ONE THING AFTER ANOTHER: DEVELOPING A DIGITAL ARCHEOLOGY FOR
THE WARREN SPECTOR COLLECTION, A CASE STUDY

ELIZABETH L. SERAMUR

INTRODUCTION

In the spring of 2009, I worked with fellow students Liza Banks and Simon Staats at the University of Texas at Austin (UT), School of Information, to design and implement a digital preservation strategy for a group of computer disks dating from 1982 to 1988. Taken together, this group of disks constitutes the "Early Works" portion of the Warren Spector Design Documentation materials, chronicling the incipient stages of his career, with files ranging from academic notes such as "PhD Ideas" from the early 1980s to his novels in the late 1980s. These records are maintained in the UT Videogame Archive at the Dolph Briscoe Center for American History (DBCAH) in Austin, Texas. These eighteen 5 ¼ in. floppy disks contain mostly word processing files related to game developer Warren Spector's (b. 1955) work in the design of pencil-and-paper role-playing games.

Retrieving and accessing files on these disks entailed mastering technology and protocols from two decades ago, as well as some creative hardware manipulation. This process involved strategies ranging from digital archeology to automated upload of the records into an online digital repository with the necessary documentation and metadata for their long-term preservation and access on current machines.

CONSIDERING OBSOLESCENCE: HARDWARE, SOFTWARE, AND INFORMATION

Preserving archival records that exist solely in electronic form is fundamentally distinct from traditional archival practice. Unlike paper-based or other physical

documents—whose contents may be quickly ascertained via visual inspection—electronic records' contents are written in binary code to machine-readable media. These materials require locating and operating the appropriate hardware in order to access them. The earlier the production date, the more challenging this task becomes because of obsolescence, media failure, and the physical preservation needs of the machines themselves.

Scarcity constituted a particularly acute setback for our project as we attempted to locate the correct equipment. Many of the disks were created using a rare type of personal computer known as a Kaypro, which ran CP/M—an operating system predating Windows DOS. In order to access the files on the disks, we needed precisely this type of machine. Rampant variability in early computer production mean that files created in one computing environment may not open in another, even very similar, machine. Lack of standardization in early personal computers also results in compatibility issues between various kinds of drives, ports, cables, and other access issues, often within the same model number. Hardware incompatibility is pernicious to all digital preservation activities using original machines, and will continue to present problems in that the smallest variations in a system can radically alter a documents significant properties. Our computer scavenger hunt took us all over campus, including the Harry Ransom Center, University of Texas at Austin, which maintains original equipment in its literary collections. Appropriate machines emerged, but they each lacked various components, precluding their use. Finally, I discovered the Goodwill Computer Museum in Austin, Texas, a volunteer-run organization affiliated with Goodwill Computerworks recycling and training program that could provide the necessary hardware (fig. 1).

With the older machines, the operating system often requires a bundle of disks, such as boot disks to start it up, and other disks to run certain operations, to even be used at all. Most early machines had no hard drives, and

Fig. 1. The Goodwill Computer Museum, Austin, Texas.

those that did had very little memory. Likewise, software packages, such as WordStar or AppleWorks, entail multiple program and boot disks to run. While, in general, software obsolescence presents a greater threat to the continued accessibility of materials—hence the argument advocating frequent file refreshing and migration operations—when dealing with records created on the earliest personal computers (trapped on their original media) the disappearance of hardware and software shares the same threat level.

Beyond machinery and software applications looms another critical form of obsolescence: human knowledge and experience. This project would have been impossible to complete without the insights of individuals familiar with these specific vintage systems. Instructional manuals were scarce and incomplete. We discovered very little relevant documentation available in professional archival or computer science literature since most articles we consulted discuss files accessible on today's computers. Apart from the first-hand knowledge provided by individuals having used machines contemporaneously,

we found most of our reference material from so-called hobbyists' web sites and other online forums. These too must be preserved in order to maintain access to important information, since sites themselves are becoming obsolete as their creators cease hosting them.

CASE HISTORY

Warren Spector is a video game designer and developer best known for his work with role-playing and first person shooter games. His early work developing pencil-and-paper role-playing games produced such hits as *TOON* (1984), *Top Secret/S.I* (1987), *Rocky & Bullwinkle Game* (1988), and *Spelljammer* (1989). In addition to games, he wrote a novel entitled *One Thing After Another* (1987) and a spy novel, *Hollow Earth Affair* (1988). Spector presciently transitioned to the computer gaming industry in 1989, and he continues to contribute actively to the development of successful computer games.

In 2007, Spector donated his papers to the fledgling UT Videogame Archive. The Archive's stated mission is to ". . . preserve and protect the records of videogame developers, publishers, and artists for use by a wide array of researchers." Its collections pose myriad preservation challenges due to the wide variety of file formats and software. Because they currently have little period equipment, until we completed this project, these files were inaccessible to archivists and researchers.

SIGNIFICANT PROPERTIES: DECIDING WHAT TO PRESERVE

A critical juncture in any digital preservation project entails identifying which aspects constitute the files significant properties, and hence the properties of the files or objects that ought to be preserved. Hedstrom and Lee (2002, 218–227) define significant properties as ". . . those properties of digital objects that affect their quality, usability, rendering, and behavior." These are largely determined by the functionality of the program in which the files are created. Whilst perusing paper records from Spector's collection, we came across his complaints about a new word processing software, that

was ". . . causing lots of problems . . . lack of headers, page numbers and other standard elements of professional submissions" (Spector 1986). As a writer, Spector obviously considered these "elements of professional submissions" significant properties of his work. For the objects in the Warren Spector Design Documentation having standard file formats, namely word processing documents written in WordStar and AppleWorks (which we identified using the JHOVE Harvard Object Validation Environment database), we can determine the functionality by ascertaining what affordances were available through these programs at the time of creation.

For example, in WordStar 3.30—included on the 1983 Kaypro IV computer—all formatting had to be added after the document was written. The file extension DOC indicated files with formatting, as opposed to unformatted ASCII text, TXT files. Add-ons included an early version of MailMerge and the SpellStar spell-checker. Files created on the Apple disks were created using AppleWorks version 1.0–2.0 (1986). AppleWorks' distinguishing feature was its graphical user interface (GUI), allowing users to navigate with a menu bar rather than a command-line interface. There were various add-ons, such as MacroWorks, which automated user-designated commands from the keyboard, and FontWorks, which facilitated printing in different fonts and sizes. Other available features included a spell-checker and a thesaurus.

Considering significant properties recalls Marshall McLuhan's motto "the medium is the message" (McCluhan and Gordon 2003). In the case of Warren Spector's files, the textual information they contain *could* have been generated on a typewriter, or even printed out. Still, the method of composing, editing, and storing files in a computer interface dictated the experience of creation, and to a large extent, the results. The constraints imposed by storage media determined the number of files that could be held in direct relationship with one another. Fascinatingly, Spector frequently used custom file extensions to convey information about the contents of the files, since

the operating system (OS) permitted only a small number of characters in the file name. Even if we had print-outs of all of the files, the information derived from viewing these disks in their intended environment adds a dimension that is impossible to convey in any other manner. This argument is analogous to the experience of handling a rare book in person, versus merely the microfilmed or digitized text; copies or surrogate formats are inherently incomplete records.

METHODOLOGY

We commenced this project by inventorying and describing the physical properties of the media in our collection. We imposed an ad hoc labeling system to preserve the order and created a spreadsheet to record observations such as handwritten and manufacturer's labels on disks, brand, notes, and physical condition (e.g., dust, scratches, and write-protection). Next, we documented the disks and their storage sleeves front and back photographically. Prior to locating machines capable of reading the disks, these notes provided valuable hints; inscriptions included details such as titles, computer models, and software used in their creation.

Based on visual inspection of the disks, we speculated that the majority worked with Apple II and Kaypro personal computers. In order to know for sure, we needed access to these old computers with the corresponding operating system and software. Enter the Goodwill Computer Museum and Russ Corley, the museum director. Corley, himself a retired software engineer volunteering at Goodwill, noticed many rare personal computing specimens intermingled with the rest of the donated equipment slated for recycling. He began curating, and now maintains an illuminating chronology of the evolution of personal computers.

From his collection, he located working Apple IIc, Kaypro 10, and Kaypro IV machines, facilitating the opening of most disks. Once we were able to view the disks' contents, we went about creating a file inventory. We photographed the computer screen displaying the disk directory, creating ersatz screenshots, and information from each disk directory was cataloged. We also added fields to the disk inventory indicating whether or not we were able to open the disk, and if so in which computing environment.

The Apple disk directory listed filenames, sizes in "blocks," (which we calculated to be 1:2 ratio to bytes, 22 blocks = 44 bytes), and date modified. These files are formatted as AppleWorks word processor files (AWP). For the Kaypro disks, we initially used the native CP/M directory command (DIR) to view files. The resulting display only listed file name and extension. Subsequently, we discovered that the more robust directory available through the Kermit file transfer program included additional information. The directories revealed myriad file

Fig. 2. File extensions in Kermit on a Kaypro machine.

extensions: some appeared to be standard (e.g., DOC, TXT, and BAK), along with many user-defined (e.g., ICI, BAR, CON, LET, MSS, RES, and TST) (fig. 2).

In order to ingest copies of the electronic records on the floppies into the digital repository, we had to extract them from their original hardware and software environments to those accessible by current computers. Files were migrated in two senses: in the first, we transferred the original bitstreams—literally the stream of raw binary code that constitutes the electronic files—from their hardware and operating system environments; secondly we migrated the actual files as such from their original application-specific format to more open and widely available formats in order to create access copies. We used different mechanisms for the disks created on the Apple II versus those created on the Kaypro.

To migrate the original bitstreams from the disks created on an Apple II computer we created disk images using ADTPro, a bootstrap program with an Apple IIe with a Super Serial Card connected to a Macintosh PowerBook G4 running OSX via a serial cable. ADTPro installs a client program to the Apple II, enabling it to communicate with the server client running on the PowerBook. We also used a program called Apple Commander to open the disk images and export the individual binary files to the desired location on the PowerBook. To create access copies of the bitstreams from our Apple II disks, we used the export filters in Apple Commander to create ASCII text and Rich Text (RTF) copies of the files directly from the disk images. Since many Apple II emulators, (including AppleWin) rely upon them, we exported the disk images (DSK), allowing users to approximate the look and feel of the original environment.

Creating access copies from the Kaypro disks was more difficult. Eventually, we discovered that files created using CP/M and Microsoft DOS versions of WordStar are interchangeable, meaning we could utilize a DOS program called wsasc.com, available on the WordStar

website. WordStar files were converted to ASCII text files using this program in a DOS emulator, DOSEMU, on a Linux laptop. Non-WordStar files were exported as ASCII text files with the .txt file extension following the original file extension. After the original formatted and access copy files were made, the Kaypro files were transferred to a laptop running Linux by connecting a Kaypro 10 via a custom (aka rigged) mechanism we fashioned from a modified serial cable, null modem adapter, and breakout box (fig. 3).

Fig. 3. Transferring files from Kaypro to a laptop.

DIGITAL ARCHIVAL ARRANGEMENT

Once bitstreams, disk images, and individual files were transferred from original to current machines, they were uploaded to the DSpace digital repository and put into a meaningful arrangement so as to provide context and accessibility for researchers. The notion of archival order has perpetually occupied archivists' minds. Deciding just how documents should be grouped tends to hinge on physical order. Intellectual concerns such as *provenance* and *pertinence* determine where a document is physically situated, and this can only be in one place.

For born-digital archival records, these ideas are still relevant, but an archivist's ability to reconcile them is both complicated and liberated by the fact that while the electronic objects are stored in one place per se,

they can theoretically have infinite points of access, either via alternative virtual arrangements, or via external search engine results. These decisions have preservation implications as well, since preservation without access renders our efforts irrelevant. We decided on an arrangement based on the assumption that the "order-as-received" coincided with the original order of the materials. In order to mirror the hierarchical organization of the paper records, we took advantage of the options afforded by the DSpace platform to create collections, series, sub-series, and so on.

CONCLUSION AND RECOMMENDATIONS

This case study highlights many critical issues facing those charged with preserving and providing perpetual access to electronic materials. This includes archives, in the form of born-digital records, and museums faced with the daunting task of preserving digitally produced time-based artworks. The prevailing theme of this project was its puzzle-like nature; each component (hardware, software, documentation, and human knowledge) were effectively pieces of the puzzle, and the absence of any one of them would render the end-product incomplete or impossible to achieve. In light of this, it is imperative that we not only document and preserve the files, software, and equipment, but also the acumen of those having direct experience with older computing systems through written and oral histories, along with demonstrations of useful techniques. This human knowledge is even more ephemeral than the media itself. Because many institutions lack some or all of these components, along with the effective physical media in most collections, un-cataloged computer disks tucked away in "restricted" categories, constitute a backlog of immeasurable proportions. In contrast to paper records, electronic materials cannot simply languish until archivists get around to them.

The united problems of hardware, software, and knowledge obsolescence should serve as an impetus for knowledge-sharing and robust training. Strengthening materials conservation training particular to electronic media and hardware, as well as anthropological preservation such as oral histories are essential elements in the effort to retain the knowledge and technical skills of the original users of obsolete technologies.

Often it seems that museums seek to preserve objects *as such*, whereas archives tend to preserve objects for their information. Digital preservation demands a combination of the two. Open dialog in the form of jointly-authored articles, crossover conferences and symposia, workshops led by experts from multiple allied fields, and easy-access information sharing such as blogs and wikis are all necessary to maximize the diverse expertise of various cultural heritage fields in order to staunch the loss of important works created in digital environments. For more in-depth description of our process, please refer to the documentation available on DSpace at https://pacer.ischool.utexas.edu/handle/2081/20159.

ACKNOWLEDGEMENTS

REFERENCES

Hedstrom, M. and C. A. Lee. 2002. Significant properties of digital objects: Definitions, applications, implications. *Proceedings of the DLM-Forum 2002*. Luxembourg: Office for Official Publications of the European Communities. 218–27.

Spector, W. 1986. Letter to B. Heard. October 15. UT Video Game Archives, Dolph Briscoe Center for American History, Warren Spector Papers.

McLuhan, M. and W. T. Gordon. 2003. *Understanding media: the extensions of man*. Corte Madera, California: Gingko Press.

FURTHER READING

Apple ProDOS. 2009. Wikipedia, the Free Encyclopedia. http://en.wikipedia.org/w/index.php?title=Apple_ ProDOS&oldid=446656635 (accessed 05/02/09).

Business Desktops/Calcs. 2009. HP Computer Museum. www.hpmuseum.net/display_item.php?hw=41 (accessed 05/02/09).

Digital Preservation. 2010. Library of Congress. www.digitalpreservation.gov (accessed 08/13/10).

DSpace. 2009. About DSpace. DSpace Foundation. www.dspace.org/introducing (accessed 05/02/09).

Galloway, P. 2009. Structure in DSpace: Building the structure to receive your bitstreams. Lecture presented at INF 392G: Problems in Permanent Retention of Electronic Records, The University of Texas at Austin, April 8, 2009.

GDFR (Global Digital Format Registry). 2010. Harvard University Library, National Archives and Records Administration, and OCLC. www.gdfr.info (accessed 09/13/10).

Wayback Machine. 2010. Internet Archive. www.archive.org/ web/web.php (accessed 09/13/10).

ITNS. 2009. Homebase. http://apple2.org.za/gswv/a2zine/ GS.WorldView/Resources/ProDOS.File.Types.v2.0.txt (accessed 05/02/09).

ITNS. 1999. ProDOS File Types 2.0. http://apple2.org.za/ gswv/a2zine/GS.WorldView/Resources/ProDOS.File .Types.v2.0.txt (accessed 05/02/09).

JHOVE. 2010. Documentation. Harvard University Library. http://hul.harvard.edu/jhove/documentation.html (accessed 09/13/10).

Cruz, F. D. 2011. What is Kermit? The Kermit Project, Columbia University, New York. www.columbia.edu/ kermit/kermit.html (accessed 05/02/09).

Software Preservation Society. 2010. www.softpres.org (accessed 09/13/10).

Weyhrich, S. 2011. Apple II history. http://apple2history.org/ history/ah19 (accessed 05/04/09).

WordStar 2011. Resource site. www.wordstar.org (accessed 05/04/09).

Elizabeth Seramur
Independent archivist
seramur@gmail.com