

# The Electronic Media Review

Electronic Media Group

Volume Three 2015

Papers presented at the Electronic Media Group session of the 41st AIC Annual Meeting, Indianapolis, Indiana, 2013, and the 42nd AIC Annual Meeting, San Francisco, California, 2014.

Jeffery Warda and Briana Feston-Brunet, Managing Editors

Edited by Helen Bailey, Briana Feston-Brunet, Karen Pavelka, and Jeffrey Warda

Volume Three Copyright © 2015  
Electronic Media Group  
American Institute for Conservation of Historic and Artistic Works  
All rights reserved by the individual authors

Layout by Amber Hares  
(Original design by Jon Rosenthal, JonRosenthalDesign.com)  
Typeset in Trade Gothic LT and Myriad Pro

American Institute for Conservation of Historic and Artistic Works  
Washington DC

*The Electronic Media Review* was published once every two years in print format by the Electronic Media Group (EMG), a specialty group of the American Institute for Conservation of Historic and Artistic Works (AIC), until 2013 and published online only thereafter. *The Electronic Media Review* is distributed as a benefit to members of EMG who held membership during the year of the issue. Additional copies or back issues are available from AIC. All correspondence concerning subscriptions, membership, back issues, and address changes should be addressed to:

American Institute for Conservation of Historic and Artistic Works  
727 15th Street NW, Ste. 500  
Washington, DC 20005  
info@conservation-us.org  
<http://www.conservation-us.org>

*The Electronic Media Review* is a non-juried publication. Papers presented at the EMG session of the AIC Annual Meeting are selected by committee based on abstracts. After presentation, authors have the opportunity to revise their papers before submitting them for publication in *The Electronic Media Review*. There is no further selection review of these papers. Independent submissions are published at the discretion of the EMG Publications Committee. Authors are responsible for the content and accuracy of their submissions and for the methods and materials they present. Publication in *The Electronic Media Review* does not constitute official statements or endorsement by the EMG or by the AIC.



## **CREATING A PRESERVATION AND ACCESS FRAMEWORK FOR DIGITAL ART OBJECTS**

**DESIREE ALEXANDER, MADELEINE CASAD, DIANNE DIETRICH**

### **ABSTRACT**

In February of 2013, with the support of the National Endowment for the Humanities, Cornell University Library and the Cornell Society for the Humanities began a 2-year project to develop a preservation and access framework for complex, interactive born-digital media art objects. The test collection for this project comes from the Rose Goldsen Archive of New Media Art, part of the Library's Division of Rare and Manuscript Collections, and includes more than 300 interactive born-digital artworks created for CD-ROM and DVD-ROM optical media, and web distribution. Though vitally important to understanding the development of media art and aesthetics over the past two decades, these materials are at serious risk of degradation or obsolescence. Many date back to the early 1990s, and are currently unreadable without legacy computers and software. The project's goals are to create preservation frameworks, workflows, and access strategies that will be scalable and transferrable to other kinds of complex digital collections, yet grounded in a thorough understanding of media art researchers' needs and priorities. This paper describes some major findings of the project so far, focusing on disk imaging practices, investigation of emulation environments, and survey of users of media art archives.

### **BACKGROUND OF THE TEST COLLECTION**

Though we anticipate that our findings will be of value to other kinds of complex digital media collections, this project focuses on interactive born-digital artworks in the Rose Goldsen Archive of New Media Art. As a research archive located in

the special collections unit of a major academic library, the Goldsen Archive was founded with a clear mission to offer the broadest possible access to artworks and documents of their historical contexts, for research and teaching purposes. The Goldsen collections now include extensive holdings of analog videotape in various formats, paper archives from major media art foundations, exhibition catalogues, monographs, and ephemera, in addition to the interactive born-digital materials that comprise the test collection for our preservation project. The test collection includes artworks created for personal computers as well as interactive small screen adaptations of installation works, research compendia of the works of major media artists, and offline versions of artworks created for web distribution.

#### NATURE AND SIGNIFICANCE OF THE ARTWORKS

Even before the advent of advanced, high-bandwidth digital networks, the interactive, multimedia, nonlinear capabilities of digital media technologies made them an attractive medium for artistic expression. Yet the same elements that make such works so rich aesthetically also make them complex to preserve—far more so than assets that adhere to a uniform digital file type. An interactive digital work usually comprises an entire range of digital objects, each with their own risks and dependencies; these may include media files, applications to coordinate between them, operating systems to run the applications, hardware on which to mount the operating systems. If any part of this complex system fails, the entire asset can become unreadable.

As an example of the complexity of the works in our test collection, consider Norie Neumark's (b. 1947) *Shock in the Ear* (1998, CD-ROM). Created for small-screen individual interaction, this artwork engages all the senses of the user as they navigate through stories, painterly visuals (by Maria Miranda, b. 1956), and immersive soundscapes. The screen's responsiveness to cursor movements offers the user some degree of agency in orchestrating their own experience. At the same time, the

work invokes two notable modes of arbitrariness: first, in the algorithmically random appearance of sequences from different storylines, and second, in the occasional imposition of waiting periods on a user before they may advance to a new screen, which obliges the user to experience the work at a circumscribed pace. Challenging the presumptions of productive clicking or rapid surfing from one link to the next, *Shock in the Ear* slows down the motions to better engage the user in a contemplative experience. These techniques enable Neumark to explore thematic concepts of trauma, dislocation, and the eerie calm after the shock of traumatic experience—a calm that is shattered when memories of trauma present themselves once more.

Access to such artworks is necessary for an appropriately nuanced understanding of contemporary media art history. The Goldsen Archive's complex born-digital holdings span 20 crucial years of development of such interactive interfaces, from about 1993 to the present day. At the start of our grant project, however, approximately seventy percent of the CD-ROM artworks in the Goldsen collection could not be viewed at all without access to now-obsolete hardware and operating systems.

#### THE PRESERVATION WORKFLOW

Preserving these artworks can be viewed as two separate, yet interconnected, actions. The first action is preserving data from the work's original storage media. The second action is preserving (or restoring) access to the work, which may include running the original software (or viewing the work's original files) on an emulated or virtual machine set up to match the work's stated system requirements. These actions are interrelated, because the method of data transfer can directly affect the ability to provide proper access to the work.

The most time-sensitive action in our preservation workflow was migration of the data from optical media to more stable storage media. CD-ROMs are inherently fragile: the lifespan of a prerecorded CD-ROM is dependent

on multiple factors, including environmental conditions of storage and handling and even the original manufacturing process (Shahani et al., 2009). While many of the CDs in the Goldsen collection fall into this category and are pressed silver CDs, a significant number are burned retail CD-Rs, which may be even more fragile and susceptible to decay than silver CDs. Leaving the works in their original form—or even making additional copies on the same type of media—was not a viable or sustainable plan for long-term preservation.

The OCLC report *You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media* (Erway 2012) outlines two distinct strategies for migrating data from fragile media. The first method is to copy files from the source to a suitable destination (for long-term storage); the second is to create a sector-by-sector image of the source media. Our initial review of the stated system requirements for works in our test collection immediately indicated that the first strategy—simply copying files—would not adequately capture the data we would need in order to support restoration of access by emulators or virtual machines. Many works in the collection were compatible on both Apple and Microsoft Windows operating systems. Subsequent analysis confirmed that these discs supported multiple file systems concurrently. From a user's perspective, this means that when viewing the work on a PC running Windows, one subset of the disc's total files would be visible, and when viewing the work on an Apple computer, another slightly different subset of the disc's total files would be visible. In either case, making logical copies of the source media would result in an incomplete representation of the original source material, since only one file system's representation of the files could be copied at a time. Sector-by-sector imaging of source media is the only transfer method that captures the full structure of the original disc, including all supported file systems present on that disc.

We evaluated a number of different disk imaging programs to determine which was most suitable for our

collection. We evaluated only those tools that would generate a sector-by-sector image of the original optical disc. In the end, we opted to make disk images of data-only CD-ROMs (which constitute the majority of our collection) using Guymager forensic imager, and the Microsoft Windows-based tool IsoBuster for more complex discs that have multiple sessions or include both data- and audio-reader formatted content. The Guymager software automatically writes out an extensive log file for each disk imaged, an enormous benefit from an archival and preservation perspective. The log file includes the time and date of imaging, technical specifications of the source (in our case, the CD drive used), and checksums of the image file and source file to prove that the image file created is an accurate copy of the original CD-ROM.

#### THE UNEXPECTED IMPORTANCE OF EMULATION

One of our original assumptions at the outset of the project was that emulation would not be a viable strategy for preservation of access to these materials. We soon discovered that emulation is a far more viable and scalable option than we originally suspected. Initially, we tested out two emulators to simulate obsolete Apple hardware, BasiliskII and SheepShaver. After some trial and error, we were able to build the software successfully. We then loaded various disc images in emulated environments (simulating a number of obsolete Apple Macintosh operating systems, from OS 7.6 through OS 9) and found that it was possible to interact with the works in a manner that closely approximated our “control” interaction on a legacy machine that ran the older OS natively. We noted some patterns in emulation quirks, including works that performed much faster in emulation than on the original (due to faster processor speeds on modern machines) and some infidelities in colors (often a product of viewing the work on an LCD screen rather than a CRT screen).

We further tested emulation by creating virtual machines with Windows-based environments using QEMU and Oracle VM VirtualBox and loaded disc images of works that ran under Windows. We noted some of the same is-

sues observed when testing the Macintosh emulators, but again observed that, overall, many of the works ran smoothly under emulations without extensive fine-tuning.

It is important to note the level of technical knowledge needed to fully test these emulators. Specifically, for BasiliskII and SheepShaver, it was extremely important to read through documentation and bug reports on the project's GitHub repository in order to understand the current limitations of the software. As our project continues, it will be important for us to keep track of any changes made to the software—especially changes that fix previous problems or issues with the software—and update our emulators accordingly. In our experience, while it would not be necessary to understand every component of the source code for this software, some level of technical understanding was necessary to compile it for our specific environment. Recognizing this, we have been excited to note the development of programs like the bwFLA project (Baden-Wuerttemberg Functional Long-term Archiving and Access), Emulation as a Service. We have learned much from developing an emulation program and archiving strategy, but expect that a model like that of the bwFLA project is likely to be an attractive option for the future, especially for smaller institutions without digital archivists or digital conservation staff expertise.

Though we initially considered emulation as an access strategy, we have also discovered that emulation has functioned as a valuable conservation tool. To give an example, discs in our collection that are formatted for Macintosh computers most often are formatted using the HFS file system. One notable feature of this file system is the existence of a resource fork associated with a file, which can include information about how a file is displayed in a Finder window on the Macintosh operating system. Specifically, this can include a custom icon to display with the files, and the coordinates for where the file should appear in a Finder window. The technical metadata that we create for files on an HFS-formatted

volume includes the size of a file's resource fork, but not what that resource fork contains. On one disc, we noticed the presence of many files whose filenames were a unique combination of whitespace characters, and were empty except for a resource fork. This pattern typically suggests decorative icon files (since the resource fork can specify an icon's placement on the screen) but the presence of multiple files with essentially no visible name puzzled us. Only when we viewed the work in emulation did we see that the "whitespace" files were meant to function as a mosaic when viewed in Icon View in a Finder window. Similarly, through emulation we have been able to identify additional aesthetic components of a work that would not have been visible through technical metadata alone.

#### **TECHNICAL CLASSIFICATIONS: RECOMMENDATIONS AND DISCOVERIES**

Identifying technical classifications with clear preservation recommendations is one key element of our proposed scalable preservation framework. Over the course of our project, we noticed that many of our initial classifications were not necessarily mutually exclusive, and that their preservation and access implications were more nuanced than we originally perceived. In our initial test of works in the collection, we noted several key technical categories. One of the most striking technical distinctions was the presence (or absence) of executable files on a disc. It was tempting to suggest that access to works with executable files would, as a rule, constitute the more challenging use case, and that works consisting of web-related files (e.g., HTML, image files, audio, or video) would be more straightforward. In theory, web-related files could be run in a standard browser and would not be reliant on specific hardware, as, for example, executable files often are.

As we became more comfortable and familiar with the process of setting up emulators and virtual machines, however, it became apparent that, in many cases, works that involved executable files with no other software

dependencies were quite straightforward to access. In contrast, a work consisting of web-related files could be quite challenging. In one case, we found a work that was intended to be run in Microsoft Internet Explorer, but also included embedded video, audio, and used “world files” written in Virtual Reality Markup Language (VRML) for navigation. While the plugins for the video and audio were still available on the web, the recommended VRML plugin ceased development in 1998. We could technically still install the older plugin on a modern system, but this caused conflicts with other plugins on our system—to such an extent that launching the web browser completely froze the machine. Ultimately, this one plugin determined our conservation and access strategy for the entire artwork; to run the work, it became easier to install all necessary plugins on a virtual machine running a much older version of Windows.

Noting that web-based works could be considerably more complicated than executable-based works, we looked deeper at the source for the web-based works. Even without any external plugin requirements, web-based works may be coded in a way that is no longer standard. These works may not properly render anymore on a modern browser. Even if emulator dependence is not immediately indicated by an initial technical analysis of a digital asset, providing patrons with access to an older browser in a virtual machine may provide a closer aesthetic experience to the original.

We have begun drafting classification documents to address both broad and specific properties and recommendations discovered through our analysis and experimentation. Often, works will reside in several categories, where implications for access of one classification (e.g., an executable-based work) will be modified and informed by another (e.g., dependence on multiple external plugins in order to access the work’s audiovisual content). As our investigations continue, we plan to refine, alter, and add new categories and classifications as needed.

## EMULATION AND OUR PRESERVATION FRAMEWORK

We have discovered that emulation may be part of a first-pass technical analysis very early in the preservation workflow, as in the scenario described above. It may also, of course, be a way to provide access to digital materials. In either case, we recognize that our increasing commitment to a preservation and access strategy that involves the emulation of obsolete operating systems will profoundly impact our preservation metadata framework, which aims to capture essential information needed to provide access to artworks as well as requirements for ingesting them to the Cornell University Library Archival Repository (CULAR), for long-term preservation.

In some sense, we have expanded the scope of our preservation project to include emulators themselves, as well as media artworks. This means that we must, for example, capture preservation metadata specific to emulators and we must document the compiling environment and compiling process used to set them up, noting any adjustments made by a digital conservator. We must confirm rights information, and also conduct a long-term risk assessment for the emulator itself. At the same time, the technical metadata we capture from the artworks will be informed, in part, by emulator requirements or dependencies. The descriptive metadata we record for artworks will also need to note recommended emulation environments for each work, as well as any rendering infelicities a user might encounter when accessing the artwork in such environments.

We consider it a given that such rendering infelicities *will* occur in emulation environments. Indeed, as we readily acknowledge, some audiences may consider the fact of embracing emulation as an access framework to be infelicitous from the start. By its very nature, however, our project embraces the real limits of large-scale institutional archival work. Our goal is not to painstakingly approximate a perfect rendering for a select few artworks, but to offer the best-possible *imperfect* rendering for the widest possible range of works in our test collection. To

do so effectively requires that we both acknowledge and document the distance between the emulation and a more ideal experience “authentic” to any given artwork.

### **SIGNIFICANT PROPERTIES, AUTHENTICITY, AND ASSESSING USER NEEDS**

In order to better understand and anticipate the needs of researchers, and address them adequately in our descriptive metadata and preservation priorities, we created and distributed a questionnaire in January 2014 to assess the interests, preferences, and needs of media art researchers and curators. The online questionnaire was disseminated via email to media art, art history, digital library, and digital humanities listservs as well as to artists and researchers with personal connections to the Goldsen Archive. The questionnaire was very qualitative, designed to take a “sounding” of a diverse group of media art users representing a variety of disciplines and contexts. Questions tended to be open-ended and aimed at eliciting information about respondents’ priorities, preferences, frustrations, and desired ideal scenarios when working with interactive digital media artworks in archival settings.

We initially hoped that our results would enable us to create user studies or profiles that would, in turn, help to guide our metadata framework and access strategies. Instead, our results confirmed a range of concerns and priorities that did not necessarily resolve into tidy profiles, but that were nonetheless extremely instructive (Rieger and Casad 2014).<sup>1</sup>

On the broadest and most general level, questionnaire responses indicated a tension between desire for accessibility and desire for authenticity. This may seem to be stating the obvious, but our responses led us to a much more informed, and much more nuanced, sense of what goes in to creating a sense of authenticity for media art users. Authenticity can be a contentious or problematic concept, and is especially so when discussing born-digital artwork. It is, however, an important and useful way of denoting an archive patron’s faith that the

curating authority has taken into account a wide range of considerations. These considerations include, for example, accurate preservation of the work’s significant properties, fidelity to the artist’s vision or intention in creating the work, and acknowledgement of the work’s historical contexts.

During our initial project planning stages, we had naively understood authenticity largely in terms of the need to preserve, if only through documentation, an artwork’s most significant properties. Our survey results were an excellent reminder that a user’s sense of authenticity in fact derives from many overlapping factors, and that different kinds of cultural heritage institutions will have different potentials and limitations in their capacity to provide a strong sense of authenticity to visitors. Because of the Goldsen Archive’s unique position as a media art collection within a major research library, we knew from the start that we would be limited in our capacity to provide perfectly accurate renderings of artworks’ significant properties to users. At the same time, we were confident in our capacity to provide excellent access to artworks’ historical, cultural, and media-technological contexts.

Our questionnaire results alerted us to the fact that we had not taken enough action to include input from artists in our preservation and access framework. With this in mind, we recognized the need to add artist interviews to our preservation workflow. Toward this end, we are developing a simple questionnaire and follow-up interview procedure.<sup>2</sup> In addition to asking for the artist’s own evaluation of the artwork’s most significant properties, we will inquire about:

- Original hardware and system requirements
- Original compiling environments and software
- Whether the artist has already adapted the artwork for contemporary operating systems
- Whether the artist would share (or has already shared with another institution) original working files or source code from the artwork

- Whether the artist has archived any online aspects of the artwork or would have an interest in doing so

The artist survey and interview will provide an opportunity to disclose our intention of using emulation as an access strategy, and inquire about the artist's position on emulation, especially with regard to specific emulator shortcomings that might be anticipated (for example, alterations of color fields, changed animation speed, or a zero-sum relationship between sound and video quality). We will also revisit deposit agreements with our new access framework and conservation strategies in mind.

### NEXT STEPS AND SHARING RESULTS

Our hope is that the scalable preservation and access framework and workflows we develop through this project may be of value for other institutions and other kinds of complex born-media collections. Toward that end, we will publish a full white paper at the close of the project term. This will include final versions of our preservation and access framework and workflow, full versions of our questionnaires, and recommendations for similar initiatives.

### ACKNOWLEDGEMENTS

This project has been made possible by a grant from the National Endowment for the Humanities (NEH). Any views, findings, conclusions, or recommendations expressed in this paper do not necessarily represent those of the NEH.

The authors also wish to thank the project advisors, partners, and staff, and acknowledge their invaluable support throughout this investigation: Alex Duryee, Dragan Espenshied, Ben Fino-Radin, Jean Gagnon, Rebecca Guenther, Matthew Kirschenbaum, Jason Kovari, Jon Ippolito, Chris Lacinak, Danielle Mericle, Liz Muller, Timothy Murray, Norie Neumark, Michelle Paolillo, Christiane Paul, Oya Rieger, Richard Rinehart, Kara Van Malssen, Simeon Warner.

### NOTES

1. A more complete report to Rieger and Casad 2014 is forthcoming. Contact the authors for more information.
2. Models like those of the Variable Media Network and Variable Media Questionnaire have been enormously instructive during this process. See [www.variablemedia.net](http://www.variablemedia.net).

### REFERENCES

- Erway, R. 2012. You've got to walk before you can run: First steps for managing born-digital content received on physical media. Dublin, Ohio: OCLC. Available at <http://oclc.org/research/publications/library/2012/2012-06r.html> (accessed 09/15/14).
- Rieger O. and M. Casad. 2014. "Interactive digital media art survey: Key findings and observations," *DSPS Press* (blog), Cornell University. <http://blogs.cornell.edu/dsps/2014/07/30/interactive-digital-media-art-survey-key-findings-and-observations/> (accessed 07/16/15).
- Shahani, C.J., M.H. Youket, and N. Weberg. 2009. Compact Disc Service Life: An investigation of the estimated life of prerecorded compact discs (CD-ROM). [http://www.loc.gov/preservation/resources/rt/CDservicelife\\_rev.pdf](http://www.loc.gov/preservation/resources/rt/CDservicelife_rev.pdf) (accessed 02/03/2016).

Desiree Alexander  
Collections Analysis Assistant  
B76 Kroch Library  
Cornell University  
Ithaca, NY 14853

Madeleine Casad, PhD  
Curator for Digital Scholarship  
Associate Curator, the Rose Goldsen Archive of New Media Art  
2B Kroch Library  
Cornell University  
Ithaca, NY 14853

Dianne Dietrich  
Associate Librarian  
DSPA Fellow for Digital Forensics  
283 Clark Hall  
Cornell University  
Ithaca, NY 14853