2018 ASEE ANNUAL CONFERENCE & EXPOSITION
JUNE 24–27, 2018
SALT PALACE CONVENTION CENTER
SALT LAKE CITY, UT

Paper ID #21833

Long-term Preservation of Deprecated Media: How Can Libraries Provide Information From Today's CD-ROMs in the Future?

Jamie M. Niehof, University of Michigan

Engineering Librarian Computer Science & Engineering, Nuclear Energy & Radiological Sciences, Climate & Space Sciences & Engineering University of Michigan — Ann Arbor

Lance Stuchell, University of Michigan

Lance is Head of the Digital Preservation Unit at the University of Michigan Library. In this role he oversees policy and workflow development aimed at preserving the Library's digital assets. Lance received his undergraduate degree in History from Northwestern University and an MSI with a concentration in Archives and Records Management from the School of Information at the University of Michigan.

Ms. Leena N. Lalwani, University of Michigan

Leena Lalwani is an Engineering librarian and the Coordinator for Engineering Collection at the Art, Architecture and Engineering Library (AAEL) at the University of Michigan. She is also the liaison Librarian for Biomedical Engineering, Chemical Engineering, Materials Science, Naval Architecture and Marine Engineering and Entrepreneurship. Leena has been a librarian at University of Michigan since 1995 in various ranks. Prior to joining University of Michigan, Leena has worked as Librarian at Gelman Sciences and American Tobacco Company. Leena has a M.L.S. degree from Catholic University of America and M.S. in Chemistry from the University of Mumbai.

Mr. Paul F. Grochowski, University of Michigan

Paul Grochowski is an engineering librarian at the University of Michigan. He holds a bachelor's degree from the University of Michigan and a MLIS degree from Wayne State University.

Long-term Preservation of Deprecated Media: How can libraries maintain access to information from yesterday's CD-ROMs in the future?

Abstract

New data storage media advances in the 1990s brought changes to publishing practices. Storage media had gone through a series of progressions, and with falling costs of digital media, publishers now had new and affordable options for supplementing or publishing new works. Some print books included supplementary information on computer disks, and in other cases publishers made complete works available on computer disk instead of in print format. These changes in publishing are apparent in the University of Michigan (U-M) Library's engineering collection, which now includes a large collection of resources acquired in CD-ROM format. Today, the library faces concerns about the long-term viability and access to these resources. Computer storage media have already evolved from 5.25" floppy disks to 3.5" floppy disks to CD-ROMs to DVD-ROMs to USB drives. Since computer manufacturers are phasing out optical drives, users now have few options for using library materials that are formatted on disks of any kind. This "technological obsolescence" has prompted the engineering librarians at U-M to investigate how to continue providing access to materials that are published on computer disk in a future age where computer users will not have the required hardware or software available for reading the disks. Working with the library's Digital Preservation Librarian, the engineering librarians will determine which of the engineering resources that are published in disk format must be preserved, and they will plan for best practices for preservation of, and access to, the selected resources. Only complete works published on CD-ROM are reviewed in this project. This paper will report on methods used to evaluate and decisions about long-term retention and preservation of these resources, as well as strategies for avoiding this problem in the future.

Background

The University of Michigan (U-M) Library's collection of materials has been undergoing a drastic shift in the last decade. Engineering students and faculty now vastly prefer electronic versions of their textbooks and research materials, and physical space for print books and media is at a premium. As part of the ongoing process of weeding and inventory, cataloging discrepancies and missing items have been found in the stand-alone CD-ROM collection. Faced with evaluating the collection of 1,935 CD-ROMs, the engineering librarians have decided to initiate a comprehensive data preservation project. These CD-ROMs are not companion data to print books, but rather proceedings, standards, technical reports, data from federal agencies, and more. The library no longer actively seeks out CD-ROMs to add to its collection. Anecdotal

evidence shows circulation for the CD-ROM collection, which has never been very high, has dropped precipitously in the past two years.

In addition to not being actively collected or used by students, the format of CD-ROM media is dying. Apple was the first computer company to remove optical disk readers from consumer computer models starting in 2008 [1], stating that disk readers are "anchors" weighing down smaller and lighter designs [2]. Other computer manufacturers followed suit as the industry began to rely on cloud-based options for delivering software and storage, and users shifted to services like Google Docs, Dropbox, and Facebook for file storage and photograph sharing [3]. CD-ROM media are now "deprecated" in the sense that use of them requires computer hardware and software that are no longer uniformly available. Publications contained on the CD-ROMs should be evaluated and preserved where necessary.

This library's approach to proper digital preservation follows community practice [4] by storing material in actively managed server-based storage. Material deemed worthy of long-term preservation found on optical media should be transferred to more appropriate server storage.

The goal of this preservation project is to comprehensively evaluate each stand-alone CD-ROM in the engineering collection, and either withdraw it permanently, attempt to preserve its data for the future, or allow it to remain on the shelf until it is no longer useful. CD-ROMs accompanying print books are outside the purview of this project. Stand-alone CD-ROMs are a defined collection within the U-M Art, Architecture & Engineering Library and can easily be identified and evaluated, whereas the CD-ROMs accompanying print books are kept together with the books.

Literature Review

The CD-ROM format was introduced in the 1980s, and publishers and librarians quickly saw its potential. McCarthy wrote about the long life and storage capacity of the CD-ROM, saying the "CD-ROM has a longer life than microfilm or fiche. A disc can hold information that would require hundreds of fiche [5]." Murphy saw promise as well. "A very promising and cost-effective laser data storage medium is the compact disc." And "CD-ROM is ideal for storing and distributing archives and other 'static' databases [6]." The use of CD-ROMs for publishing databases prompted Williams to write, "[CD-ROM] media have exciting possibilities in publishing, the high page capacity linked to fast automatic access to the contents makes them particularly suitable for publishing large volumes of data [7]."

Yet by 1989, one author was already questioning the new technology. "A further possible disadvantage of the medium for storage purposes is that the actual life of a CD disc is unknown in practice [8]."

By the middle of the 1990s, CD-ROM was considered a well established technology. The book *CD-ROM in Libraries* was published to "…present in a more systematic manner the collective experience of the pioneer managers of CD-ROM in libraries [9]." Yet, one author questioned whether the CD-ROM would be here for the long term: "There remains the possibility that some new data storage medium will emerge to leave CD-ROM trailing in its wake and it is for this reason that the long-term future of CD-ROM cannot be predicted with any certainty [10]."

Later in the 1990s, publishers began offering CD-ROM equivalents to print reference resources, such as the *McGraw-Hill Encyclopedia of Science and Technology* [11].

In 2011, Iraci studied the life expectancy of a computer disk to determine whether cold storage could extend the life-span of the storage material. He reported, "At standard conditions of 23 [degrees C] and 50% relative humidity…projected lifetimes for CDs and DVDs range from 2 to 100 years…" Iraci also took a broader view, and continued, "Regardless of the longevity of electronic media, the technologies associated with reading or playing them have even shorter lifetimes; this is known as equipment obsolescence. Therefore, to ensure survival of the information stored on electronic media, it is essential to transfer it to new formats [12]."

More recently, archivists, librarians, and other users of CD-ROMs have begun asking about the preservation of information stored on CD-ROM and other digital media. The National Archives of Australia wrote in 2006 that "Digital records are subject to three types of obsolescence: 1) The physical carrier…becomes obsolete; 2) The hardware needed…becomes obsolete; 3) The software needed…becomes obsolete [13]." Lin et al. observed in 2003 that "an electronic record is preserved if and only if it continues to exist in a form that allows it to be retrieved…[14]," and del Pozo, Long, and Pearson, writing about digital objects stored on CD-ROM, said in 2010, "Before any action can take place on the image, there must be an appropriate medium for reading the stored form of the digital object. So, a computer with a CD-ROM drive will be required [15]." And, as Hogenboom and Hayslett stated in 2017, "In recent years, fewer computers come with CD drives at all…[16]"

Piascik, Shrauger, and Ritzert [17] wrote about a project undertaken to determine the disposition of CD-ROMs that were included as supplemental material in books, but no literature on studies about weeding or preservation of stand-alone CD-ROMs was found.

Methods

A spreadsheet of 1,935 CD-ROMs was provided to engineering subject liaisons in February 2016 by circulation staff. Instead of cleaning the errors on the list, engineering subject liaisons elected to make decisions on whether these CD-ROMs should be archived for future use or withdrawn.

Subject liaisons made the following decisions on the spreadsheet for each CD-ROM:

W = Withdraw from the collection
P = Preserve data if possible
K = Keep where it is until it is no longer useful

Decisions were based on the library's collection statement coupled with subject liaisons' knowledge of the material and available formats.

In some cases, conference proceedings or journal issues that the library has collected on CD-ROM have been duplicated by online holdings in archival format. These CD-ROMs are marked to be withdrawn. Other items that the librarians have determined could be withdrawn include superseded editions of engineering handbooks and government publications that have become available online at the government agencies' websites. In this way, it is possible to identify the remaining CD-ROMs requiring evaluation for preservation or replacement.

Where there was a determination that the materials on CD-ROM needed to be retained in the library collection, subject liaisons first looked for online availability. Materials were identified for purchase in an online format when possible. The remaining CD-ROMs were then evaluated for preservation.

Once decisions had been made on the first 52 titles on the spreadsheet, that batch of disks were sent to the Digital Preservation Librarian for evaluation and workflow development based on possibility of preservation.

Archiving and Preservation

In the U-M Library, preservation of material identified through the librarians' process requires that files be transferred off optical media to server storage to facilitate important preservation actions. Ongoing monitoring of a file's bit-level integrity and the creation of multiple backup copies are core preservation actions [18] and are much more effective using spinning-disk server storage. Transferring content to server storage also removes challenges posed to preservation and access due to the obsolescence of optical media and media reading equipment.

Files are transferred from CD-ROMs using the following workflow. Actions in the workflow include:

1. Creation of an .iso image of the entire disk using Forensic Toolkit (FTK) software
2. Extraction of necessary files from the .iso image using FTK
3. Migration of problematic file formats, e.g. old versions of .pdf files are converted to the more stable PDF/A version

4. Temporary storage of image and extracted files in a zipped package using the BagIt specification

The plan is to add files extracted (and migrated when applicable) from the CD-ROMs to an online accessible repository. Corresponding catalog records will provide a link so university authenticated users can access the material. This online access service will also include any needed licensing or copyright reviews before files are made accessible.

An important caveat is that the scope of the project is limited to disks containing individual files, such as .pdf and .html, and not executable content or software. Substantial changes to operating systems over the years make it likely that older executables (usually .exe files) are incompatible with computers in use today. However, there is currently no policy governing the long-term preservation and management of executable content.

Implications for the Future

Now that the engineering librarians and the digital preservation librarian have established a workflow for evaluating and preserving CD-ROM data, they will work through the remainder of the disks, with an estimated completion time of six to nine months. Other libraries on campus also have CD-ROMs in their collections; the new workflow will be shared with these libraries, with the hope that they can integrate it into their established collection development procedures.

Looking forward, the engineering liaisons will make collection development decisions based in part on whether materials are available in preservable formats. However, one society whose conference proceedings are collected by the library still actively publishes those proceedings exclusively on CD-ROM. That society is working on developing an online platform for dissemination, and the library will continue to collect and catalog their CD-ROMs until online access is available.

The U-M Library is in the process of implementing a new library-wide policy which will state that all digital content be stored and disseminated via server whenever possible, regardless of the medium in which it is acquired. This policy is intended to address the issue of content residing on unmanaged media for extended periods of time. Acquisitions staff is also adapting the current purchasing system to give liaison librarians the necessary information to determine the preservability of a resource before making a purchase.

Conclusion

The CD-ROM was an important technological advancement in data storage, and perhaps it was natural that publishers would publish handbooks, conference proceedings, and other complete

works on CD-ROM. Once the CD-ROM became an accepted format for publication, it was expected that libraries would adapt their collections policies and procedures to include the CD-ROM format. However, libraries with CD-ROM collections are now dealing with the fact that they have added resources to their collections in a format whose long-term viability is uncertain. Where there is valuable content on these CD-ROMs that should be retained, it is imperative to plan for the time when access to the files in their original format will be difficult or impossible. The project described in this paper has been successful due to a robust collaboration between the subject librarians and the digital preservation librarian. As a result, this team of librarians has taken a concrete step towards preserving the engineering collection for future generations.

Acknowledgements

The authors thank David Carter and Lauren Briggs for their review and comments on the draft version of this paper.

References

[1] J. Lowensohn, "Apple's plan to wipe out disc drives is nearly complete," CNET. [Online]. Available: https://www.cnet.com/news/apples-plan-to-wipe-out-disc-drives-is-nearly-complete [Accessed January 14, 2018].

[2] H. McCracken, "Apple's Schiller: 'Old technologies are holding us back,'" CNN. [Online]. Available: http://www.cnn.com/2012/10/26/tech/gaming-gadgets/apple-mac-schiller/index.html [Accessed January 15, 2018].

[3] S. Grobart, "Daddy, what were compact discs?," New York Times, 2012. [Online]. Available: http://www.nytimes.com/2012/05/31/technology/personaltech/daddy-what-are-turntables-eight-tracks-and-floppy-disks.html [Accessed January 14, 2018].

[4] "Digital Preservation Handbook, 2nd Edition," *Digital Preservation Coalition*, 2015. [Online]. Available: http://handbook.dpconline.org [Accessed January 15, 2018].

[5] P. McCarthy, "CD ROM—An Alternative Information Distribution Medium," in 9th *International Online Information Meeting, London, 3-5 December 1985*, Oxford: Learned Information (Europe) Ltd., 1985. pp. 163-169.

[6] B. Murphy, "CD-ROM and Libraries," *Library Hi Tech*, vol. 3 no. 2, pp. 21-26, 1985.

[7] B.J.S. Williams, "CD-ROM: a New Medium for Publishing," *Information Media & Technology*, vol. 19 no.3, pp.107-109, 1986.

[8] S. Peniston, "Market Potential of CD ROM for Text Applications," *Journal of Information Science*, vol. 15, pp. 21-26, 1989.

[9] T. Hanson and J. Day Ed., *CD-ROM in Libraries: Management Issues*. London: Bowker-Saur, 1994.

[10] J. Cox, "The CD-ROM Market," in *CD-ROM in Libraries: Management Issues*, T. Hanson and J. Day, Ed. London: Bowker-Saur, 1994, p.19.

[11] S. Mantell, "Looking it Up Is Looking Up: Improved Graphics and Increased Use of Electronics Are Making Authoritative Tomes More Appealing and Accessible," *Publishers Weekly*, vol. 244 no. 40, 1997, pp. 51-58.

[12] J. Iraci, "The Cold Storage of CD, DVD, and VHS Tape Media," *Restaurator*, vol. 32, pp.81-109, 2011.

[13] National Archives of Australia, *Digital Preservation, Illuminating the Past, Guiding the Future*, [Online] Available at: http://www.naa.gov.au/Images/XENA_brochure%5B1%5D_tcm16-88754.pdf [Accessed January 15, 2018].

[14] L.S. Lin, C.K. Ramaiah, P.K. Wal, "Problems in the Preservation of Electronic Records," *Library Review*, vol. 52 no. 3, pp. 117-125, 2003.

[15] N. del Pozo, A.S. Long, D. Pearson, "'Land of the Lost': a Discussion of What Can be Preserved through Digital Preservation," *Library Hi Tech*, vol. 28 no. 2, pp. 290-300, 2010.

[16] K. Hogenboom, M. Hayslett, "Pioneers in the Wild West: Managing Data Collections," *portal: Libraries and the Academy*, vol. 17 no. 2, pp. 295-319, 2017.

[17] J. Piascik, K.J. Shrauger, L. Ritzert, "CD-ROM's Acquaintances: What to Do with the Skeletons in Our Academic Libraries," *Journal of Access Services*, vol. 14 no. 3, pp. 114-117, 2017.

[18] National Digital Stewardship Alliance, "Levels of Digital Preservation," National Digital Stewardship Alliance - Digital Library Federation, 2013. [Online]. Available: http://ndsa.org/activities/levels-of-digital-preservation/ [Accessed January 15, 2018].