

Configuration Workflows

EaaSII Webinar Series

October 2019

Interviewer: Welcome everyone. Thank you for joining us for today's webinar. My name is Seth Anderson and I'm the software preservation program manager at Yale University. Yale University is the host of the EaaSII Program of Work. I am filling in today as the host since Ethan, our usual webinar host, is participating today. We are excited to hear from him and his work configurations. Today we are presenting EaaSII Configuration Workflows, a round table discussion with Lauren Work, Zachary Firsty, and Ethan Gates. Just a little housekeeping before we get started. If you have any questions during the presentation, please type them into the chat box in your zoom control panel, and I will bring them up during the presentation or we'll have some time for your questions at the end. Also, please mute yourself and turn off your video to maximize the quality of our recording. This webinar recording as well as the recordings from all of the previous webinars will be made available on the EaaSII Website with transcripts as soon as possible. Now, let's introduce our guest speakers and members of the EaaSII Network.

First, we have Lauren Work who is the digital preservation librarian at the University of Virginia where she is responsible for the implementations of preservation strategy and systems for university digital resources. She is also the project lead for the fostering a community of practice, FCOP Grant at UVA. The digital preservation strategist for UVA's EaaSII Node and a member of the Sphin Research Work-in Group.

Then we have Zachary Firsty from Carnegie Mellon University who has received a PhD in film and visual studies from Harvard University. As a clear post-doctoral fellow in software curation for the sciences and social sciences, he collaborates with faculty, staff, and students to develop strategies for preserving and curating executable archival materials such as research software, digital scholarship, and code-based artworks. He plays a key role in developing software preservation and curation workflows policies and strategies in universities libraries. Previously Firsty held a Mellon post-doctoral scholarship in digital humanities at the University of South California.

Ethan Gates is our software preservation and analyst at Yale University library where he coordinates the creation and documentation of emulated computer environments for the EaaSII network as well as the development of training materials.

Sorry just going to mute a couple of lines here. Where was I? Ethan leads the creation documentation of emulated computer environments for the EaaSII Network as well as the development of training materials for EaaSII users and the digital preservation community. Ethan is a graduate of NYU's masters program in moving image and archiving and preservation and has worked with organizations including Metro Transfer Collective and the Association of Moving Image Archivists Open Source Committee to develop workshops and resources that demystify digital technology for archivist. So, without further ado it gives me pleasure to hand it

Configuration Workflows

EaaSII Webinar Series

October 2019

over to Ethan for an overview of EaaSII configuration workflows that we implemented here at Yale University.

Respondent: Wonderful. Thanks Seth and I'm just going to hand it, can you all see slides at this point?

Interviewer: Yes.

Respondent: Someone confirmed that. Thank you very much. Great to have Seth. I'm going to, as we have been doing with all of these webinars give an overview of the topic from the EaaSII staff perspective. So today, since we are talking about configuration workflows and basically actually filling the EaaSII network with content with emulated environments and how we do that. I'm just basically going to go over what a workflow looks like here at Yale University Library, the host for the EaaSII Program to give an idea of how we're using Yale's legacy software's collections and moving them from, into the EaaSII network and what the process looks like.

So, to sort of give an overview, there's a foundational principle here of the emulation as a service platform and the EaaSII Program of Work in the reduction of variability and redundancy. When we're creating emulating environments for the EaaSII network, it's really important for us to start from a base of well-known, well described, stable environments that are going to be broadly applicable across many domains and user communities. If we start from that point in our creation of emulated environments the way that the derivative system and the emulation of the service platform works, that allows EaaSII users to then go off and craft their more specific, custom environments according to their needs and their software collections without everyone needing to start from scratch of putting together hardware settings and following an operating system every single time that they need emulation for an accessory purpose.

What that looks like in practice is to first create environments based essentially on major operating system releases. One of our, sort of looking back through the whole history of software and of computer systems, we have to narrow it down to okay what are we actually going to target? What are we actually going to put into the system? Where do we start from? So, you can see some of the work it took to start charting out not just what major operating release to target, but even possibly appropriate updates and the hardware settings to match up with those releases when we consider what is even based environment. What do we want to start from? In the case of say Windows 95 for instance, later updates or service packs for that operating system might be actually a higher target than the initial release because they allowed for USB support or bigger hard drive sizes. Features that give us more flexibility down the line installing more software or adding collection content to a more broadly applicable range of cases. Excuse me, just struggling

Configuration Workflows

EaaSII Webinar Series

October 2019

with my google a little bit. Meanwhile, there's parallel work on prioritization that has to get done in Yale's collection of legacy software.

Okay, once we have that base operating system, those base environments in place, what's the next step? What are the actual applications that we want to start loading in the environment that will be of use to a number of users right off the bat? It's a little hard to pull data on the number of unique titles in Yale's software collection but we can certainly say tens of thousands of applications are available for us to start working with. Again, one of the reasons we're starting this work hosting at Yale. We're given this perhaps giant CSC, giant list of applications that you can start with covering tens of thousands personal applications starting from the era of the late eighties up through the twenty tens. Where do you start with that? So, we had to do some prioritization work first through a public survey that we put out in February which was based on basically what are the sort of popular software genres that we could work with that people would be interested in seeing EaaSII working with. That returned mostly some predictable results. I would say given the software preservation network and digital preservation community it was marketed to interested in Cad, architectural software, office productivity, etc. Web archiving is clearly on everyone's mind because there was a lot of interest in web browsers. But also, the prioritization also came out of a lot of internal conversations about what applications would be useful for the other areas of EaaSII's Programs of Work. Generating sample data for the metadata model that Seth described last month or for the work that Kat is doing with wikidata and wikiBP.

What would be data that or applications that we should work with that would help prove use case for what EaaSII is doing? The universal interactor, the API work being done by Klaus and Olega in UN on taking a file and matching it up with file format capabilities and software hiding recommendations for emulated environments. What would, again, provide good demonstrations or use cases to prove the value of that work? And then also, of course we're always open to request from the nodes that we're working with or internal stake holders at Yale in terms of just what are individual applications that we should be move forward with.

In terms of actually taking all of this priority and target work and applying it to configuration, I handle most of the base creation. That is like the work of those foundational operating system releases, places of starting work. I do this mostly outside, actually, of emulation as a service. Outside of the EaaSII platform in local versions of QMU or other relevant underlying emulators just for right now while there is ongoing development work in the back end and the front end of the EaaSII platform while we're essentially in the beta release still of the network. Because working locally allows me to do some thorough trouble shooting and documentation. There are environments that I know need to be stable and I need to just know their features and their configuration back and front so that when I'm helping other users at other nodes to troubleshoot

Configuration Workflows

EaaSII Webinar Series

October 2019

environments, I can say very confidently what is perhaps inherent to the disconese or inherent to that the environment and what is not. Every base environment that I go through is subjected to a checklist of certain baseline guarantee features to match up with the EAS platform. That includes TCP, IP networking capability, postscript printing, CD rom drivers if appropriate. Exceptions always made if it's not appropriate given the timeline of the operating system of course. At the end each base is created and packaged up and backed up locally here at Yale with screenshots of the installation process so that at any point later on in our EaaSII platform development as that work is ongoing we can always return to the stable disc images no matter what and reset if necessary. On the software front, items are then, individual software applications are either uploaded manually through the EaaSII demo administration UI. As you can see here, which is generally from our one off cases or boutique needs, like. Oh, are you not seeing my slides advancing Seth?

Interviewer: No.

Respondent: Not at all? Are we still stuck on...

Interviewer: Yeah, you're on slide one still.

Respondent: Still on slide one?

Interviewer: Yeah.

Respondent: Okay. Sorry. That is a problem. Sorry folks. How about now, do you see slide movement or advancement?

Interviewer: Yes.

Respondent: Moving around. Okay. Sorry about that. I didn't realize that you weren't seeing any of that work. So, here's some of the base environment creation spreadsheets. Here's our legacy software collection sorting through those thousands of applications that I mentioned. Here is an idea of the base configuration work that I do locally in terms of making notes, going through checklist, packaging up disc images, and this brings us back to upload of individual software applications into the EaaSII platform demonstration UI. Again, this is for more one off boutique needs so this might be like crafting together the exact stack file requirements to recreate say a mid two thousands archivist tool kit instance with my SQL demonstration, SQL database. Or for the bulk of our configuration work we actually use a python script that was written by one of our student employees her at Yale to take existing metadata from the legacy software collection and turn it into easy mets xml records that will be actually make that item usable on the EaaSII back end. This allows us to batch import dozens of items potentially at a time that I can then handoff to that team of student workers for catalog and configuration. Again, this is the point at which I

Configuration Workflows

EaaSII Webinar Series

October 2019

actually handoff to our student employees the work of, excuse me, I do the work of actually handing off the cataloging of items to our student employees which varies from semester to semester; the size of our team. At times we've had two employees like over the summer. At times we have nine coming in and out. I'd say during the academic year we basically get the equivalent of one full time employee altogether out of their hours. They basically work in spreadsheets using the metadata model that Seth, again, described in more detail last month for the kind of information we're looking up for about these applications to drive our system and capturing as much of that information as possible about each item. That includes release states for applications, operating system requirements, language settings, file format capabilities, all of those. Each column here is going to drive some piece of functionality in the EaaSII platform ultimately whether it's discoverability in the new browse and search interfaces of the UI work that is being done by portal media, or it might drive recommendations crafted by the universal interactor. Again, each one of these has a purpose, will get mapped into our database and into our system. Gathering this information just takes a lot of back and forth between that spreadsheet you were just seeing and running the title and emulation on the EaaSII platform. So, for our student workers this takes a lot of time not just learning our platform which they've certainly never seen before, but just how to interact with legacy computer environments that, again given we're talking about frequently undergraduate college students, legacy operating systems and environments and applications that they've probably never encountered before. There's also a lot of variability in that information in that spreadsheet where they might find that. While there have certainly been certain patterns in the history of software design, publish date or manufacturer listed on an about page, there's no guarantee from one item to the next that that's going to be in the same place or even present at all. So, it takes a lot of flexibility and curiosity, critical thinking really to get to the level of that thorough documentation that we're asking for. I know we're running, I want to get to our discussion and not take up too much of our time here so just quickly summarizing the challenges of this configuration workflow. Training is an issue, especially with student turnover from one semester to the next. How do you very quickly get students who may not have had a lot of experience cataloging or with using legacy software to understand what we're doing and how to do it quickly? There's a constant balance in that prioritization work and configuration between sort of balancing bulk and boutique work between how do we upload and create a lot of environments that are useful to a lot of people while also addressing specific use cases and very custom code or extremely far down the line, in terms of lot system requirements in very specific version application requirements. There's some challenging quality assurance and control both between our platform and some of the technical imitations during the beta. And otherwise, in terms of what does it mean to say that environment is configured or has been cataloged and should be published now to the network. It's easy to say that an environment is created and done when a software application is installed but does that mean that every single

Configuration Workflows

EaaS Webinar Series

October 2019

feature contained within that application works and is there a way we can test that? There's a lot of exploration still to be done in making that aspect of our work more efficient. Then I say there's going to be troubles with advanced configurations. The ESA platform bundles together a lot of underlying open source emulators and while the vast majority of software and use cases that we're working with currently are covered by a very small number of emulators and template settings for those emulators, ninety to ninety five percent of our work so far has been intel compatible PC emulation offered by QEMU. The further we go and the further we look into not just creating these environments but optimizing them for particular use cases. When we're not just talking about cranking out every version of Microsoft office ever, but really looking at custom code written for the Amiga by an artist in an interactive exhibition in a gallery. The more we look at EaaS for those kinds of very particular use cases, the more it's going to require learning not just how EaaS works and moving pieces together on the EaaS platform, but how these underlying emulators work and how to troubleshoot between those different layers of the emulator code, our code, the legacy code, and figure out what the best possible experience is for a curator or a researcher or a patron. That's my summary. I apologize for the slight glitch in the slides there and I'm sure we'll make these available later on but I want to wrap this up now so we can start hearing from configuration work at the nodes as well. Hear how this is going there.

Interviewer: Thank you Ethan. That was great and yes, we will I'm sure post these slides along with the transcript so you can get a closer glimpse at all of the spreadsheets that Ethan has been putting together over the last year. So, yeah, as someone said earlier this week at a conference, it's turtles all the way down for software application. Layers upon layers of complexity. With that, we will now transition to our guest speakers and dive into discussion. Zach and Lauren, hope you're ready. Alright. And just a reminder, if you have any questions that come up please feel free to put them in the chat box. I'm just going to alert everyone where that is by sending a message now. Alright, so on to moderate discussion with Ethan and Lauren and Zach, and hopefully this will ground the overview and the practice and context of other institutions besides Yale. First, Lauren, can you maybe give us some background on what your previous experience in using emulation software was prior to EaaS FCOP. Were there any emulation programs that you had already incorporated into your institutional workflows which predate EaaS?

Respondent: Sure. And I apologize if my connection is a little wonky. I may have to turn off the video. I'm at home so apologies ahead of time for that. To answer your question. Maybe this is heartening to folks on the call, but I actually didn't have a ton of emulation experience. I'm a digital preservation librarian. My role covers a lot of things. I think like a lot of folks, I had messed around with emulation a little bit in courses both in school and then in my role at UVA. Was fortunate enough to take some things. So, I messed around with things like SheepShaver you know, but never really had a problem set distinct enough to say I need to get this

Configuration Workflows

EaaS Webinar Series

October 2019

implemented and running on my own. So, it's something that I had had some experience with but not a lot prior to something like EaaS which made it a lot easier to scale and test without having to necessarily build things out on my own which was great or support things on my own. Or even with Doves, with things like UVA. I know that's been a significant factor moving forward. As far as emulation programs incorporated into institutional workflows, we hadn't had anything yet. There are some small things like using Amazon App streams for things like AON if any other folks are using that. Other than that, we haven't really implemented things that scaled, so that's to start. I'll throw things over to Zach who may have a very different experience with this but we'll do that.

Interviewer: Yeah Zach. What was your previous experience with emulation prior to starting work on EaaS?

Respondent: It was actually pretty similar to Lauren in the sense that I didn't have any really formal context in which to use emulation, but I had as kind of an inveterate tinkerer and sort of techno explorer, I had certainly booted up virtual machines mainly times before. Kind of toyed around often. It was oh hey there's this windows program and I only have a Mac ready. How could I run a Windows on my Mac? Something that I was thinking about a little bit in preparation for this discussion was that thinking back on these experiences there was a severe limit. I remember I've started a lot of virtual machines but they end up taking up so much of my hard drive, so much of my resources that it's not the kind of thing that, you know, during the triage of oh no I need to free up some space on the hard drive. That big forty giga bytes of Windows XP is going to be the first one to go. Certainly, a way in which I was really excited to learn about EaaS. I think one way that offers a lot to people that do want to work with emulation. In terms of the, just very quickly, the emulation work flows at the institution at Carnegie Mellon I'm not aware of any other, at the libraries at least, any other use of emulation outside of EaaS. There's some work with containerization and some new projects using Doc to containerize scholarly work and digital projects but not emulation.

Interviewer: Right. I would point out before I move on to the next question. Carnegie Mellon is at an interesting and I think unique position in they have previously attempted a system similar to EaaS so the Olive Project which has been dormant for a little while, but I know that has at least informed the continuing development of emulation services at Carnegie Mellon that EaaS is now sort of the bases for.

Respondent: Yeah that's right. It's useful to note on an institutional level that that's a project really run, or it was. Olive was something run by a couple of computer science, one computer science professor in particular. I think a lot of the work of EaaS and bringing that to the wider

Configuration Workflows

EaaS Webinar Series

October 2019

campus has been figuring out that interface between a library and the school of computer science. Especially with that history.

Interviewer: Great, thank you. So, Lauren or Zach, I know Ethan mentioned that we have many student workers who are creating environments here. We have yet to see too much activity from other users but we haven't really opened up the ability to configure environments too much. But have you seen other users in your node whether that's library staff or anyone really at the university who have created environments? Let us know what did they create, what challenges may they have faced, and did you actually spend any time working with them on that?

Respondent: I'll go ahead and go first again if that's okay Zach because my answer is pretty short. So, at our node that UVA currently, it's myself, and Christian Dopleson who's our node developer who's done a lot of work on actually setting up EaaS who have mostly worked in creating environments. So, we haven't really done a lot of testing out with other users. However, through our kind of dual work on the FCOP project, we've had a lot starting to work within the environment. This is with some preloaded things that I've loaded in, setting up environments, putting software in particular operating environments, and things like that. So, we have had people start to even get in, get used to the environment, understand how the menu and navigation and those things work. I think that's an important kind of grounding concept to all of this as well is the introduction to how emulation works. So, we've been working on that as well at UVA through as many outreach things as possible. But in short, nothing really to report yet in terms of setting up and creating environments, but I do have some comments for kind of all the work that goes into preparing everything that will go into an environment and I can talk about that a little bit. For right now, not a lot to report yet on things like the computing concepts, but things like the interface, that has taken some explanation and I have to say the documentation that you all have created, we've been leveraging that. So that's a really good resource to kind of highlight on the call. Also, use that to help build out our internal workflows for things that maybe different from environments we're creating as well as really thinking through. So, we're working with some particular software right now for the FCOP project that is a Cad bim software that is nuance like a lot of things and really thinking through what we're including in environments to allow people to use those more easily. So, whether that's a readme about the nuances of some of the work that's done. I'm sure you all have already really thought closely about these things, but all those kind of granular things that come up as you actually start working in the environment and how you could better leverage and scale peoples understanding who aren't going to come to all of these environments with any inherent knowledge. Even if they've been working in computers for a while. So, I'll leave it at that for now and toss it over to Zach.

Configuration Workflows

EaaS Webinar Series

October 2019

Interviewer: Well I'll follow up really quick before we turn it over to Zach. I'm wondering since you have been one of the primary users at UVA, what have you found was the I guess surprise or challenge? What's something that you didn't expect to have to learn or what skills did you need to improve?

Respondent: I think something that Ethan kind of highlighted a little. So not necessarily anything within the environment and I think that speaks to how well Klaus and you all have done to make it usable and functional and have documentation. But some of the things around licensing and dependencies for softwares. So, getting licensed keys, getting software running, making sure that those dependencies are well known ahead of time, that was something for me that was new. You had mentioned earlier of those be interested in web archiving; I'm really interested around the roles of documentation for software. So, the example that we have right now that we're working on is a commercial software. There's a ton of user boards and things like that online that actually helped me navigate through understanding how to actually get some things to run. I see Ethan nodding here so that's something I've been thinking about. We also have a web archiving program here. Just trying to connect all these dots between how do we actually understand how to obviously technically make things run, but to have users really understand how to make and use the projects and collections we're taking all this effort for and understand how to use them in the future. That's something for me that was kind of elevated in this whole process.

Interviewer: Great. Thank you. Zach, just a reminder so the question was are there other users at your node or at the institution who are creating environments and if you have any experience you can share or feedback they provided as far as challenges or skills they had to learn. Curious to hear about those.

Respondent: Sure. So, similarly to Lauren, while we don't have other users creating environments, what we do have and it just really began this week actually, there's a group in the computational design at the College of Fine Arts at Carnegie Mellon that is doing a course on the history of computerized design. We've been working with the professor in the school of Architecture there to get a bunch of old Cad environments basically, up and running in EaaS. This week on Monday I'm hopefully going to be hearing back tomorrow on a call about the experience the students had poking around in EaaS. We let them log right in and open up the different environments. There are some in Windows 98, Windows XP, Windows 3.1, and DOS. So, it should be really interesting to hear from a perspective. I would also say that I kind of just joined a couple a months ago so in some ways I'm jumping into the EaaS workflow. Kind of personally reflecting about sort of the biggest learning curve and also as I explain what EaaS is to faculty and other collaborators, I do think this terminology, what's so interesting about EaaS that turtles all the way down the emulations stacked on virtual machine on an instance on a

Configuration Workflows

EaaS Webinar Series

October 2019

server. I think this kind of slippage of terms, when we talk about the server that runs the instances that have containers that contain virtual machines that can spin environments can be a bit dizzying. I think you know it's easy to say the wrong thing. To talk about an instance or a server when you mean not the emulation. Just kind of clarifying some of those things. I would second Lauren's kiddos on the documentation which does have some useful graphics about this and kind of breaking it down. I'm not sure really what kind of feedback this is. I think it's really inherent to the issue of emulation so it's not necessarily something that we need to work on but that has been I would say big bit of work. Breaking down those different interrelated kinds of nested concepts that are invoked by EaaS.

Interviewer: Right. Go ahead Ethan.

Respondent: I'm wondering if I can jump in at this point because both of what you're talking about is bringing up a question I think that I get a lot which is that question of training and how do you get students to wrap their heads around these concepts and what resources are most useful. We're still sort of figuring that out because I think as both Lauren and Zach, you point to there's this dizzying amount of possible information to wrap your head around both in terms of using emulators and all these possible applications. People are surprised that a lot of this just boils down to internet searching as Lauren sort of pointed out and I think that the web archiving and how we bring in and consolidate all of that documentation. That does exist out there and that would be really exhausting to recreate ourselves how we can leverage what already exists that's really interesting. I guess I'm interested in hearing from both Zach and Lauren. Are there any particular places that you go to off the bat for sort of explaining EaaS and emulation to first time users?

Respondent: I'll say that my experience has been talking about emulation, the first reference point that many people have. And is probably the first way that I really thought about it too is emulating old gaming systems. There's something about that that is a really useful kind of shoristic to say oh yeah you know how you can run that super NES on whatever, on your computer or something like that. It helps people kind of think about oh so you're kind of virtually assembling the component parts of that legacy system and that's what you do. I think that's pretty easily extrapolated. So, I would say that as kind of an example to explain what's going on. There really are, like we're all saying, this tremendous resource. These kind of hobbyist communities about emulation and legacy systems. There are various boards. I've been recently spending a lot of time in the world of enaculation which is you know the emulation of old MacIntosh environments which has a really robust community instead of documentation and stuff like that. There are many others. Sometimes I've noticed the trick can be, actually the search term indexing like Google does not believe that you are really searching for a certain Windows

Configuration Workflows

EaaS Webinar Series

October 2019

95 driver. Their like there's no way. You must be looking for Windows 7. So that's kind of an interesting secondary feature. You have to kind of fight against search engine optimization sometimes to work with this legacy stuff.

Interviewer: Yeah. That's an interesting, yeah. I do notice that you often get indexed support sites from the developers that are actually get 404's when you actually click through. So, it's always an interesting...

Respondent: Then you go to the wayback machine.

Interviewer: Right, well...

Respondent: On the internet archive and you see if that manual or help page from 2004 ended up in there. Sometimes it did, sometimes it didn't. That is an interesting thing of could we borrow off them of that work that's been done. Again, it exists but can we bring it together and how do we present it to a user in a useful way to Laurens point.

Interviewer: Yeah. Moving along as it's about time to open it up to audience Q and A. Maybe these are short answers given that we're still early in rolling out the service and everything. Zach or Lauren, how formal at this point are your workflows for configuration? Have you started documenting any of this work, creating manuals, putting together guidelines, etc.?

Respondent: Yeah. One thing I wanted to highlight just going back to what Ethan was talking about at the very beginning. This idea of well known and well described, that comes into kind of the workflows that we're establishing for EaaS. Some of this right now is centered around some of our FCOP work. So, one thing I wanted to say to folks because I know there's a lot and people have questions, I was going to drop as part of the FCOP project that we did at UVA, we had a workshop where we pretty much published everything including some of our in progress work flows and OSF's. I'm going to drop that in the chat so people can reference that if they want or if you've come and attended, hi again. It was a great way to talk through some of the work that we're doing so far. So, I just dropped it in the chat there. The one thing I wanted to highlight that I think folks have to think about and the way that we've tried to start highlighting this infrastructurally, thinking about leveraging existing workflows. Thinking about how people can understand emulation and actually understand how it may come into their day to day workflows across the library. So, I'm in an academic institution. We've been using things like the best practice for fair use to do things like update our (inaudible), to do things like create checklists and questionnaires around soliciting questions and getting archivist to start asking questions about software and software dependent materials. That's something that we have started to work into you know workflows and how that actually is going to go. I wanted to highlight, I was reviewing the EaaS documentation that you all had written up about publishing and replicating

Configuration Workflows

EaaS Webinar Series

October 2019

resources and you have great questions in here about what could or should you be publishing. Those questions should be answered up the line. So, we're really trying to get those configured into our workflows as soon as possible because it's going to help everything in the future. Same goes for digital preservation and processing workflows. Do we know what we have? One thing I wanted to highlight was that through this work, I connected with Nick Cravenhoft. I saw that we had a question from the folks at CCA about things that have come around Cad bim software. Factor works which is the name of the software that we're working with specifically right now is our one use case for FCOP and it relates to EaaS in the fact that pronum has some signatures but now all so again this is an opportunity to take work that can benefit the community as a whole. Nick and I both have some of these file types in our collections and there's some questions about what's being identified and how we can best have the signature be accurate. There are some distinctions in the header between identifications of file type is doctor works. Doctor works was this saying that people have used to migrate file formats. Sometimes there's two potential software versions that are highlighted there. Things like that. these granular things that come up way down the line that eventually will allow us to say yes render this, yes contribute this to wikidata to update this, contribute this to the signature and pronouns so that everybody running it, whether it's an archive maca or other ways that they're doing file format. Identification can better and more accurately describe what you are trying to preserve and potentially emulate in the first place. All of these things. Again, that (inaudible) like the donor checklist for questions. Even start asking people do you have licenses? Do you have documentation and manuals for these things that you are potentially giving to us? Things like leveraging existing metadata. Again, our case was commercial software so we were like let's write a very quick way of maybe if there's a mark metadata about this in the system. See if we want to transfer this into something we can leverage to start. Again, not using mark but leveraging existing work and data. I think I'll cut it off there so we have time and Zach can chip in too. That's where we are right now.

Interviewer: Thank you.

Respondent: About the documentation for workflows, its been something that we've been in the process of working through. We did drop as a group though before individuals within the libraries have been working on the EaaS project. I put together some documentation about imaging and basically adding images to our collection so that they can then be added as software within EaaS. Then we realized we were having trouble actually using some of these images so I had a big stack of a couple of dozen expert witness format, to get technical, files that contain the images of various discs then we weren't able to mount them. Something that I noticed was kind of these competing tensions where, in my background as someone who's just often doing kind of a DIY sort of artistic project, I just want to pack something together that will work given the constraints of the current system. So my first impulse is oh I'll just have to write a script that

Configuration Workflows

EaaS Webinar Series

October 2019

converts these to a format that is accepted, but then stepping back a little bit and thinking about the broader project, noticing that's intentional a little bit with kind of building up under the robust systematic and spending that time working on documentation and writing out the big reporting. Saying we expected this, here's the steps that we went through and here's the unexpected results. That is a note on documentation. I definitely second what Lauren was saying about kind of importing, at least conceptually importing, existing workflows that's extinct concepts whether they're metadata or other kinds of things into the EaaS project. Are we talking already about the questions in the chat?

Interviewer: Sorry, it spilled over. I want to come back because I do want to focus on Stefana's questions about Cad bim in detail since it's been a big use case for us. Before we do that, just really quickly maybe we can get yes or no answers to this one. Have you thought about QC criteria yet? Ethan have you, maybe you can fill this one in a little bit. What are some of the methods you are thinking about at this point that you might pursue?

Respondent: Thought about it quite a bit. It is what guys that thinking a lot it becomes the QC criteria for each environment becomes so specific depending on what the intended use of that environment is. Whether it is just broadly like this environment is just supposed to have Adobe Reader 4.0 in it and that works, the QC criteria for it opens PDS relatively simply, but QC criteria for a more advanced Cad application or even something else might be much more complicated. The thinking is how do we scale that thinking at that moment and what are the platforms that would be necessary to allow for that variability and even establishing the criteria. It's been interesting because I think it has involved maybe stepping back. As Zach was even just saying with imaging workflows, from what I expected in terms of where we might be of oh it'd be easy to with a base environment creation checklist. That was relatively easy because there are certain standards, study features that have to work with emulation as a service that have to be offered by the operating system that are known that it was not that difficult to come up with. Again, thinking that that could maybe step to software environments, it's had to actually be a step back to how do we build a system that allows for people to create their own criteria basically. For that to vary from one institution to the next and one collection to the next, one item to the next.

Interviewer: Right. I think the challenge as well is to be explicit about what the baseline is for quality control because an application may work with whatever our control variable is. So, a sample pdf file for instance that we might load in. but it may not work for somebody else's pdf. Whether that's due to issues with the file or some specific dependency that we're not aware of. There are edge cases always that we can't deal with. Yeah. As we continue...

Configuration Workflows

EaaS Webinar Series

October 2019

Respondent: And a lot of possible points of failure to even Zach's point of oh all these expert witness format disc images that everyone in the digital preservation community has been making because they offer great documentation of disc images. Guess what? Almost no emulators take those as input formats and that's down to the underlying projects that we have to rely on. So how do we think through that? There's a lot of different points to have to work through.

Interviewer: Right. Okay. Well thank you everyone. Lauren, Zach and Ethan. So, we're a little behind schedule so we only have ten minutes for audience questions but hopefully we can get through the two or three that we have. If anyone has any that have come up over the last discussion please post them in the chat box now. So Stefana asked are you finding any challenges particular to configuring environments that include Cad bim software? Lauren you touched on this a bit but Zach, maybe since you've been working on this classroom project has anything come up in particular that you have had troubles with?

Respondent: Well, one very particular as in specific thing, the licensing can be pretty tricky. For example, Carnegie Mellon had a few old copies of Archacad, an important architectural software. These required a hardware key, a little USB key to be plugged in and those of course have slipped through the cracks of whatever department office that the CD's were in. So that's one specific thing. It seems like when you have a stack of CD's that's one thing. They have a momentum and inertia that will keep those together. When you start adding dongles just from my experience that starts to become something that slips through the cracks. That's one thing. I have actually, I'll say, I've been rather impressed with how well these virtual machines are able to actually run this Cad software. I was pretty concerned, I mean, depending on the environment obviously there are issues of just kind of carserlag and things like that that make it less than optimal. But in terms of the actual rendering kind of through put I've been surprised. Something that has been absolutely essential in this classroom project has been insuring to bring in all the demos or samples. Not just install the base software but make sure if there's another folder on that disc that has demos and sample, that's something that's able to kind of illustrate the utility and the history of these applications a lot better. Especially for me. I don't know how to use Cad so I can immediately make an interesting image appear by opening the spaceship or whatever in the Cad program.

Interviewer: Great, thank you. Ethan, I know you've...

Respondent: To echo on that because we did have a considered effort to get as much, we knew this request was coming down from Carnegie Mellon that they were going to be using this class. So, I'll say we did try to process as much as we could out of Yale's collection. Just echo the problem with the licensing. We were either missing license keys. There were either physical dongle devices that were missing or sometimes with some versions of auto cad, especially in the

Configuration Workflows

EaaS Webinar Series

October 2019

mid two thousands Windows XP era, get to a point here we had the license key, we had a registration code and we're just still missing an activation code. It was like four layers of licensing that you could get to. Given some of the work that a cyberlog clinic and software preservation network have done with the DMCA exception for cracking that software, we are confident enough that we could do that to circumvent those measures but it would just so much extra work for me to even learn how to do that. To even go find whatever legacy piece of ND circumvention cracking software would be. Find that, apply it. Then extra concerns with the DMC exemption rather than the code of best practices for fair use given the software that is circumvented has possibly more restrictions on what we can publish what we can give access to than any other items. The licensing is obviously not as much of a problem as it was even a year or two ago before the publishing of the code, but it's still an issue.

Interviewer: Yeah and that's a good point just to direct everyone once the recording is posted. Just to remind you all that we did have accession on legal and policy issues related to emulation a couple months ago. There is more information on that of course out there.

Respondent: Yeah I should have said so, so Brandon doesn't get mad at me. It's just licensing almost as a technical issue, not even as a philosophical or legal issue anymore. It's the technical measures that are still in place there.

Interviewer: Right. So, before we have to wrap up, I want to get to this other question because I think it is pretty interesting. Rick from Smithsonian asked in thinking about the challenge of assembling app user documentation, is there a thought to identify priority apps going end of life and archiving their websites before they disappear? Just to pile on to that, there's another question in the chat ask are we ever concerned about the persistence of this information on the web and how that might affect the longevity of the network or our ability to use the software that we've been configuring? Who wants to take a crack at it?

Respondent: I mean I don't have much to say...

Respondent: Are we concerned? Yes.

Respondent: Certainly. I do think this is one reason why it's very important that we build strong ties to projects like the wayback machine as well so that we can say hey. From my perspective it seems rather unfeasible to have EaaS itself start doing a separate archiving project. Maybe that is in the works but if we have good channels of communication with other individuals that are doing web archiving say hey oh please make sure to save this resource that's got a lot of good tips. Maybe if there was a kind of forum for where those of us working on EaaS could kind of collect that important material that would be good.

Configuration Workflows

EaaSII Webinar Series

October 2019

Respondent: Yeah, this is Lauren. I'll just add yes concerned. To add to Zach, I think the strong community ties as well. We had kind of given shout outs to the retro and gaming communities too and the folks that have been doing this kind of work for a long time I think that's another place to really continue to work on relationships. Be part of that community. Work with folks, share information as well as the kind of web archiving. It'd be a lot to kind of expect the EaaSII to kind of take that on. It brings up interesting questions when thinking about the role of academic libraries, other organizations who are concerned with this. If we're talking about software and I managed to catch some of the maintainers comp parts where I know that some of these discussions are being had, you know what does that look like. The role of documentation around software, whether that's code or commercial software, whatever it is is huge as far as preservation over time. I'd be really interested in pursuing what that actually looks like to alleviate some of all of our concerns that we're clearly concerned about.

Respondent: I do want to chime in in terms from the EaaSII perspective. Yes, we are concerned about the persistence of this information. I think we would agree from our own scope of work perspective with the idea of community outreach is going to be the most efficient way of insuring that persistence of this information long term. We are also, the discovery of the need for a lot of consolidation and leveraging of online information has informed a lot of the user interface work that portal has been doing and our vision for the EaaSII platform in terms of the design. And how again, we have ideas now for how help text or readme texts that Lauren was talking about earlier, how once you find this information, whether it's on the live web or in a web archive, whatever platform that turns out to be, again how do you link that to an EaaSII environment so it gets presented in the place that it's most useful for the end user. We are definitely thinking about that and in that way at least, as long as that information gets to our platform and we continue working on the sustainability and persistence of the EaaSII platform itself, that information will continue.

Interviewer: And I would also point out that one of the fields that the student configuration workers are capturing is a link to if they do any research or find documentation they are recording where those are located on the web. Our semantic architect Kat Thorton who is doing a lot of work in publishing that information on wikidata or adding it to the wikidata knowledge base, she herself or they have also been. When finding documentation whether that's a scan of a manual or an existing web version of that, including that as an item or property in that item so that we don't lose those connections to those existing documentation. And that question of especially for older applications of digitizing manuals has come up quite a bit. We've been discussing that as far as in the future efforts go and subsequent rounds of funding. Yeah, something that we are keeping tabs on. And with that it looks like we've run out of time. I don't think we have anymore time to address questions but as always if you have questions feel free to email any of today's participants or contact us through the website or hit us up on twitter. We

Configuration Workflows

EaaS Webinar Series

October 2019

really appreciate everyone being here. Keep an eye out for a feedback survey that will be sent around to all of the attendees and for the release of all four webinar recordings with transcripts any supplementary materials. We will hopefully have those up very soon on the EaaS website. So, thank you all again and I hope everyone had a great day.

Respondent: Thanks everyone.

Respondent: Thanks Seth.

End of audio.