2-9-2018

The Copyright Permissions Culture in Software Preservation and Its Implications for the Cultural Record

Patricia Aufderheide
American University

Brandon Butler
University of Virginia

Krista Cox
Association for Research Libraries

Peter Jaszi
American University Law School

The Copyright Permissions Culture in Software Preservation and Its Implications for the Cultural Record

ASSOCIATION OF RESEARCH LIBRARIES

Table of Contents

By Patricia Aufderheide, Brandon Butler, Krista Cox, and Peter Jaszi[1]

Funding provided by the Alfred P. Sloan Foundation

Abstract

This report summarizes results from research with the professionals who make up the software preservation community about how their understanding of copyright intersects with their preservation mission. Professionals typically face significant challenges from perceived copyright barriers. They tend to assume that a license or other express permission from a copyright holder is required before embarking on a wide variety of preservation activities, and typically find that such permissions are difficult or impossible to obtain. In the absence of reliable information to guide informed risk assessment, professionals act on the reasonable assumption that high levels of legal risk could be associated with activities that potentially implicate copyright and related doctrines. As a result, they often forego and postpone essential preservation activities, and establish access policies for collection materials that strictly limit scholarship. Preservation professionals have actively explored opportunities for collaboration and resource-sharing, but their prospects are clouded by legal uncertainty. At the same time, professionals are frustrated and deeply concerned that over-conservative approaches are limiting access to software and software-dependent works, imperiling the future of digital memory. The community has so far had little access to information or expert advice about alternatives to seeking permission, and in particular about the fair use doctrine, which allows the use of copyrighted materials without permission from the copyright holder under certain circumstances. Developing a shared understanding among preservation professionals of best practices around employing fair use to achieve their preservation and access mission will facilitate their work.

---

1. Patricia Aufderheide is university professor in the School of Communication at American University, and coauthor with Peter Jaszi of *Reclaiming Fair Use: How to Put Balance Back in Copyright* (University of Chicago Press, 2011); Brandon Butler is the director of information policy at the University of Virginia Library; Krista Cox is director of public policy initiatives at the Association of Research Libraries; Peter Jaszi is an emeritus professor at American University Law School. The authors would like to acknowledge the Alfred P. Sloan Foundation for its generous funding of this research, as well as the Software Preservation Network, and in particular Jessica Meyerson, for support and insight into this dynamic community. And of course, we thank the professionals who gave generously of their time and expertise to inform this work.

Introduction

Memory institutions share a common, fundamental mission to preserve and share knowledge, usually in combination with providing access to information. Christian Dupont has suggested that libraries, archives, and museums can be grouped conceptually around the theme of memory because they all exist to "make a better future by helping us remember and understand the past."[2] As Helena Robinson has put it, these institutions are "aligned in the basic function of accumulation and preservation of information, much (but not all) of which concerns the past."[3] Notably, this mission entails not only the accumulation of materials, but also the task of "producing sophisticated ways of selecting, classifying, organising and enabling streamlined user access to collection information."[4] With the spread of digital access technologies, the specialist and generalist researchers who depend on memory institutions expect to consult that information in more convenient and less mediated ways. And yet, the complexity of digital objects necessitates new, specialized forms of intervention and support to ensure access.

In the words of Jessica Meyerson, "Our cultural record is increasingly made up of complex digital objects."[5] In the digital age, much information, knowledge, and culture is "born digital" and relies on software in its creation and to render it accessible. Software preservation has become a critical link to ensure that knowledge created today is not lost to future generations—or even to those

2. Christian Dupont, "Libraries, Archives, and Museums in the Twenty-First Century: Intersecting Missions, Converging Futures?," *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 8, no. 1 (2007): 13, https://doi.org/10.5860/rbm.8.1.271.
3. Helena Robinson, "Remembering Things Differently: Museums, Libraries and Archives as Memory Institutions and the Implications for Convergence," *Museum Management and Curatorship* 27, no. 4 (2012): 416, https://doi.org/10.1080/09647775.2012.720188.
4. Robinson, "Remembering Things Differently," 416.
5. Jessica Meyerson, "Software Preservation Network: The Access Breakdown and System Level Change" (slides presented at ADE Summit, Library of Congress, Washington, DC, November 16, 2017), http://digitalpreservation.gov/meetings/ade/slides/session4_speaker2_Meyerson.pdf.

working a few years from now. Preservation of software is essential, not only for those studying and researching the history, culture, and technology of software itself, but for those interacting with materials that are software dependent, i.e., texts, images, videos, audio files, or other records that can only be accessed with the aid of specific software. In addition to ensuring that future generations can interact with a work in the way it was intended, software preservation allows for reproducibility of research, which is critical to validate research results and continue exploration in various fields.

Software preservation faces key challenges, including media degradation and technological change. Because of the rates at which digital media deteriorates and digital technology shifts, our ability to access software just a few years after its release is highly endangered—and the knowledge that depends on that software is therefore at risk.[6] A field of professionals with specialized skills and tools is emerging to meet this challenge. The Software Preservation Network (SPN), which represents some of these professionals, describes its mission as: "Preserving software through community engagement, infrastructure support, and knowledge generation." One of SPN's five core values is access: "We enable long-term access to software and software-dependent cultural heritage through standards development, documentation, software development, and training."[7] The professionals who make up SPN recognize that preservation today must be responsive to technologies that are changing how the world shares knowledge.

6. For a classic and readable overview of the vulnerabilities of digital media, see Jeff Rothenberg, *Ensuring the Longevity of Digital Information*, revised February 22, 1999, http://www.clir.org/wp-content/uploads/sites/6/ensuring.pdf. A more recent, in-depth treatment of the challenge of digital media preservation is Richard Rinehart and Jon Ippolito, *Re-collection: Art, New Media, and Social Memory* (Cambridge, MA: MIT Press, 2014).
7. "About," Software Preservation Network, accessed February 1, 2018, http://www.softwarepreservationnetwork.org/about/.

Professionals who contributed to this report described software preservation as a multistage (and often cyclical, iterative) process, entailing:

- The acquisition of the original digital media[8] representing a particular software program
- The making of a faithful representation of the contents of that media on the collecting institution's servers, or another medium more stable than the original media
- The professional assessment of those contents, and the creation of appropriate collection records
- The creation of access copies or other derivatives, including redundant copies in different geographic locations
- The provision of access (either on-site or online) to the records themselves
- The development of tools that permit researchers to render the original software within a configured software/computational environment (on the institution's premises or beyond)

Within this workflow, questions about whether a contemplated step is legally permissible inevitably arise. In particular, copyright protection is a ubiquitous background condition for software created since the 1950s. Although not all programs enjoy such protection, most do, so all must be assumed to until determined otherwise. As a result, the potential for legal problems with preserving copyright-protected software is widely recognized in the field.[9] There has,

8. Most software preservation processes we discussed dealt with software distributed on installation media (floppy disks, optical disks, etc.), though of course software preservation does not always involve such media.
9. In 1995 Rothenberg suggested optimistically (and incorrectly) that "with luck, copyright and royalty restrictions for proprietary programs may expire when those programs become obsolete, making them available for future access to historical documents." (Rothenberg, *Ensuring the Longevity of Digital Information*, 14.) By the early 2000s, however, concerns about copyright law were well known in the digital preservation community. See, for example, Catherine Ayre and Adrienne Muir, *Right to Preserve? Copyright and Licensing for Digital Preservation Project: Final Report* (Loughborough, England: Loughborough University, 2004), https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/343.

however, been virtually no discussion of the potential of the copyright fair use doctrine to help solve those problems[10]—this despite the fact that in other domains of preservation, memory institutions recently have embraced fair use as a valuable tool.[11] The fair use doctrine provides important exceptions that allow the use of copyrighted materials without permission from the copyright holder under certain circumstances. The relevance of fair use to software preservation is discussed in detail in the "Copyright Law and Software Preservation" section below.

This study explores how software preservation professionals currently understand copyright issues that arise in their work, and the effect of copyright on their ability to meet their mission. The study concludes that the software preservation community experiences a high level of frustration in their activities due to perceived copyright barriers; these frustrations could be substantially reduced by engaging in a process that will result in a *Code of Best Practices in Fair Use for Software Preservation*. Such a code, like those created by other communities of practice,[12] will provide reasoning to assist in decision-making around fair use in software preservation for the most common situations

10. See, for example, *Preserving.exe: Toward a National Strategy for Software Preservation* (Washington, DC: National Digital Information Infrastructure and Preservation Program, Library of Congress, 2013), http://hdl.loc.gov/loc.gdc/lcpub.2013655114.1 (describing copyright and Digital Millennium Copyright Act concerns, suggesting seeking permission and new legislation as the only possible solutions).
11. See, for example, *Code of Best Practices in Fair Use for Academic and Research Libraries* (Washington, DC: Association of Research Libraries; Center for Social Media, School of Communication, American University; and Program on Information Justice and Intellectual Property, Washington College of Law, American University, 2012), http://www.arl.org/storage/documents/publications/code-of-best-practices-fair-use.pdf; *Statement of Best Practices in Fair Use of Collections Containing Orphan Works for Libraries, Archives, and Other Memory Institutions* (Washington, DC; Berkeley, CA: Program on Information Justice and Intellectual Property, Washington College of Law, American University; Center for Media & Social Impact, School of Communication, American University; and Berkeley Digital Library Copyright Project, University of California, Berkeley, School of Law, 2014), http://cmsimpact.org/code/statement-best-practices-fair-use-collections-containing-orphan-works-libraries-archives-memory-institutions/.
12. See "Codes of Best Practices," Center for Media & Social Impact, accessed February 1, 2018, http://cmsimpact.org/codes-of-best-practices/.

that currently arise. For other communities of practice, such codes have been remarkably effective in enabling the core mission of the community.[13] The community should also engage actively in the Copyright Office's current rulemaking to establish exemptions from the Digital Millennium Copyright Act's provisions regarding digital locks.

Methodology

In preparing this study, we conducted more than 40 in-depth, long-form interviews with practitioners in the field. Each hour-long interview was conducted with at least two of this project's co–principal investigators. Interviewees were chosen by recommendations from the Software Preservation Network and through snowball sampling. Interviews followed a general protocol which included:

- What is the nature of software preservation, as you understand and practice it?
- What are the goals of software preservation, and what constituencies does it serve?
- What problems do you face in your job with software preservation?
- What problems do you have specifically with copyright, if any?
- Would it be easier to do [a range of things, including probably some the interviewee has already brought up] if you could copy some of the software without licensing it?
- What problems do you see as challenging around copyright in the field more widely?
- Who else should we talk to? What else should we read? Which meetings should we go to, to meet more people?

We listened and recorded what the interviewees had to say about the mission and methods of software preservation, and to their accounts

13. See generally, Patricia Aufderheide and Peter Jaszi, *Reclaiming Fair Use: How to Put Balance Back in Copyright* (Chicago: University of Chicago Press, 2011).

of copyright-related challenges they had encountered. In our review, we sought to identify both characteristic, recurrent problems and idiosyncratic ones. In preparing this account of the field and its engagement with the copyright system, we also drew on a survey of the scholarly literature.[14]

Copyright Law and Software Preservation

Copyright affects a range of software preservation activities, and this project is premised on the likelihood that these activities are more permissible under the law than is widely recognized in the software preservation community. This section explains the relationship of copyright law and some related legal provisions to software, including the doctrine with the greatest potential to enable software preservation activity: fair use.

Copyright extends some degree of protection to all kinds of software programs as "literary works," without regard to the medium or language in which they are represented. As case law relating to software copyright has developed since the mid-1980s, however, the actual extent of such protection is relatively superficial, applying in many cases only to the literal code in which the software is expressed (rather than to the underlying structure of programs, let alone the algorithms that they implement). From the perspective of software preservation, however, even this relatively "thin" protection is potentially problematic.

Copyright protection of software is problematic because the production and making available of faithful copies, and enabling researchers to interact with the programs those copies represent, is central to the work of software preservation. A disk image, for instance, is a bit-for-

14. The Software Preservation Network's helpful literature review provided a crucial starting point. Jessica Meyerson, *Software Preservation Literature Review*, last updated fall 2017, http://osf.io/qdsyc.

bit copy of the data on a piece of storage media (hard drive, optical disc, etc.); if any aspect of the software is protected by copyright, that aspect will be copied in the process of creating an image. Similarly, the process of running a copy of a program, whether on old hardware or in a new environment, typically generates additional copies; unless it is otherwise justified, doing such acts without authorization could infringe the exclusive right of reproduction conferred on copyright owners by Section 106 of the Copyright Act. Likewise, making archived software available for off-site use, whether in other memory institutions or to the public more generally (including online), necessarily implicates the right of "distribution." And the process of emulating computing environments, described at greater length below, may involve a distribution of software or may in some cases constitute a "public performance" of the software, depending on how the emulation technology works (see "New Technological and Community-Building Opportunities" below). Although these may be "technical" violations of the law, in the sense that they do not generate direct harms to the copyright owner, they are nevertheless regulated activities, about which ethical professionals are correct to be concerned.

Indeed, even technical violations can have real consequences, particularly where the protected works in question were once sold on a commercial scale. This is because producers and vendors of works such as operating systems and software applications are likely to have registered their rights with the Copyright Office, which leaves even a technical infringer open to potentially significant money judgments, or "statutory damages" (under 17 U.S.C. Sec. 502), as well as attorneys' fees.

Court-ordered injunctive relief—a court order to do something (like take offline or destroy a digital collection)—is another possible area of concern. Thus, a software preservation professional could fear that, at least potentially, an aggrieved software copyright owner could successfully sue to shut down activities in which substantial time and money have been invested.

In practice, these concerns about the high stakes of possible infringement may be exaggerated. Specific statutory protections for nonprofit educational institutions acting in good faith make the practical threat less severe for many memory institutions. And recent changes in interpretation of the law, especially after the Supreme Court's decision in *eBay Inc. v. MercExchange, LLC*, 547 U.S. 388 (2006), make injunctions significantly less likely as a practical matter—even in the low-probability event of a dispute about software preservation leading to litigation.

Turning from theory to practice, it is worth noting that there is no evidence that software preservation activities have ever given rise to an actual lawsuit. While litigation is certainly not uncommon in the software industry, it is focused on commercially significant activity—allegedly unfair practices by competitors or commercial sellers, for instance. We heard anecdotal evidence that large software companies appear to tolerate memory institution practices, although we also were told about threats of litigation by smaller developers—some serious (or at least intimidating) enough to cause software preservation professionals to cease and desist.

Licensing Issues

Restrictive licensing is a related legal issue of concern to software preservation professionals, although a review of current law suggests that it too may be less of a problem than it often is perceived to be. Vendors often seek to impose contractual terms on the use of purchased software. These terms are sometimes negotiated but more commonly imposed (in the form of "shrink-wrap" licenses accompanying boxed software products, or "click-wrap" licenses associated with downloaded ones). Some licenses merely track (and thus reinforce) restrictions on use imposed by copyright itself. Other licenses purport to limit use in ways that copyright alone would not (restricting otherwise permissible resale of purchased software, for example, or barring reverse engineering). The proliferation of licenses

in the software realm may be one reason so many of our interviewees focused on seeking additional permissions as a necessary precursor to certain preservation activities. Other items in memory collections were sold outright, so physical possession may lead to a natural assumption of certain rights to use these materials. By contrast, the software industry long has asserted that its wares are merely licensed, leaving collections with a sense of having fewer use rights.

Several aspects of the law of contract might mitigate these fears, however. First, the terms of software licenses, like other kinds of contracts, generally apply only to those who have agreed to them, and not (for instance) to an institution preserving software it bought in the resale market or received by donation. Even if the requisite contractual "privity" exists,[15] there are limits on how far courts will go in honoring one-sided deals. Strong public and constitutional policy supports the idea that mass-market licenses should not wall copyrighted works off from all kinds of future access—especially by students and scholars. Faced with the argument that the terms of a license designed to regulate consumer behavior should be enforced against culturally significant preservation activities by nonprofit memory institutions, a court might well decline to do so.

Nor is it clear how most software licenses could plausibly be read to exclude preservation, even if they were technically applicable. Although there may be agreements so clear that they must be read as broadly exclusionary—even to the extent of barring preservation—this has not been typical in the industry. At least where legacy software is concerned, it has been more common for vendor licenses to authorize some uses affirmatively, without specifically excluding others authorized by law (such as fair uses). The same is true of so-called end user license agreements (EULAs) where mass-market software is concerned. If called upon to consider whether an agreement associated

15. "Privity" exists when a court finds that someone who is not technically a party to a contract nevertheless should be bound by its terms.

with obsolete commercial software should be read broadly or narrowly, a contemporary court would be more likely to interpret it in ways that support rather than defeat the goals of the copyright system. Indeed, courts typically construe ambiguity in such contracts against the drafter, as they were in the best position to make the provision clear. No test case has clarified these issues in the preservation context because no software company has ever brought a claim that software preservation activities were barred by license terms.

Copy Protection and the Digital Millennium Copyright Act

Another common copyright-related concern is that software preservation could violate the Digital Millennium Copyright Act's (DMCA) prohibitions (codified in 17 U.S.C. Sec. 1201(a)) against bypassing "technical protection measures" (such as encryption or passwords).[16] These "paracopyright" prohibitions and related penalties, which were introduced in 1998, are not part of the Copyright Act, so they are not literally subject to copyright's ordinary flexibilities (including the fair use doctrine).[17] Although Section 1201 incorporates certain exceptions of its own (including one for libraries and archives), they do not apply to software preservation. As of 2015, the exceptions granted by the Librarian of Congress offer only narrow relief for "video games for which outside server support has been discontinued, to allow individual play by gamers and preservation of games by libraries, archives and museum[s] (as well as necessary jailbreaking of console

16. A particularly poignant account of this provision's implications for preservation is to be found in Jerome McDonough et al., *Preserving Virtual Worlds: Final Report* (Washington, DC: National Digital Information Infrastructure Program, Library of Congress, 2010), http://hdl.handle.net/1903/14734. Another striking story is Dan Bricklin's account of nearly losing VisiCalc, the first spreadsheet program for personal computers (which he coauthored, with Bob Frankston), due to copy protection: Dan Bricklin, *Copy Protection Robs the Future*, accessed February 2, 2018, http://www.bricklin.com/robfuture.htm.
17. Nonetheless, the Federal Circuit's decision in *Chamberlain Group, Inc. v. Skylink Technologies, Inc.*, 381 F.3d 1178 (Fed. Cir. 2004), suggests, according to some scholars' interpretation, a potential basis for fair use defenses to liability under Section 1201(a).

computer code for preservation uses only)."[18] In the current round of rulemaking, the Software Preservation Network and others have petitioned for a broader exception, applicable to software preservation in general.[19] It is possible to advocate powerfully for such an exemption both because it is important to the mission of software preservation, and because preservation uses are arguably lawful, in themselves, under copyright law.

Exceptions and Limitations to Copyright, Especially Fair Use

The specific exceptions in the Copyright Act allowing for certain uses by archives, libraries, and educational institutions are of real but severely limited utility for software preservation. The language of 17 U.S.C. 108(c), which provides exceptions for libraries and archives, is far too narrow to be of substantial benefit to software preservation activities. It only allows for the reproduction of three copies of a program in an "obsolete" format, and only if they are not made available "outside the premises" of the institution that owns the original.[20] Section 117(a) affords owners of software the right to make an

18. See United States Copyright Office, "Understanding the Section 1201 Rulemaking," accessed February 2, 2018, https://www.copyright.gov/1201/2015/2015_1201_FAQ_final.pdf.
19. This Proposed Class Nine is described in the Copyright Office's Notice of Proposed Rulemaking on "Exemptions to Permit Circumvention of Access Controls on Copyrighted Works" at 82 Fed. Reg. 49550, 49562 (Oct. 26, 2017), https://www.gpo.gov/fdsys/pkg/FR-2017-10-26/pdf/2017-23038.pdf.
20. Putting to one side the difficulties surrounding the definition of obsolescence in Section 108, the geographical limits on the circulation of copies created under the authority of the provision would assure that most scholars and students of digital culture would continue to be barred from effective access to archival software collections. As discussed below, the minimal preservation activity of making a copy to be kept on premises was rarely a source of concern among our interviewees. Their frustrations centered more on questions of resource-sharing and access. For more on the practical shortcomings of Section 108 in the new digital environment, see United States Copyright Office, *Section 108 of Title 17: A Discussion Document of the Register of Copyrights*, (Washington, DC: US Copyright Office, September 2017), https://www.copyright.gov/policy/section108/discussion-document.pdf. On the other hand, the complementary relationship between fair use and Section 108 is clear both from the terms of the Copyright Act itself and from recent court decisions interpreting it: Although Section 108 provides a "safe harbor" for certain library and archival preservation activities by declaring them to be categor-

archival copy and to make whatever copies are necessary as "essential steps" in using the software. This statute was drafted with ordinary software consumers in mind, however, and may not be sufficiently flexible or capacious to protect the wide range of specialized software preservation practices, or downstream research uses.[21] Likewise, the educational exceptions contained in 17 U.S.C. 110 are of little use, as they are limited to performances and displays in the course of teaching, and do not apply to reproduction, distribution, or activities outside the teaching context.

Fair use, by contrast, is subject to no such categorical limitations. The Copyright Act of 1976 provides that "fair use of a copyrighted work... is not an infringement of copyright."[22] In our time, this 175-year-old but still dynamic doctrine is effectively the most important limit on

---

ically permissible, other forms of preservation that these institutions undertake, with authorization from copyright owners, will be considered fair uses under Section 107 if they qualify under the balancing test set forth in that section. In other words, the existence of specific exceptions does not limit or circumscribe the potential scope of preservation under fair use.

21. The statute's limitation to "owners" could also present a challenge, given the software industry's frequent characterization of software transactions as "licenses."

22. 17 U.S.C. § 107. Limitations on exclusive rights: Fair use

    Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyright ed work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—

    1. the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;

    2. the nature of the copyrighted work;

    3. the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and

    4. the effect of the use upon the potential market for or value of the copyrighted work.

    The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.

    Note that although the factors are often viewed as representing the four corners of fair use analysis, the list is made explicitly nonexclusive; thus, courts can and (from time to time) do take other considerations into account, including the "public interest," in allowing the use under consideration to go forward.

copyright monopoly rights. Rather than following a rigid formula, lawyers and judges assess whether a particular use of copyrighted material is "fair" according to an "equitable rule of reason." The statute itself identifies some illustrative categories of uses to which the doctrine may apply, and provides a non-exhaustive list of factors that should be considered in fair use decision-making. Because copyright law describes fair use only in general terms, the doctrine can adjust to evolving circumstances. And where it applies, fair use is a user's right and not a mere privilege.

Since the early 1990s, fair use case law has taken a dramatic and user-friendly turn. In one decision after another, federal courts have indicated that a critical consideration in evaluating the fair use factors is whether the purpose of a use can be considered "transformative"—whether it "adds something new, with a further purpose or different character," as the Supreme Court put it in *Campbell v. Acuff-Rose Music*, 510 U.S. 569 (1994). This central emphasis on the importance of protecting transformative uses[23] continues to characterize fair use jurisprudence today.

In the history of US copyright, there has never been such a strong judicial consensus about the nature of fair use as exists today. Since *Campbell*, decisions at every level of the federal court system consistently have confirmed the proposition that for a use to be considered "transformative," it need not—in fact, it usually does not—entail a literal modification or revision of the original material. Instead, it is crucial that the use has put that material in a new context where it performs a new function.

Fair use offers members of the software preservation community opportunities to pursue their important mission without undue concern about legal pitfalls. The doctrine directly addresses issues of

23. In 2003, the Supreme Court affirmed the strong connection between fair use and First Amendment guarantees of freedom of expression in *Eldred v. Ashcroft*, 537 U.S. 186 (2003).

potential copyright liability for preservation activities, and (because agreements that limit fair use are disfavored) offers a significant refuge against license-based claims as well; moreover, establishing the fair use basis for software preservation will be a key to obtaining an exception to the DMCA's prohibitions on circumvention that is broad enough to support the community's work. A first step toward taking advantage of these opportunities is to explore how software preservation professionals currently understand their rights, and the effects of that understanding on practice.

The Context for Software Preservation

Despite its urgency, copyright law is only one of the issues facing software preservation professionals, and efforts to reduce the frustrations associated with copyright will benefit from awareness of the general landscape of challenges and opportunities facing software preservation today. To help us understand the context for their practice, interviewees discussed with us:

- How software preservation works
- What preserved software is typically useful for
- What kinds of users are interested in software collections
- New opportunities on the horizon for software preservation
- The non-legal challenges collections face

Software and Digital Preservation

Software preservation is part of the general field of digital preservation, a field of study and practice dedicated to ensuring continued access to digital materials of enduring value and interest. The key challenges that digital preservation must overcome are media failure and technological change. Media failure is a challenge across many preservation fields—information stored on paper, on film, and on any number of analog recording formats are all subject to loss or degradation due to the media on which they are stored. However, as many television viewers

learned when broadcast television moved from analog to digital, the loss of information in a digital format can be much more disruptive than in analog format: a weak analog signal could still produce a recognizable, if distorted, video program, while an incomplete digital television signal results simply in a blank screen. Similar problems plague digital media in general. So, despite the ease with which they are copied and shared, digital files are in some senses more fragile than information stored in analog media. Digital media formats can fail much more quickly and catastrophically than analog ones,[24] and they rely on a complex constellation of supporting technology.

Technological change has created increasingly acute preservation challenges in the digital era. Unlike some analog media, like paper and film, which can provide unmediated access to stored information, a digital format—whether a website, a multitrack music recording session, or an architect's design files—requires reconstruction of a constellation of supporting systems and tools, including hardware, an operating system, and a software application (sometimes multiple applications, plug-ins, or scripts). Rendering a digital file in a different software environment can change the information available from the file in significant ways.[25] Every element of commercially available support systems is subject to change as hardware manufacturers and software publishers compete to entice consumers with new products and features and as companies shift priorities, merge, are acquired, or go out of business. Obsolescence sets in quickly as technology advances. One interviewee told us that there is typically a 25-year lag between creation and donation of electronic records, so collections are almost never working with materials created using contemporary

24. Perhaps the most striking expression of this rate of failure is Jeff Rothenberg's observation that "digital information lasts forever—or five years, whichever comes first." Rothenberg, *Ensuring the Longevity of Digital Information*. The argument underlying Rothenberg's quip is that many digital media formats can fail in as few as five years, so it is dangerous to plan preservation activities based on more generous estimates that may not hold up in individual cases.
25. See Euan Cochrane, *Rendering Matters: Report on the Results of Research into Digital Object Rendering* (Wellington, New Zealand: Archives New Zealand, January 2012), http://archives.govt.nz/sites/default/files/Rendering_Matters.pdf.

hardware or software environments. Ensuring continued access to digital materials, therefore, requires preserving not only individual digital files or records created by authors or institutions, but also the software environments necessary to render the files meaningful to a reader. Another interviewee invoked investor Marc Andreessen's observation that "software is eating the world,"[26] arguing that software is so crucial for access to digital information that in some sense all digital preservation is predicated on software preservation; software preservation is eating digital preservation.

Two Key Purposes

One key purpose for software preservation, then, is to support ongoing access to digital materials (texts, images, movies, artworks, games) that depend on software for access to their contents. Interviewees repeatedly said that providing access to archival and special collections materials in legacy digital formats requires access to legacy software, including operating systems. In some cases (relatively simple textual material with minimal formatting), files can be migrated to modern formats, or the textual information in the files can be viewed in specialized modern viewing software. For files that are more complex than basic text, however, original software is required for full access. Digital documents are thus highly "software dependent." Software dependency has been recognized as a defining challenge for digital preservation since at least 1984, when Trudy Peterson observed that, "A software-dependent file will print out as gibberish unless it is processed on a computer that has the right software."[27]

An example from the world of research helps show the importance of software preservation in support of access to digital information. A relatively new, but fast-growing, cadre of researchers is interested

26. Marc Andreessen, "Why Software Is Eating the World," *Wall Street Journal*, August 20, 2011, https://www.wsj.com/articles/SB10001424053111903480904576512250915629460.
27. Trudy Peterson, "Archival Principles and Records of the New Technology," *American Archivist* 47, no. 4 (Fall 1984), 383–393, https://doi.org/10.17723/aarc.47.4.30u45640617n2184.

in research reproducibility, i.e., whether a particular research activity can be repeated in order to confirm the original results. Some pivotal research has been called into question as independent labs have tried and failed to reproduce published results.[28] Robust reproducibility requires access not only to the data from the original experiment but also, in most modern cases, to the software (often highly customized with scripts and plug-ins) used to process the data to generate results. Researchers committed to reproducibility need to work with software preservation professionals to develop strategies for documenting and preserving the software they use as well as their data and other important aspects of the research process.[29]

Software preservation has another key purpose, as software is the subject of serious scholarship and study in itself.[30] Software studies[31]

28. Benedict Carey, "Many Psychology Findings Not as Strong as Claimed, Study Says," *New York Times*, August 27, 2015, https://www.nytimes.com/2015/08/28/science/many-social-science-findings-not-as-strong-as-claimed-study-says.html.
29. Major research funders have recognized the importance of sharing software as an element of reproducibility. See, for example, Wellcome Trust, "Policy on Data, Software and Materials Management and Sharing," updated July 10, 2017, https://wellcome.ac.uk/funding/managing-grant/policy-data-software-materials-management-and-sharing. ("As a minimum, the data underpinning research papers should be made available to other researchers at the time of publication, as well as any original software that is required to view datasets or to replicate analyses.")
30. See, for example, Matthew Kirschenbaum, "Software: It's a Thing," *Medium*, July 24, 2014, https://medium.com/@mkirschenbaum/software-its-a-thing-a550448d0ed3 (describing a series of "reference points" for considering software as an object of preservation); Nathan Ensmenger, "Software as History Embodied," *IEEE Annals of the History of Computing* 31, no. 1 (January—March 2009): 88—91, http://doi.ieeecomputersociety.org/10.1109/MAHC.2009.16 ("Software is history, organization, and social relationships made tangible"); John G. Zabolitzky, "Preserving Software: Why and How," *Iterations: An Interdisciplinary Journal of Software History*, 1, no. 13 (September 2002): 1—8, http://www.cbi.umn.edu/iterations/zabolitzky.html; J.W. Cortada, "Researching the History of Software from the 1960s," *IEEE Annals of the History of Computing* 24, no. 1 (2002): 72—79, http://doi.org/10.1109/85.988584. The new discipline also embraces a number of emerging subfields, such as "platform studies," as represented by the volumes published since 2009 in the MIT Press series of the same name edited by Nick Montfort and Ian Bogost, at https://mitpress.mit.edu/books/series/platform-studies.
31. See, for example, Nick Montfort et al., *10 PRINT CHR$(205.5+RND①); : GOTO 10* (Cambridge, MA: MIT Press, 2013), https://mitpress.mit.edu/sites/default/files/9780262018463.pdf; Lev Manovich, *Software Takes Command* (New York: Bloomsbury Academic, 2013); Matthew Fuller, "Behind the Blip: Software as Culture," 2002, 7, http://noemalab.eu/wp-content/

is its own field, with a rapidly growing ecosystem of scholars, and the study of software is becoming important in a wide variety of established disciplines. The affordances of software can change the way software users work, making new kinds of creations possible and shaping the evolution of entire fields. Word processing affects textual authorship; 3-D modeling affects animation, architecture, art, and design; digital multitrack recording affects music.[32] One interviewee told us that the adoption of CAD software was perhaps the most significant change in architectural practice in a thousand years.[33] To study the nature and impact of that evolution, scholars will need access to legacy software. The other key purpose of software preservation, then, is to ensure that the software itself is available for future access and study, despite technological change and media failure.

Again and again in our conversations, we heard these two rationales for preservation: software as a tool for access to digital files, and software as digital material worthy of study in its own right.

A Variety of Constituencies Served

Given these dual purposes, it is clear that software collections serve at least as wide a variety of users as any other memory institution collections, and software preservation professionals consider these users as they plan their activities. Users and uses for legacy software are as various as human inquiry, and will surely multiply over time. Among the many potential constituencies (some as yet unimagined), interviewees noted:

---

uploads/2011/09/fuller_sw_as_culture.pdf; Matthew Fuller, ed., *Software Studies: A Lexicon* (Cambridge, MA: MIT Press, 2008).

32. For some striking examples, see Matthew G. Kirschenbaum, *Track Changes: A Literary History of Word Processing* (Cambridge, MA: Harvard University Press, 2016); and Janet H. Murray, *Hamlet on the Holodeck: The Future of Narrative in Cyberspace* (Cambridge, MA: MIT Press, 1997).
33. Archivists began wrestling with the challenges posed by CAD programs almost as soon as the software came into wider use. See William J. Mitchell, "Architectural Archives in the Digital Era," *American Archivist 59*, no. 2 (Spring 1996): 200–204, http://www.jstor.org/stable/40293973.

- Scholars studying the software itself[34]
- Scholars working with digital materials that have software dependencies
- Fans and users of software, especially games, but also electronic literature and art
- Businesses such as design firms and animation studios, which need access to their own archives or to older files in their field as part of their business operations
- Owners of buildings, industrial equipment, and other complex, designed objects who may need to consult original plans to diagnose problems, anticipate repairs, or plan upgrades and modifications
- Memory institution professionals themselves, who may need access to software for tasks such as appraisal and selection of materials for retention or processing
- Future users and uses yet unimagined

New Technological and Community-Building Opportunities

Many interviewees believed that the software preservation community is at an inflection point in its historical development, with promising technological and social opportunities emerging. On the technology side, many were excited by the promise of "emulation-as-a-service" or EaaS. Socially, software preservation professionals are beginning to connect and collaborate through the newly formed SPN, which promises to help coordinate and align diverse activities across the profession.

Just as digital files are software dependent, software is, in turn, hardware dependent. Software is written to be run by a particular kind of machine (an x86 PC, an Apple PowerPC machine, or a Silicon Graphics IRIS 2000 workstation, for example). Like software, hardware evolves quickly and obsolete systems can be challenging

34. See materials in footnote 31, above.

to preserve. Even when limited numbers of hardware machines can be preserved in working order (and some institutions are making valiant efforts to do so),[35] access to digital content will be dramatically curtailed if it requires in-person access to obsolete hardware as well as legacy software. Digital archives would become largely inaccessible as the machines on which they were created become obsolete and scarce. Even in the best case, they would be accessible only where such machines survived.

Emulation promises a way around these access barriers. An emulator (usually embodied in software) allows newer hardware and software environments to mimic the behavior of older hardware, enabling legacy software to run on contemporary machines. A software file that was originally written for the Apple II system can be run on a modern Windows PC laptop with the help of an Apple II emulator. Emulators for arcade games, console gaming systems, and some older computer systems have been widely available on the internet for years.[36] This is far less true, however, of emulators for a variety of specialized computing environments, such as those adapted to support data analysis or graphic design. Finding and running an appropriate emulator can be a technically daunting task for specialists as well as laypeople. A researcher consulting digital files created across even a 10-year span, assuming she also had access to legacy software (a feat in itself), could easily be daunted by the challenge of emulating multiple underlying hardware systems.

EaaS vastly simplifies matters for users by locating all of the required technology (an emulated hardware environment, legacy operating system software, and even legacy software applications in some cases) on a centrally maintained server and providing an emulated

35. See, for example, Living Computers Museum + Labs, accessed February 2, 2018, http://www.livingcomputers.org/.
36. See, for example, the MAME Project, accessed February 2, 2018, http://mamedev.org/, (a "multi-purpose emulation environment" that grew out of efforts to emulate vintage arcade games on modern PC hardware).

environment on demand to appropriate end users inside a normal web browser on the user's internet-connected device.[37] A user can view remote collections material or her own files within the environment (in her browser), and the environment can be configured to prevent downloading or altering software or files accessed within it. Once the emulated session ends, changes in the environment can be wiped out and the configured emulator and viewed content restored to its original state. The result is a technology that could permit collaborative preservation and collection development across institutions, and the provision of meaningful access to a wide variety of legacy software to remote researchers.[38]

This technical capacity for collaboration arises at the same time as software preservation professionals are advocating for a "coordinated preservation strategy that would avoid duplication of effort, and potentially result in resources available to all organizations."[39] SPN is taking important steps to help define and nurture a community of practice around software preservation that crosses traditional disciplinary and professional lines. Although still in its early stages of development, SPN is already helping to build this community through a network of working groups, virtual meetings, and physical convenings.

---

37. For a more detailed explanation of EaaS and its use in preservation workflows, see Euan Cochrane, Jonathan Tilbury, and Oleg Stobbe, *Adding Emulation Functionality to Existing Digital Preservation Infrastructure*, accessed February 2, 2018, https://ipres2017.jp/wp-content/uploads/45Euan-Cochrane.pdf.
38. Other strategies for meeting this challenge include "virtualization," pursuant to which a legacy software application would be made available as part of a package, along with full versions of the operating system and dependencies needed to run it, and "containerization," in which the package includes only those aspects of the environment essential for the specific purpose. See Mathijs Jeroen Scheepers, *Virtualization and Containerization of Application Infrastructure: A Comparison*, accessed February 2, 2018, http://referaat.cs.utwente.nl/conference/21/paper/7449/virtualization-andcontainerization-of-application-infrastructure-a-comparison.pdf. Notably, all of these strategies involve software preservation professionals in activities that fall within the regulatory scope of copyright law.
39. See Jessica Meyerson et al., "The Software Preservation Network (SPN): A Community Effort to Ensure Long Term Access to Digital Cultural Heritage," *D-Lib Magazine* 23, no. 5/6 (May/June 2017), http://www.dlib.org/dlib/may17/meyerson/05meyerson.html.

Challenges for Software Preservation

Like motion pictures and television before it, software has followed a trajectory of slowly building cultural interest and, eventually, scholarly interest, resulting in challenges to collection and preservation. For example, cultural memory institutions did not typically take these media seriously as worth collecting and preserving in their earliest days. Some of the largest collections of software in traditional institutions have been acquired from private collectors and secondary sources, rather than collected contemporaneously by institutions themselves. Software often enters digital library collections alongside born-digital documents and records, as part of the acquisition of literary papers or institutional archives. It is thus often collected opportunistically rather than intentionally. Likewise, many software creators and publishers did not maintain careful archives of their past work, focusing instead on developing the next product for market. Like the early movie studios, which destroyed or neglected negatives and prints at the end of their films' commercial release cycles, or the early television studios, which recorded over the only copies of early programs to save money on expensive video tape, software companies often lost track of their copies of old software, and their records of its development, in their rush to bring out the next thing.

The software industry has been extremely volatile, with small firms starting up, going out of business, merging, and being acquired at a rapid clip. Many software publishers who were in business a decade ago no longer exist, or are under new ownership. Interviewees told us that, in their experience researching software ownership, the business arrangements between software companies and the employees and contractors who wrote software could be informal, or, if they were formalized, the relevant paperwork is often lost, leaving present-day business owners unable to vouch definitively for their ownership of legacy titles. This volatile environment makes it very difficult to determine who, if anyone, is the lawful owner of legacy software.

Competitive pressures also shaped software creation in ways that affect preservation. Firms often guard the source code for their titles as trade secrets, making research and reconstruction of the software creation process difficult. Firms have sought competitive advantage by locking users in to proprietary formats that are not interoperable with third-party software. Even when apparent standards for file formats emerge, firms differentiate themselves by adding features and functionality that competing software cannot replicate. Firms typically don't sell or license older versions of software once a new, improved version is available, and backward compatibility is not an industry standard. In some sectors it is common for newer versions of a software program to be incompatible with files and applications created by older versions of the same programs.

In academia, software and software-dependent digital archives are still a relative novelty to researchers. The perception of low demand for access (together with the high cost of providing it) has limited the resources institutions are willing to devote to supporting software collections. Interviewees suggested, however, that limited research interest to date may be a function of the limited availability of these materials beyond a few, specialized collections that are typically restricted to on-site use. Intense popular interest in cutting-edge efforts like the Internet Archive's Internet Arcade suggest that wider availability may trigger wider interest.[40] The explosive growth in software studies and related scholarly and pedagogical developments suggests that software collections could see substantially more use if access were less constrained.

Finally, interviewees said that software preservation requires resources and expertise that are not evenly distributed across institutions. One key resource required for software preservation is the software

40. Dante D'Orazio, "The Internet Arcade Puts 900 Classic Games Right in Your Web Browser," *The Verge*, November 2, 2014, https://www.theverge.com/2014/11/2/7147505/the-internet-arcade-puts-900-classic-games-right-in-your-web-browser.

itself, and interviewees anticipated that the need to consult software-dependent material will be much, much more widely distributed than physical copies of legacy software. Even as more and more institutions find themselves in possession of collections containing software or digital artifacts that require legacy software for access, only a few have staff and technical support in-house that is focused on the challenges raised by software preservation. The roles and responsibilities of relevant staff vary widely across institutions, and some feel that the challenges associated with software preservation are still underappreciated in parts of the community. SPN has identified this wide, uneven distribution of expertise and collections as a core challenge for the software preservation community.

Many of the challenges just described can be met through providing more general access to software collections, and by more active resource-sharing among collecting institutions. The success of these initiatives, in turn, depends on how those institutions operate in the face of copyright-related constraints on preservation practice. As memory institutions recognize the importance of software to their core missions, meeting these challenges will inevitably be bound up with resolving the copyright uncertainties that surround software preservation.

Findings

Overview

Software preservation professionals perceive correctly that the interpretation and application of copyright law is crucial to their work, and unanimously find copyright a significant barrier to the fulfillment of the preservation mission. But preservation professionals' lack of consensus or access to reliable information about copyright leads them to choices that limit their ability to fulfill their mission. In addition to concerns about copyright compliance, they also focus on terms of licensing agreements under which commercial software products are

marketed, and the prohibitions against circumvention of technological protection measures (TPMs) that were introduced into US law by the 1998 Digital Millennium Copyright Act—two sources of constraint that ultimately derive from basic copyright principles. They also underestimate or misunderstand the benefits of the fair use doctrine, and how they might take advantage of it. Even though there has never been a lawsuit involving software preservation, professionals in the field fear litigation. Not only are they concerned about damages and injunctions, but also they fear repercussions from potential donors and partners and reputational harm to themselves and their organizations.

Permissions Culture and the Search for Prior Authorization

Generally, interviewees believed it was strongly preferable, if not always strictly necessary, to obtain licenses or other permission from rightsholders to reproduce or distribute the legacy software they collect. Both the literature and our interviews reflect a shared preference among preservationists for securing advance authorization for their activities from existing software firms that claim rights to older programs (either as their developers or as successors to other, now defunct, companies). Professionals' conviction that it is risky to proceed with preservation activities in the absence of permission is sometimes extended to the work of researchers using software collections: one interviewee said her patrons were warned to seek permission from game publishers before using screenshots from legacy games in a scholarly article. Another interviewee informed us that because they lacked such permissions, scholars using the collection were permitted to run and then describe the on-screen behavior of software, but prohibited from illustrating it directly with screenshots. In this, memory institutions devoted to software preservation share a widespread bias on the part of cultural professionals in favor of permissions.[41] The acceptance of this very cautious approach to

41. See Susan M. Bielstein, *Permissions, A Survival Guide: Blunt Talk about Art as Intellectual Property* (Chicago: University of Chicago Press, 2006).

copyright in software preservation also demonstrates the community's lack of awareness of, or comfort with, fair use—a doctrine designed to support culturally valuable uses of copyrighted material in the absence of permission.[42]

In practice, software preservation professionals have been frustrated in their efforts to secure advance permissions. Software vendors have not engaged them, usually ignoring inquiries and requests. One interviewee reported that it is a "nightmare" to get in touch with the "right person" at software companies when seeking permission for cultural memory institution uses. Another told us that on the rare occasions when companies could be reached, they could only provide permission to use current versions of software titles, and were not interested in supporting older titles or versions. Interviewees variously suggested these firms may (1) lack economic motivation, (2) believe licensed preservation activities could complicate their efforts to commercialize their current product lines, or (3) doubt they have the legal authority to license some older software products (given the uncertainty about past agreements with developers, sublicensed third-party software components, and the like). Even were some companies amenable, the number of different companies and vendors (or their successors) required to participate in a bespoke licensing solution for preservation would probably be insurmountable, interviewees said.

Interviewees also reported problems with the relatively large number of legacy software programs whose owners cannot be identified or found. A number of interviewees explained how difficult it is to trace records of corporate ownership in the shifting environment of the software industry; they also expressed concerned that under the complex copyright "work-for-hire" doctrine, the remaining rights to some legacy software programs might belong to individual developers rather than the firms that employed them decades ago.

---

42. There are notable exceptions to this generalization in recent writings, such as Trevor Owens, *The Theory and Craft of Digital Preservation* (Baltimore: Johns Hopkins University Press, forthcoming), 108, http://www.trevorowens.org/2017/06/full-draft-of-theory-craft-of-digital-preservation/.

Licensing Concerns

Interviewees were concerned that their preservation activities might be in conflict with the terms of license agreements that accompanied commercial software packages when they were originally sold. Most seemed not to recognize that such contracts were applicable only to the original purchasers, or worried that without a license of their own they had no rights whatsoever to use the software. To compound the problem, interviewees said that collecting institutions often lack reliable information about the specific licensing terms originally associated with commercial software products they have acquired second-hand. Original license terms are, in many cases, lost to time. Preservation professionals are acutely aware, however, of the range of restrictive terms that have been employed by particular vendors from time to time. So they often assume the worst, so to speak, making highly conservative assumptions about which essential preservation activities can be undertaken without violating license terms.

Anti-circumvention Concerns

A number of interviewees were concerned about violating the special protections for TPMs provided under the Digital Millennium Copyright Act and codified in 17 U.S.C. Sec. 1201. As interviewees noted, TPMs of various kinds have been part of the landscape of software commerce almost since its beginnings; as a result, most archival activities relating to software involve efforts to "avoid, bypass, remove, deactivate, or impair" TPMs. Most of the interviewees who shared this concern were unaware that if software preservation activities qualify as fair use, it would be possible to secure a DMCA exemption for the purpose of software preservation. Only a few were aware that through the efforts of the SPN and others, a petition for such an exception already is pending.

Inflated Risk Assessment and Its Sources

Interviewees participated in or observed poorly grounded decisions about risk that stemmed both from specific misunderstandings of copyright and a general lack of reliable shared information within the software preservation community. Interviewees were generally eager to be in legal compliance, and they expressed a substantial fear of lawsuits founded in part on a dubious analogy to the recording industry's now-defunct and failed campaign against small, noncommercial users' file-sharing in the early 2000s. Several interviewees said that major software companies were both highly protective and potentially litigious, although there is, as noted above, no record of any lawsuit or threat of lawsuit targeting software preservation activities. As important to interviewees as potential financial costs is the specter of reputational harm, including the loss of goodwill and funding.

The conservatism of interviewees resonates with that of other communities of practice. Specifically, many interviewees told us that they had no regular access to specialized legal professionals, and relied primarily on advice from their non-lawyer peers. We were also told of occasional legal reviews of software preservation projects that had been performed by institutional general counsel, legally trained board members, voluntary legal advisors, and others, with predictably conservative results. Our interviewees' comments reinforced the proposition that legal generalists who work only occasionally with a specific practice community often tend to emphasize risk-avoidance (or at least, risk-minimization) in their advice, both because they may be unfamiliar with the objectives and methods of the work in question, and because they may feel cautious about authorizing activities in a field for which no specific legal precedents exist.

Copyright Conservatism's Effect on Preservation and Access Practices

The strategies interviewees used to implement this hyper-conservative approach to managing copyright risk ranged widely:

- At one extreme, we saw software preservation professionals avoiding any preservation activities that involve reproducing software. A few limited themselves to acquiring and storing original media and documenting the behavior of the software when run on legacy hardware. Many were willing to create images—a bit-by-bit copy in a more stable storage format to guard against media volatility—but avoided preservation activities that involved further copying or distribution of the software.
- Others avoided the problem of legacy software where possible by migrating the contents of legacy files to new formats compatible with contemporary off-the-shelf systems. It was generally noted that this was not a feasible solution in the majority of cases, however.
- Still others were comfortable engaging in internal analysis of titles acquired for purposes of selection and description, even without explicit copyright authorization, but they were not comfortable with providing significant scholarly or public access.
- To accommodate scholars' research needs under these circumstances, some allowed limited numbers of qualified scholars to interact on-site with software from the collection, using original media running on donated or purchased legacy hardware. Several professionals adopting this approach told us, however, that they felt compelled to limit the ways in which even trusted scholars could reference these materials.
- More liberally, some created emulation environments and other tools that allow scholars working at dedicated terminals on the physical premises of the institution fuller and more realistic access to legacy software, without the need for legacy hardware. These terminals were typically bespoke and one-of-a-kind,

designed to highlight a particularly high-profile collection or as part of an exhibition, for example, and were not employed systematically.

- No interviewees based in conventional memory institutions facilitated access by offering remote users online access to legacy software environments; specifically, none was willing to engage in EaaS, an approach that would allow for greater institutional information sharing based on specialization.[43]
- We found the boldest approach, making archived legacy software generally available online, only outside mainstream institutions. We did not hear any suggestion that mainstream, nonprofit, preservation programs, particularly those affiliated with universities or research libraries, are currently comfortable imitating this strategy.

The stated rationales for these various approaches to risk assessment varied accordingly. Those who took more conservative positions tended to speak in formalistic terms, explaining that they were authorized by law to engage only in certain specific on-site preservation activities.[44] More liberal practitioners argued pragmatically, believing their activities were unlikely to be challenged if they remained internal to the institution. Some of these expressed varying levels of confidence that the "educational" nature of their activities might help shield them from liability. Interestingly, few viewed the copyright fair use doctrine as a meaningful option, and those who did referenced it in general terms only, as interchangeable with "educational use"—a generic category of safe uses. The boldest professionals trusted to a combination of practical and informal considerations. They hoped that

43. Many interviewees based in conventional memory institutions acknowledged relying on informal networks of enthusiasts and private scholars to undertake some otherwise "off-limits" software preservation tasks, including creation of emulators, and development of tools and techniques to circumvent TPMs.
44. This understanding may derive from a misunderstanding of how the Section 108 exceptions for libraries and archives operate, although preservation professionals did not say as much. See footnote 20, above.

commercial software companies, with their aggressive focus on the bottom line, may not notice or care enough to take action. If software companies do take action, they are likely to signal their concerns before actually filing suit, which would allow for takedown of the potentially offending material.

Mission Failure

As a result of unduly conservative practices, a range of mission failures ensues:

- Items acquired by some collections cannot be inventoried fully, because the software tools required to assess them are unavailable.
- Consequently, finding aids omit electronic holdings that may be of interest to researchers.
- Archivists cannot diagnose errors or understand what may be missing or otherwise misrepresented when they open files in software environments that differ from the original environment where files were created.
- Museums struggle to determine the long-term value of acquiring digital artwork that relies on third-party software without confidence that future use of the software will be lawful.
- Collections of programs and other born-digital objects remain, for the most part, inaccessible online or at physical sites other than that of the collecting institution.
- Digital objects in migrated formats sometimes can be investigated on the premises of memory institutions, but not as fully or authentically as would be possible if emulation techniques could be employed more generally.[45]
- At least one ambitious software preservation initiative was denied funding due to copyright concerns.

45. The same is true where open source tools are employed, in lieu of emulation, to provide approximations of legacy environments.

- Even on-site, scholars are frequently denied effective access to digital works that were designed to operate in complex, customized, computing environments with multiple software dependencies unavailable or unusable by the collecting institution.
- The software preservation community, which has so far been discouraged from embracing the most advanced current solutions to archival challenges, is even more poorly situated to innovate (or adopt) the next generation of new technological approaches.

Next Steps

Software preservation professionals can better educate themselves about current law on copyright, fair use, licensing terms, and anti-circumvention. To do so, they need to develop a community-wide consensus around best practices relating to copyright law, and particularly fair use. In the next stage of this project, we will undertake the following:

- Convening small deliberative groups of professionals in the field to discuss best practices in fair use for preservation of and access to software
- Shaping of a fair use best practices code from the conclusions of these convenings
- Developing informational backgrounders for the field on closely related issues, such as the effect of mass market licenses on preservation activity

In addition, we recommend that concerned members of the field take an active role in supporting the SPN's pending petition for a broad DMCA exemption for software preservation, about which more information is available from Harvard Law School's Cyberlaw Clinic.[46]

46. Harvard Law School Cyberlaw Clinic staff, "Software Preservation Comments Filed in 1201 Rulemaking," *Cyberlaw Clinic Blog*, January 2, 2018, http://blogs.harvard.edu/cyberlawclinic/2018/01/02/software-preservation-comments-filed-in-1201-rulemaking/.

ASSOCIATION OF RESEARCH LIBRARIES

Association of Research Libraries
21 Dupont Circle, NW
Suite 800
Washington, DC 20036
T 202.296.2296
F 202.872.0884

ARL.org
pubs@arl.org