reference Library run by Doug White, which I described as the national forensics, software forensics laboratory.

Now as for situation two, documenting software in operation, I'm not going to say too much about that directly, even though it pains me greatly not to, that particular thing is something that's occupied me quite a bit, both as a historian and as a curator. I've written about it quite a bit, my one sentence description of what I

would say is that documentation of that sort is at least as important to game historians as access to operating software of the past.

I'm going to leave it there. If you have questions, if you want to talk about it in Q&A, I can certainly do that. So back to the Cabrinety Collection, the full name of which is the Stephen M. Cabrinety Collection, the history of micro computing. We acquired this collection at Stanford in 1998 and 1999, that's 20 years ago, a little over 20 years ago.

I will state right now, I believe this is true that it was the first acquisition of a software collection by any repository. I've written a little bit about that, you can look for an article called software archives and software libraries that I wrote in a recent book in the Smithsonian studies in the history of science and technology series, basically, about the history of software collecting.

And so again, the Cabrinety Collection has been around at Stanford for over 20 years and we are still working through projects to deal with the workflow that leads you from acquisition to full access. The current project, being the easy project, we're finally at access. The project, of course, is one that the Software Preservation Network has set up for a number of institutions to participate in including Stanford.

I'll be talking about, when I say software, I'm talking about packaged, PC software, productivity software, game software, edutainment, all those sorts of things. I'm not talking about mainframe or bundled software as it used to be called, unpublished software, scientific or research software, academic software, if you will. I'm not talking about non-PC software, newer things like mobile software, things like that. Some of what I say I think is applicable to streamed and downloadable software.

Some of it's not. And we actually do have a little bit of downloaded software from bulletin boards and things like that in the Cabrinety Collection.

Now, if I were talking about some of these other topics, like for example, academic software, some of what I would say would be a little bit different, in terms of agreements and things like that. And again, I'm not going to dwell on that. I'm just going to say if you have questions about those categories of software, some of which I definitely have worked on, feel free to ask later.

So in terms of workflow, we've been working on the Cabrinety Collection now for over 20 years as I've said. This work has largely been carried out through a string of funded projects, some of which Jessica mentioned in the introduction, included the two preserving virtual worlds projects funded by Library of Congress and IMLS. The NIST Cabrinety capture project, which I'll be talking about, the game citation project, also funded by IMLS and now the easy project.

Noticed in that sequence, started with what's out there, what could you acquire project, followed by a capture migration project, followed by a description project, followed by an access project, and here we are 20 years later. These have all been multi-institutional projects, including Stanford.

So back to the point about fair use and the specific points where materials have been donated, their preservation should be undertaken in light of the terms of donor agreements, which may limit reuse and access. I'm going to put this point in a slightly different way. Software preservation involves kind of a complicated game. And the players in that game include copyright law, the Digital Millennium Copyright Act and so forth, the various provisions and things there. Fair use, which

of course, we're talking about today. Contracts, in the sense of shrink wrap agreements, and things like that. And then specific agreements with donors and rights holders.

So all of those things can come into play and interact in different ways. Sometimes you have something in one category, but nothing in another category. Sometimes you have multiple agreements and concerns about copyright law and all sorts of things playing together. It's kind of like a complicated game of rock, paper, scissors, figuring out that sometimes fair use maybe beats copyright law here, while maybe in another occasion, a donor agreement would trump fair use and so on and so forth. So it's quite complicated.

Can we have the next slide, please? Let's see if this works. Or did I put Brandon to sleep? Oh, there it is. Okay, great.

Okay, so here's the deed of gift, of course, the acquisitions process in my area of curatorial practice. Sometimes I do buy individual software titles, we do have a media center, where we do that sort of thing. With historical software, it's been mostly around collections, acquiring collections. And these have mostly been gifts, beginning with this instrument called the deed of gift.

By the way, the Cabrinety Collection, alas, the one that we've been doing all this work on was acquired in 1998. You can imagine, you're looking at the current template for deed of gift at the Stanford Libraries. You can imagine the horror that you will experience when you look at a deed of gift done in 1998, in terms of its applicability to the projects that we're doing today. It really involves another layer of

translation in that game that I described of figuring out how the terms of a 20 year old agreement will apply.

Next slide, please.

So one of the most important conversations whenever you talk to donors concerns how to handle copyright in the materials that are donated. In our current template, a donor gets to choose from three options, transferring copyrighted to Stanford, granting Stanford a license, which you'll see in a minute, or just saying nothing about rights.

The key point, however, is that these choices are only applicable to the extent that the donor owns copyright and the materials that they're giving to you. Or other IP rights potentially, like patent rights, which has come up on occasion.

In the case of the Cabrinity Collection, this did apply to a portion of the collection we were given. Exactly three titles out of the circa 15,000 titles in the collection were copyright to Stephen Cabrinity. 14,997, let's say, were not. So this portion of the agreement only applied to those three titles and also to his personal papers that were included. So that's one thing right off the bat, these agreements often don't address the copyright issues, because the donor doesn't own copyright in the materials.

Next slide, please.

So here for completeness are the other two options that are available to the donor. Take a few seconds to browse option B, which is our preferred choice. This option

grants Stanford a license to carry out pretty much any migration or reformatting we'd like to do, as well as granting us the right to provide what we call world access via the web.

But again, the key point, the donor can only grant this license if he or she is the copyright holder for the material. And sadly, this is generally not the case, as I just said. And of course, in the case of the Cabrinety Collection, this option was not stated, was not available at all, because well, 1998. We just didn't think about these things back then.

Next slide, please.

This is here, again, pretty much just for completeness. The same options that I've mentioned before, the same three options would also be available in the case of collections that we buy, as opposed to acquire as gifts. And, yes, we have bought collections on occasion and we've acquired copyright on occasion.

That's another thing I wanted just to mention briefly, is there is an option in a sale, as well as in a gift, to transfer copyright. And there's even the option of acquiring copyright straight away, say, for a collection that you already own and we've done that on occasion. I just wanted to put those on the table.

Wow, that was amazing. That was a mind read slide advance. Okay, that's fine.

That's where we want to be.

So we did this project with NIST, which created a big collection of disc images. And just as Brandon said, sometimes when you're doing preservation, you need to be

thinking about what's going to happen down the road. In fact, you probably won't even do the preservation, if you're not thinking about what's going to happen down the road. You need to think about access, even while you're supposedly focusing just on preservation.

So the focus of the project with NIST was to capture software from original media, create portable disc images, then that could be stored in the Stanford Digital Repository, and theoretically could be seen and downloaded by researchers. In addition to the disc images, we also created photographic images, thus complicating the word image for us forever, when we can't refer to images now and know whether we're talking about discs or photographs. By photographic images, I'm referring to photographs of the physical media, the carrier media, photographs of the boxes, the box covers from all sides and photographs of the inserts, such as manuals, and other things that were inside the box.

So we anticipated research access to the software right from the beginning as we were designed the project and had lots of discussion about what we thought we would be able to do. This involved, remember those players I was talking about, thinking about copyright law, thinking about fair use, this was 2012, 2013. We didn't have the fair use document that you have now. So we were pretty much guessing. We didn't really even have a lot of the documentation that ARL has compiled for other kinds of materials that we might have used in a kind of a transfer to software, we were pretty much guessing.

And finally, we got tired of guessing, and said in this case, we're going to mount a parallel project to contact the rights holders for the software we were migrating. And this letter that you see here in the slide is the letter that we wrote to the rights

holders we contacted. We began with the rights holders who held the most titles, so your Activisions, Microsofts and so forth.

We didn't go very far down the tail, we still plan to do that. But if you can imagine the collection of software from the 1970s, '80s and early '90s, many of the companies on the long end of the tail don't exist anymore and we'll have to think about how we do that.

Next slide, please.

So this is what we asked of the rights holders. We said we're contacting you for guidance about the level of access you will allow us to provide to your materials. And we would ask them that question, as you'll see in a second, we provided them with some options, that would then be documented. And we wouldn't need to care about copyright law or fair use or any of that stuff after doing this, because we heard from the rights holders, and they said exactly what we could do. That was the hope. We felt that this would eliminate this game that I was talking about to enable us to proceed without ambiguity.

The next slide.

So here's what we sent, we sent something like this to every rights holder we contacted. This is from the letter to strategic studies group. We listed the software titles we had identified, which stated that they owned copyright, so that could be on the disc or on the box or something inside the box, it says copyright SSG, we're

contacting you about these titles. And we first of all, we asked them to verify indeed that they did own copyright.

Then, next slide.

And then we asked them for permission, according to this simple grid that you see here, both for the disc images and for the photographic images, world access is unlimited over the web allowing download and all that sort of thing. Research use only would be some sort of access with no permission to redistribute copy such as no downloads. And then restricted research use only would also include reading room access, I should mention. Restricted research use was if you've got something that you've got a problem with, let's talk about it and figure out a special case here.

Okay, next slide, please.

I'll conclude here. We actually can go back to the last one, we can just stare at it while I'm talking, it's a little more to look at. So again, we contacted rights holders about titles we thought they owned copyright to. And there was every reason to think that due to statements on software boxes, and so forth. Here's what we learned from their responses.

The first thing we learned, I think was the major thing was, we had discovered a new category of orphan software, which is if you think about that SSG list, there were 10 titles there or say a Microsoft and Activision to whom we might have had 200 or 300 titles. Typically, we received back confirmation that they felt they owned on copyright or were willing to assert copyright to half, two-thirds, three-quarters.

Many titles for which we were certain they owned copyright, the purported holder said they did not own it. At least they were not willing to assert it.

There were a variety of reasons for that. If you want to know some of those reasons, let's save that for the Q&A. Secondly, we learned that we're not going to acquire world access for very much. The total right now is up to about 15 titles out of 15,000. So less than 1% for which we have unrestricted access to the disc images. However, in those cases, we're mostly dealing with what we call reading room access.

However, for photographic images, we've received permission for world access that is unrestricted, almost in every case. So this suggests that the rights holders are maybe less concerned about certain kinds of documentation around the software, they're less concerned about restricting access to them, than they are about restricting access to the software itself.

I'll conclude on this last sentence, kind of circles back to the second case that we discussed, concerning documentation and its importance. That's a little bit of a light at the end of the tunnel, indicating that probably with documentation, we're not going to have very many problems with rights holders. Okay, I'll stop there, and hand over to Leslie.

Brandon Butler:

Awesome. Thanks, Henry and I'm just switching over to Leslie slides here.

Leslie Johnston:

Thank you, Brandon.

Brandon Butler:

All right. All go.

Leslie Johnston:

All right. Next slide.

All right. So we need to start with sort of some context for what we do at the National Archives. And the most important question is, do we actually collect software? We don't explicitly collect software, but we collect the permanently valuable records, permanent records of the federal government. And if an agency identifies code or software that they have developed as a permanent record, then it does come to the archives.

This is uncommon so far, but it is not unknown. I have a variety of different types of code that we have gotten from agencies. Some of it Java, that's actually the largest category of code that we have gotten in. Most of the code that's created by the federal government is in the public domain. So it is different from Henry's situation where he is bringing in primarily commercial software or open source software that was created through some sort of license. Most of what we get is public domain, unless it was created under a contract that had terms that overrode that status. So

it's really incredibly rare that we get the same sort of proprietary software that Henry gets.

What we do have is over 1 billion files, actually, it's over 1.5 billion files in our permanent record holdings. So federal and presidential, that are born digital, that date back to our first transfers in 1968. So we've been bringing files in for over 40 years, which means we have over 200 versions of file formats created in a variety of packages, in different operating environments, that have come into our collection.

Next slide.

So the context formats for us, is that we issue guidance, which we call transfer guidance for the agencies that have to send materials to us. So it's about the media types, it's about the file formats, it's the metadata. Some of this is actually in federal code, but most of this is guidance. We don't actually have a record type for software or code yet. We have record types for textual, for GIS, for databases, for email, but not yet for code, because we have received it so irregularly.

We're not able to be prescriptive about what we receive. We have concepts of preferred and acceptable formats and that's approximately 50 formats across all the record types. Like we prefer PDF A to other forms of PDFs. We prefer open standards, to proprietary standards, such as the Microsoft suite.

But this is real life, so the agencies do the work in the environments where they do the work. And as you can imagine, agencies, the work of those agencies goes from, we just do email or documents or spreadsheets or presentations, to the scientific

agencies such as NOAA or NASA, where they have observational data, as well as code that they have written to work with that data at those agencies.

And because we have the variety of work and we have the longitudinal question about what we're getting in, we are always going to have to have flexibility. And we're always going to say take the record, even though it doesn't meet our guidance, versus we don't want to preserve the record in our holdings.

Next slide.

So the way that federal agency transfers to NARA work, are that agencies identify records, because they know their records better than us. But they do consult with NARA on which of their records are temporary, which means it has temporary business value to the agency, no permanent value, these are working files, they're not going to come to us. And then a schedule is agreed upon for the disposal, where they're given the authority to dispose of their temporary records, and then are required to transfer their permanent records to NARA.

This, again is where we come into some interesting questions about not only the records, and I swear I'm getting to it, the software, because they could hold onto it for five years, 50 years, or as we heard a couple of weeks ago, 500 years. We had an agency tell us that their records have value for the life of the physical structures that they are responsible for. And that one of those structures, the Hoover Dam, the records related to the building and maintenance of the Hoover Dam, will have business value for as long as the dam is in existence. And they apply this same standard to permanent records and retention for all of the structures that they are responsible for. So as they told us in this call, we will not be getting most of the

Hoover Dam records for 500 years, plus 20. Because they add a plus 22 to everything just in case.

Next slide.

So what does any of this have to do with software preservation and fair use? So as Brandon mentioned, we have two use cases. The smaller use case is that we do receive code from agencies. But the more prevalent use case for us is that given that potential for lengthy periods of retention by agencies, the uncertainty that they will be able to migrate files over time, because I will say that federal records managers, that category of position in the federal government has to be one of the most underfunded and understaffed areas of the federal government. And there being such a variety and vintage and formats, as well as the software and operating environments, we need older software packages to be able to validate, process, described and migrate the files that we have in our holdings now, will have into the future, and we'll get in the future.

Next slide.

So the workflows for bringing in code or any type of born digital record is the same. We have a single workflow for accessioning, processing, ingest of born digital records. So agencies let us know that there is a transfer that they would like to schedule, they have to tell us what schedule it is, and what type of materials these are, both in terms of record types, but also the intellectual content. Are these emails, are these press releases, are these project management records? So that

we can actually confirm that what they're sending is what we expect to receive, and that what we've received is what they claimed they were sending.

So our workflows are not unlike any other digital accessioning, ingest and preservation workflow. We need to validate that they conform to the format that they purport to be. Are they really PDFs? Are they really drawing files? Are they really Java code? Is it compressed or uncompressed? We want everything to come to us uncompressed. Is it compiled or under compiled? If we get code, we want it to come to us uncompiled and the confirmation that the intellectual content meets the requirements of the record schedule. If they send us things in these transfers and they don't meet that, we do not take them into the permanent collections.

As I mentioned, with compression, we also require that any materials be transferred to us without encryption. So things must be uncompressed, and unencrypted when they come to us, and that includes code.

Next slide.

So, agencies are expected to transfer supporting documentation along with the files that go into that transfer dossier. Not surprisingly, this can either be present or absent or be highly variable, in terms of the granularity of documentation that they send us for things like datasets, databases, spreadsheets. We hope for some sort of XML, JSON, we expect for some sort of documentation of the data schema or the markup scheme. We don't always get that.

So our processing archivists have to work with whatever it is that we have gotten. So we need to make copies for ingest, if the files have come to us on media. And

that can range from coming in on a USB stick, to coming in on an entirely racked server environment, depending on the scale of the transfer that is coming to us.

We run format characterization tools to identify if they are or aren't what they say they are. We attempt to open and or run them and I will be circling back to that activity. We have to review them for PII, because we will get in particular, datasets that come in with PII. And as a separate activity, we do also receive classified materials. And that can include code that comes along with things from say, the Department of Defense.

We need to describe them and we need to create processing notes about the state of the files, and the associated environmental requirements for not only us, but for researchers to interact with them. And from our point of view, even though our records and our code are public domain, we have created processes that we believe are in line with the code of best practices that we're talking about today, especially principle one, in terms of how we actually process code for our holdings.

We do not make any recordings of the software in use, if we have received it, or any other interactive materials that we have received. We don't record user interactions, we don't generally ever receive packaging. So we don't create any sorts of images that would relate to principle two.

Next slide.

So even though we're focusing today on principles one and two, I need to talk a little bit about principles three and five, about the work that's necessary for preservation and the work that's necessary to provide access. So the Federal

Records Act requires that agencies send us their records, including code, in a manner that actually, explicitly allows NARA to provide the access.

So we do indeed retain and preserve the original format of the files, we create public use copies and we provide access to the holdings in common formats. But we also provide the original formats, as requested. And we believe that our activities are in line with the code of best practices.

Next slide.

But as Brandon mentioned, what this means for us if we have 1.6 billion files in well over 200 variants, this means that we must have software to process the records, as well as potentially provide access to them. And that's slightly out of the scope of the core principles, but if you look at the appendix two in the code best practices, that's a section that I recommend to every archivist who deals with born digital materials, and take to heart when you discover inevitably that you will need legacy or vintage licensed software and operating systems that are required for the processing and preservation of your collections.

Next slide.

So that's it for what I wanted to make sure that I covered, as sort of the introduction to how we do things and the issues that we have come across in our work. And now I throw it back to Brandon, and everybody so we could have a discussion about this.

Jess Meyerson:

Yeah, thank you so much, Leslie, Henry, Brandon and Peter. We do have a few questions queued up for today's Q&A. So we'll take some time to review those now. I will take a moment to encourage everyone to continue to paste their questions into the chat. We may be developing a backlog of questions. If we don't have time to answer them today, I'll continue to repeat that we will address them over the course of the series. And they may be addressed explicitly in writing on the post that includes the publication of the recording.

So the questions we have for today, there was some follow-up about Henry's presentation. Most museums provide images of their objects in their collection, including copyrighted materials. So just a follow-up question for Henry and Stanford library policy, in terms of what the concern was about providing images of the physical materials.

Henry Lowood:

Well, the image of the carrier format would probably be analogous to what the museums do and that was not the part we were concerned about. The part we were concerned about were things like the manuals and the boxes themselves, which the manuals are text, and are certainly covered by the copyright that the publisher owns over the software title.

Box covers, a similar argument could apply, although as we know, in the age of Amazon and so forth, box covers and things like that fly around quite easily. It would be very unlikely that a publisher would have a problem with that.

But we were pretty concerned about things like manuals, basically, the booklet inside the box. Maybe I should backtrack and remind people that there was a time when software included a printed manual. And in the era we're talking about of the '70s, '80s and early '90s, that was quite common. Some of them, I'd say, went up to about 200 pages, there are games in particular and even productivity software that have quite lengthy manuals. And so certainly those would be covered under copyright. We wanted to have clearance on them.

Peter Jaszi:

If I could just jump in for a moment, it's such a good question. The reason that museums provide images of things in their collection is because they feel, correctly, that they can rely on the fair use document to do so. So the question and the project of today are very closely related.

Jess Meyerson:

Perfect. That's our first question. We also have a, this is also follow-on from Henry's presentation. And so Henry, if you could speak to this, and maybe Peter, as you did just now or Brandon or Krista or Pat, might speak a little bit to the broader context for this.

So this was a question about why particular donors maybe think that they don't own the rights or particular software companies think that they don't own the rights. So this was maybe hearkening back, Henry, to your information gathering phase, when you were initially doing permissions. And I would like for the attendee

that asked that question to please step in and correct me if I've gotten the thrust of your question incorrect.

Henry Lowood:

Okay, I'll just go ahead and if that questioner wants to interrupt and say I'm going in the wrong direction, feel free. Well, donors, of course, rarely have copyright over everything they give a library. You can imagine somebody gives their collection of magazines or software to the library, they generally will own the copyright, and none of it. And even in the case of their papers, very often, they're donating things that they don't own copyright to.

In the case of the publishers, that's the more interesting thing, the publishers we contacted, for whom we were relying on copyright statements in the materials that we had, that stated that they owned copyright. And now we go to the publisher, and they tell us we're only willing to assert rights on about half the titles or two-thirds, whatever the number would be. Why is that?

A bunch of different reasons. Keep in mind here, that we're talking about software that's at least in the youngest case is 25 years old, in the Cabrinety Collection. But that doesn't explain everything. It might be that the company has a policy of not answering a question like that unless they can locate their contracts. And guess what? They can't find the contracts for the software from 1982.

It might be that licenses reverted. This is a thing that, I work on film and media as well. It's a thing that I don't see very much discussion about in library land, about how rights sometimes, due to contracts, will change. A very famous example of

that, that has caused me no limit of grief over the years is the famous Macintosh commercial, the 1984 Super Bowl ad, from Apple, where rights reverted to Ridley Scott from Apple, after a certain number of years. And it's been very difficult to get Ridley Scott's attention to let us deal with some things there.

So rights revert. Sometimes they're sub licenses, sometimes there might be, well, a good example of that would be music soundtracks. That's true of games software, for example, just as much as it might be for a television show. If any of you have seen the TV show, SC TV, Second City TV, the DVD of it, you'll notice there's blackouts on the DVD. And that's because some of the musical performances, they couldn't go back and revisit the rights on that, they had to black it out.

It can be the same with a game. An example of that would be Doom, where the musical rights, there was a sub license involved with that, that affected the distribution of the version of the game for which the source code was released. And that might be a reason that a company like Electronic Arts is not willing to assert their rights, because they're maybe not sure about music rights or something else that's underneath, or it might even be a piece of software that is within the software that they've distributed.

So there are lots of reasons that, it turns out, can orphan a piece of software, as far as permission goes, that you thought was unambiguous in terms of the ownership of copyright. I think, I may even have left out some other factors that came up, but I think those were the most common.

Jess Meyerson:

Thank you so much Henry. I want to open that up to our research team, Peter, Brandon, Krista and Pat to respond. And then due to time, I think we'll have to wrap it up after that. However, we do have a queue of two to three other questions that we will pull forward to episode three.

Brandon Butler:

So I'm about to jump through the screen if you didn't notice, and I want to get this out there before we get to the last thing. The great thing about fair use is that this is what it was meant to do is to solve this problem. If you are a startup that wants to cash in on the vintage gaming trend by rebooting Doom and selling it for the iPhone, then copyright makes you go and get permission, and that's good. You should, you're going to make a lot of money, you should go find whoever wrote that music and give them a piece of it. And if you can't find them, you can take it out and that's okay. And that's the way copyright works and that's good.

Copyright was never since 1790 supposed to discourage research and learning. And so when you get people like Henry having to go through this process, that is not what copyright intended to do. For the reasons Henry described, getting permission can be wonderful. It's not a waste of time if you think you're going to get it, it's great to get it.

But if you hit a brick wall, copyright is never supposed to be a thing that prevents research and teaching in this way. And fair use is the safety valve that lets you do this. The principles we've been describing today are the reasons, are the principles

that will let you do the things that, when you hit that orphan work brick wall, fair use saves the day. So go forth and fair use in peace.

Jess Meyerson:

That's excellent and what I'll do is I'll be sure to highlight that, that last portion, the actual like minute time, while Brandon says, and here's fair use, this is the problem that it solves.

That was a wonderful episode, as always, just a huge thanks to the entire research team. That's Brandon, Peter, Pat and Krista and warm thanks to our esteemed guests today, Henry and Leslie. Also, sincerest words of appreciation to each of our attendees today. Thank you so much for joining us.

And join us next week, same time, same place for episode three, Access Within Institutions and Across Networks. This will be featuring Jonathan Farbowitz of the Guggenheim Museum and Euan Cochrane of Yale University Libraries.

So next week's episode will be facilitated by Krista Cox from the Association of Research Libraries and Peter Jaszi of the Washington School of Law at American University. Thank you again to all of you and we look forward to next week.

Peter Jaszi:

Bye, everybody. Thank you.

Leslie Johnston:

Thank you.

Peter Jaszi:

And thanks, Henry and Leslie.

Leslie Johnston:

Thank you for including us, Brandon. You're welcome.

Henry Lowood:

You're welcome.