Erfgoed 2.0

Nieuwe perspectieven voor digitaal erfgoed

Bart De Nil & Jeroen Walterus (red.)

PHARO PUBLISHING

Erfgoed 2.0
Nieuwe perspectieven voor digitaal erfgoed

Priemstraat 51, BE-1000 Brussel
www.faronet.be

www.pharopublishing.be

Concept en redactie: Bart De Nil & Jeroen Walterus
Eindredactie: Hilde Schoefs
Vormgeving: Typeface (Leuven)
Lettertype & papier: Cronos pro & Briem Akademi op maco mat
Beeldredactie: Bart De Nil & Jeroen Walterus
Tabbladillustraties: U.S. Army Photo's (Public domain)
Tekstcorrectie: Berty Goudriaan
Druk: Leën Offsetdruk (Hasselt)

ISBN: 978-90-8992-004-1
Wettelijk depot: D/2009/11.524/2

NUR : 612

faro
VLAAMS STEUNPUNT VOOR CULTUREEL ERFGOED

Met steun van de Vlaamse overheid

Inhoud

How will we preserve virtual worlds?

Jerome P. McDonough

Introduction

In 1961, the Digital Equipment Corporation donated a PDP-1, their new computer, to MIT. The PDP-1 was in many ways a direct expression of the beliefs of DEC's founders that computing should be both interactive and accessible. It provided users with a variety of mechanisms for interaction, including a paper tape punch and tape reader, as well as a typewriter for keyboard input along with a vector graphics cathode ray display with a light-pen. While the PDP-1's relatively low cost and advanced capabilities made it an attractive platform for scientific computing, one of the most lasting contributions of the PDP-1 to computing history would have nothing to do with scientific research. Steve Russell, Martin Graetz and Wayne Witanen at MIT, contemplating how they might demonstrate the computing power and vector graphic capabilities of the new system, decided that a game based on the Lensmen series of books by E. E. 'Doc' Smith might provide just what was needed. By February of 1962, Russell, along with contributions from Alan Kotok, completed a first version of their game, *Spacewar!* , a two person game simulating combat between two spacecraft maneuvering around a star. There had been computer games authored prior to *Spacewar!*; *Noughts & Crosses* was written for the EDSAC computer at Cambridge University in 1949. There was also at least one game to make use of a vector display, the game *Tennis for Two*, a tennis simulation designed to be played on an analog oscilloscope. *Spacewar!*, however, was the first game to combine digital computing with high quality (for the time) graphics and interactive controls. It was, in short, the first modern computer video game.

Figure 1. A PDP-1 Computer running Spacewar! Image provided by Marcin Wichary under Creative Commons License (www.creativecommons.org/licenses/by/2.0)

The PDP-1 did not have a long product life span. DEC released the first PDP-4 in 1962, the same year that *Spacewar!* was completed, and by 1965, the PDP-8, one of DEC's

major successes, was released. *Spacewar!*, in contrast, had a much longer life span. It was the basis for several arcade games in the 1970s, including the *Galaxy Game* released in 1971, and Cinematronics' *Space Wars* game released in 1977. As the first space combat computer simulation, it can be considered the inspiration for an entire genre of games, including such notables as *Star Raiders* on the Atari platform, Acornsoft's well-known *Elite* computer game of the 1980s, and contemporary massive multiplayer games such as *EVE Online* and *Star Wars Galaxies*. A version of the original game, running under a Java emulation of the PDP-1, can still be found online today.

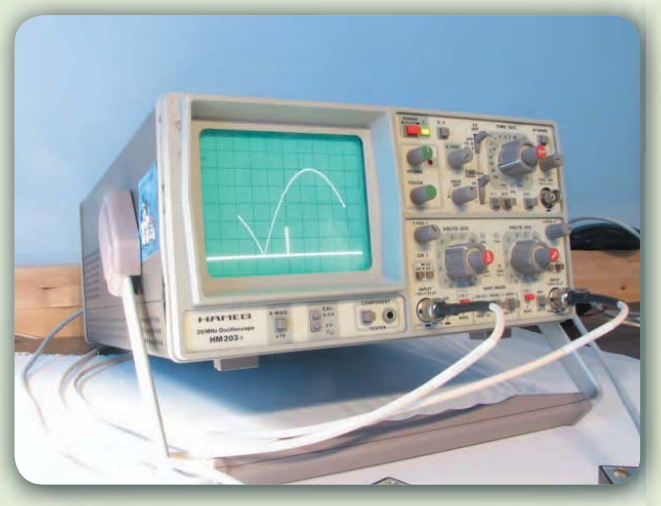

Figure 2. Tennis for Two, image provided by Windell H. Oskay (www.evilmadscientist.com) under Creative Commons License (www.creativecommons.org/licenses/by/2.0)

Despite the enduring popularity of *Spacewar!* and its continuing availability in derivative forms, the ability to experience the game in its original form is very much in doubt. At this point, the only PDP-1 known to still be in operating condition is at the Computer History Museum in Mountain View, California. Restoration of that PDP-1 to operating condition was a lengthy and difficult undertaking, requiring cannibalization of parts from other systems when available and jury-rigged repairs to existing parts when they were not. Maintenance of such antiquated equipment is likely to prove more difficult and expensive as time goes on, and replacement parts, and the expertise to apply them, become increasingly rare. While visitors to the Computer History Museum can experience *Spacewar!* on its original platform now, the inevitable deterioration of the equipment on which it runs means that *Spacewar!*, at least as it was originally conceived and played, will eventually cease to exist.

Spacewar! spawned a genre of work that has become a critical part of our cultural heritage. Computer games have achieved a tremendous level of social and economic importance. The research firm Screen Digest has estimated that the retail market for computer game software will reach $24.7 billion in sales in 2009, and online gaming may add another $14 billion to that amount[1]. Estimates of the number of online games globally go as high as 217 million people[2], or nearly one in four Internet users. If we are to understand the evolution of this cultural form, it is essential that we preserve the contributions that game designers make to its on-going development. Just as it would be very difficult to analyze a painting such as Manet's *Olympia*

1 Thomas & Mullen, 2007.

2 Estimate obtained from comScore. See: *www.comscore.com/press/release.asp?press=1521.*

without reference to Titian's *Venus of Urbino*, it will prove very difficult in the future to discuss games such as *Star Wars Galaxies* without reference to *Spacewar!*. It will also prove difficult to provide a full accounting of our cultural heritage outside the realm of gaming without preserving games as well. In an age where games such as *Tomb Raider*, *Resident Evil*, *Final Fantasy* and *DOOM* spawn media franchises including films, novelizations, comic books and a variety of action figures and collectibles, and where games such as Microsoft's *Halo* and Blizzard's *World of Warcraft* provide media production platforms for the creation of machinima videos, it is impossible to provide a proper contextualization for much of our existing popular culture without preserving games.

Unfortunately, the case of *Spacewar!* demonstrates much of the difficulty in preserving games. Unlike a book in a library, computer games have very poorly defined boundaries which make it difficult to determine exactly what the object of preservation should be. Is it the source code for the program? The binary executable version of the program? Is it the executable program along with the operating system under which the program runs? Should the hardware the operating system runs on be included? Ultimately, a computer game cannot be played without a complex and interconnected set of programs and hardware. Is the preservationist's job maintaining a particular, operating combination of elements, or is to preserve the capability to produce an operating combination using existing software and hardware? Is it both? Once these questions of the boundaries of the preservation object are addressed, there are a host of other difficulties presenting the would-be preservationist. What information, beyond the game itself, will we need to insure continuing access to the game? How should librarians, archivists and preservationists go about organizing the body of information needed to preserve a game? What strategy should we adopt to preserve software in a technological environment in which computing hardware and operating systems are undergoing constant and rapid evolution? Given the costs of preservation of normal library and archival materials, how can we possibly sustain the additional costs of preserving these complex and fragile technological artifacts?

Under the auspices of the Library of Congress's National Digital Information Infrastructure for Preservation Program, researchers at the University of Illinois, the Rochester Institute of Technology, Stanford University and the University of Maryland have begun an investigation into the preservation of computer games and interactive, computer-based fiction. Our hope is to try to build upon prior work within the digital preservation community on the preservation of data and extend it to address the preservation of software. While our research is on-going, we have proceeded far enough to have an understanding of the major problems, and at least some grasp on some solutions to those problems. The remainder of this chapter will outline some of the prior work in digital preservation upon which our research is based, describe

some of the significant problems in the preservation of computer games, and try to provide at least an outline of possible solutions to the problem of preserving virtual worlds.

The OAIS Reference Model & Problems in Preserving Virtual Worlds

Reference Model for an Open Archival Information System

One of the most influential pieces of work within the digital preservation community in the last decade has the *Reference Model for an Open Archival Information System (OAIS)* and standardized as ISO 14721:2003, *Space Data and Information Transfer Systems – Open Archival Information System – Reference Model* (International Organization for Standardization, 2003). The OAIS reference model was developed as a response to the growing problem of preserving digital information being produced by the world's space agencies. Audits by the U.S. Government Accountability Office have shown that much of the data collected by NASA over the years has been at great risk of loss, and there have been a few high profile instances of space data which has gone missing, including the original video transmissions of Neil Armstrong's first moonwalk during the Apollo 11 mission.[3] The OAIS reference model sets forth the responsibilities that any digital archive must fulfill, a functional model of a digital archive's operations, and an information model describing the types of information which must be gathered and maintained to preserve a digital object and their relationship to one another.

Figure 3. Destroyer-class Spacecraft from EVE Online, image provided by Tom Francis (a.k.a. Pentadact) under Creative Commons License (www.creativecommons.org/licenses/by/2.0.)

Several ideas from the OAIS reference model are essential to any effort to preserve computer games. The first is the model's notion of a 'designated community.' The model defines a designated community as "an identified group of potential Consumers who should be able to understand a particular set of information". This notion of a designated community is key to understanding the full import of another idea within the reference model, that of 'representation information.' The reference model asserts that to preserve a digital object, you need to preserve not only the object itself, but additional information necessary to decode that digital object. This additional information is called representation information and can be of two types: structure information and semantic information. Structure information documents

3 Sarkissian, 2006.

the basic data formats comprising the digital object, the rules which allow a series of bits to be interpreted as character data, numeric data, pixels, et cetera. Semantic information places an interpretative frame around the information in the digital object and allows the person examining the digital object to understand its meaning and importance. Structure information allows you to know that a particular string of bits in a data file represents a decimal number; semantic information allows you to know that the decimal number represents the atmospheric pressure of Mars measured in millibar.

Representation information itself may require additional representation information to render it interpretable to an archive's designated community. In theory, such recursive chains of representation information could become quite extensive. Understanding the PDP-1's Manual[4] (Digital Equipment Corporation, 1961), for example, requires a knowledge of the English language, so theoretically, a complete set of semantic representation information for the PDP-1 manual should include an English dictionary and grammar (preferably in one or more languages other than English). Recognizing the difficulty inherent in completely specifying (let alone gathering) the set of information necessary to render a digital object interpretable to a designated community, the reference model states that a digital archive need only collect sufficient representation information to be confident that a member of its designated community can understand the representation information and interpret the digital object. Those running the digital archive should monitor the state of knowledge of their designated community, and base the amount and kind of representation information they collect upon their designated community's knowledge base. An English-speaking community will not need an English dictionary and grammar to interpret documents written in English, and planetary scientists investigating the Martian atmosphere will not need a copy of *Atmospheric Science: An Introductory Survey*[5] to interpret data sets from the MARS Pathfinder mission.

Problems with Game Preservation

Like many recent efforts in the digital preservation arena, The Preserving Virtual Worlds research project has taken the OAIS reference model as a grounding point for our investigations. While the model provides an excellent basis for thinking through the issues surrounding the preservation of computer games, our project has encountered several obstacles in trying to make software preservation generally, and game preservation in particular, fit within the OAIS framework. These problems reflect the nature of computers games as both complex technological artifacts and complex cultural artifacts.

4 Digital Equipment Corporation, 1961.
5 Wallace, 2006.

The Object Organization Problem

A major problem in archiving computer games is that they have very poorly defined boundaries, even in comparison to other types of digital objects. The OAIS reference model was developed with a strong bias towards the archiving of *data*, and data files tend to have relatively fixed and stable boundaries. The set of data files comprising the United States Census results for the year 2000 is finite and discrete; the number of files is fixed and limited, and they stand separate and easily distinguished from other data files. Software, particularly modern software, does not have such clearly defined borders. If we look at an early piece of interactive fiction such as William Crowther's widely disseminated text game *Adventure*, we may believe that games can be as clearly and easily delineated as datasets. On first examination, it appears to consist of two files, one containing the game's FORTRAN source code, and the other containing a collection of data used by the software, consisting primarily of text the game will display to a player and a map of the cave system that the player will explore. On closer inspection, however, the situation becomes less clear. The source code, after all does not constitute a playable version of the game. To do that, you must compile the source code and create a binary executable. Binaries, however, are specific to a particular operating hardware and operating system platform. The original *Adventure* was developed for a PDP-10. A binary created for that platform will only run today under an emulator. Is our game then the source code? The source code plus the compiler and linker needed to turn it into a modern executable? An executable file for the PDP-10 platform? An executable for the original platform and an emulator to enable it to run on current hardware and software? All of the above? These problems may seem relatively trivial in the case of *Adventure*, but if we consider more modern games, they rapidly escalate. A game such as *DOOM 3* relies on the ability to draw upon dynamic link libraries included with the Windows operating system; the line between the game and the operating system it runs on has become more blurred than in the case of *Adventure*. *Doom 3* is also more tightly bound to a particular hardware platform than *Adventure* was. The minimum system requirements for the game advertised at id Software's website include a "100% DirectX® 9.0b compatible 64MB Hardware Accelerated video card with the latest drivers" as well as a "100% DirectX® 9.0b compatible 16-bit sound card and the latest drivers". The video card must employ one of a limited number of chipsets from the ATI® Radeon™ and nVidia® GeForce™ families. In the end, a *functioning* computer game is a program executing with a particular computing environment composed of hardware and additional software. Trying to separate out particular components of the complete system as representing 'the game' as an entity separate and distinct from the rest of the operating platform is in many ways a somewhat artificial and futile enterprise.

There is a temptation when trying to resolve questions regarding the exact boundaries of a computer game to try to draw upon the OAIS reference model's notion

of a 'designated community.' If we take the point of view that the exact boundaries of an artifact such as a computer game are determined by a process of social construction rather than reflective of some underlying reality, than we might be better off trying to determine how our designated community draws the boundary lines that determines what constitutes a game, and use their criteria in deciding what exactly the object of preservation for a given computer game should be. While attractive in theory, this approach runs up against several practical barriers. The first is that, to the extent that we can identify a designated community for game preservation, they appear to be a rather diverse group of people who exhibit a wide range of interests and knowledge with respect to games. There are members of the gaming community who are interested in playing older games, but may not care about having a perfect recreation of the original gaming experience. There are cultural and media studies researchers who may want to gain a better understanding of how the game was experienced when it was originally released. There are researchers interested in the history of technology who may want to examine the underlying technological components that went into constructing a game. Each of these communities might have different views of what constitutes 'the game.' For the gaming community, preserving an intact version of the complete original gaming platform may not be significant. A version which has been migrated to modern computing platforms, or even a re-creation of the original, may be sufficient; their definition of what constitutes 'the game' can be rather abstract and might be fulfilled by something which provides a reasonable recreation of the experience of the game, even if it is not exact (and perhaps not derived from the original artifact at all). For cultural and media studies researchers, seeing the game as it was originally experienced on older hardware may be significant. While they, like the gaming community, may be focused on the experiential aspects of the game, their demands for authenticity can be much higher. Researchers in the area of history of technology, however, might not care about having a functioning version of the game; they might care intensely, though, about having access to the source code used to create the game, something that might not be of concern to researchers in cultural studies.

Figure 4. DOOM 3, image provided by Alex Hopkinson under Creative Commons License (www.creativecommons.org/licenses/by/2.0)

The potential diversity of the members of the designated community for an archive of computer games makes it difficult to use their world views as the basis for iden-

tifying the exact boundaries of the object of preservation. This diversity is also problematic when trying to determine appropriate amounts of representation information to store. Some individuals interested in a game such as *Spacewar!* may very well have backgrounds in electrical engineering and computer science; the gaming community has a number of such individuals who have worked on creating software emulators for early gaming platforms. Individuals with this level of knowledge and sophistication with regards to computer platforms and operating systems might not need much more than a copy of the *Programmed Data Processor-1 Manual* to be able to decipher the game's original assembly language source code. For others without that level of knowledge, however, the manual for the PDP-1 is hardly a transparent explanation of the MACRO assembly language used for *Spacewar!*. A significant amount of additional information would need to be maintained to render that particular piece of representation information interpretable. Unlike a scientific data archive, where it can be safely assumed that the designated community consists of a relatively homogeneous community of research scientists, a computer game archive is likely to have a designated community consisting of individuals with myriad reasons for their interest in games and widely varying levels of technical knowledge.

Figure 5. A Second Life Museum Exhibit on the Declaration of Independence, image provided by John Lester under Creative Commons License (www.creativecommons.org/licenses/by/2.0)

Assuming that game archivists manage to successfully negotiate the difficult issues surrounding what information they need to archive to preserve a game, they must still deal with a variety of technical, legal and economic issues surrounding game preservation. On the technical side, game archivists will confront issues both with respect to organizing the information they have collected, and with insuring the ability for users to experience the game. While the digital preservation community has made progress in both of these areas, most of the work to date has focused on the preservation of various kinds of data, not software. Whether the approaches to preservation being contemplated for digital data will be adequate to deal with the scale and complexity of software is an open question.

With regard to the organization of data, a single modern game can comprise thousands of files of differing kinds; depending on the boundaries established for what to include in the object of preservation, the files contained in the game might include textual source code files, binary executables, shared object libraries, manuals in any of a number of possible document formats, still image files, sound files, mo-

ving image files, 3D geometry files, and others. Every different format will require its own structural and semantic representation information. Moreover, a single game may have multiple versions; versions may differ from each other with regard to some files, but not others. And in some cases, the game as originally issued represents only a segment of the content likely to be of interest to a designated community. Many recent games allow for significant amounts of user-contributed content. Game scholars in the future studying *DOOM* will certainly be interested in the various user created WAD files which substituted new levels and monsters (such as the purple dinosaur character Barney) for the original ones. It is hard to imagine anyone studying *Second Life* without examining the huge quantity of 3D content generated by its users. In addition to tracking all of this content, in all of its various versions, archives will also need to keep track of intellectual property rights information, descriptive information needed to inform users of what materials the archive holds, and provenance information regarding the archive's own efforts to keep the games viable and accessible.

The Preservation Strategy Problem

The processes necessary to keep games functioning are also a problematic issue for game archivists. There are four basic strategies for preservation of digital information: storage, migration, emulation and reinterpretation. Storage, simply storing the original bits as they are first received by the archive, is the most feasible and also the least helpful strategy. With the rapid evolution of computing technology, the odds of any game still be playable after a period of ten years will be small, and after twenty years, non-existent. Migration, the process of making changes to digital information in order to maintain access to it on modern computing platforms, has been used as a strategy for the preservation of software within the computing industry for quite some time. However, what works perfectly well for a business application is not always suitable for a highly interactive and creative work such as a computer game. Subtle changes in timing and display can alter game play significantly. Moreover, this strategy, unless combined with the storage strategy, destroys the original artifact; it seeks to preserve the experience of the digital object at the expense of preserving the object itself. Emulation, the process of creating a software-based recreation of a specific computing platform to essentially fool a piece of software into thinking it is running on the equipment for which it was designed, has also been used by the software industry as a preservation strategy. As with migration, emulation is not guaranteed to provide an exact recreation of the original experience. Unlike migration, however, emulation does at least try to preserve both the viability of the game and the integrity of the original artifact. The final strategy, reinterpretation, has emerged from the museum community as they have sought to confront the issues

around preserving artistic works in new media. Reinterpretation involves the recreation of an artistic work based on instructions from the artist and/or curator. While reinterpretation may have the potential to provide some experience of a work which cannot be maintained in its original form, it is obviously a very risky strategy, and its viability for complex games has never been tested.

None of these strategies appears to provide a perfect answer to the preservation of games in all instances. Storage and maintenance of the original bit streams for games may be useful for those who wish simply to examine the files without playing the game, but the utility of this approach is obviously limited. Individuals in the gaming community have managed to extract the contents of older game cartridges to make copies of games such as Atari's *Star Raiders*, but storing a copy of this material does not by itself provide any meaningful access to the game. Research conducted at the University of Michigan[6] on emulation and migration as preservation strategies for game software revealed that neither technique provided an exact replication of the original gaming experience; users reported a number of differences between the original game and both a migrated version and a copy of the game running under an emulator, with neither preservation strategy being preferred by a statistically significant number of their subjects. Moreover, for certain modern, multiplayer games such as *World of Warcraft* and *Second Life*, preservation strategies focused on the game as artifact are to a certain degree missing the point. The primary attraction of these systems is the ability to interact with other users, and while the capacity to interact may be preserved via a technological strategy, the unique culture of the game world cannot.

Legal and Economic Problems

There are also legal and economic impediments to these various preservation strategies. The Digital Millennium Copyright Act makes it illegal to defeat a technological protection mechanism used to prevent copying of software. While an exemption to allow libraries and archives to circumvent such mechanisms to preserve copies of obsolete software and video games was granted by the Librarian of Congress, this exemption is temporary and will lapse in October of 2009. Many computer games have employed some form of technological protection measure, and unless this exemption is renewed, even storage may not be a viable (or more precisely, legal) strategy for preserving some games. Another problem that our project has encountered is that the exact copyright status and ownership of many games is less clear than it might appear. A great number of people contributed to the development of games such as *Spacewar!* and *Adventure*, and the exact terms under which this work was done were never formalized, to say the least. It is not clear whom, exactly, an archive would need to negotiate with to obtain permission to engage in preservati-

6 Hedstrom & Lampe, 2001.

on actions such as copying a resource or preparing a derivative work through migration. There are also many commercial games that qualify as orphan works; while they were copyrighted to the company which produced them when they were first released, that company has gone out of business, and there is no clear record indicating to whom the copyright devolved.

Figure 6. International Spaceflight Museum in Second Life

The economics of game preservation are another significant obstacle. Migrating a modern computer game to a new platform is a non-trivial task requiring multiple programmers and potentially thousands of hours of labor. Writing an emulator for a modern operating system is also a very difficult and expensive task, as those involved in projects such as the Wine emulator (which provides an emulation of Microsoft Windows under Linux and other Unix-derived operating systems) can attest. While there has been research on ways these costs might be reduced, such as the work by IBM and the Koninklijke Bibliotheek in the Netherlands to produce a Universal Virtual Computer[7], to date there does not appear to be any universally applicable solution. Keeping any data other than simple text alive in the long term is an expensive proposition; keeping software alive is likely to be significantly more so.

Despite these rather significant obstacles, however, our project believes that it is possible to make headway on the preservation of computer games and interactive fiction. While we do not claim to have anything like a full solution to the range of problems confronting those wishing to preserve games as part of our cultural heritage, we do believe that there are some strategies which will make the problems more tractable. How then, can we go about trying to preserve these virtual worlds?

Some Possible Solutions to Preserving Virtual Worlds

The impediments to preserving computer games are both numerous and significant, and the technological obstacles are probably the most manageable part of the problem. If we are to preserve computer games, the relationship between those responsible for preserving our cultural heritage and the communities they serve is going to need to undergo a radical restructuring. Libraries, museums and archives have

7 van der Hoeven, van Diessen & van der Meer, 2005.

traditionally seen, if not a yawning gap, at least a fairly clear line between the staff sitting behind the desk and the patrons in front of the desk. The boundaries between patron and preservationist are going to have to become far *less* clear if we wish to succeed in preservation of virtual worlds.

Changing Relationships

One of the most essential tasks for institutions supporting the preservation of computer games will be to establish infrastructures that will assist their patrons in contributing to the work necessary to preserve virtual worlds. As an example of what this might entail, consider the creation of emulators for obsolete gaming platforms. The gaming community has invested quite a bit of creative energy in creating emulators for older videogame systems such as the Atari 2600. These emulators are typically distributed as freeware, and are often open-source; three well-known emulators for the Atari 2600 system, z26, Stella and PC Atari Emulator, are all distributed as open source packages under the GNU Public License and are freely available. Creating an emulator, even for some of the older, simpler gaming platforms, is a difficult undertaking, and beyond the resources of most archives. One of the essentials for creating a good emulator is access to the specifications for the CPU used in the original game, including the chip's opcodes and instruction details, the chip speed, and details on memory addressing and interrupts. Some of this information is available through informal channels such as gaming community sites, but it is often incomplete and can be difficult to obtain. As it happens, much of this type of technical information is invaluable as representation information if you are confronted with the need to preserve a binary executable and need to understand the hardware architecture upon which the executable is intended to run. If the archival world were to create an online repository of freely available technical documents and specifications for various types of computer hardware that the gaming community could draw upon, it would ease some of the work of those in the gaming community involved in developing emulators. Libraries, archives and museums are not really in the position to become developers of emulators for a variety of obsolete platforms, but if they can use their expertise in collection management to provide the knowledge needed to assist others in that task, they may be able to draw upon a vast pool of technical talent willing to volunteer their labor to assist in game preservation efforts.

The gaming community, and other communities of practice interested in games, can make other contributions to the preservation of virtual worlds as well. As mentioned previously, multiplayer online games such as *World of Warcraft* and *Second Life* present a major problem for preservationists in that preserving the software does not adequately preserve what is important about the game. There is an entire online

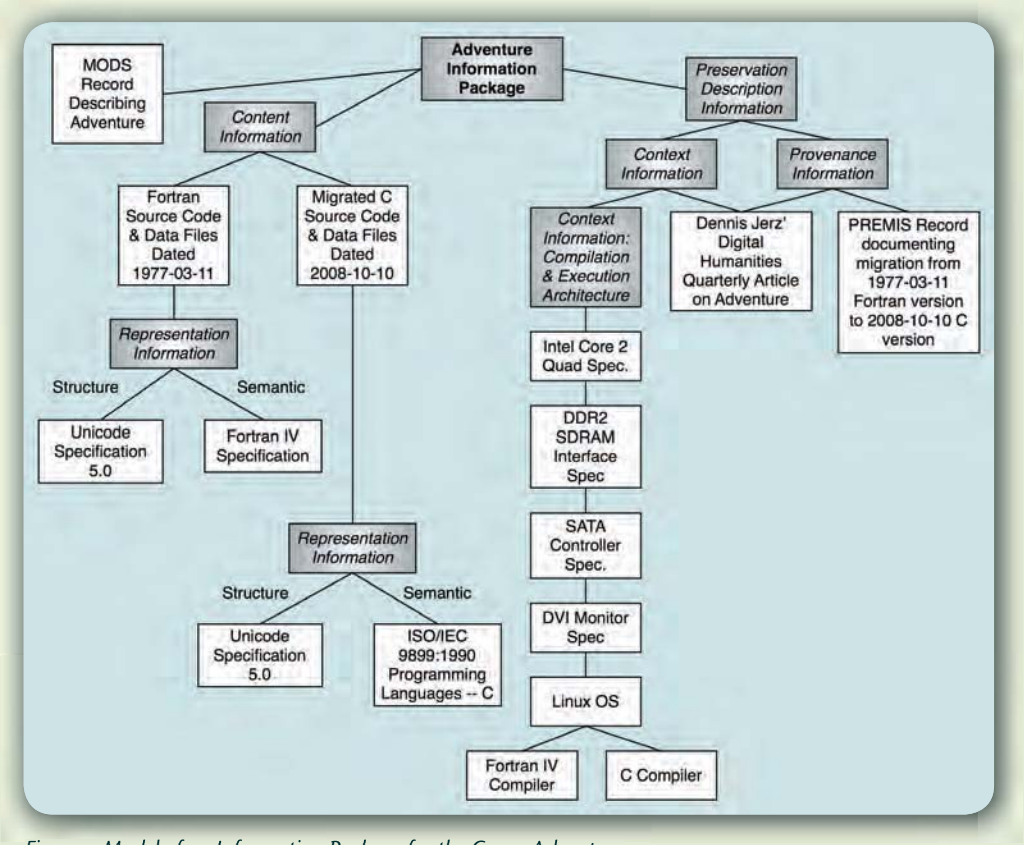

Figure 7. Model of an Information Package for the Game Adventure

culture surrounding such games, carried on both within the virtual world and in electronic venues outside of it, and simply preserving the game software and its virtual terrain will not protect that. As with the production of emulators, the gaming community has done a great deal of work in gathering together information documenting the culture of various virtual worlds and making that information available on the Internet. There has also been a good deal of scholarly research on virtual worlds which has gathered significant documentary material. Unfortunately, much of the material from these sources which has been made public suffers from the same problem as all material placed on the web; it has a half-life of approximately two years. If archivists were to provide a stable storage infrastructure for this kind of material to which the gaming community and scholars could actively contribute, it might vastly improve our ability to preserve not only these games but knowledge of the cultures which have emerged with them. As part of our research, the Stanford Humanities Lab and the Internet Archive have created such a space within the Internet Archive's Moving Image collection. The 'Archiving Virtual Worlds' collection[8] within the Internet Archive has received over 200 contributions to date, includ-

8 The collection can be found at *www.archive.org/details/virtual_worlds*.

ing footage documenting the closure of Electronic Arts' *EA-Land*, various areas and events within *Second Life*, raiding parties in *World of Warcraft,* and early multiplayer online systems such as *Onlive Traveller* and the *Starbright Pediatric Network*. The reaction among members of the gaming community so far has been positive and a more extensive version of such an archive, with support for a wider range of formats and the assistance of curators to do outreach to denizens of different gaming worlds, would appear to offer an excellent opportunity to promote collaboration between game archivists and the wider world interested in preserving knowledge of the history of these games' development.

In addition to taking steps to enlist patrons as active collaborators in preserving virtual worlds, archivists working on the preservation of computer games need to rethink their relationship with each other. As the previous examples of possible collaborative infrastructures to support preservation suggest, there are a variety of tasks that will need to be undertaken to insure the preservation of virtual worlds, and no single institution, not even the Library of Congress, will be able to address all of them successfully. A *coordinated* strategy for game preservation, with different institutions dedicating themselves to different roles, will be essential if games are to be preserved. Some institutions may want to specialize in trying to provide collaborative infrastructures such as the above. Others may want to focus on providing digital repository services for particular kinds of content, with some focused on console games, others focused on interactive fiction works, et cetera. Given the limited financial resources of the cultural heritage sector, we will need to develop mechanisms for leveraging our collaboration with each other to accomplish more than we could acting independently.

Reforming the Legal Environment for Preservation

We will also need to channel our efforts into trying to create a legal infrastructure that makes the preservation of virtual worlds feasible. Within the United States, the current exemption to the Digital Millennium Copyright Act to allow defeating technological protection measures in order to preserve video games is certainly better than nothing, but its temporary life span drastically limits its potential impact. This is exacerbated by the fact that the exemption applies only to works in formats where "the machine or system necessary to render perceptible a work stored in that format is no longer manufactured or is no longer reasonably available in the commercial marketplace"[9]. You can purchase an Apple II through online vendors such as eBay; would a game such as *Mindwheel* on the Apple II platform then not qualify under this exemption? The language of 'reasonably available' is subject to a great deal of interpretation and may leave libraries, archives and museums feeling that they are risking legal exposure if they take action to rescue material from a dying format while

9 Copyright Office, 2006.

a reading device is still available somewhere. As custodians of cultural heritage materials, librarians, archivists and curators will need to continue to press for a legal system that allows for virtual worlds to be preserved just like any other material that we have handled in the past.

Organizing Virtual Worlds for Preservation

In addition to insuring that the social and legal dimensions of preserving virtual worlds are handled correctly, we must also take steps to address the technological dimensions. The two most significant technological issues to be confronted in archiving computer games are organizing the large amounts of data and metadata that must be gathered and generated in the process of generating games, and insuring viable access to games as computing platforms change and evolve. Our project is investigating both of these issues.

The digital library and digital preservation communities have settled on XML as their preferred format for recording metadata, due to both its simplicity and its standardization. An XML document is fundamentally a text document, and experience has shown that plain text is an extremely stable and sustainable form of digital information. Our project has thus settled on XML as our basis for trying to manage both the digital files comprising computer games and metadata that describes them.

There are a variety of XML formats that have been developed specifically to provide generic and extensible mechanisms for wrapping together content and metadata for digital library and digital curation applications. Some of the better known ones within the digital library community include the Metadata Encoding & Transmission Standard (METS), the MPEG-21 Digital Item Declaration Language (DIDL), the XML Formatted Data Units (XFDU) specification, and the newly-emerging Object Reuse & Exchange specification from the Open Archives Initiative (OAI-ORE). All of them provide the capacity to arrange a set of content files into a hierarchical arrangement of subsets, and link individual files or subsets with both descriptive metadata (necessary to provide intellectual access to the materials) and administrative metadata (necessary to insure proper long-term custodial care, and including representation information, intellectual property rights management information, and provenance information).

Figure 7 provides a graphic depiction of the arrangement of files and metadata we might need to handle a case such as the classic interactive text, A*dventure*. For the content files, we would certainly want to include the earliest available version, the original Fortran IV source code authored by William Crowther in 1977. We might also want to have a migrated version of that code available, written in a somewhat more recent language, to simplify access to a playable version of the game. For both of these versions we will require structural and semantic representation information;

as they are all text files, for the structural information we will want a copy of the Unicode character specification. For the semantic representation information, we will need the Fortran IV and C programming language specification documents. We will also want to provide archive users some contextual information about how these files might be used, including the specifications of a computing platform capable of compiling and running either the Fortran IV or C language versions. The Digital Humanities Quarterly article[10] provides additional contextualizing information regarding the game to scholars. We will want to document the provenance of our various source code incarnations. The article by Dennis Jerz provides information regarding the provenance of the original source files, but we will want additional documentation regarding the migration to the C language version. PREMIS is a metadata standard developed by the digital library community specifically to document the provenance of digital objects, so we can use a metadata record in this format to document the migration event. Finally, we will want a link to a descriptive metadata record (in this case, in the MODS XML format maintained by the Library of Congress) to inform users of the contents of this package.

What this graphical model demonstrates is that even for a game consisting of a single source code file and a single data file, a large amount of ancillary information must be tracked and maintained. Much of that information consists of documents which might not all be in digital formats (and in fact, the OAIS reference model dictates that some representation information must not be in digital form). There is also a certain amount of replication present; both the Fortran version of *Adventure* and our C version reference the Unicode specification as structural representation information. We anticipate that this sort of replication will also be present across digital objects; probably every text file on a modern computing platform is likely to have Unicode as structural representation information. This does not mean that an archive will need hundreds of copies, obviously. A single copy referenced by the various digital objects will do.

The data/metadata management component of game preservation then, requires an XML wrapper which can:

1.1. provide links (preferably through some persistent identifier) to each of the digital content files comprising a version;

1.2. sort content files into subsets hierarchically as needed to indicate different versions (and components of versions);

1.3. link a content file to representation information, either directly (in the case of digital representation information) or via a bibliographic reference to an analog copy;

10 Jerz, 2007.

1.4. link content file(s) to additional contextual information (in both digital and analog forms) beyond the representation information needed to understand and employ the content files;

1.5. track the provenance of all of the various content files associated with the object; and

1.6. link the entire contents of the package to a descriptive metadata record that provides bibliographic access to the game by the archive's users.

Fortunately, this is all relatively easy to achieve with the existing XML wrapper formats mentioned above, supplemented by additional XML-based metadata formats for description (such as MODS) and provenance (such as PREMIS). This is not to say that the production of such a package is a simple (or quick) process, but existing standards for data and metadata management seem capable of being extended to handle software preservation.

Strategies for Preserving Virtual Worlds

The issues around insuring the viability of a game as a functioning piece of software are much more difficult, and our project is still exploring them. However, from our investigations of various older games to date, we suspect that there is not going to be a single solution which can be applied equally to all games in all cases. Much will depend on the technological nature of the original artifact and what materials are available for the archives to collect. A cartridge for a console-based game such as *Star Raiders* is going to be a poor candidate for migration. Even though Atari still exists, the odds of original source code for the game being available as the basis for a migrated version at this point are small; game companies have historically done a remarkably poor job of preserving their own products. Copying the ROM image from the game cartridge and running it under an emulation program for the original game platform is a much more feasible solution for such a case. For a game like *DOOM*, however, where the company has released the source code for the Linux version under the GPL open source license, migration is a distinct possibility and may in fact be easier and cheaper than trying to maintain a Linux emulator in working condition in the future.

Conclusion

Computer games and interactive fiction form an essential part of our cultural heritage, and their importance as both artistic works and cultural artifacts only seems likely to increase. Their status as complex, functioning virtual machines, however, makes them exceptionally vulnerable to deterioration and loss. Only by reconcep-

tualizing our role as librarians, archivists and curators and our relationships with our users (and each other) will we be able to attempt the preservation of these materials. Fortunately, the technological impediments to the preservation of virtual worlds do not seem insurmountable. With enough effort and cooperation, *Spacewar!* may last long enough to be enjoyed by passengers on a ship E. E. Smith only dreamed of.

Jerome McDonough is Assistant Professor aan de Graduate School of Library & Information Science van de University of Illinois. Voordien was McDonough teamleider van het digitale bibliotheekprogramma van de New York University. Hij was lid van NISO's Standards Development Committee en was betrokken bij verschillende initiatieven rond metadatastandaarden, zoals METS, PREMIS and the Open Digital Rights Language. Zijn onderzoek focust zich op het gebruik van metadata in digitale bibliotheeksystemen en de bewaring van digitale informatie.