

>> So the 2016 National Digital Stewardship Residency Cohort of DC. We welcome you to the second annual NDSR Symposium. So I was here – well, I was there in 2014 when the first cohort did their symposium; I was a graduate student at the time. And I was really curious about what the whole NDSR thing was and if it was a viable option for me after graduation. So it's really interesting to be on the other side now, hosting you with all of the battle scars of event planning. But I will say that the thing that I didn't realize sitting there is that from our perspective it's kind of a party. I mean, I know it doesn't look like a party. But it's kind of an excuse for us to bring people that we've been admiring in our field from afar and hear them speak and meet them – all the rock stars of digital preservation. And it's basically just a celebration, I would say, and a meeting of minds. So it's much more fun than I expected it would be. Okay. So a couple housekeeping things. First, if you're planning on tweeting about the event or following it, it's #NDSR16. And let's see, bathrooms, if you didn't see them, if you go out of the exit and through the entry way, I believe they're on your left. A reminder that no refreshments are allowed inside of the auditorium. The audio for this event is being recorded. The recording with the transcription and slides will be made available on the Library of Congress website. And you will receive an email when this happens. We will have a summary panel at the end that will be moderated by NDSR alumni Julia Kim. And we asked you to submit questions when you registered for the event. And we've chosen some questions that will be asked during the panel. And then we may have a chance for to you speak ad hoc at the end. It's kind of like the UN in here. So if you do want to speak to make sure that it's able to be recorded, there's a button on your microphone. Press it and when it's on, it turns red. Make sure to press it again to turn it off in case you're talking under your breath and you don't want it recorded. Let's see, what else? I think that's it. So just an overview of the day. Hopefully everybody got a folder that has the agenda inside of it. So we're going to start with some introductions first. And then before lunch we're going to have Jason Scott speak. And a mini break. And then we'll start at 11:15 again and the cohort is going to present on our projects kind of in a lightning presentation format. Lunch will be on your own from 12:00 until 1:15. And I don't know if anybody had a problem getting here, but don't worry, we'll have Nicole explain to you how to get to where you can eat. And then after lunch we're going to hear from [inaudible]. And finally, Caroline Catchpole. And then we'll finish up with the panel and bring everybody back together. So first I'd like to introduce George Coulbourne. So he is the chief of internships and fellowships in the Office of National and International Outreach at the Library of Congress. Mr. Coulbourne was a cofounder of pilot NDSR program in Washington, D.C. and he currently leads the third iteration of the Library of Congress IMLS National Digital Stewardship residency program and serves as an NDSR Advisory Board member for the American Association of Public Broadcasting and the newly-awarded NDSR Philadelphia Museum of Art and ARLIS North America and DSR programs. He was cofounder of the Library's Nationwide Digital Preservation, Education, and Outreach initiative and serves as the agency lead for the Hispanic Association of Colleges and

Universities internship program. He is also a coworker of mine now. Welcome, George.

[ Applause ]

>> Betsy Humphreys: Before George starts – that’s all right. Don’t worry about this. Yes, it’s a party, exactly. Before George starts I would like to welcome you to the National Library of Medicine. We’re delighted. My name is Betsy Humphreys. I am the deputy director and currently also the acting director of the National Library of Medicine. We had the great pleasure of hosting this event last year, and we were delighted that it must have been a good enough venue that you decided to come back. And I love the title of your program this year, Digital Frenemies. The National Library of Medicine is very, very interested in both born digital and made digital collections. We have quite a substantial body of both, as I mentioned last year. We have the distinction of having preserved information that was born digital more than 50 years ago. We’re proud of that. Of course, we were working with the easy stuff, that that was the stuff that we actually created ourselves and made – created in some form of standard to start with, which, of course, is a much easier proposition in terms of long-term preservation and access than the kind of general material that you have been dealing with. In thinking about this event today, it occurred to me that you all – everyone who’s working in this field today – thinking about long-term preservation of born and made digital information, you all have a lot in common with John Shaw Billings. John Shaw Billings was a tremendous figure in the development of information systems, medical information systems in the 19th century. He is the person who started the Index Medicus, which is the ancestor of PubMed today, and did many other things in a wide variety of areas. But he has a lot of notable quotations. And since he was responsible for starting what became the National Library of Medicine, we’re used to quoting him on many occasions. And I have one here, and it applies to all of you: There’s nothing really difficult if you only begin. Some people contemplate a task until it looms so big it seems impossible, but I just begin and it gets done somehow. There would be no coral islands if the first bug sat down and began to wonder how the job was to be done. So we’re all grateful to you because I don’t want to compare you to bugs. But you didn’t sit down and wonder how the job was to be gotten done, you just started working on it. And I think that that’s the way something as important and as large and looming as ensuring that we preserve key elements of our digital heritage is going to happen. So thank you all for taking on this interesting and important problem. And I’m sure we’re all going to be interested to hear what progress is being made during this time today. I am not going to be able to stay with you this morning. But I hope I’ll be able to come back and hear more about this, this afternoon. So welcome to the National Library of Medicine.

[ Applause ]

>> George Coulbourn: Thank you, Betsy, for so graciously offering NLM once again. The inaugural year symposium was actually held here, as you just heard.

They've been such a gracious host. The students have experienced such wonderful support, and the mentors have been wonderful here, really. The NLM has been one of the best hosts and actually awarded on the second time not only for the work, for their project that they proposed but also how they support the residents and the program. Clearly, I'm delighted to be part of this program today. It's the second cohort that the DC IMLS Library of Congress collaboration has gotten together. And I would like just to spend a little bit of time on some of the changes and some of the new programs that are going to come forth from this NDSR experience. We're very excited. Two of them are just getting ready to begin and another is being proposed. As I said, this is a very important time in the history of NDSR. We've come a long way in the 20 years since the publication of the 1996 Garrett Waters report on preserving digital information. Just as the zero milestone that you see on the ellipse below the White House marks the starting point for measuring distances along roadways radiating from the nation's capital, the NDSR DC program marks the starting point for the first digital stewardship field model. And it has been and will continue to be replicated across the country to meet the needs of current and future digital stewards. In 2013 the Library of Congress and IMLS rolled out the pilot of the National Digital Stewardship Residency program or as you will hear it, NDSR, to cultivate a nationwide talent in the field of digital stewardship. And I'm very happy to have a previous IMLS project officer, program officer, associate director –Joyce Ray – here today. She's now with Hopkins. We hatched this program out, as we were just telling Betsy, in a sidewalk table at the [foreign language], the Spanish restaurant downtown. It was not called NDSR then. Now, over the last five years NDSR has become the Library's flagship educational program in digital preservation. The program was designed to enable recent post-graduates to experience anywhere from nine months to a year hands-on training in real-world environments based on the premise that learning comes from doing and doing leads to success. Now, the initial development of the NDSR program started back, as I said, in 2010 when the library convened an advisory panel comprised mostly of people local, although it was some national members. But they were thought leaders and educators from across the country. And they helped us to develop the baseline curriculum that is in use today. Key competencies were defined and resulted in a baseline curriculum, as I said, for classroom component of the program, which has served as their curricular basis for all other programs that have followed. While the classroom learning is important, the real value of the program clearly lies with the resident's participation in the cohort and the many professional networking opportunities that the program brings. Professional conferences that they attend and they often present at, this symposium, which we have designed, planned, and are now executing – and thank you for recognizing that special event planning is not an easy job – educational enrichment sessions that each of the residents and their host institution has put together. And all of these bring together an experience that is somewhat unique to this program. I'm very proud of how NDSR has launched the careers of many residents. You'll see some of them from the first cohort here. And through leveraging their experiential

experiences and real-world challenges that they faced at their host institution, they developed a unique and strong business acumen and technical expertise. We're also happy to announce that some of this class' current residents have already been offered positions and accepted the positions. And currently on a national level, I think, it takes anywhere from about – the most it's taken is a month to a month and a half for someone to gain a pretty significant position in the digital preservation community. However, not only am I proud of the residents, I'm extremely proud of the host institutions. They've increased an awareness of an array of digital preservation challenges that we all face. And it's through participation in the program by the residents that they help bring it to the forefront, as they brought some of the top speakers today in the field. Now, since the inaugural year, NDSR has engaged 14 host institutions in the DC metro area alone – that's not counting the other cohorts in Boston and New York. And an advisory committee will be selecting another five next week. We start on-site visits, I believe, the 16th. Now, moving forward, NDSR – and notice we used the metro here because, you know, the metro doesn't always move forward. So we do pause and have issues every now and then, just like this program. But moving forward, we will continue to go. We're going to focus on educating the next generation of digital stewards by placing them in real-world experiences not just around urban centers as it's been in the past but across the country. The goals are to increase the program's impact, continually improve the program, and to ensure the efficacy and relevancy and grow the program so that it can reach even more of the digital preservation community. We've already made strides in improving NDSR. Each program is independently evaluated. And we take that evaluation to tweak and often make significant changes. One of the major changes we made was the first cohort actually had a two-week immersion program in digital preservation. It was too much for these residents to try to absorb in a week. And also, some of them were fresh out of school so they thought it was redundant. So that's been changed. And it's going to be changed even again for the next cohort that will be launching at the beginning of October 2016. We are also looking at doing a prototype program for the regional model at the end of 2017. We don't have anything firm on that. We've had discussions on it. But the regional model is a model that has not yet been addressed. It's a distributed model, which I'll talk about just very briefly. And then as I said, the cohort model's centered around urban centers, which is fairly easy. It's not that challenging to do – I don't mean the program's easy – but transportation is easy for the residents. We've got good metro systems, Boston has good metro systems, New York has good metro systems. But we've yet to – except for the two that's going to be coming up that are distributed, we've never done regional model where they may be in Virginia, Maryland, and DC, but we've convened them in certain places throughout the year so that they can still have that strong cohort. So that's something we're going to look at. I can actually – I'm actually happy that the American Archive of Public Broadcasting and the Philadelphia Museum of Art and ARLIS of America – there may be one other. Emily is here from IMLS, so I don't know, she might want to bring that up later on. These two are going to be your distributed model. And what

that's going to be is that these residents are going to be located throughout the country. Now, this will be the first time that there – as I said, that there has not been a cohort. So that cohort is going to have to be maintained virtually with the exception of I think they're doing two weeks in the very beginning to start the formation of the cohort and bring them up to a baseline as far as some of the particular skills and how the program will be managed. And that will be held in Boston. Now, ARLIS and the Philadelphia Museum of Art, that's going to be unique because it is the first residency that's going to address the specific needs of the museum community. And that's going to require also that we have to address the curriculum. Now, this curriculum is refreshed yearly. Nancy McGovern at MIT has been helpful on that, Helen Tibbo from UNC. These were original people on the group. We now have brought another individual on as a member of the staff at the Library of Congress, Mary Alice Ball. And she'll be helping us as well. But we're going to be convening a committee mid-summer and we're going to reaching out to some of the mentors and the host institution to serve on this working group to look at the curriculum both for the type of program that we've had here in these urban centers, as well as a distributed model and potentially a regional model. So it will be a true working group. So we're really looking forward to some unique opportunities as we move forward with NDSR. Now, I could speak in more depth about the opportunities; however, there's a number of great speakers lined up behind me. I do want to thank a member of the staff that's here today, Barry Howard. We've gone through changes at the library over the last few years and since the founder of [inaudible] Laura Campbell left, and it has been clear that that's been challenging, but we've kept these programs move forward. Not without flaw. You know, you learn from flaws. But it's very clear that, you know, it's the support of staff that I've had and the vision of my colleagues in the community that have been critical to keep this program moving forward. If you'd like to talk to me or any member of the staff, please do. My contact information is up on the screen. We're always looking for partners in different programs. We manage mostly the educational outreach programs. We're under a new division now called National and International Outreach. And so there's going to be a lot of opportunities coming in the future. Now, I want to end by saying thank you all for attending. It's a nice turnout. But I want to give a special thank you, a sincere thank you to the residents and the host that have joined us here today and have supported this program during the past year. This program would not exist without you. Thank you all very much.

[ Applause ]

>> Okay. So when we were planning this symposium however long ago, eight months ago, we were pretty much unanimous that we wanted to put together a theme that focused on our perspective of being new professionals in this field. And we talked a lot about different options. And we came up with Digital Frenemies because we were looking at jobs. And we saw this kind of weird bifurcation in professional duties of jobs related to digital preservation. So, for example, we'd see some jobs that said, you know, "We have this digitization

project and we want to you digitalize 20,000 pages of the so and so collection, etc.” And we’d see other jobs that said, “We have a lot of born digital records and we want you to choose a content management system to make them accessible.” So that’s kind of why we came up with this theme. And inevitably, obviously, making things digital and curating them – curating born digital are linked. And we think in the future that hopefully these job duties that we see will become a little bit more symbiotic and not so bifurcated. So we looked around in the field of people that were working in the born digital and the made digital and both and one of the first names that was on our list was Jason Scott. His extensive, extensive experience with software preservation and his work rescuing materials related to technological history both analog and digital makes him what I would call an all-around preservation Renaissance rebel. He is a historian, a filmmaker, an archivist, and a cat caretaker. He works with the software collections at the Internet Archive. And I’m happy to welcome him today to give a talk titled The Walking Dead. Welcome, Jason.

[ Applause ]

[ Inaudible ]

>> Jason Scott: And they work really hard [inaudible] magically making this happen. So thank you to both of you.

[ Applause ]

Also because I am in Washington, D.C., this is a recording device. So don’t get worried. Because I have had many people go, “What’s the thing [inaudible]?” I’ve given speeches. I always record my own speeches because I always like to have evidence I actually did say that. And I also have people go, “What’s the thing at the base of the Jason’s feet? What’s going on? Are we okay?” Anyway. Thank you all so very much for coming to this. I am delighted to be invited. Today this particular year was one where I tried to cut back on my speeches because I’ve been work on some projects, which you’ll see. And I however was just really honored to be asked to go to this. So here I am, one of the few speeches. And if you’ve been to a lot of archiving speeches, which some of you may have, some of you may not, it only takes a little while before you see this slide. And I’ve decided that you should always figure out when you go to this event how long it is before you get time to warehouse. Time to warehouse is how long before somebody throws out this cliché. Because if you’re going to do it, at least try to go a little bit out of the realm. Like, this is the one from Citizen Kane , which is much more impressive to me because it’s not just, you know, there but completely a mess. But this is the one I’m really proud of because this is the Internet Archive warehouse. And a lot of people say, like, “The Internet Archive? What’s that?” It’s this place, archive.org. And believe it or not, we have this warehouse because you actually, it turns out, can’t make a website that only collects digital materials; you’re going to have physical materials up in your life. So that’s kind of what happens, you know, even though my boss, Brewster Kahle, thought that he was going to have

a purely digital experience. It didn't take him that long before he realized that, "Oh, I better buy some warehouses." And luckily he was able to do so. It's a place that has this building, which he bought because it looked like his logo. And inside of it it's got a beautiful meeting room. I really like this place. And inside the meeting room we actually have some of our servers. Each one of those servers holds between one and two petabytes of data. So when people connect to the Internet Archive, you're actually connecting to this data going back and forth inside of this church, which is now not a church. But it was a Christian Science Church that they bought and renovated. And there's even little, tiny statues inside of it. There's me. If you work at the Internet Archive for three years, they build a little statue of you because, as Brewster says, we don't have stock options. So there you go. And so why does my little guy have wings is, of course, one question? And the reason he has wings is because – how many people here have actually heard of me before as a thing? Just a thing? Okay, good. So I'm known in some circles as the Angel of Death. And the reason why is because I am really, really active about preserving materials and stopping websites from disappearing into nothingness. And as a result, when I show up, things are bad. You know? So it's like, "Oh, man, we're not getting our funding. We're going to have to shut this back. Oh, God, Jason Scott's here." So that's one of those things of like, "Oh, Jason Scott's here. It's the Angel of Death." So anyway, he's got little wings and more importantly, he's clutching little floppy discs to his heart. So that's also what's going on there, right? Just a little sideline, here's what I was working on recently. This is in the Westminster Mall up in Westminster, Maryland. This is a coffee shop that is now not a coffee shop. It actually contains tons and tons of manuals, which I got from here. So the deal with this place is this is called Manuals Plus, and it was purchased from another company in Utah. And then after a number of years the owner, Nick, decided "The manuals business isn't just really working out for me. So I think I'm going to throw everything out." And there's different ways people could react to that upon hearing that news. And some people might say, "Well, that's the way of the world; paper is gone and now it's digital." Other people might say, "I wish I could do something, but I feel inherently powerless. I don't even know where to begin." And I decided to drive down and have a talk with Nick about this place. This is a very large collection. And they were selling them on eBay and you could order them. And it was a pretty cool place in terms of if you wanted this obscure Hewlett Packard manual, you could get it. You talk to Nick here. And here he is walking me through. And I was alerted by one of the people who worked there. And she said, "Yeah, I think he's throwing it out soon." And I immediately drove down two days later, down here to Maryland. I live in New York. And I said, "So, you know, I could probably do something with this; when are you throwing it out?" And it was Friday. And he said, "Oh, Wednesday." Now, the thing is, is that I'm used to this kind of turnaround, right? As you're going to find out, I'm not really a pretty standard-issue archivist. And I think of archivists and archiving in many ways like nursing homes and storage of things. But the stuff I do is like an EMT, which means that it's not always pretty, but we try to

get the job done and try to save things. So I made a big noise on Friday and then told people about this. I took a bunch of pictures and said, "These are old electronics testing manuals." Now, I don't care less about electronic testing manuals versus anything else; it was just a matter of man, what a unique and interesting collection. And it was a unique and interesting collection, almost complete in some cases of some companies that made this equipment. And so I looked through it and now you hit a point here. And at that point you say to yourself, "Well, that looks worthless," or "That looks very interesting, I wish that would be preserved." And it's not really my job to convince you if you're in the "I don't really think this should be preserved" crowd to help me preserve it. I'll find enough people on my own. And I did. Sent out for people and on Monday, people started to show up. And we tried to take one unique one of each manual. And people were working on this as fast as they can. These were just volunteers wandering, showing up, looking at these beautiful old manuals. We had to put them in bankers boxes – that turned out to be a mistake – but banker's boxes and people just started to sort through, do their best, put them in, and start to gather up these banker boxes until we had 1,600. Now, what are you going to do with 1,600 banker boxes? Well, it turns out you want to hire some movers. And so these movers moved them into three storage units down the street from this place. This is all happening between Monday and Wednesday. And we had 60 people show up. And we did this through social media, through a Slashdot story or Reddit story. And, you know, hey, he's available. I'm just saying. And what we ended up with was a big pile of manuals. Now, the thing is, is that this big pile of manuals was costing me \$1,000 a month to store in the three storage units. And we had taken what we could. We had to leave this like this. Everything you see here was thrown away, thrown into a big [inaudible]. The owner begged me not to put a picture of that up. So I don't have a picture of it actually in the trash. He didn't want people to come after him. Death threats were coming in from people. But we got a lot of it. And we figured out it's something between 60,000 and 75,000 unique manuals. And I left a beautiful set of roses for the woman who first told me about this happening. And off it went. And we now are storing it in a old coffee shop inside of a mall where I am now going to be sending it to multiple homes to be saved. That's what I do. Like, I just get called into these things. I suddenly went from, "Geez, I sure wish I had nothing to do this week" to I'm driving down and we're loading 1,600 boxes into storage units. Because fundamentally – fundamentally – it's a belief that this material has persistent historic and cultural meaning. And I can't convince people – I think – to jump into finding meaning into it. Some people might be convinced, but I just don't get into that job. And I find that a lot of people who work in this area have to do that, this constant, unending justification for their very good, very intense, and very thankless work. And so I always feel very blessed that this is the situation. But this is one of the other things that I'm working on, which is part of why they brought me in today. These are a bunch of CD-ROM's. I have a lot of CD-ROM's. I have a ton of CD-ROM's, things that people have been sending me by the thousands. And I've been taking them, imaging them,

scanning their fronts, putting them into a box. This is my room. And so there's floppy discs here, there's CD-ROM's, there's parts that are being worked on. This is what I work on all day. At the Internet Archive in California in our warehouse are these two guys. The scale isn't very good with this photo. So here's a better one when it has a pallet jack. And this contains CD-ROM's. And so there's a room there where somebody in Los Angeles would run an educational software review program, had all of the software that he had ever been given for ten years. He estimates it's about \$1 million retail value of the software, and he gave it all to me. So this issues of old educational newsletters and the actual software. And there's some pretty ones. I just like this one just because wow, why – I'm sure when you opened up, you felt like you had really made a good purchase. But anyway, these are all CD-ROM's. They're actually clip art CD-ROM's, believe it or not. But there you go. And I like this box because it gives me good life advice, which I try to follow. Anyway, so I got all these things. And they date – you know, CD-ROM's particularly are of interest to me because right now they're a really good tradeoff. At anywhere between 30 and 600 or 700 megabytes per plastic disc, there was between let's say 1991 to be honest and about 2002 – that's the way that we would do large-scale data transfer. So a lot of material got pulled through CD-ROM's. So if I rip a CD-ROM of its material, I can have, you know, a very large collection in no time at all. Now, floppy discs are a little more of a pain in the ass. They have to be put together in a reader and you end up with a whopping 100K or a whopping 300K or a whopping meg of data. So I do those as well. But these CD's, they are unbelievably, unbelievably useful. Now the Archive has up what I believe to be the largest online accessible software collection in the world. Find me a bigger one, and I'll download it and I will add it to ours. But right now it's pretty huge, among things 7,000 CD-ROM's; there's tens of thousands of software packages; there are zip files of FTP sites because we've been going to every FTP site on the net and grabbing a copy because I think they're doomed; we have Android applications, iOS applications; we have hard drive images. We have been putting these things up – incredibly large. And when Brewster hired me, he said, "It will be nice that you collect things, but can you please do something where people can use them?" If you haven't been to the Archive – I'm sorry but that will change after today – if you go to archive.org and you want to read a book and you click on a book, you get a book. You can flip through the pages instantly. This visceral reaction from "I really want to look at this knitting catalogue from 1968," you're looking at the knitting catalog within a few seconds. If you want to play some music, you click on it and you're hearing the music. If you want to see a video, you can see the video. And we're so used to it now it's boring. Like, if you say to your friend, "Hey, go look at this YouTube link," they're not freaking out over text going like, "I'm playing it and it's just going in my computer. It's in stereo. Oh my God, how did you do this? This is black magic." Instead, they go, "Wow, you're right, the bassist does look like he's drunk," or "Oh my God, I can't believe they put a kitten in that." Or whatever. You end up with a reaction. Your response and commentary moves to that. Somebody puts up some terrible rantyism speech

and you react to what the person says and make comments on it or maybe even link to a portion of it to go “This was particularly heinous.” But you absolutely stop thinking about that transport. So I spent a number of years working with an incredible group of volunteers on something which we call the Emularity. And that’s what this thing is, the Emularity. And the Emularity is the ability to play software instantly in your browser. Now, this exists. It’s done. I don’t need funding. I don’t need your permission. I don’t need you to, like, get behind me on this. It’s done. You go to the browser at the Internet Archive and in the same way that there’s a picture of an image or a video or music, an Apple II will boot up and you will be in an Apple II playing that software. It’s just a viewer. And we released this about two years ago. And I’ll talk a little bit more about what that did when that hit the world. But the thing is, is that the Emularity dropped into the world. And what happens is that you get to our site and there will be a collection of programs here. Now, we just got Windows 3.1 working in the browser, so you’re actually able to go to these Windows 3.1 programs, click on them, and be in a Windows 3.1 box playing these Windows programs. And these are terrible, let’s be clear about that. I am not rescuing a long-lost classic. I am not pulling from nowhere pieces of the Torah that we didn’t know exist or a Dead Sea Scroll. These are terrible. Some of them of them are minor pieces of work, obviously experiments. Some of them don’t do anything. They all have various aspects. But you can react to them viscerally by going right to them, right? And the Emularity works in pretty much any browser that supports JavaScript. So for instance, here’s Doom. Doom, which I have actually running on a New York Linksys terminal that they’re now installing in the city for you to be able to walk up and check the web and use Wi-Fi at. So it’s running Doom there. I can run Doom anywhere. Here’s an Atari Jaguar running on a random laptop at a Best Buy that I walked into. Here is Doom running on a random Sprint phone that I walked into a Sprint store and just started it, running it up. By turning all of computer history into an embeddable object, we are now in a world where this is all back. This is the first thing that I mean by the Walking Dead, okay? Software is no longer dead and buried. We now live in a world where pretty much any piece of software written for DOS, for Windows, for the Atari 800, for the Apple II, for the Commodore 64, for the TRS-80, for a whole variety of machines can instantly come back into the world as if it never left. There’s certainly a lot of refinement that needs to be done to have certain aspects better about emulation. And I’ll mention that in a moment as well. But we now live in this world – it’s not coming, it’s been here. And we as a species, as a culture, as everything need to realize that it’s all back. And what does that mean to us as a world, right? Before I go further, though, I want to give a real shout-out. You know, one of the reasons that this works is because – for this one, this is a JavaScript sport of an emulator. That is to say there’s literally a full emulation system ported into JavaScript whose primary purpose initially was to make it that menus sparkled or that your mouse changed color. Now it’s being used to run fully running virtual machines inside of embeddable objects inside of your browser in a Canvas tag. And so we’re there. But the reason that we converted to JavaScript is because there are some amazing people out there

who are working in the emulation field, who are working really hard to emulate every aspect of old computers that they can find. They are constantly trying to seek out large amounts of old circuit boards or obscure ROM chips and pull them apart. The most prominent project is the Multiarcade Machine Emulator Project, MAME, which now does 32,000 different arcade machines and 1,600 different computer platforms to various degrees of accuracy and support. And some of them are well supported indeed. And it's expanding. It now does a popcorn machine. It actually does a synthesizer, the Korg M1 because it's pulling in anything that has a ROM now. So they've already move away from like, "Yeah, computers, let's move into anything that's ever had a chip." Will they do heating systems? Will they do walkie-talkies, HAM radios? Probably. The system is completely open source – fully opened in a license. It took them eight months last year to fully push it to a full open license. So people are now adding modification daily. By my estimate they do between 100 and 200 feature adds or bug fixes every month and release every month. So we port that to the Internet Archive within 12 hours using our system. So if they put in a whole bunch of improvements to how Apple II's boot or Apple III's boot, it will be reflected on the Internet Archive's software player within about 12 hours to 14 hours. There are volunteers who have done this amazing work in terms of finding this stuff, in terms of making sure that it all works, in terms of – I've had blind users who go in and listen and they say, "It doesn't sound like you said it right." And they've gone in and repaired all of these things. So there's people out there who are just working to say, "Where's this gap that I remember? Can you fix it?" And they do it. And there are people who they are not just working on the project but going after the parts of the project that nobody wants to do. And that's the hardest part of any open source part is because it's really easy to do the sexy entrancing parts that make the little glittery parts go. But then to have somebody run through every month and say, "Did we break anything?" There's actually a group called the main testers; all they do is test the thing and run it through a test harness. There are people who are saying, "You know, even though you have an Apple II swift card, I feel like there were other versions." The emulator actually emulates a Texas Instruments Speak and Spell, which you might recall. Turns out the Texas Instruments Speak and Spell had five different ROM revisions across its life. And so there's been incredible debate as to which one is the canonical one. And they've ended up using the one that's mentioned in the patent. And they can do the others and you can choose them, but there are variations. And these kind of choices are unbelievably tedious and boring and utterly necessary. And I love all those people. And of course, the original creators of this software – I burned one out in 20 minutes. Not bad. It's like batteries, just oh my God. And so the creators of the original software worked on their items in a way that they were trying to bring, you know, information, and interaction, and all the explorations that they did, they put them into formats that turn out to be pretty easy to bring back. And they made them available either as public domain or as commercial products. And they created amazing things. And it pushes the whole thing forward because they really are truly gems. It's not all terrible. There are some

beautiful, obscure pieces of work inside of these software collections that are years and decades gone that have been amazing. And of course, a shout out to my boss because he lets me do this. I have a story about why I think Brewster Kahle is such an interesting fellow. I don't know how many people know about Prattkeeping. So this was a meme that transferred out there because in Jurassic World at one scene Christopher Pratt is trying to train three velociraptors to not attack. And so he moves into this maneuver where he's educating them and moving. And people really liked this. So zoo keepers around the world started to make pictures where they started to work with their animals and do them. And people were just screwing around with it. So I was like, "We need that." A, I thought we need that. And I ran – I happened to be in town. I transfer over to the San Francisco office for a few weeks every few months. And I went into Brewster's office. And Brewster works tirelessly. He's there at 8:00 in the morning; he leaves at 8:00 at night. I mean, this guy, he's worth a significant amount of money. And there are a lot of contemporary millionaires who would just build boats out of caviar and sail them to islands that are not on maps because they paid to have them not be there. But he has chosen to spend his life and his fortune working on a library. And that is an amazing thing. And I went down there and said, "Brewster, I can't even begin to explain this to you, but we need to do Prattkeeping with the servers." And he was like, "Okay, is it offensive to this group or this group or this group?" And I said no. And he goes, "All right, let's do this." And so he actually walked up and did it. And the fact that he just put that amount of faith in me is just an indication of what he does. And I always appreciate that. He gives me an incredible amount of leverage to do these things. Anyway, so back to the Emularity, right? So the thing about the Emularity is that the Emularity is a actually a framework of multiple emulators that run in the browser, and as time goes on more specific emulators that perhaps do one service better than the others will be added to it. And it's an open source framework to just make it that you go in. I will, of course, let you know that there are other emulation frameworks. I wouldn't quite call us competitors, but we are in the same space. It's not quite an Avis/Hertz thing, it's more like a, I don't know, Florida/Mississippi thing. It's like they're both there. It's not really in competition. One of them is called Olive, which is being done by an engineer name Sacha. I think it's one of the most brilliant pieces of code written. The fact that it's being used to do this, this emulation, I have read his papers on the technology behind it – it's like he's taken a Hairier jet and made it deliver pizza. It is an amazing thing. His ideas about how he put this together will have repercussions 10, 20 years down the line in the same way that people are looking at papers by Turing and Rosenbaum and saying, "Wow, I guess they figured that out in 1948." They're going to do that with Sacha. So Olive. The other one is called BWFLA. And BWFLA is a emulation system that works different than mine – they both work different. In mine there's a Canvas and its executing locally on your browser. That means it can scale because each person, it turns out their own machine's doing the work. So it's not so hard. But the others are having external machines that kind of Netflix over the computer experience. This means their emulations tend to run

a little better, it means that it runs at much better speed and consistently, and it means there's a lot of control for rights issues and other things. We don't do that. So they do that very well. But they can only have a number of seats. And the more people who connect, the machines they need to add. So there's pros and cons. You know, I don't think of us as competitors. I think of us as colleagues who all believe that emulation is the way to go. So we've released a number of things to get people's attention about this. And about two years ago I released something called the Console Living Room in which we put up 16 or 17 consoles. Now, this is like the Magnavox Odyssey, the Atari 2600, the Bally Astrocade – a machine called the Entex Adventure Vision, which I promise you, you will never want to see in your entire life. It is the worst machine ever made. It is a Faustian bargain of epic proportions. It released four cartridges. It has no screen. What it has instead – I am not kidding – it has a rotating mirror that has a string of LED's that shine off the mirror against a reflective piece of plastic to simulate a video screen. It is the worst thing ever made. But if you want to experience the wonder and the power of the Adventure Vision, you can give it the 17 seconds it deserves at the Internet Archive. We released the Console Living Room and people went a little nuts. I put it out on the day after Christmas and called it a second Christmas morning. And it was something on the order of two or three thousand console games brought back. That got a lot of press. But what really got the press was a few months later we put up 900 and then 1,100 video arcade machines. So we're just talking straight up Pac Man, Asteroids, Birdie King – games that some of you may have played, some of them you might not have played. But it was something like 1,100. And I knew this might get a little attention – oh, it got some attention. And it was interesting to me how we had also put up collections of business software and we had put up collections of historical pieces. Like, we have the first VisiCalc up, we have original word processors, we have original calculators. But video games, oh, I get it. This is what's so funny. Again, it's like a Hairier jet delivering pizza. It can do all these things and people are like, "Video games. Oh my god. That's amazing. You can play it all these places. Video games." It became a question on Chris Hardwick's At Midnight . It actually made Fox News . You can play video games. And our server hated it. I had a slight inefficiency. So this is our normal day and then this is our system being pounded into submission. And Ralf Muehlen, our network tech – bless you, Ralf, wherever you are – working hard to bring all of this machinery back up online. Because it murdered us. We figured out that over the weekend that it got announced we had something on the order of 3.5 million new visitors – individual visitors – to play the video games. For them it didn't matter that there was this whole emulation piece that was running or that it was a part of a historical software piece or anything. To them it was just free video games. And I get it. I totally get it. After we got our 3 million users, we put the – you know, I had put up all these other pieces, right? But I knew something about arcade machines just triggers within the human psyche, especially in this country. So I knew this was emulation D-Day. That was my internal name for it. I knew that many brave emulators would go up and hit the net, hit the beaches of the net and many would not survive. I knew

that when you put Super Street Fighter up and Pac Man up, that awakens the elder gods. So I wasn't so worried too much about the elder gods. You know, people use this great phrase – sue into oblivion. Aren't you afraid of being sued into oblivion? Well, we're a nonprofit library. We have no ads. We lose money every time you click. So if you want us to lose more money, click harder. And this is all being done with a move towards historical replacement of otherwise not available materials. And it was very hard to determine what's really available and what isn't. So let's find out what's there. And so we started to get, you know, the usual letters, right, with people saying, "So you have this up; we'd rather you didn't." And they all had different attitudes. The Atari lawyer Kristin and I are on a first-name basis. And Nintendo had something to say. In fact, there were all these companies that kind of came forward. Most of these, Atari's fake – there's a company that actually owns Atari now that – let's just say it's four people in room that license it. But the rest of them are sort of here. Sega's sort of around. And then there's something called the Hamster Corporation. And the Hamster Corporation bought up five Japanese video companies. And so they kind of own it. And so they all wrote to us. And so when we were done we had 607 and the rest went away. And to be honest, that's about right. There were games that everybody knows, right, in the software area. We all know in some way about perhaps Microsoft Word and PowerPoint. Some people who are older might remember Word Perfect, some of them might remember FrEDWriter. Some of them maybe if you're lucky will remember Personal Editor, you know? Word Star is such a goulash of terrible keys that people were in it, with the exception perhaps of the writer of Game of Thrones George R. R. Martin, who still uses an original DOS box to write. These are obsolete pieces of equipment run by companies that no longer exist. But some games still matter. We are still shoving Mario into everyone's face. We can be assured historically that when the species that comes after us lands on earth, there will be nothing left of the grey goo, except for an obsidian monolithic of Super Mario. Like, we're not worried about the historical saving of Super Mario. It's going to be fine. It's the other stuff that I'm worried about, it's the stuff that has no advocate. I call it advocateless material. Because you either end up with material that has somebody who says this is why it's important. You know, like, why is this important? Well, the guy who did it also wrote this. So that's an important book. So let's get his previous works. This woman who is a famous philosopher, she actually learned these things. She would always quote from these certain musicians. So let's get all the work of these musicians. And in doing that a little light is shown on these various tangential pieces. But let's get every Windows program; let's get every CD-ROM – that doesn't really happen because they tend to not have an advocate. And in the arcade biz – we'll just go back with that for a moment – there are some terrible video games or pretty good ones that just through various degrees – for your video game history, there's video games start coming really fast and furious around the 1981 to 1983 era. And then there's kind of a crash because they don't really keep making money. And so they come back slowly and they become really genre. But between '81 and '83, everybody wants a piece of the

arcade action. So unbelievably obscure companies that didn't do anything in arcade games make one arcade game, drop the millions, hate it, and get out of the business. And those video games are being preserved now, like Jack Rabbit, a game which I promise you've never played, or Mr. Poo, which was a flatulent miner who fights various creatures for treasure. We've got it emulated. You're welcome.

[ Laughter ]

And so, yeah, I mean, we're absolutely trying to go after these weird things, right? So fine, the games that people care about still – the Street Fighters, the Marios, the Sonics – yes, they're being actively kept track of. But a lot aren't, right? Now I fundamentally believe that emulation is the way to go. And you might be surprised or not to hear that to some people this is still a controversial position. That to some extent why would you waste your time on emulation when we have the original materials available, the original hardware? And the reason why is because emulation, through the use of embeddable objects, through the use of consistent playback, through the use of high configuration can provide it to a near infinite amount of people quickly and easy. And the barrier – you know, even let's say in an ideal world your staff has provided you with a vintage piece of equipment and the nine pieces of software you wish to study and you bounce through them, you're still going to have to wait for them to boot, hope that everything works, and bring it to term, and then you can't provide that experience to others in your reports because they don't have access to that. So I know that emulation is going to win in the long term. Fifty years from now when there's a reduced number of Pac Man machines – Pac Man is actually very easy to get right now – you know, it's going to be emulations all the way down. So I can't argue with that. People who don't believe that, I'll just wait because eventually they'll be emulated. Anyway, so the point is I'm really big into emulation. Right? And I get people who are like. “It can't happen.” And I'm like, “Sad trombone, it's going to happen.” So the work I'm doing doesn't try to convince anybody of that. I'm not there anymore. That's not where I am. I'm not in the advocacy position. I might be in the example. And by the way, for you, I'm more than happy to be a negative example. So if some of you tweet like, “I'm watching this guy – don't agree with him, but here, this is one point that I can drag out of this, fine.” I am emotionally, ego-wise, and personally fine with being somebody's bad example. I don't mind that at all. But I do believe that emulation is the way to go in where we're going with software preservation. There are programs that would otherwise disappear. Here is a 3-D benchmark program for the 486 that enables you to benchmark the virtual machine that I have running. And I love these. I have a personal kind of interest in these; I love disc-checking programs, memory test programs because they don't know where they are, right? You have, like, a disc floppy checker and it shows up. And I'm, like, the architect from the Matrix. And I'm like, “Well, your first question, while the most pertinent, is the least relevant.” And it's like, “How's the disc doing?” And it's like the disc, which is actually a virtual machine running inside of a machine 1 million times more powerful than

you currently think you're in and which is connected to a million more machines through a worldwide span is doing fine. There is no disc. So I'm quite fine with that. Here's a program, again, it lets you look at how the 486 processor inside of this virtual machine running inside of a browser, running inside of a Windows box – about the only thing that would shock this is that Microsoft is still around. But other than that, it's doing all of this analysis on a chip that does not exist. There are unbelievable amounts of programs like this, these one-off applications that are instantly available. And by putting them up – I must point out, I've not played all these. Instead, I have a bot that goes in and plays them for me. So there's a machine that does nothing but play video games all day and play computer programs. It takes random screenshots while playing it, then it tries to choose the most interesting one, and it makes that the screenshot. And if you go in, it will show you some of the screenshots it had. So there are 50,000 programs on the Internet Archive that are playable, and I've not really played any of them beyond the ones that I'm kind of testing or somebody asks me about or so on. So it's my own QA department. It's a machine. Baby for You, pregnancy software, 1991, done by some doctors. And it walks you through. It's like Clippy – “It looks like you're pregnant. Would you like to talk about it?” And you can put through all this information that walks you through your expected trimesters. This is a proto app. It's an app. It's a kind of just – this company made this baby program for – nobody is going to go out of their way to download this, get an emulator running, get all the ROM's in place, put it all together, boot it up, and then go, “Yep, worthless baby program,” unless you're a person who was trying to study the cultural treatment of women and one of the pieces you would like is how is this program going to treat you as somebody who is doing this very personal medical thing? What language does it use? What does the documentation say? Is it well written? Is it not well written? What did these people move on to do? What other programs have they done? And, you know, that's where you find all this value. And I love the graphics. You know? Right here, I mean, look at the use of shadow within the controls. You know, this implication that they have to do in two dimensions that there's a three dimensional floating aspect to those directional buttons. That's crazy. I love this. This is a disc that boots to tell you it's not going to boot. So when it boots up, it just says, “Sorry, I'm not actually here. I'm not booting.” I love that. So the year that this happened we had 12 million new people show up over the year. That's 12 million people who otherwise wouldn't have experienced this historical software, who either come to it for nostalgia reasons, or to learn something, or to go through, right? So this is all I care about now. This is all I care about. I'm not even working hard to improve the emulator. That's kind of happening on its own. I've already helped with this group to make the pipeline as smooth as possible so that new improvements come in – better keyboard support, better chip support. What I care about now is bridging the error gap. The opening speech mentioned this, but the amount of software that has been turned into a digital object beyond the original analog recording of the digital object is vanishingly small. It is a small percentage of what was out there. If it's lucky, it was commercially released and went out into

various canonical groups and we can get our hands on it. But there's software that's just hard to get your hands on because it was proprietary or it's on a format that people aren't comfortable with. So a lot of what I do in my showy weirdness is I'm a kind of talisman for an idea that doesn't hit the public very well, which is, "Oh, there turns out to be a use for this old pile of junk that I have." So they're sending me the junk, right? They're sending me the floppies that I'm imaging – piles and piles of floppies. This person sent it to me with a little note like, "Look, I don't even know where this came from. Enjoy." I will take in collections of floppies in all formats and I'm turning them all into these things. This is most of the collection of the largest CD-ROM producer for commercial software, I should say commercial versions of CD-ROM's – Walnut Creek CD-ROM. And I made some noise because of this. This is where I wrote a thing called, "Yes, I want your AOL CD's," in which I said, "Please send me all the AOL CD's." That made the news because that's weird. Why would you want that? Well, AOL CD's were 50% of all CD's produced for a number of years. They are in their own way a cultural reflection. Luckily each one has an individual serial or a part number, which I discovered because they were using so many CD-ROM processing plants that I'm able to get them. So this is all AOL CD-ROM's. That's all this is. And the creator of Walnut Creek contacted me and said, "Hey, I saw you doing that. Do you want all the Walnut Creek CD-ROM's?" I was like, "Yes." So I drove down to his house where he had 19 boxes. And he said, "My wife has been on me for the last five years about this, but I knew – I knew someone like you would come along." And I said, "I'm glad to be that little piece of domestic bliss." So all of these have been imaged and put online. And they are grabs of lots of operating systems, public domain software, images. And they're just being, you know, thrown up there. And I can't do it alone. I'm really working with a lot of people who are imaging discs, imaging floppies because we didn't do this because there was so much fear. There were concerns about rights, they were concerns about how do we do it, there were concerns about what if we get it wrong? And we let it [inaudible]. And really, imaging in a real true sense of something that is not a funded project or one that had 15 forms put out to indicate how important it was is tiny. So it's like we didn't do any collecting of art until 1955 – any art on paintings or sculpture. And then what would you do then? Well, the answer is drive around and go to every warehouse and group and say, "Do you have any art? We'll take it." And just take piles. No provenance, no exacting – just grab it as fast as we can. Bridging that error gap will enable robots, machines, scripts. and others to go through them and make better analysis. We are now to the point – we started out in 1995 with 30-second delay to render a JPEG. And then within a few months or I should say a few years we had thumbnails as the default. By 2005 they can tell if there's a person in the photo. Now we can tell the emotion of the person and who it is. And we are now rapidly getting to the point that it can pretty well describe a photo. This is a radio next to a dog. So I'm not worried about that part. It just needs to go into a part that the machine likes. And there's one piece of software that has had – this is actually inaccurate. It's 1.4 million when I took this screenshot; it's now up to 2 million – one piece of

software has had 2 million plays in the year and a half that it's been up, and it's the Oregon Trail. So the Oregon Trail is booted up at the Internet Archive every four seconds 24 hours a day. It is now – the link to it is now a default in dozens of school curriculums. I've been sent pictures of – if you look in the back – entire classrooms where they're able to put up 30 or 40 instances of it. And the person who gave this to me said, "We used to have one that we would boot up and then 30 kids would gather around and shout." That does not favor the girls and it doesn't favor the people who don't want to interact with other people; it helps them get to this thing, learn it on their own basis and their own speed. So whatever else happens, I already know it's doing good work. So one other piece to all this, this is a little negative and there's profanity in one slide. So please prepare yourself. You might recall these manuals that I did, which are now sitting in a coffee house. It made *The Atlantic*. Now, I get in the press a bit here and there. And it made *The Atlantic* and, of course, a lot of people linked to it. And there were a lot of people who linked to it who were archivists and other groups. And they said, "Oh my God, look, there's all this stuff." And then there was this, like, terms like hoarding, sarcasm. You know, people – and the reason I even knew about these was because I happened to search on the link, right? And because of that I could see that magical thing called the subtweet. And a subtweet is when you tweet about somebody, but you don't tell them so they can't tell they're being talked about. But you're doing it in a public forum, so surprise. And this person's a professional archivist. And so this guy wrote, you know, "Why would you promote this to the Society of American Archivists who link to my project?" And there were eight people who agreed. And all of them are professional archivists who are more than happy to sign on to say, "Yeah, why are you supporting people like this?" So I've thought about this. I've thought about what brings that in. And, you know, you can go after people who are professionals and whatever and make them feel whatever and call them out like I've done or whatever. But I've spent some time trying to realize, like, where did this backlash come from? Why are people angry at me, that I exist? Like, they were very unhappy that we had nonprofessionals going into archive and we had – you know, one of them was unhappy about the fact that some of the manuals were placed on the floor in the process of rescuing them in three days. You know? And there was a lot of anger about, like, who is this guy? I mean, let's be clear, I have a film degree. I make documentaries. I fell into archiving backwards and flailed because it was something that needed to be done. And I realized that this happened with Linux. See, when Linux came out an entire range, an entire industry of software engineering and developers and hardware makers were very angry about Linux. Because here they had spent 20, 30 years getting certifications, getting professional grade, going to conferences, and then some 14-year-old snot writes a device driver and everyone holds him up as if he cured cancer. And the reason why is because in Linux there's no payment so people would accolade you. That's the payment. But it was also true that maybe device drivers of that sort didn't need to be done by professionals. And so they were angry that I was able to get people to come in to field the archive. And I feel like there's two things for that reason. One of them

is that in archiving, the case hasn't been made for certain things being needed by professionals. If I had said, "Hey, everyone, in this warehouse in two days we're going to do some tracheotomies. I got some guys with apnea, about six of them, and we're just going to try it," people would be horrified. Because why – that's crazy. I mean, there are some subreddits that would think that was great and they'd do it. But there was – you know, there would be a horror. And there wasn't a horror of like, "What are you doing? Leave this to the professionals." And so the reason for that, I've decided, is because people don't give archivists enough agency. People who are in this or who are listening who are managers of archivists, I have found that archivists go wild on the parties at night; they go wild on the Twitter that their institution doesn't know about because they feel constricted that they can't make choices; that when they propose something, they're shot down; that their voices are papered over. And I happen to be very lucky with an incredible amount of agency to be able to do this. And I think that this industry as a whole needs to recognize that they have incredibly smart people who are very well trained and will do work that people like me can't do. Like, yeah, I can do some of this. But I can't do this properly. And the case will be made if you give those people that freedom. Don't be afraid. The thing that always concerns me when I do this work is how much fear there is when I talk to people. They say, "Oh my God, what's going to happen to you? Are you going to end up in Guantanamo Bay?" And the answer is no, but thank you for your concern. But don't be afraid. There's a lot of dread and concern about where we're going to go, right? So people – like I said, I feel very lucky in this work that I'm doing. This is how I approach software preservation – to make it ubiquitous, available, low-cost. I do not think that I replace a lot of institutions; I'm just the use case. We are in a situation where people have a lot of work ahead – real work – that needs to be done, boring work. I may have made some strides in here and others have made strides with me, and a lot of my work is dependent on others. I am the loud voice you can throw to, to say, "Well, yeah, he's a little bit annoying, but he has a good point. And let me make my case." And I hope – I hope that the institutions listen. Because we have a lot of work left in software preservation. Thanks a lot.

[ Applause ]

I think questions are later or people can ask me now, I guess. If anybody has a – does anybody have a burning desire to ask a question of me? Yes?

>> Yeah, I'm sorry [inaudible].

>> Jason Scott: Absolutely.

[ Inaudible ]

Yes.

[ Inaudible ]

Were we able to side step it? Well, so the number one question after any time I give this speech is IP copyright and why aren't you in jail? And the answer is, is

this was a exploratory move to see what would happen. Throw a hammer into the darkness and find out if it's a minefield or a zoo. If you hear a lion roar, it's a zoo; if you hear an explosion, it's a mine field. And what we discovered was what we discovered, which was a certain percentage of these companies said, "What are you doing? Stop." And then we stopped. And we went through, we tried to find every program or product that was still being sold and removed those. And every once in a while someone says, "Hey, I – whatever, please don't." I see what you're doing. You know? This is being done by a legitimate nonprofit – a registered library with their no ads. And what we've discovered is that the two things that companies really don't like is you being ahead of the search of their product and they don't like major confusion, like, where it goes. Like, if you look for Pac Man, we are the hit for Pac Man. And those are the biggest things that happen. Nobody has ever come after me for an Apple II program. Too long, 30 years. And in fact, the original creators often come in and write reviews of their own work about how awesome it is and what great work they did. Or they ask for help in how to get these other discs they have up or their original source code. You know? It's like a honey pot because it mentions them. So I can't propose that every place do this. We're not lawyers. We deal with what comes. But if you believe that this stuff is important, waiting any longer is now going to vitally destroy this material. Like, it is hard to read a floppy disc now. You don't always get it. We have made stronger and stronger readers and we have done technological solutions to make it even more involved. But on the whole, it's becoming harder and harder for people to just read an old floppy, to just look at an old cassette, to just look at an old cartridge. So we're moving and we're seeing what happens. We would like there to be copyright reform. Now, I could talk about copyright for the next 20 minutes and it would be unpleasant. Like, it would really be unpleasant. So I'll just summarize it to say we have perpetual copyright in this country. We have decided that intellectual property and its theft is a worse crime than murder or terrorism. And we are never going to fix it, ever. I don't care who they stick in as Library of Congress. I don't care who they say, what move they do, we're not going to do it. America has decided its two greatest exports are war and entertainment – not in that order. And we're never going to move away from it. And in fact, we are working aggressively to go after other countries that show anything away from a copyright regime. So that's where we live. Now, what do you in a world of perpetual copyright, of never-ending, where [inaudible] eggs will live for a thousand years, where Mario will once again have his obelisk? You move carefully, you try to make a change because you believe this needs to be done. That's why. Sorry to be a bummer. But it's always the way that moves in. And that position is not everybody's; that is certainly not the official Internet Archive position. Please think of me of the Internet Archive as Noam Chomsky is to MIT. But absolutely I'm like, "Okay, maybe the result of this will be I guess down the line software will never be allowed to be copy and it will all disappear." But I'm not going to go to bed and go to my death bed saying, "Well, at least I didn't make a fuss." Anyway. Yes? Yes, you, and then you, and then you.

>> So it's sort of a follow up. For all of that software that you pulled down when they said stop, stop, stop, does the Internet Archive have it somewhere and if 20 years later Nintendo goes away, 50 years later Nintendo goes away, they go, "Oh, hey, look. This stuff's not available," will you put it on now?

>> Jason Scott: Maybe. Next question. Was the question up there? Yes. Oh. Yes, you. Oh, yes. You're the winner.

>> Yeah, I see a Cloud model here. So when you say institutions, what are you expecting from them? Do you want them to work with you, or do you want them to handle the stuff by themselves?

>> Jason Scott: If somebody comes to an institution with paintings and you are a medicine archive and they're submitting the work of this great doctor who got a Nobel Prize and with it are 47 oil paintings the doctor collected because the widow gave you their entire estate, the first reaction isn't trash those. You know? You either keep them or you find a painting institution and you say like, "Look, we want to give the – we can't store paintings here. Can we give the Forester collection to you and we'll mark it all down as provenance?" But with software, I see institutions who are like, "Well, whatever these are. These aren't printed or what do we do with this?" And they do what they do with the printed stuff, which is they put it into a really nice box and put it up on a shelf and then let it become goo. And I'm trying to have institutions recognize that software is a valid thing and that there's a use case through this to make the stuff playable through BWFLA, through all these other projects. We can take software and we can make it live again, come back as if it was always there and be able to have people interact with it. So the use case is there. Because I've discovered institutions want use cases. Because I've discovered institutions want use cases. And to recognize software, CD-ROM's, and other media as a valid thing, the same as paintings, objects, art, tapes, you know? That's what I'm looking for personally. And yes, [inaudible].

>> It's not on.

>> Jason Scott: Now it's on.

>> Okay, great. Thanks so much for the presentation. That was wonderful. My question is not quite on software but the other digital content that you're working toward keeping. So I'm wondering if you have, like, a standard operating procedure are for the Geocities scenario. There's a lot of content on a third-party website that's suddenly at risk. Like, say Yahoo!'s going to be sold and Flickr's going to be shut down in 30 days or something like that.

>> Jason Scott: So I didn't want to go into that because I wanted to, you know, punch some haters in the face, but there's a whole variety of projects that I've been involved with. And the one that some people might now or not know of is Archive Team. ArchiveTeam.org is an activist preservationist group that goes in when sites are going to be shut down and we grab them as fast as we can. And we're very good at it now. So when we hear a warning, the Angel

of Death shows up in and this low-orbiting ion canon will suck out terabytes of data out of it. Because we found out that the users are the least important things to the programmer press complex of Silicon Valley, that they basically are like, “Well, we tried to open up an apartment building, it didn’t work, so we just blew up the apartment building. We’re going to try another building.” And they’re just forgetting all of this user-generated content that has valid meaning. And we don’t have any legal framework for it. We don’t have any tenancy laws. Like, if you give them data, it’s their data forever and a day. And so we go in and we grab it. And so we have a lot of use cases. And it’s a huge debate and everything else. But ArchiveTeam.org has all of our current projects. There’s about 200 people who work with that. And on any given day we’re downloading between 100 and 400 gigabytes of data off the Internet. This is separate from the Internet Archive that we’re going after. So Ted Cruz announces he’s not running anymore, we will run out, we will grab every Ted Cruz site. Somebody dies, we grab everything. A new feature is announced that’s amazing, we go after that site. If there’s protests or somebody writes something that is so far out you’re like, “This is doomed,” we’ll go out and grab a copy. For instance, last night at midnight the CBC always puts up its shows. And they put up a show that was a financial discussion program and they did not turn off the mics during the commercial break. So on their streaming service and the replay is some pretty randy discussion about how are we going to deal with this crap. And look, as soon as we can get this guest out of here and out to a martini, the better. So we grabbed it. It’s there. But it’s not just about embarrassment, it’s about saying, “Oh my God, the Space Jam site that’s been around for 20 years, let’s grab it.” Or discovering that there’s some obscure sets of technical documents. Let’s go grab this. Oh my God, it’s part of a college site that has 500 websites. And we grab them just to be sure that – because otherwise the conversation ends. Archive Team is a whole deal. We’ve been grabbing billions of URL shorteners for years because they’re the worst-single idea of the web. They’re one-time cryptographic pads for URLs. In 20 years if there’s Bit.ly.ab92 and there’s no Bit.ly, game is over for you. So we’re grabbing every Bit.ly URL to ensure that it survives. So someone’s doing it, we’re doing it. That’s what Archive Team is. So we’re going to rescue your crap – that’s our slogan. And I think this is the last one just because I don’t want – I mean, people have to –

[ Inaudible ]

Is it? Okay. If people get up to leave because they thought [inaudible] no hard feelings. Exactly.

>> Two more questions.

>> Jason Scott: Okay. Hit me up.

>> Back to software. You’re probably familiar that with scientific software one of the recent approaches is to put the software in a docker container of information, right?

>> Jason Scott: Right.

>> But usually the docker container to produce the environment, the word flow, the pipeline around later software itself. So I could see that's similar to the emulator approach, you know? In some ways.

>> Jason Scott: Oh yeah, later it's docker in JavaScript or what comes after, which is web assembly, which is the current thing, which nobody knows about. But it's out there. That's where we're headed. We're headed towards virtual machines that are meant to emulate these things.

>> Right. And so you think that is the next step from the docker container, then?

>> Jason Scott: Something that emulates a framework for the docker container to live on. I mean, it's always going to be – it's always going to be a nightmare when you deal with a work. Because these servers, these things – especially scientific – I mean, commercial products, you have a 143k image of an Apple II disc. And generally, you're going to be fine. Like, you're going to be able to get 95% of the experience of the Apple II program at this point. Pretty good. But some of these things, like scientific studies that were done on a specific hardware configuration and done with certain expectations, they're like desks. You know, they're almost like a human desk. So you're like, "Well, how do you preserve the desk?" Well, you took a picture of the desk. You took the components off it, you marked them. But maybe that person says, "Oh, yeah, never hit switch six," and nobody knows that. And it's gone. So they'll pull it apart a different way and discover, you know, switch six doesn't do anything. In fact, I wouldn't use that at all. And I've seen this. I've seen somebody take Prince of Persia source code that I helped rescue and reverse engineer how Prince of Persia works as a memory system, enough that the creator was like, "Oh, I do that. I was pretty smart at 19." Like, he doesn't remember. And the same way we're now getting to the point of, like, with art of, like, running MRI's against it or running other kinds of stuff to go, "Oh, look at the brush strokes. Oh, he really screwed this one up. Oh, this one looks like 200 years later someone decided no genitals today and decided to get rid of it." In the same way, I think software's going to have that attribute. Like, it reflects the human being. And while commercial software is relatively easy, scientific software, industrial software, and proprietary one-off custom software are going to be unbelievably hard to capture. They're going to be – they're going to need somebody within that group who has the mind virus to go, "Wow, we should think about 20 years from now and not just our funding to the end of fall. We have to do something. Can we just do a DV raw image of the hard drive and shove it somewhere and hope for the best?" We rescued a Cray operating system disc. It was 143 megs. And the way that it was done was somebody took – we had no Cray. This guy, Chris Fenton, let's give him credit. Chris Fenton took a Raspberry Pi and a magnetic reader and did a 20-gigabyte magnetic image of this disc and then was like, "I don't know what happens next." And I said, "Let's put it up on the archive. Why not? Let's make a little noise about it." For people who want the kind of Machiavellian trickery I do, I put up the wrong picture of the disc – the

wrong model – and people showed up going that’s the wrong model. And I was like you have a project. So be wrong and make people correct you; they’ll show up. And somebody in Sweden reverse engineered it in 48 hours and sent the files to him. So we didn’t have a Cray. We didn’t have a disc reader. We didn’t have – it was the platter that you shoved into the reader and somebody did it. But it needed nine different things to go its way. And that’s a lottery. That is essentially a lottery. So it’s tough. I would not pretend this isn’t tough. I just want it to be as tough as transporting a fragile piece of art or a painting that can’t hit sunlight. I want it to be something we do when we need to, which we don’t do now. And the last question. Yes, go ahead, sir.

>> Do you get the original source for a lot of these items? And if so, how do you make it available or if you make it available to the masochists who like to look through that stuff?

>> Jason Scott: So as time goes on – so the answer is we do get source code. And the way we get it is we’ll often get as both a print out or a text file. And if it’s a print out, we’ll put it up on a – because sometimes it’s an old piece of line feed paper or somebody gives us a disc image and goes, “This was my disc of source code.” And we pull it out and we’re like, “Wow, we don’t even know where the compiler is for this.” And the answer is sometimes it is useful and sometimes it is not. But it is digital now in some form. I’ve discovered that people flip you the when you use terms like saving, like, “We saved it, we rescued it.” They don’t like that. It’s heroic speech and it ticks them off. But we do our best to make it that the amount of effort needed to get to the next level is reduced. But all stuff needs in some way an advocate. I think bots will do some of the work, but ultimately it will take a random person going, “Well, how does this work?” You know? Sometimes, you know, people – especially as people are getting away from software engineering as a thing, like, now it’s very middlewarey. Like, I have up on my site in a collection oh, like, 80 Doom level CD’s that came out. And it is possible people have written JavaScript and other readers that will boot up and show you the layout, enable you to kind of see it. And in theory you could play it. Some of them are awful. Like, you’re talking about 10,000 Doom level CD’s. Now, contrast that with a Doom level CD that one of the Columbine shooters made before he was killed. He made a Doom level CD that was released. People will maybe want to see that in 30 years because of that ancillary thing. So that will drag along a bunch of Doom CD rescuing and stuff because people will be like, “How do we make this work and so on?” So it’s a lottery. Like I said, it’s really a lottery. We’re really behind. I really want to see us get better. But you’ll find edge cases all the way down the line, right? I mean, people start using terms. I’ve tried to learn all the archivist moon language that I can. So fixity, and, you know, provenance and being able to, like, understand where it came from. But sometimes it’s just an old guy gives me a thousand floppies and I go, “Thank you, old guy.” I don’t go, “Where’s my receipt?” I don’t go, “What’s the provenance?” I just go, “Thanks, old guy.” And up it goes. And a thousand floppies is, like, 80 megs. And then I, you know, put it as an attachment to somebody else and go look

at this. And that's that. Again, thanks to the two signers for doing the tough mutter marathon version of signing. People like me should not be given people whose job is to figure out what I'm saying at any given time. Thanks again.

[ Applause ]

All right.

>> So perfect [inaudible]. We're going to take a ten-minute break. And when we come back, we'll hear [inaudible].

[ Inaudible Chatter ]