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

Timestamp [0:00]
Jessica Meyerson: I'm also the research program officer at a 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 brand and Leslie and Henry we will queue those up and they'll be ready for Q&A; we'll go in the order that they were presented so every 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 2 beginning the preservation workflow and this is a discussion with members of the coda 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 miss 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 here at her 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 combined interests in history technological innovation and the history of digital games and simulations had several long-term projects at Stanford including how they got game the history and culture of interactive simulations and video games in the Stanford communities 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 we've 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 his 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

Timestamp [3:47]

Brandon Butler: great thanks Jess so we're really excited to join you in the second episode in our 7-7 webinar series you know 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 code of best practices and fair use or software preservation and let me give you a little overview a roadmap of the talk so I'm gonna I'm going to start off by talking a little bit about principle one and then Peter Jaszi will talk a little bit about principle two and then we're going to it over to our guest Henry and lovely 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 the 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 so then we've 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 whatever is on your mind related to the code but do keep in mind we're only going to have Henry and lovely today so 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 you know follows this roughly chronologically a workflow of software preservation starting with you know a box of media on your desk and ending with you know the cyber utopia where everybody can read everything no matter how old it is right and so the first principle is getting from box of
stuff on your desk to software in your collection right or software that's part of your toolset for managing your collection as Leslie will talk about later and so principle one covers accessioning stabilizing evaluating and describing those digital objects that software objects and that includes the kind 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 right like Henry and Leslie in confidence you know and in small groups as well as individually about what they do and so these this process tries to mirror what people told us they do when they're - when they're trying to preserve software and so that includes things like making multiple disk 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 right when you're processing the collection you're figuring out what's in it you're getting it from unstable unmanaged to stable manage described it's ready to be a thing that you're caring for and so that's why I chose our little currents at computer pal over there as the image I know my colleague Lauren Work has one of those things and it costs more than my first car so sure you'd better let you use that thing because it's way too expensive to just sit there on your desk

each principal has a set of limitations that are associated with that principal and these this is the way that all of these codes work in the focus group discussions that we led in the in the process of developing the codes we proved 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 as part of working your way through a situation like this one and so you know 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 know you really need to have preservation activities that relate to your mission you know this was some thing that anyone you know the first thing that any group it you know within 10 minutes of our first focus group discussion this emerged as you know we 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 you know as you're conducting this activity you're protected by fair use but you know 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 continent that's sensitive for other record non copyright reasons this is part of being a good actor so the sort of 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 you know 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 description should be created expressed and shared to facilitate discovery within and where possible beyond the institution and this is another one that where we heard over and over again you know 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 were thinking about. we want to describe it in a way that it’s findable to people so that was an important part of you know stage one when you’re getting started consider your user and then finally at this stage and this is important thanking chronologically right for the purposes of principle one the people who are handling software for this purpose should be personnel including staff volunteers contractors so it’s not you know don’t we don’t need to see your badge but there needs to be an affiliation whether at the home institution are at a partner institution or a vendor who are directly engaged in this kind of activity you know 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 you know fairly like robust and interesting but not limiting in a way that I think going to constrain anyone from doing what they see is 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 -

Timestamp: [12:00]

Peter Jaszi: I do Brandon and I want to start by emphasizing something that that you have just said and then 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 and 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 online doesn't make any long-term sense and certainly not as a as Oh 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 in our code are really all about different varieties of access to preserve material and here yes indeed just a brief request is there any way that you could speak slightly closer to the mic okay perfect thank you so much what I had said a moment ago is that one of the first things we learned is that the preservation mission and the 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 ass of the access side of the 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 and operation or software and 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 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 very use question last time we talked a little bit about the ruling concept in contemporary u.s. fair use jurisprudence which is this idea of using something for a transformative purpose and this is a classic example obviously when I when I present documentation of the 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 meets 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 in our context is concerned is that the first limitation fair use transformative purpose are always easier to demonstrate when you are showing and telling more about the context of the thing that is being demonstrated won't always be possible to provide rich context but when it is it should be done the second Prince limiting limiting proposition here is really just a restatement of a general theory use concept that I explained a little bit last time everyone cares including the courts that when you use something without having expressed permission to do so the extent of your use should be commend appropriate proportional pick your word they all mean more or less the same thing to the purpose of the more the more extensive your goals the more documentation is justified and then finally there's the last limitation which which came up and and the number of our groups thought it was important to to include I have to say that this one is for the moment I fervently hope that that will change more 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 of 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 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 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 gonna see with respect to any of the fair use propositions that are contained in the code itself.

Transcript: [19:35]
Brandon hanks there we go thanks Peter [Music] 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 y'all to hear a little bit from Henry and locally and especially I think and you know they will correct me if I'm wrong and their illustrations will bear me out or not but I think they have a nice complementary aspect to their two use cases because Henry is really someone who's collecting software for software sake and lovely is collecting digital documents and she needs that software to make sense of the documents as she'll tell you more about later and those were really in a way two mega overarching use cases that we heard about over and over again and so starting with Henry I'm really excited to hear from you all about what it's really like to do this stuff

Timestamp: [20:25]

Henry okay um 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 gonna focus I think mostly on situation one about accessioning, stabilizing, evaluating and describing digital objects specifically around a project to create disk images from original media and in doing that I'm gonna focus on the second limitation the limitation be 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 but of course donor agreements can also I'll argue help you with access it can augment access in some ways a use case again will be this the Cabrinety collection and in particular project one of the projects we carried out with the cabinet collection with the National Institute for Standards and Technology specifically the what's called the NS RL the national software reference library run by Doug white which I described as the National Forensic Software forensics laboratory now as for situation to documenting software and operation I'm not going to say too much about that directly even though it pains me greatly not to it's that 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 operation operating software of the past I'm gonna leave it there if you have questions if you want to talk about it in Q&A; well I can certainly do that so back to the Cabrinety collection the full name of which is the Steven M Cabrienty Attic Collection in the History of Microcomputing 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 collect collecting and so again the cabin attic 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 now I'll be talking about when I say software I'm talking about packaged PC so where productivity software game software education all those sorts of things I'm not talking about mainframe or bundled software's it used to be called unpublished software scientific or research software there's academic software if you will I'm not talking about non PC software newer things like mobile you know 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 cabinet attic 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 gonna say if you have questions about that those categories of software some some of which I definitely have worked on feel free to ask later so in terms of workflow we've been working on the cabernet collection now for over 20 years as I've said this works largely been carried out through a string of funded projects some of which Jessica mentioned in the introduction included the NIST/Cabrinety capture project which I'll be talking about the game citation project also funded by IMLS and now the easy project notice in that sequence started with sort of a just a quick can you acquire project followed by a capture migration project followed by a description project followed by an access project in here we are 20 years later these have all been multi institutional projects including Stanford so back to the point about fair use in the specific point that we're materials have been donated their preservation should be undertaken in light of the term 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 a of a complicated game and the players in that game include copyright law the Digital Millennium Copyright Act and so forth the very various provisions and things that their 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 in different and in different way 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 game of a complicated game of rock-paper-scissors figuring out that sometimes you know fair use maybe beats copyright law here well maybe in another occasion a donor agreement would trump fair use you know and so on and so forth so it's quite complicated could 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 a deed of gift of course the acquisitions process in my area of curatorial practice you know it's sometimes I do buy individual software titles we do have a media center where we do that sort of thing but with the historical software it's been mostly around collecting collections acquiring collections and these have mostly been gifts beginning with this
instrument called the deed of gift by the way the cabinet 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 a 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's 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 you know we offer 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 copyright stanford granting stanford a license which you'll see on the net 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 in the materials that they're giving to you you know our other IP rights potentially like patent rights which has come up on occasion in the case of the Cabrinety collection this did apply to a portion of the collection we were given exactly three titles out of the circle fifteen thousand titles in the collection were a copyright to Stephen Cabrinety any fourteen thousand nine hundred and ninety-seven let's say we're not so this portion of the agreement only applied to those three titles and to the 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 you know this is generally not the case as I just said and of course in case the cabin Cabrinety Collection this option was not stated I was not available at all because well you know 1998 right 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 gift 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 sale as well as in a gift to transfer copyright and there's even the option of acquiring copyright straight away say for some for a collection that you already own and we've done that on occasion I just I just want to put those on the table wow that was amazing that was a mind-read slide advance okay that's fine that's that's where we want to be so we did this project with NIST which created a big collection of disk images and just as Brandon said sometimes when you're doing preservation you need to be thinking about what's gonna happen down the road in fact you probably won't even do the preservation if you're not thinking about what's gonna 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 disk images then that could be stored in the Stanford digital repository and theoretically could be seen and downloaded
by researchers in addition to the disk images we also created photographic images that's 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 you know 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 Association of Research Libraries (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 Activision's Microsoft and so forth we didn't go very far down the tail we still plan to do that but you would there are lectures you can imagine what the collection of software from the 1978 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 now 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 him that that question as you'll see in a second will we provided him 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 and enable us to proceed without ambiguity on the next slide so here's what we sent we sent something like this to every rights holder recontact 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 disk or on the box or something inside the box says copyright SSG we're contacting about these titles and we first of all we said 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 weblog download and all of that sort of thing research use only would be some sort of access with no permission to redistribute or copy such as no downloads and then restricted research act research use only is basically would also include reading room access I should mention restricted research used was if you've got some something that you've got a problem with let's talk about it and figure out a special case here and okay next slide please oh it's all conclude here I 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 - 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 ten titles there or say a Microsoft reenact
division to whom he might have had 200 or 300 titles typically we received back
confirmation that they owned felt they own 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 QA secondly we were 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% do we
for which we have unrestricted access to the disc images however you know 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 and about we're less concerned about restricting
access to them then they are about restricting access to the software itself which and I'll
conclude on this last sentence kind of circles back to the second 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 ok I'll stop there and hand over to Leslie

Transcript: [38:31]
Brandon Butler: awesome thanks Henry and I'll just I'm switching over to Leslie slides here

Transcript: [38:36]
Leslie Johnston: Thank You Brandon

Transcript: [38:42]
Brandon Butler: all right I'll go

Transcript: [38:43]
Leslie Johnston: alright 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 different types of
code that we have gotten from agencies some of it Java that's actually the largest category
of you know 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 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 heard 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 the context for 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 you know 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 verses 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 in to some interesting questions about not only the records but I swear I'm getting to it the software because it can be they could hold on to it for five years 50 years or as we heard a couple of weeks ago five hundred 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 five hundred years plus 20 because they add a plus 20 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 my great files over time because I will say that federal records managers that category of position and the federal government has to be one of the most underfunded and under staffed areas of the federal government and there being
such a variety and vintage of formats as well as the software in operating environments we need older software packages to be able to validate process describe and migrate the files that we have in our holdings now we'll have into the future and we'll get in the future next slide so the workflows for bringing in code or any type of boring digital record is the same we have a single workflow for accession impressing ingest of boring 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 you know project management records so that we can actually confirm that what they're sending is what we expect to receive and that what we received is what they claimed they were sending so our workflows are not unlike any other digital accession 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 uncompelled if we get code we want it to come to us uncompelled and the confirmation that the intellectual content meets the requirements of the record schedule if they send us they in these transfers then 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 us without encryption so things must be uncompressed and unencrypted when they come to us and that includes 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 you know XML JSON we expect for some sort of documentation of the data schema or the markup schema 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 received 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 personally identifiable information 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 enter with them and from our point of view even though our records in 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 you know 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 generate 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 you know explicitly allows narrative 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 but as requested and we believe that our activities are in line with the code of best practices next slide but 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 2 in the code of 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 lysis 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 cover it 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 can have a discussion about this

Timestamp: [53:00]

Jessica 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 pace 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?”

Timestamp: [53:58]

Henry Lowood: Well the image of the carrier format would probably be analogous to what what 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 similar argument could apply although as we know in the 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 that a that 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 so and in the era were talking about of that 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 that had in particular that had an even productivity software that quite lengthy manuals and so certainly those would be covered under copyright and we wanted to have clearance on them

Timestamp: [55:23]

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 doctrine to do so so the the question in the end and the project of today are very closely related

Timestamp: [55:50]

Jessica Meyerson: perfect so that's that's our first question we also have this is also a follow-on from Henry's presentation and so 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 hardenking 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

Timestamp: [56:40]

Henry Lowood: ok I'll just go ahead and then you know 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 he can imagine somebody gives their collection of magazines or software to the library they generally will own copyright and none of it and in the even in the case of their papers very often they'll they're donating things that they don't want copyright to in the case of the publishers that's the more interesting thing the publishers we contacted for whom you know 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 certain rights on about half the titles or two-thirds or whatever the number would be why is that a bunch of different reasons and keep in mind here that we're talking about software that's at least in the youngest case is 25 years old okay in the cabernet t collection but that's not 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 license is 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 know the limit of grief over the years is the famous Macintosh commercial the 1984 Superbowl ad from Apple where the 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 their sub licenses
sometimes there might be well a good example of that would be music soundtracks that's
true of games game software for example just as as much much as it might be for a
television show if any of you have seen the TV show at CTV second city TV the DVD of it
you'll notice there's blackouts in in the on the DVD and that's because some of the musical
performances could not they couldn't go back and revisit the rights on that mad 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 game 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 that about music rights or something else that's underneath or it
might even be a piece of software that is within the software that 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 its in terms of the
ownership of copyright and and that I think I may have even left out some other factors
that came up but I think that that those were the most common

Timestamp: [1:00:21]
Jessica Meyerson: and 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

Timestamp: [1:00:37]
Brandon Butler: so I'm about to jump through the screen if you don't if you didn't notice
and I want to get this out there before we get to the last the great thing about fair use is
that this is what it's this is what it was meant to do is to solve this problem if you're you
know if you are a startup that wants to cash in on the vintage gaming trend by rebooting
doom and selling it up for the iPhone then copyright makes you go and get permission and
that's good 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 you know it's for the reasons Henry described getting permission can be wonderful
it's not a waste of time if you think you're gonna 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 spare you saves the day so go forth and fair use

Timestamp: [1:02:12]
Jessica Meyerson: 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 solved so 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 warm thanks to our esteemed guest 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 bye everybody