Phantoms of the Digital Opera:
The need for long term preservation of born-digital actors and multimedia objects using methods that permit ongoing new creations

Dena L. Strong
Technology Services
University of Illinois at Urbana-Champaign
1304 W. Springfield, Urbana, IL 61820
dlstrong@illinois.edu

ABSTRACT
While the rapid improvement in computing power and
digital tools have offered a vast new realm of creativity to
more independent creators than ever before, that same
rapidly changing pace also offers a unique threat to those
digital creators: What can you do when the length of time
you need to finish your project is longer than the length of
time your computer’s OS or software is supported? How
can a small creative or research group best ensure that their
work can still be actively worked with in five years – or
twenty-five years?
In this paper, I’ve assessed digital preservation
recommendations over time, discussing the gap between
the recommendations for preserving a completed work and
digital creators’ ongoing needs to be able to create further
work. I have three case studies created from longitudinal
interviews with digital creators to determine how their own
creative ecosystems have held up to the pressure of time,
where their systems have begun to degrade, and how they
continue their digital work across years or decades.
Currently, either maintaining an aging system or recreating
entire worlds with different, newer systems are the best
available options for the digital creators I interviewed,
though neither is optimal. Hardware and software vendors’
tentional dropping of backwards compatibility prohibits
accurate forward migration for many digital media creators.
The Pericles model seems promising, but Pericles itself is
not open to individuals. Future improvements in emulation,
including cloud-based virtualization, may offer a path
forward if these services can improve their user friendliness
and resolve the version compatibility issues that create
obstacles today.

Keywords
Long term digital preservation, digital video creation,
digital animation, translation, version compatibility,
emulation, digital lifecycle

1. INTRODUCTION
In 2016, Makoto Shinkai made waves throughout the
digital animation world by releasing a record-breaking
feature film called *Kimi no Na wa (Your Name)*; he was the
director, the writer, the cinematographer, and the editor,
and it became Japan’s highest-grossing animation title and
fourth-highest grossing film of all time [28].
In many ways, those waves originated 14 years earlier,
with Shinkai’s *Hoshi no Koe (Voices of a Distant Star)*, a
short film for which he had personally created everything
but the soundtrack. And he did it with what now sounds
like an antique: A 400 Mhz Mac G4 with 1 Gb memory and
a 300 Gb hard drive, along with Lightwave 3d 6.5,
Photoshop 5, and AfterEffects 4.[31].
In the years since, digital technology has opened doors
for creators worldwide by allowing them to create a
complete work single-handed. But digital creators are also
at the mercy of the software and hardware upgrade cycle in
ways physical creators aren’t. A sequel to *Hoshi no Koe*
would likely need to rebuild the entire digital world, unable
to reuse any of the character or scenery models or
animations involved in its original creation.
I’ve interviewed three digital creators with similar long-
term digital media projects ranging from six to twenty
years in duration. All the artists I interviewed have the
same problem: The working lifespan of their projects is
significantly longer than that of their digital creative
infrastructure.
Physical arts can be recreated and extended after a
multi-year gap, but when the film production is born
digital, the passage of time can destroy the digital video
creators’ entire artistic infrastructure – and sometimes it
only takes a year or two if a key component fails.
Shinkai’s *Hoshi no Koe* was only 25 minutes long, and he
could finish it within a year. Award-winning animator Nina
Paley also creates feature-length films by herself, but the
five to ten years a project takes her can encompass multiple
digital generations. [26],[22].
Digitally creative professionals working on long term
projects need to maintain their own personally-managed
creation-capable digital preservation systems without a
clear set of best practices, because (for the most part) the
long-term preservation best practices are determined by
library and archive professionals, not digital media
creators.
Digital creators need preservationists’ help in order to
keep their own works alive – but preservationists’ usual
goals often don’t encompass creators’ needs.
Libraries and archives emphasize read-only formats to
preserve authenticity and provenance, as well as to ensure
that content is available to end users in the future. In general, creators and researchers who are still amid the creative process are not considered the target audience for digital preservation. Creators and researchers are usually directed toward storage instead, but the problem they face is both larger and more enduring than file storage.

Preservationists are accustomed to considering how to store the equivalent of a completed roll of film, not the equivalent of an entire opera house with a full prop and costume department, lights, sound, rigging, and actors. But digital creators and researchers need to maintain the entire opera house – and the smaller their team is, the longer the span of time they’ll need it available, because that small team plays every role in the production one at a time.

In this paper, I’ll assess what the professionals in the field consider(ed) to be the best solutions at various points in time, and contrast that with three longitudinal and qualitative case studies of individual media creators whose works span years or decades of development.

If the digital preservation ‘state of the art’ could be made more inclusive of individual creators’ and small working groups’ needs, we could support the ongoing use of the creation suite (the digital opera house), not just a single static production (the roll of film).

When a digital actor can’t continue to make new interactions with its digital environment, then that actor is effectively dead, and the only remnants are its completed works stored in a read only format – a phantom cast on the screen in static form, not a dynamic and living being in an active creative space with continued potential.

If the preservation field can’t provide this creative sustenance across digital generations, who else could?

2. LITERATURE REVIEW

In the scholarly literature surrounding digital preservation, preserving digital creation and editing capacity ranges from a secondary consideration that may be an acceptable loss [11, 12] to actively prevented by converting from editable formats to read-only formats [19, 32].

One school of preservation thought emphasizes mechanically comparing a migrated file is to its original [8, 9], and another emphasizes migration to new formats on demand [18], but most of the approaches to either emulation or migration assume that the original is a fixed item that is not to be modified. The Pericles project[1, 5–7, 14] is an exception, but it’s restricted to specific content from European agencies and not structured to support independent creators’ ongoing active-use needs.

While there are quite a few papers discussing the preservation of digital playback formats, not many discuss the preservation of creation formats. The history of digital preservation recommendations for other multimedia and non-multimedia formats shows that the best guesses made are not always correct.

Looking back at 2000, while assessing the long-term preservation risk of several digital formats, Lawrence et al considered Lotus 1-2-3 an important enough format to include it in their project, but no current spreadsheet system offers the ability to read its file format. However, they considered TIFF at similar risk, but versions of TIFF still appear in recommendation standards fifteen years later [19, 32]. We can’t rely on our own current format recommendations remaining viable decades into the future; we can’t know if our guesses today will be any better than Lawrence’s were in his time.

While Lawrence and his collaborators considered leaving a file unchanged less risky than migration, Mellor et al. considered migration the best possible solution to digital preservation [18]. Their recommendation course was to preserve originals in various formats and devote the bulk of their work to a type of ‘skeleton key’ master-conversion system that could migrate content to new formats on demand.

The concept seems simple, but it also makes assumptions that others haven’t supported. One is that it’s easy to keep the source format’s bitstream intact to continue to make copies; another is that migration is an unequivocally good solution to the problem. All of the digital creators I interviewed would love to migrate their works to a newer format – but the software and hardware vendors have made accurate forward migration of editable working copies impossible for them.

Becker and Rauber wrote multiple papers on languages used in verification of files that were migrated to other formats, and are much more reticent about the value of migration on demand than Mellor et al. In the XCL language paper, Becker and Rauber present the primary options as emulation or migration, and consider emulation at least as viable as migration. Migration presents serious concerns for them, even when converting static formats, as the requirements for conversion are dependent on context.

In addition, they didn’t consider continued editing capability an unacceptable loss. But for working multimedia creators, I posit that the ability to edit files is an unacceptable loss. Here again, we see the difference between the perspectives of creators of repositories and creators of dynamic content.

One of the most closely related fields of study is digital game preservation, which also addresses the preservation of born-digital dynamic content with specific hardware and software requirements. There’s still a gap between preservationists’ and practitioners’ perspectives in this realm, though. In Bettin’s paper, she notes that “significant properties” of a game are often interpreted by preservationists to mean “metadata,” but computer scientists and gamers find other properties like the game experience itself more significant. [10]

Still, the Preserving Virtual Worlds focus doesn’t quite match this paper’s focus either. Their focus is on the players of the games and the re-creation of their experiences, not on the game creators, except when re-
creation on a different platform is considered as a preservation strategy[10, 17]. This may be a viable option, and in some cases may be the only one available [10, 17, 33], but isn’t optimal. Time a digital media creator spends in re-creating an existing work on a new platform can’t be spent creating new works.

In addition to the philosophical approaches to preservation, the difference in scale between small creative groups and the larger industry is a key consideration. Most profitable film or game companies employ a large enough staff to pay for a dedicated digital preservationist, but independent media creators rarely have these resources.

Schumacher et al. surveyed a group of library professionals from smaller Illinois universities, and notes, “Practitioners at smaller institutions often do not have time to stay abreast of the frequent developments in the field of digital preservation, may not have the expertise or technical infrastructure necessary to install and maintain complex software solutions, and frequently lack the funds to pay for complete, ready-to-use solutions that may exist.” [21] (p. 4-5)

Schumacher’s statement could apply to researchers, digital media and content creators, independent video professionals, and other groups not traditionally considered full “digital repositories”, but editable-repository needs are critical to many of them. The ability to reuse files and environments saves time and effort that can be applied to creating new works.

Schumacher’s team takes nearly the opposite approach from Mellor’s. Rather than using limited resources building and maintaining a single tool, Schumacher’s team leveraged tools built by others to focus on preserving the materials unique to their institutions. In the process, they concluded that digital preservation is a continually shifting gradient scale, and a solution ‘good enough for now’ fits their needs better than a perfect solution in the future.

They investigated seventy tools, and their top recommendations for end-to-end preservation tools included Archivematica, DuraCloud, and Preservica. They concluded there was no clear one-stop solution, but provided a range of recommendations for different-sized institutions.

Their assessment of the needs of very small institutions came close to addressing the situation of independent digital media creators; still, their focus was on preserving read-only rather than editable files, they didn’t address unique hardware preservation, and the prices of all the systems but Archivematica would put them out of reach of the digital creators I interviewed.

In some cases, expensive research gathered over a longer time than a hardware or software generation faces similar issues, particularly for preserving data. Smit et al. undertook a project called PARSE.insight to assess the state of preservation among major publishers; they found that the publishers are fairly well positioned to preserve their own published papers, but that the research behind those papers is much more precarious, with 69% of publishers not offering any research preservation options [24]. Authors would like to be able to digitally preserve their research along with the publications, but there is debate over whose responsibility it is, due to the larger, more varied, and less-standardized nature of the data [24].

This gap is very similar to the digital media creators’. If the goal of digital preservation is purely defined as readability, the content distribution formats would be sufficient – but researchers need content creation formats to be able to add to research data that has been generated over years or decades.

A digital preservation strategy for content consumers is very different than one for active content creators and maintainers. I was glad to see references to the need to talk to a system’s user community in several papers [5, 8, 10, 11]. But the ability to preserve content changes -- let alone maintaining a specialized environment to continue making new creations on an old system -- is often given short shrift.

The Pericles project is a notable exception, looking at digital preservation methods for dynamic digital creations for both artists and researchers[5–7, 33], but has drawbacks here and now. First, it only preserves content from European space agencies [1]. Second, it’s still under development. Third, even if it were opened to independent media creators, there’s no way to estimate participation costs. Fourth, the Pericles model is complex, multi-layered, and not reproducible by an independent creator. So, while Pericles’ theoretical approach could help independent creators over the long term, they can’t benefit from it today.

3. METHODS

To explore the digital preservation needs of media creators over a span of time, I performed a literature review including scholastic publications about digital preservation. I also conducted personal, qualitative interviews with three local digital media creators across a span of decades to see whether and, if so, how their creative processes and environments had changed.

I interviewed digital 2D animator and filmmaker Nina Paley, digital 3D animator and filmmaker Shaun Mills, and multilingual translator David Fleming. In Paley’s case, our formal interviews began in 2014. In the case of Mills and Fleming, I’ve been learning about their digital creations and working infrastructure since we first met in 1997, and I performed formal interviews with them in 2014. I followed up in 2017 to see whether and how the passage of years have changed their working environments.

I chose a small group of known interviewees to follow up with the same people across years and see whether their own assessments of their best available options held true over time.

4. FINDINGS
Interviews with the digital creators and findings have been trimmed for length. More information is available at http://go.illinois.edu/digitaloperasupplements.

4.1 Nina Paley, digital and stop-motion filmmaker, Sita Sings the Blues, This Land is Mine, and more

Paley has the most urgent digital preservation needs of the three creative professionals I interviewed, since she has no other collaborators for her most enduring projects.

From her perspective, “computer hardware really reached its best level of price-to-functionality in about 2010, and software in about 2005” (personal conversation, 2014). Changes Adobe made when acquiring Flash in 2005 and that Apple made to Final Cut Pro in 2011 removed both functionality and backwards compatibility, leaving her and many other creators with the choice of updating at the cost of recreating their previous work, or not updating and facing increased obsolescence and difficulty in consultation.

Because Paley was a studio of one, she had originally decided to stay with her old tools. In 2014, Paley didn’t see an alternative to preserving the old hardware and software. But by 2017 she had split her digital “opera house” in half. For her own decade-long projects, she depends on aging versions of software released by companies with no interest in releasing open source versions, and which might sue those that do. But to earn money and work with collaborators on commercial projects, some of her newer work is done in an entirely different and newer environment, capable of 4K output but incompatible with her ten-year project and the rest of her creative works.

4.2 Shaun Mills, integrated live action and CGI filmmaker

Shaun Mills is an independent filmmaker with a degree in film production. He’s been working on completing his own science fiction movie, The Gatherers, since 2009. He’s already captured the actors’ footage and has spent the past several years creating digital effects.

Mills has been unable to find a working video content preservation system that fits his needs; he hasn’t been able to find a version control system that would capture the interactions between Lightwave and AfterEffects files, rough cut footage, and digitally manipulated .PNG files needed to assemble the final frames.

After losing hundreds of hours of work because the 3D rendering and video effect files he created in AfterEffects in 2009 were no longer usable after a software update dropped backwards compatibility five years later, he’s created his own “brute force” digital preservation method: He exports every single frame of his two-hour movie to PNG, requiring 24 separate files for each second of film, then re-importing the PNGs for the next rendering pass. It takes a long time to render a scene, to the point where he’s considering reducing his paid work hours to have more time to work on his film (personal conversation, 2014). Follow-up in 2017 reveals that he’s still using his frame by frame approach; he’s kept his own backwards compatibility, but at the price of excruciatingly slow progress.

4.3 David Fleming, professional film, game, and anime translator

David Fleming, a professional translator working in the media translation industry since 1997, has faced the digital preservation dilemma many times. When a long-running series he’s translated 10 to 20 years earlier is be released in a new edition or format, if he can’t access his earlier digital materials because of version incompatibility, then he has to reproduce hundreds of hours of work for each season of a series.

Many of the files he works with are provided to him from international sources in compressed formats that don’t fit US media authoring standards, meaning that much of his billable work time is spent in reformatting video, in addition to translating and subtitle timing. He stores terabytes worth of working video footage on a home-built RAID server.

Changes to hardware and software have required Fleming to retain outdated versions of operating systems to continue to use his open source and custom commissioned tools; the modern corporate alternatives would block his backwards compatibility and cost tens of thousands of dollars, prohibitive for an individual freelance worker. In the mid-2000s, Fleming hired programmer Keith Hays to write a customized Excel plugin that would make it possible for Fleming to use open software packages VirtualDub and SubStation Alpha to capture subtitle timing information in an Excel spreadsheet. Hays’ plugin worked with Excel 2003, but wasn’t compatible with later versions, and Hays has been unable to create an updated version that works precisely enough for Fleming’s split-second subtitle timing needs.

5. DISCUSSION

I agree most with the approach taken by Schumacher’s [21] group, looking for inexpensive “good enough” solutions for the moment, while acknowledging that they may not be sufficient in the future. Lawrence et al. [16] follow the same principle. Mellor’s [18] monolithic solution seems unlikely to be viable for small and independent users.

The Pericles project, seems to be a larger-scale version of Schumacher’s approach – their scenarios involve assessing individual projects’ needs and customizing the solutions with the aid of digital toolkits [14, 33].

Cost and suitability continue to be a prohibitive factor for most of the available solutions. Pericles has potential, but isn’t available to independent creators yet, and under its current focus may never be.
Of the specific systems explored by Schumacher et al., three have price points of thousands of dollars per year, and the fourth requires technical knowledge and produces read-oriented packages. By comparison, online retailers offer considerable storage space for under $1000. Neither solution solves the problem of software compatibility.

On the video creation side, no good answer has been offered for Paley and Fleming’s frequent issues with vendor-planned obsolescence. Mills’ manual approach is less version-dependent, but at the cost of a considerable time investment that prevents him from making progress at sufficient speed to be financially supported by his work. A free, open-source equivalent of their tools would make long-term access safer from planned obsolescence, but they would require open source versions of many software packages. Corporations have little motivation to create open alternatives to their products, and independent media creators tend to focus more on making their art than on learning to program new tools for themselves. (Fleming, by far the most technical of the creators I interviewed, still chose to hire a programmer for the Excel plugin – and an Excel plugin is vastly simpler than a video manipulation program like AfterEffects.)

Currently, either maintaining the older system or recreating their works in a newer system are the best available options for the digital creators I interviewed. Neither emulation nor forward-migration have offered them viable options.

In a future where emulation will become more accurate and less expensive, that may become a more viable option to extend the lifespan of digital environments. However, thus far hardware-specific and graphic card-specific version incompatibilities have prevented emulation from maintaining an aging digital environment for any of the creators I interviewed.

Cloud infrastructure offerings like Amazon Web Services provide tantalizing glimpses of effectively hardware-independent emulation and scalability, but as of this writing, effective use of AWS needs more technical skill and cloud-development-specific dedication than the digital creators I interviewed have available.

In theory, it sounds ideal – a major company like Amazon lends its vast pool of expertise to the business of hosting systems for thousands of large and small companies and organizations already, and the “container-based” model of thinking is designed to be more portable than many other development types.

If an individual creator could simply replicate his or her own creative environment in a platform-independent container, charged by usage, not charged when idle, and scaled to meet his or her needs, then they could preserve their digital worlds for as long as they were needed. Unfortunately, we’re not at the stage where AWS containers offer the ease of installation, use, and maintenance or the solution for hardware-specific requirements that digital creators need. In my department, even system administrators with decades of experience need help adapting to AWS in time-and-cost-effective ways, and they want to keep their systems as well patched and secured as possible. Amazon doesn’t yet offer anything like a walled garden for secure use of out of date environments, out of date software, and specific graphic card emulation – but, conceivably, they could.

6. CONCLUSIONS

Over the fifteen years since Makoto Shinkai’s Hoshi no Koe, many more independent digital creators have begun to follow in his footsteps -- and a very specific irony stands out in conversations with them.

Shinkai’s first film considered the personal effects of time dilation on two human lives that are out of sync with each other, an artifact of light-years of intervening space, with increasing gaps of time between any communications from the past to the present and the future.

Fifteen years later, the greatest threat to each of these new creators’ works is the personal effects of time compression across digital environments that are out of sync with their creators’ human lives, an artifact of a digital generation gap across the technical equivalent of light-years’ hardware, software, and encoding differences that offers fewer and fewer points of creative compatibility between past, present, and future.

Right now, we simply don’t have a good option for preserving the digital opera houses – but we haven’t had nearly as much time to solve the problem as the people who work in physical preservation. Truly effective and simple-to-implement emulation could offer aging systems a new lease on life, but we’re not there yet.

Still, user-friendly cloud-based emulation is an area where digital preservation, computer science, and digital creativity could merge to offer a new world. I can imagine a future where preservationists and developers have worked together to offer cloud-based emulation of a range of historic and modern systems and software, easy for digital creators to access and use, without depending on the specific Final Cut Pro version and Matrox graphics card a video creator’s used since their project’s inception. We haven’t reached that easy emulation world yet. But maybe, by Hoshi no Koe’s 20th anniversary, we could.

7. ACKNOWLEDGEMENTS

I wish to thank David Fleming, Shaun Mills, and Nina Paley for sharing their time and their experiences with me over the course of several years’ investigative conversations; Rhiannon Bettivia for her support and insight on previous drafts; and Heidi Imker, Dan Tracy, Elise Dunham, and Doug Atkinson for their encouragement and help in structuring and formatting this paper.

8. REFERENCES

[2] Anthony Willis’s answer to Video Editing: What’s a


