Interview with Jeff Rothenberg

Jeff Rothenberg is an independent Information Science consultant, based in Santa Monica, California. A former researcher at The RAND Corporation (1984-2010), he has been involved in numerous projects concerning knowledge-based modeling, expert systems, interoperability, System-of-Systems acquisition, data fusion, information assurance, and the design of architectures and associated process models to support adaptive decision making in complex environments. He is internationally recognized as an early and continuing advocate of the need for active measures to ensure the long-term longevity of digital documents, records, and other artifacts.

June 17, 2013

Interviewers: Crystal Sanchez and James Smith

Please start off by talking about your involvement with time based media art and digital preservation.

My involvement goes back to the late 90s when I got involved with the National Archives and Records Administration (NARA) doing some research that was published in the *American Archivist*, looking at the impact of digital content in archives and how the digital revolution might be expected to affect archives, archivists, and the scholarly communication process that is considered the primary or most important use of archives.

So, if you think of the preservation community as consisting of three legs of a tripod—archivists, librarians, and museum curators, I started with the Archives leg, and then I did a lot of work with libraries, the Dutch Koninklijke Bibliotheek, the Library of Congress, the British Library, and a number of others, primarily deposit libraries whose responsibility is more like an archives. Then I did some work with museums, the Guggenheim, the Whitney, SFMOMA, and the Variable Media Network, which led to the *Seeing Double* show at the Guggenheim, based in part on our work with Grahame Weinbren's *Erl King* Project and what we called the 'renewal' of that artwork. So I've worked in all three areas and have some experience with the different perspectives that each of those three areas generate; they are not always quite the same, but they share a lot of concerns. In some ways, I've found the art world the most interesting and challenging because I feel that it stresses the preservation aspect of digital artifacts more than most of the other areas, unless you're talking about the preservation of software or games, which is an area that isn't always taken very seriously but is actually very similar to the preservation of artwork. I am a technologist by training, but I'm also very interested in art history. I nearly majored in art history in college, and my wife was an art history grad student, so I'm very interested in issues of long-term preservation in the art world.

That's really interesting because I come out of the archival world too and I am looking at some of the standards, technologies, and work that's been happening in digital preservation,

repositories, etc. [Crystal] What do you think about the relevance of that work and those standards when working with fine artworks?

This is an interesting question. My sense in the 25 years that I've been involved with this, which really goes back to the first era of talking about digital preservation, is that there is a certain amount of what engineers sometimes call 'looking under the light post for your keys.' It's a joke about searching for your car keys: you're looking under a light post, and someone asks, "Is that where you lost them?" and you say, "No but that's where the light is, so that's where I'm looking." There's a certain amount of that in that the libraries, archives, and museums communities: they have a tendency to look where they can get something started and get a handle on the issues of preservation. Frequently the first two places that have tended to catch peoples' attention are metadata and standards. Metadata is an obvious place for these communities, because it's an area that's familiar and that is central to the notion of curation and management of collections-records keeping, records management; so it makes sense. Similarly, once you look at the technical issues, it starts to occur to people that standards are perhaps a key issue. What has struck me all along is that leaping into those two areas has a tendency to miss some of the key technical depth of the problem, not bothering initially to try to understand what the technical issues are. One can, I think, get misled into thinking "Oh, if we just create appropriate metadata and define appropriate standards then that's all we need to do" whereas I think that can be premature.

So my focus has always been to try to back up a little bit before talking too much about either metadata or standards, and to ask what are the key technical issues involved in preserving these digital media, and only then, what are the implications for metadata, standards, and whatever else? If we're talking specifically about time-based media, which are an interesting subset of all digital artifacts, I like to lead off the discussion by suggesting that there is a notion of digital behavior which goes with digital artifacts, and I think it's easy to ignore the fact that at least some, if not all, digital artifacts—records, documents, artworks, games, everything that we create digitally—have some inherent digital qualities or attributes that make them distinct from traditional works of similar kinds. And I like to make a point that some digital things that we create are not essentially digital. That is, they don't have to be digital. They are simply page-images, that is, fixed, static content that could have been generated or recorded using traditional media—printing, or still photography, or whatever. Those things are not where the real problems lie.

The problems lie with objects that I like to call inherently digital, which means something that really relies on a program being executed by a computer. This is what a digital artifact is in its essence. And if we push the issue, I think we can say that all digital artifacts are to some extent inherently digital. Even a simple ASCII text file can be thought of as consisting of a sequence of commands that are trivial, like print a, print b, print c or display a, b, or c. So character codes themselves can be thought of as trivial commands. The point is that even the simplest digital object is really a program, and a program is a sequence of commands that are intended to be interpreted by some active process, which in this case renders the object or the artifact. Now the rendering doesn't have to be in 'real time' except for time-based media. I would make a distinction between time-based media and other media in the digital case: both require rendering, that is the interpretation of something

that is in some sense a program, a stream of digital commands, but time-based media require rendering in real time or at least capturing the rendering so that it can be played back in real time.

What does that mean for an artwork? Does that change the approach to the preservation of the work [as opposed to a different kind of digital asset] because of that necessary requirement of the real time, especially a generative work that is being created at the moment of display. Do you find that there is a different approach or any part of the approach that makes it different?

I would say that the essence of preserving anything that is inherently digital, something that requires running an interpreter to process a digital stream, is that we be able to run a computer to interpret the program that generates it, whether it's in real time or not. That certainly is a key issue because, as we know, programs and the computers that run them become obsolete, and the question is how can we guarantee the ability to interpret these digital artifacts correctly in perpetuity, or at least for the indefinite foreseeable future? That issue is there for both non-time-based and time-based media.

In the case of time-based media, the additional real time constraint comes in; we have to be able to do that interpretation and rendering in real time—or to do a rendering that is captured by something that is capable of reproducing the artifact's behavior in real time. Taking a sequence of photographs or frames of a process as it unfolds and then playing back those frames in real time would be one way of doing this for non-interactive time-based media. But for an interactive or self-generating time-based work, which many of these things that we're talking about are, I really see no alternative but to recreate or reproduce the original behavior in real time. So if we look at the Erl King project, for example, the project that I worked on with the Guggenheim, that was a wonderful test case, similar to the BBC Doomsday Book project in terms of technology and use—both were interactive, where the user could control the output and behavior of a complex set of video, still, and audio material, with no single playback path. That is, it isn't a film; it's an interactive video, so its behavior changes as the user sits there and touches the screen to determine choice points. I see no way to capture that in a meaningful way except to run the original code. There is a short video that Grahame Weinbren produced of the piece that shows a fixed sequence of one user interacting with it. That shows a few examples of branches that occur, based on a particular interaction with the piece, but if you really want to preserve the piece, you need to preserve all of its capability to interact and change in real time, and I think that's a real challenge.

When you're working on some of these pieces, for example, The Erl King*, do you find that there are things where if you were to work on a similar piece, maybe a reinstallation, that you would be able to draw strategies from* The Erl King *to a new piece? At what point does each piece then need its own specific attention?*

That's a particularly key issue, and I think it's one of the big distinctions between the museum world and the archives and libraries world, where museums tend to have the luxury of focusing attention on individual pieces to the extent that a library can't afford to do. Libraries do it to some extent, they rebind pages, they take care of the pages of old

manuscripts, but for the most part, approaches to preservation for corpuses of thousands or millions of works clearly have to be different from approaches that can be applied to single works. As a technologist, I have tended, as most technologists do, to look for more generic, universal solutions that could apply very widely, and I've made the argument that if you can find a universal solution, that would be preferable. We shouldn't give up the possibility right off the bat of finding more generic or universal solutions just because we might have the luxury or funding to focus on individual solutions. So even in the museum context, if someone could come along and say, "This is a generic preservation technique" that would be preferable even though the mind set in the museum world is "We have to understand the intricacies and the details of each individual work." It's a luxury to be able to do that, and if you can afford it and the work is worthy of that kind of attention and effort, all the better. But we shouldn't give up on wider ranging technical approaches.

So getting back to your question, yes, I think things that we did in *The Erl King*—and I wrote a lengthy summary paper on that project that gets into all these issues in some detail, including the generality and how applicable some of these approaches might be in the future—would be applicable to other cases. But you still have to evaluate each individual work to see if it conforms to the pattern of things to which these more general, universal solutions apply. I think you could say that The Erl King is close to the most general case—in that it combines digital logic and unique hardware with video and audio media. If it had involved a slide projector or audio tape in addition to video, then you would have had a true mixed media work where you have to worry about maintaining mechanical devices and materials in addition to the digital content.

In the case of *The Erl King*, the only content was video and audio and text, all of which was not originally digital, but we felt it was convertible into digital form without loss. Actually, the content was originally shot on 16mm film and then the 16mm was converted to video, but we didn't have the original 16mm A/B edit in usable form. The video was already a corruption of the 16mm because it used 3-2 pull-down to convert 24 frames to 30, and you can't really undo that in the analog video very well. Then there were issues about translating non-digital electronic media such as analog video and audio recordings into digital form; we did that in this case because we felt that digital was a much better long-term preservation mechanism than retaining the analog electronic media.

So we wound up with something that was essentially all digital. It did have a mechanical or physical interface. That is, it had a display screen, a touch screen, speakers, and an audio amplification system, but aside from that, the content of the work was entirely digital—so that simplified things to some extent. If you focus only on digital content of that sort—once you have digital content, then there are a set of fairly universal techniques that can be applied to preserving the data, in this case the data stream of video and audio and the logic or program that controls it and that allows the human interaction to control it. Of these, preserving the program is the hardest part because you need a way of running that program indefinitely; the approach we took was an emulation approach, which I've long been interested in as a means of indefinitely preserving a digital object.

As a follow-up question, you said that you like other technologists tend to gravitate to more generic solutions whenever they're available. Do you think at some point, as the amount of work being collected and in need of preservation increases, that Museums may be forced to have to accept more generic solutions even if they may be optimal from the traditional perspective of conservation?

That's more of a collections and curation policy question, in that, if you take the traditional model, museums have relatively large—not huge—numbers of artifacts of particular kinds, like paintings or textiles or whatever they collect, generally not millions, but manageable numbers compared to, say the number of documents or emails that are produced. So traditionally, museums have developed generic techniques for restoring and maintaining paintings and furniture and clothing that are in their collections. I'm not sure that the digital world is all that different. Just because it's relatively cheap to produce digital artwork—because there is no real material cost involved—the creative effort is still there, but what I'm getting at is that I don't see why museums would necessarily have larger numbers of digital artifacts in their collections than they have traditionally had no-digital objects, so the strategy for just how generic versus how individual you have to be in treating works for preservation I see more as a policy/funding decision that a museum has to make. Are you going to start collecting huge numbers of digital art or are you going to be selective about it? So far, the museums I've been in contact with—Guggenheim, SFMOMA, Whitney, Berkeley, and a couple of others—so far they tend to be quite selective in what they collect.

Again, I think generic solutions are better if you can find them, because they reduce your cost and effort and they give you a strategy that you don't have to reinvent every time you look at a new work. So to the extent that we can find universal or generic solutions, I think that's a good thing. At the same time, especially in the art world, we have to be open to looking at an individual work and asking, "Does this or doesn't this fit into our generic framework?" because there's always the possibility, particularly with artwork, that something unique is there which doesn't conform to our generic model and therefore requires some special handling. That will always be a human decision made by a curator or preservationist who knows enough about the artwork and its context to make that decision intelligently. Those are not decisions that technologists should make, at least not by themselves. So I would say that the drive for universal or generic solutions is a good one, it saves funding and moreover it provides a framework for approaching work, but we also need to be flexible to deal with things that may not fit into our framework.

Standards lurk at the edges of all of this because once you start talking about general solutions people immediately ask, "Are we talking about standards of some sort?" whether it's a standard process that we apply for preservation or a technical standard for media—the word media is so overloaded in this context that I even hesitate to use it, but media, in the sense of digital media. Let me lay out first the range of approaches that have been proposed for preserving digital things, and I emphasize that most of what's been written about digital preservation, with some important exceptions, has not taken the technical depth into account very well. I think this is because most people think of digital objects as Microsoft word files or JPEG or PDF documents that they've scanned, and all of those things

are static or what I call page image objects, that is, things that don't change and don't really rely on a computer in any serious way. They may be "born digital" and may never have seen paper, but they don't have to be digital. They could just as well have been created on paper or film. Those are what most people think of when we speak of digital objects. Preserving those isn't that challenging. The tricky part is when we get to things that are inherently digital, that really require a computer to render them in some way. Inherently digital things are more challenging because they involve behavior. This happens in games, which are considered something of a trivial example by many people but are actually a fascinating case, and similarly in digital art. Even in art and time-based media, there is often a focus on static time-based objects. A film or an audio recording is time-based, but it isn't interactive, so a digital form of that is not really inherently digital. You may be rendering an MP3 or MP4 file as an audio or a video experience but it doesn't really have to be done by a computer. You could record the output on some other non digital media and it wouldn't really lose anything. It's when you get to the interaction or the generative aspect of the work, where there isn't really any content except what's produced by a running program, that's where things start getting complicated. So we have people proposing various approaches for preservation, but they tend to be skewed toward thinking about more static page image kinds of objects, and they will often wave their hands and say, "This should work for time-based and interactive and generative works as well," but they haven't really thought that part through.

I've divided this set of proposed solutions up various ways over the years. I like a three part division. One is to convert a work into some kind of universal mathematical description, or formalism if you will—that is, to create a formal description of an artifact. That's an interesting idea that some people have taken fairly seriously. The idea would be to eliminate the work by creating a mathematical description of it that's so comprehensive and so effective that it replaces the work; then you keep that description for all time and you can use that description to regenerate the work at will. That's a nice idea, but I think it's far beyond our current state of capability, if only because digital technology is still evolving so rapidly that we really don't know how to formalize or characterize digital objects in that way. I think of this as an ultimate ideal but not very realistic for the near term future, so I dismiss it.

The second approach is one based on standards, where you use whatever existing standards there are for whatever kinds of media there are. You might use MP3 for audio and MP4 for video; you decide this is what we're going to use. If a digital artifact doesn't already happen to be in that standard form, then we have to convert it into that form in order to preserve it. Once it's in a standard form, that standard will last long enough that we can rely on it and we can rely on there being programs in the future that know how to understand and render that standard, so that's our preservation strategy. The problem with this, of course, is that standards don't last; they evolve and become obsolete. And of course, many works are not created in any standard form to begin with, so if you're going to use this approach, you're going to have convert these work into the standard form of your choice, and that conversion always involves corruption and degradation of some kind, so you no longer really have the original work. To some extent, we can be accused of having done that with the Erl King because we did convert analog audio and video into digital

form, which could be thought of as a kind of corruption, though we tried to do it very well and we worried about the outcome a lot. But there are problems with standards. Because standards evolve over time, you have to convert your standard form into a new future standard as it appears. That means that you're constantly converting things; you're basically playing the game of telephone where you're changing something each time you convert it, and it eventually becomes unrecognizable. Nobody would consider this approach for a moment in the visual art world. It would be analogous to saying that in order to preserve the Mona Lisa, you have another artist come in five years after DaVinci paints it, and they paint a copy and we throw away the original because (in the digital case) we can know longer use the original. Then five years later, we have another painter come in and make another copy, and, oh and by the way, they change media, so the original was oil but the next copy is watercolor, and then the next copy is acrylic. This would be laughable in the visual world, and yet in the digital world that's what people propose for using standards. So standards for the preservation of digital content has serious problems. Standards for preserving metadata and processes are a different story, but that's a topic for another discussion.

Finally, the third approach that has been talked about for preservation of content is emulation. The idea is to emulate the physical hardware that allows running the program logic for a digital object. By emulating hardware in software, that is writing a program that emulates a computer before it becomes obsolete, we basically turn hardware into software. So we turn a soon-to-become obsolete computer into a program, and that program hopefully runs on successive future computers, even as generation after generation of future computers becomes obsolete. The basic idea here is that preserving hardware is very difficult, computer hardware for example, but preserving computer software is relatively straightforward because it's soft, it's just a bunch of bits. Even though there are issues in doing this, theoretically it's very easy to preserve bits. So we can turn the hardware that we know is going to become obsolete into a bunch of bits. The content is already a bunch of bits, so now we just have a larger bunch of bits that we have to preserve into the future in such a way that we can still run the program. That's emulation. That of course does not address the broader context of the digital artwork, which is its interfaces. This doesn't preserve a touch screen or a speaker system or any special-purpose hardware that was designed as part of an artwork. You can't preserve that by making it digital. You can emulate that hardware to some extent, but then you really change the nature of the work. If an original work used a joystick to control something and you produce a future version of that work that uses a mouse instead of the joystick or some other device instead of the mouse, than you are in fact changing the behavior and the perception of the work.

So those are the three approaches that I see at this stage, and I can go into more depth on any of those if you like. I think people have tended to gravitate towards the second of those three approaches, thinking that if we can just get everybody to produce things in standard forms or if they aren't already in standard forms, convert them into standard forms, then we can preserve them because the standards will be self-preserving over time. The idea is that there are enough people using PDF or Microsoft Word or whatever your favorite format is that those will just live on automatically without having to do much. I think this is flawed, but it's the approach that a lot of people are taking.

That's really fascinating. It's a conceptual shift but almost back to an original art conservation approach to preserving the original object, which would be the content, and then emulating the hardware.

Part of the issue as you just hinted is that you have to define what the content of interest is. What is the work? What is it that we want to preserve? For example, in *The Erl King*, Grahame Weinbren was very clear that the hardware that implemented the work initially was never considered part of the work. This might be different from say, Nam June Paik, whose works included physical TV hardware that was presumably part of the intent. Museums that have his works go to great lengths to preserve the original CRTs and keep their color balances the same to try to make it look like the original. In the case of *The Erl King*, the hardware was instrumental but not part of the work itself. It was not visible in the original installations. The audience or user of the work (it was a single-user interactive piece) never actually saw the hardware except for the touch screen in front of them. When we renewed the work by turning the original computer into an emulator, it ran on a regular PC and at the Guggenheim, in the *Seeing Double* show, they did display the hardware: they showed the original version, which still ran on its original computer next to the modern, renewed version. The user could actually see the different sets of hardware. There was some question in the show whether we should do this because the point was not to focus on the hardware itself but rather on its behavior; but if we hadn't shown the hardware, the two versions of the work would have been virtually indistinguishable, and the point of the renewal effort would have been invisible.

So, you can ask, how much of the original machinery of a work is considered part of the work? How relevant is it? For example, if its disk drive makes noise, is that relevant? It's like asking about sprocket noise in a film projector. Is that relevant? We build soundproof projection booths and go to great lengths to avoid sprocket noise when we show a film, but some mixed-media artists like hearing the sprocket noise because it evokes the original medium. That's a decision for curators and the artist themselves. What is relevant to be preserved? Computer generated heat, is that important? Do we want the user to feel the heat of the machine that they're touching? These are interesting questions. I think for most purposes people would ignore these issues. But when you talk about digital art, you can't ignore any of these aspects. You have to at least allow any of those things to be a factor in the discussion because the artist could have cared about any of those aspects, and you could argue that a modern audience deserves to see the way the original work was presented, even if the artist didn't care. Bach was composing for instruments that had a particular flavor and sound. Modern audiences aren't used to hearing those instruments, they are used to hearing more modern strings on modern violins. I recently heard an outdoor concert on gut instruments by a group here that does early music. The performers were constantly apologizing that they had to keep retuning their instruments because they were outdoors and the humidity was affecting the tuning of the gut strings. You could argue that that's part of the experience of hearing that work, but do you want to preserve that or not? I don't know. If you were making a record of that performance, you would probably edit out the retuning between each movement.

The Guggenheim's Variable Media Initiative, led by Jon Ippolito and others, went to great lengths to try to capture information from the artist about what aspects of their works they considered important to preserve. So you try to get an artist's intent—what they consider to be important for preservation purposes. You can't always do that. If someone is dead or not communicative or not interested, you may not have that kind of information available. But somebody has to make those decisions, whether a curator, artist or whoever. Once you make those decisions, then you're in a position to ask "Do I care about preserving the mouse or the joystick, or the exact feel of the way the joystick worked, or the resolution on the screen, or the precise color?" Color is a terribly intractable area; it's very difficult to preserve color, but of course we try. Pip Laurenson of the Tate Modern once told me that even using the fancy CIElab technique for trying to preserve digital color is an uphill battle. But presumably we care about preserving color. So these are all questions that to some extent can be addressed generically but also have to be addressed in individual cases. What is important about a given work? What is relevant for preservation purposes?

You brought up the notion of the original. That's a subject that I'd like to talk about for a few minutes. When I first started thinking about digital preservation, a colleague of mine, Tora Bikson, and I discussed this at great length. What is an original in the digital sense? An original edition of a printed book is an individual copy that is presumably identical, except for incidental history—the fact that pages might be marked or the cover is coffee stained—to every other copy of a particular edition. In the case of lithographs or prints, where there are numbered editions, that's actually not true and we can distinguish between number three and three hundred. But for most printed works, an original is not a singular item, it's one of a set. Now for a manuscript or painting or other media, there is a singular original—an original artifact. So the question in the digital case is, what do we mean by "original" and is it even a meaningful term? Initially, we considered not using the term at all in the digital context because digital copies—unlike books—are truly identical: each copy consists of the same set of bits, the same bit stream. Mathematicians like to say that there is only one number two. All instances of "2" are the same instance, if you will. Of course, we can write "2" in different places—but it's always a representation of the same "2" (the Platonic ideal of "2"). In a sense, digital things are like that: a given bit stream, 11011, is 11011 no matter where it appears or how many times we copy it. So you could argue that a digital object has no original – any copy of it is exactly the same as the original. So what's the point of talking about the original? When you're talking about storage media or an actual manifestation of the bit stream, yes, there could be an original floppy disk or hard drive or CD that carried the bit stream. But the bit stream itself has no original. Many of us who have talked about digital preservation have come to the conclusion that when we talk about the digital aspect of a digital work, it is its bits, whether the bits are data or a program, that we really care about. We don't care about storage media that these bits happen to be stored on or that they were distributed on. We don't consider the CD that a digital work may have come on as a part of the original work. We might very well save that CD along with its jacket as an artifact in a museum, but the real digital content is not the CD itself, it's the bit stream that is on the CD, and the faster we get it off the CD onto something that's longer lived and easier to preserve than the CD, the better chance we'll have of preserving it.

So there's the question "What is a digital original?" An original is any copy of the object's bit stream. If we have not converted or changed that bit stream in any way, then it is still the original. This gets to the issue of conversion and standards and format change over time. For example, if you used the Word Star program 30 years ago, you might have had a Word Star document that you later converted into Word Perfect and then into Microsoft Word. Each of those conversions changed the format, so you no longer have the original. You have actually changed the bits in the conversion process. You might retain the core content that is the text of the document, but you're going to lose formatting—fonts, pagination, line breaks, justification, footnotes, cross references, embedded graphics, headings, footers, all sorts of things that may or may not be relevant for preservation purposes. Any time you convert the bit stream, you are giving up the original. After thinking about this for years, I realized that in the preservation world, we generally preserve originals even though we also produce all sorts of what librarians call surrogate copies—modern versions for general consumption. Museums have gift shops with postcards and posters that are reproductions of the artwork; these are not originals. They serve a useful function, since visitors can take them home; even scholars who study the works often use these kinds of surrogates because they can't stand in front of the painting all the time. For a lot of their research they use photographs or copies. We save the originals for very good reasons. First, there is what's often called the fetish of the original. We like to know there is an original that we can see or touch or get close to. Another reason is that we may generate a different surrogate for each generation, depending on the available technology and people's context and ability to interpret. For example, we create new versions of Homer every 20-30 years because the last version becomes less accessible to modern readers and seems old fashioned, since it doesn't use modern idioms or language. So every 20-30 years, scholars retranslate Homer into vernacular languages, like English. I refer to all such surrogates as "vernacular renditions". When scholars do this, they don't use the last vernacular rendition as their source, they go back to the original. (There is no original for Homer, but they use the earliest, most authentic version they can find.) We keep originals so that we can generate new vernacular renditions—surrogates—in the future and verify that they are faithful to the original. If we threw out the original we wouldn't be able to verify it this way. We do this with all media except digital. In the digital world, people are shockingly cavalier about giving up the original. They say, "Well the new format is just as good." Remember my Mona Lisa example: nobody would consider it a valid preservation strategy to copy the painting every few years and discard the previous version; but in the digital world, we do that routinely. My argument has been that we really need to preserve original digital works. In the case of artwork or games or anything that has complex inherently digital behavior, what we're actually talking about is software. We're talking about preserving software in a perennially executable state. This is the original core of any inherently digital object. In the case of artwork, this may not include all the peripheral interface equipment or special purpose hardware or other external devices that may be part of the work. The digital behavior or core is an executable program, and unless we're able to preserve that in an executable state for the future, we have not preserved the original. We might still generate future vernacular renditions. An example would be Grahame Weinbren's 30 minute video tape that was produced back in the 80's about The Erl King. If he wants to show somebody what this work was that he produced, he can't really reinstall it in their living room to show them, but he can show them the video, which shows someone using

the artwork and shows what a typical installation looked like. It gives a flavor of how the artwork behaved, but it is not the work. You cannot interact with it; you just watch it. I argue that in the digital world, just as in the traditional world and for exactly the same reasons, we need to both preserve originals and produce modern surrogate versions or vernacular renditions to make the preserved artifacts more accessible. If we're not able to actually run original software for these kinds of artifacts, then we have not preserved the original.

I think it's a very interesting way to frame this. I was really struck by thinking at what point can you create classes of components to try to generalize approaches to some extent? There are so many levels of fields that you bring to this discussion. You've got technology, art conservation theory and even philosophy and mathematics. How do you stay abreast of this? If you were to train somebody to think this way, what would be the key skills?

I almost triple majored in philosophy, art history, and math. Ultimately I didn't, because it was a little too impractical, but I have a strong background in art history and philosophy. I think someone who has been steeped in traditional art history and/or art preservation probably already has a lot of this perspective, but they haven't seen how to apply it in the modern digital context. It's more a matter of teasing out how traditional ideas about preservation translate into the modern digital context—as I think they do, largely. Issues include "What are you trying to preserve?" and "How much effort do you devote to individual aspects or attributes of a work?" We all recognize that some things change over time and there's nothing we can do about it. There are limits to what we can do. There's physical decay. So in the digital world, in emulation, people say—you can emulate the logic of an older system but you can't emulate the exact look and feel of the interface, the way the keyboard felt, the exact radiance of the screen. True enough, you can't do that perfectly. But you can't perfectly preserve the Mona Lisa either; you have to be willing to make some practical decisions, and I think that carries over. What is possible and feasible, and what is the intent? How do you define success or failure? If you've managed to partially recreate something, do you call that success or not? I think these are difficult questions. There are some technical things we can do. Digital artworks are the most complex case because they can include digital content, data and software as well as physical materials, including the computers they run on but also the interfaces, the physical devices or unique physical aspects the artist may include. That's sort of the worst of all worlds; it's the most complex case. You have to bring to that, not just an understanding of digital preservation but also materials science. It's a really stressing case for digital preservation, which is why I'm fascinated by it.

Games are a little similar. There's a very active gaming community out there that does try to preserve old games. Games are often not taken very seriously, but they actually are a prototype for digital artwork because they have a combination of complex interactive behavior and some kind of physical interface. These interfaces are not usually complex, but gamers worry about things like "Does this feel like the original joystick that I used with my Commodore 64?" They tend to talk about such things. Some preservationists laugh at this, but it's a good analogy to a work of art. Does using the work feel the same? I think gaming is

an interesting case for this reason. There's some interesting work in Germany on preserving games.

To me it sounds like you're making a strong case for emulation over migration. At the same time you say that migration remains the default option for a lot of people. Why is that the case? What is the obstacle against wider acceptance of the philosophy of emulation? What are the factors that cause people to default to this idea that migrating from one platform to the next to the next despite all the degradation that you've mentioned? Is it an economical issue, technical issue?

That's a question that I've thought about a lot, because for 25 years, every time I bring this up, it gets pushed back for a number of reasons. Migration, for people who haven't done it, seems quite straightforward and easy to do. People think "What's the big deal? We just take the content from one format and convert it into another." If you've done it over the years—and over my 40 year career as a computer scientist, I've done migration over and over again, on my own text files, my address book, and my own programs—you discover that migration is not at all trivial. It requires a lot of understanding of both the original and the new format in order to do it decently, and even when you do it carefully and with a lot of effort, it's never perfect. You always make compromises. And in extreme cases, we get paradigm shifts, and you can't migrate across a paradigm shift. For example, 40 years ago, the relational database model was introduced. Prior to that there were a couple of other database models, network models and hierarchical models, so people had existing databases, but the relational paradigm was so different that you could not really convert a network or hierarchical database into a relational database; it was not a meaningful thing to do. There was no way to really migrate it. You basically had to reconstruct and redesign your database from scratch, and that's what people did. When that happens, when there are paradigm shifts, migration is not really feasible, and these are show-stoppers for a preservation effort where you care about the long term, more than 10-20 years. You will encounter paradigm shifts. We currently think that word processing is what Microsoft Word does, but in 50 years we may very well have a quite different model for what we mean by documents, and it may become infeasible to take an older Microsoft Word document and convert it into whatever our new format is. People think it's straightforward--it sounds straightforward. And they also tend to think of simple page image kinds of objects, texts, or photographs and those things migrate reasonably well. The losses are not immediately obvious. I think that's why migration seems straightforward. In contrast, emulation, even though it's been in use for 50 years in computer science and is actually a very well known and well used technique, most people have never even heard of it let alone thought about it or tried to do it—though they have used it unknowingly. It sounds like smoke and mirrors, which it is, in a way. Computer scientists are very good at that; some of our greatest accomplishments involve smoke and mirrors! (laughter) You manage to take a piece of hardware and turn it into a virtual piece of hardware by writing a program that does what the hardware does, and then you run that program on some future computer, and if you're clever, you can figure out a way of writing the emulator so that any future computer can run it with minimal effort, such as writing it in a platform independent language like Java and then saying, "Write all your emulators in Java" so that all you have to do in the future is implement the Java Virtual Machine on any new computer, and you

automatically have all your old emulators running. Emulation is a well known technique, but to most people it just sounds like magic. So there is a lot of the push-back. People hear it and say, "That sounds crazy, how can you do that?" People who know a little bit more about it but haven't thought it through very far, tend to say, "Well emulation is never perfect, and you have to emulate all the interface devices and the peripherals." That is true. You also have to worry about speed, to make sure the original speed is the same—especially for time-based media—getting back to the real-time issue. So yes, there are some technical issues, but I think they are at least worth trying to address. There are some efforts out there that have been working on this. The KB [Koninklijke Bibliotheek national library] in Holland has taken emulation quite seriously. The British Library has done some work on it, as well, so there are some efforts out there. In the art world, I think it's even more important because the behavior is so crucial to at least some digital artworks.

It also seems like a point of departure from the standardized practices some other tangential fields like archival and library fields, maybe because of the nature of the materials?

In terms of the inherently digital attributes of their corpuses and what they care about, I place these three fields in the ascending order, archives, libraries, and museums. Archivists, particularly national archivists, have a very precise and relatively simple definition of what they mean by records, and archives are places that maintain records. A record is a document of some sort that gives evidence of the behavior of some organization in the performance of its mandate or its function. That's what a record is. The incidental aspects of the record, what ink it's printed in, the font, what it looks like, are all profoundly irrelevant to the notion of a record for most archivists. Migration of records for an archivist is no big deal, so long as you've retained the provenance of the record—that is, the history of its origin and stewardship, which is key for retaining authenticity—and also its content. That's all they really care about. They don't care about look and feel for the most part. Libraries are different; particularly deposit libraries, which are a little more of a museum. They not only have books and publications in them but they have a historical function of retaining the originals as deposited with them by publishers. And of course national libraries also have something of a museum quality. So do national archives. There's a bit of a blurring--some artifacts, like the Declaration of Independence, which is in the Library of

Congress, are certainly considered historical objects. But generally, libraries are in the middle and museums are on the far end and are concerned with artifacts, many more of whose attributes are potentially relevant. Yet another reason for preserving originals is that some of the attributes of a work may not even be definable until the future. For example, if we talk about the chemical content or the radioactive signature of the paper in a book, those concepts didn't even exist a couple of hundred years ago so nobody could have worried about preserving those particular attributes. This gets into the metadata issue and standards. People talk about describing the objects in their corpuses in metadata and capturing all of their relevant attributes, but you can capture only those that you know about. Somebody may come along in the future and ask what the Carbon-14 signature of this paper was and if you didn't know what Carbon-14 was when you created the metadata, you're in trouble. Preserving the original preserves as many of its attributes as possible. Not all of them, because some of them do decay over time (Carbon-14 for example, literally

decays over time) so you can't be perfect, but you can at least preserve as much of the original as you can. That's the safest thing you can do if you're talking about preservation. This includes things like subtlety of color and resolution and time sequence. If we really cared about preserving The Erl King in its original form, which used video, we would have tried to preserve the effects of the 3-2 pulldown process, which generated the video from its original film source--that horrible, absurd process! It creates subtle anomalies that you can sometimes perceive when you look at a movie, occasionally you can see a slight jitter that is caused by the 3-2 pulldown: it's usually invisible, but not always. You can argue that we should preserve this jitter because it was part of the original work; but it wasn't part of the artist's intent. The artist shot the original material in 16mm and might have been delighted if that could have been used directly, but it couldn't, because the work required random access to small snippets of the content, which film cannot do. So do we care about preserving these anomalies or not? We decided not to. You would worry about that if you thought it was relevant. If you clean a work up too much, you may be distorting it in some way.

I have friends who like to collect old 78 rpm records and say that the scratches add to the authenticity. It depends on what you're after. The scratches weren't part of the original. They're part of the history that the original went through to get from its original state to you. They're the patina that grew on the original over time. You have to decide which of these things you're going to try to preserve and what your rationale is.