VOLUME 69 • NUMBER 1 • SPRING/SUMMER 2006

the

AMERICAN ARCHIVIST

 ISSN 0361-9081

"The Old Version Flickers More": Digital Preservation from the User's Perspective

Margaret L. Hedstrom, Christopher A. Lee, Judith S. Olson, Clifford A. Lampe

Abstract

Most criteria for evaluating digital preservation strategies are based on needs and requirements from the archivist's perspective. In the CAMiLEON Project, we wanted to learn what *significant properties* users consider worth preserving. In this article, we present the results of two experiments that used human subjects to learn about user preferences for different formats of preserved digital objects. We tested subjects' reactions to digital materials that were preserved using three common methods: 1) conversion to a "software-independent" format; 2) migration; and 3) presenting the original bitstream using emulation. The results of this exploratory study suggest directions for further research and help archivists understand how user needs and preferences may inform selection of preservation methods. Further research on the effectiveness of emulation and migration needs to account for the quality of the emulator, the impact of specific approaches to migration on document attributes and behaviors, and numerous aspects of the original computing environment that may affect the user experience. Research on the importance of authenticity should consider how users judge authenticity and the tradeoffs they are willing to make between using the original source and the ease of access and use.

For more than forty years, archivists have debated the best strategies and methods for preserving digital information. They hold strong opinions about the feasibility and effectiveness of different digital preservation strategies, whether preserving data in software-independent formats, migration, or emulation. Most evaluations of these strategies have been based on the relative merits of theoretical models, pilot projects, or practical experience in

a small number of archives.[1] Most criteria for evaluating digital preservation strategies rely on requirements from the archival perspective that emphasize ease of accessioning, simplicity of long-term maintenance, and authenticity. Users' needs and preferences are rarely considered when evaluating digital preservation strategies or when choosing which methods to apply. In the CAMiLEON Project, we conducted research that differs in two important respects from previous research on digital preservation strategies.[2] First, our primary interest was to learn about digital preservation from the user's perspective. We wanted to understand which features users consider worth preserving, rather than what archivists believe is important or what theoretical models would predict. Second, we used experimental methods and gathered empirical data from human subjects to evaluate how well different preservation methods retained the original "look and feel" of digital objects.

Researchers at the University of Michigan and at the University of Leeds conducted the CAMiLEON Project jointly. Its broad goals were to test the technical feasibility of using emulation for preserving digital information and to evaluate the effectiveness of digital preservation strategies from the user's perspective. The research team at the University of Leeds was primarily responsible for evaluating the technical and practical aspects of emulation. The University of Michigan researchers, whose results are presented in this article, evaluated users' perceptions and responses to digital objects in their original format and to those same objects preserved using emulation and migration.

Emulation uses computer programs, called emulators, to imitate the functionality of obsolete computer platforms. Emulators make it possible to run obsolete software on current computer platforms and to present users with an exact copy of the original material.[3] Jeff Rothenberg introduced the concept of

[1] Philip Bantin, "The Indiana University Electronic Records Project Revisited," *American Archivist* 62 (Spring 1999): 153–63; Luciana Duranti and Heather MacNeil, "The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project," *Archivaria* 42 (Fall 1996): 46–67; Charles Dollar, *Authentic Electronic Records: Strategies for Long-Term Access* (Chicago, Ill.: Cohasset Associates, Inc., 2000), 47–50; Jeff Rothenberg, *Avoiding Technological Quicksand: Finding a Viable Technological Foundation for Digital Preservation* (Washington, D.C.: Council on Library and Information Resources, 1999), at http://www.clir.org/pubs/reports/rothenberg/contents.html; David Bearman, "Reality and Chimeras in the Preservation of Electronic Records," *D-Lib Magazine* 5 (April 1999), at http://www.dlib.org/dlib/april99/bearman/04bearman.html; Margaret Hedstrom, "Digital Preservation: A Time Bomb for Digital Libraries, "*Computers and the Humanities* 31, no. 3 (1997–98): 189–202, preprint version available at http://www.uky.edu/~kiernan/DL/hedstrom.html.

[2] CAMiLEON was an International Digital Library Initiative Project funded by the National Science Foundation, award number 99-05935 and the Joint Information Systems Committee from 1999 to 2003. CAMiLEON (pronounced like *chameleon*) stands for Creative Archiving at Michigan and Leeds: Emulating the Old on the New. Additional papers and reports from the CAMiLEON project are archived at www.si.umich.edu/CAMILEON/.

[3] There is an extended debate in the literature about what constitutes "original" digital information. For the purposes of our research, "original" means the bitstream as it was created or an exact copy of that bitstream.

emulation as a potential method for long-term preservation of digital materials in 1995.[4] Advocates of emulation contend that emulation would make it possible to retrieve, display, and use digital documents with their original software. This strategy could have the advantage of preserving the "look and feel" and authenticity of digital objects because the original bitstream is never altered and the objects can be used in a software environment identical or similar to the original one.[5] Some digital preservation experts, however, have reservations about emulation, asserting that this approach is complicated, expensive, unreliable, and unnecessary to satisfy users' needs.[6] Migration, a more widely implemented preservation method, involves the periodic transfer of data from an obsolete computing platform, medium, or format to current platforms, media, or formats. Migration almost always changes the original bitstream, and it can have a significant impact on the appearance and performance of preserved digital objects. Creating hardware- and software-independent versions means either creating digital information in a format that all widely available hardware and software can handle or converting older digital information to such a format.

All preservation strategies can alter the presentation, appearance, behavior, and even content of digital objects, whether converting digital information from one format to another, migrating it to current generations of hardware and software, or using emulation. Any of these strategies may affect hundreds of features, attributes, or properties, such as document layout, font styles, color matching, or the speed of input and output devices, in ways that range from extremely obvious to almost imperceptible. We wanted to learn which features users notice and which make a difference in their perceptions of the digital objects, their interactions with them, and the ease or difficulty of using the objects in different formats. We refer to features, attributes, or properties that impinge upon future use and understanding as *significant properties.*[7] Significant properties warrant ongoing preservation due to their demonstrated or

[4] Jeff Rothenberg, "Ensuring the Longevity of Digital Documents," *Scientific American* 272, no. 1 (1995): 24–29.

[5] Jeff Rothenberg and Tora K. Bikson, *Carrying Authentic, Understandable and Usable Digital Records through Time* (RAND Europe, 1999); David Holdsworth and Paul Wheatley, "*Emulation, Preservation, and Abstraction,*" *RLG DigiNews* 5 (15 August 2001); Rothenberg, *Avoiding Technological Quicksand.*

[6] Raymond A. Lorie, "*Long Term Preservation of Digital Information,*" in *Proceedings of the First ACM/IEEE-CS Joint Conference on Digital Libraries,* Roanoke, Va, 24–28 June 2001 (New York: ACM, 2001), 346–52; Kenneth Thibodeau, "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years," in *The State of Digital Preservation: An International Perspective,* Conference Proceedings (Washington, D.C.: Council on Library and Information Resources, July 2002), 4–31; Bearman, "Reality and Chimeras."

[7] Members of the Center of Excellence for Document Analysis and Recognition (CEDAR) project introduced the concept of significant properties. David Holdsworth and Derek M. Sergeant, "A Blueprint for Representation Information in the OAIS Model," paper presented at the 17th IEEE Symposium on Mass Storage Systems, College Park, Maryland, 27–30 March 2000.

predicted contributions to the appearance, interpretation, or usability of digital objects. We first identified thousands of properties that may be affected when copying, converting, or reformatting digital information, and we developed an abstract model to represent these properties.[8] Then we tested our *a priori* assumptions about significant properties by comparing our very large set of significant properties with the properties that human subjects detected in two different experiments.

In this article, we present the results of two experiments, the first evaluating responses to different versions of a computer game, and the second examining responses to different versions of documents. In both experiments, we gathered data from human subjects and observed them performing tasks in order to gain insights into users' responses to digital materials that were preserved using different methods. Following a review of the literature and prior work, we briefly describe the methodology and design for the two experiments. We then present the findings and results of the two experiments. Finally, we draw some general conclusions from both studies and suggest areas for further research.

Literature Review

We drew on literature from five different areas that address some aspect of "look and feel," significant properties, authenticity, and usability of physical and digital objects: 1) past experience with defining significant properties; 2) theoretical approaches; 3) institutional policies; 4) user considerations and requirements; and 5) human-computer interaction and usability studies. Across this diverse literature is a consensus that preserving useful and usable digital objects entails more than simply preserving the content. Context, structure, and appearance may also be important for future reuse. The literature offers some guidance based on prior experience with physical materials, institutional policies, and theoretical models, but little is known about the specific properties of digital objects that users will consider significant. The Human-Computer Interaction (HCI) literature includes rich empirical research on user needs and usability, but most of the research focuses on current users of current materials, not on future users of obsolete systems and noncurrent materials.

Past experience with preserving physical materials offers several criteria to guide decisions about preserving properties of physical artifacts in their original form. Factors such as age, evidential value, aesthetic value, scarcity, associational value, market value, and exhibition value help archivists, librarians, and curators decide when to preserve materials in their original form rather than,

[8] Margaret Hedstrom and Christopher A. Lee, "Significant properties of digital objects: definitions, applications, implications," in *Proceedings of the DLM-Forum 2002, Barcelona, 6–8 May 2002* (Luxembourg: Office for Official Publications of the European Communities, 2002), 218–27.

or in addition to, reformatting them.[9] The archival literature has contributed the concepts of "intrinsic value" and "symbolic value" as factors that may call for preservation of originals.[10] Several studies have investigated the role of specific characteristics of documents, such as color of text, layout, and annotations.[11] Guidelines for microfilm and digital reformatting are also useful when considering which significant properties warrant preservation.[12] Past experience provides useful guidance when identifying the significant properties of physical objects, but there is little research and only limited experience with applying these concepts to born-digital objects.

Archivists have contributed formal models and functional requirements that identify the characteristics of digital objects that may warrant preservation. The models and requirements, typically, have been deduced from the principles of diplomatics and archival science, or generalized from requirements proposed by individuals responsible for the administration, management, and oversight of information or recordkeeping systems.[13] Authenticity has been analyzed

[9] *The Evidence in Hand: The Report of the Task Force on the Artifact in Library Collections* (Washington, D.C.: Council on Library and Information Resources, 2001).

[10] National Archives and Records Service, "Intrinsic Value in Archival Material," Staff Information Paper 21 (Washington, D.C.: NARS, 1982); Angelika Menne-Haritz and Nils Brübach, *The Intrinsic Value of Archive and Library Material* (Marburg, Germany: Archivschule Marburg, 1997); and James M. O'Toole, "The Symbolic Significance of Archives," *American Archivist* 56 (Spring 1993): 234–55.

[11] Catherine C. Marshall, "Annotation: from paper books to the digital library," *Second ACM International Conference on Digital Libraries* (Philadelphia: ACM Press, 1997): 131–40; Else Pettersson, "Automatic Information Processes in Document Reading: A Study of Information Handling in Two Intensive Care Units," paper presented at the 1st European Conference on Computer Supported Cooperative Work, London, U.K., 13–15 September 1989; J. J. Cadiz, Anoop Gupta, and Jonathan Grudin, "Using web annotations for asynchronous collaboration around documents," in *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (New York: ACM Press, 2000), 309–18; and Elaine G. Toms and D. Grant Campbell, "Genre as interface metaphor: exploiting form and function in digital environments," in *Proceedings of the 32nd Hawaii International Conference on System Sciences: HICSS 32* (IEEE, 1999).

[12] Anne R. Kenney and Oya Y. Reiger, eds., *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Mountain View, Calif.: Research Libraries Group, 2000); Robin Dale, et al., "Reformatting and Transfer Re-recording," in *Audio Preservation: A Selective Annotated Bibliography and Brief Summary of Current Practices* (Chicago: American Library Association, 1998), 18–19; Nancy E. Elkington, ed., *Archives Microfilming Manual* (Mountain View, Calif.: Research Libraries Group, 1994); Benchmark Working Group, "Benchmark for Faithful Digital Reproductions of Monographs and Serials" (Digital Library Federation, 2002); and Anne R. Kenney and Louis H. Sharpe II, Illustrated Book Study: Digital Conversion Requirements Printed Illustrations, Report to the Library of Congress—Preservation Directorate Contract #IN97C-22/97CLCCT7021, July 1999, at http://www.loc.gov/preserv/rt/illbk/ibs.htm.

[13] Heather MacNeil, *Trusting Records: Legal, Historical, and Diplomatic Perspectives* (Dordrecht, Germany: Kluwer Academic Publishers, 2000); Anne J. Gilliland-Swetland, *Enduring Paradigms, New Opportunities: The Value of the Archival Perspective in the Digital Environment* (Washington, D.C.: Council on Library and Information Resources, 2000); Peter B. Hirtle, "Archival Authenticity in a Digital Age," in *Authenticity in a Digital Environment* (Washington, D.C.: Council on Library and Information Resources, 2000), 8–23; Wendy Duff, "Ensuring the Preservation of Reliable Evidence: A Research Project Funded by the NHPRC," *Archivaria* 42 (Fall 1996): 28–45; Duranti and MacNeil, "The Protection of the Integrity of Electronic Records"; and *Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (Vancouver: The InterPARES Project, 2002).

extensively from theoretical, legal, and procedural perspectives, but there has been little effort to understand how *users* of digital objects judge their authenticity.[14] Typically, theory-based models have been developed or evaluated by administrators, archivists, records managers, and information technology professionals, with little input from researchers or other potential users of preserved digital objects.

Institutions faced with the need to preserve digital materials have also developed policies and guidelines that attempt to balance the ideal of preserving all of the attributes and functionality of original materials with concerns for ongoing preservation costs, technical feasibility, and institutional capabilities.[15] According to current U.S. federal policy, for example, agencies "may not transfer to the National Archives electronic records that are in a format dependent on specific hardware and/or software."[16] As a consequence, NARA accepts permanent electronic records in a limited, albeit increasing, number of formats. Researchers at Cornell University applied risk management methods to identify properties that might be altered or lost during migration of several different types of files. They found that the original bitstream could be altered by bugs in conversion software, mishandling or failure of storage media, incompatibilities

[14] David Bearman and Jennifer Trant, "Authenticity of Digital Resources: Towards a Statement of Requirements in the Research Process," *D-Lib Magazine* 4, no. 6 (June 1998); Dollar, *Authentic Electronic Records*; Nancy Brodie, "Authenticity, Preservation and Access in Digital Collections," paper presented at Preservation 2000: An International Conference on the Preservation and Long Term Accessibility of Digital Materials, York, U.K., 7–8 December 2000; Heather MacNeil, "Providing Grounds for Trust: Developing Conceptual Requirements for the Long-Term Preservation of Authentic Electronic Records," *Archivaria* 50 (Fall 2000): 52–78; Abby Smith, ed., *Authenticity in a Digital Environment* (Washington, D.C.: Council on Library and Information Resources, 2000); Anne J. Gilliland-Swetland and Philip B. Eppard, "Preserving the Authenticity of Contingent Digital Objects: The InterPARES Project," *D-Lib Magazine* 6 (July–August 2000); Eun G. Park, "Understanding 'Authenticity' in Records and Information Management: Analyzing Practitioner Constructs," *American Archivist* 64 (Fall–Winter 2001): 270–91; "Long-Term Preservation of Authentic Electronic Records; Integrity and Authenticity of Digital Cultural Heritage Objects," *DigiCULT Thematic Issue* 1 (August 2002); Anne J. Gilliland-Swetland, "Testing Our Truths: Delineating the Parameters of the Authentic Archival Electronic Record," *American Archivist* 65 (Fall–Winter 2002): 196–215; Claes Granstrom, et al., *Authenticity of Electronic Records: A Report Prepared for UNESCO* (Paris: International Council on Archives, 2002); Laura Millar, *Authenticity of Electronic Records: A Report Prepared for UNESCO and the International Council on Archives* (Paris: International Council on Archives, 2004). One pioneering work that collected users' perceptions of the authenticity of documents that were reproduced on paper from a digital database is Mark D. Giguere, *Electronic Document Description Standards: A Technical Feasibility Study of Their Use in the Microform Preservation of Contextual Cues Embedded in Structured Electronic Documents During Digital/Analog/Digital Reformatting*, PhD diss. State University of New York at Albany, December 1995.

[15] Mark D. Giguere, "Automating Electronic Records Management in a Transactional Environment: The Philadelphia Story," *Bulletin of the American Society for Information Science* 23 (June–July 1997): 1–19; Philip C. Bantin, "The Indiana University Electronic Records Project Revisited," *American Archivist* 62, no. 1 (1999): 153–63; Timothy A. Slavin, "Ensuring Authentic Electronic Records: From Requirements to Demonstration," *Proceedings of the DLM-Forum on Electronic Records, Brussels, 18–19 October 1999* (Luxembourg: Office for Official Publications of the European Community, 2000), 90–101.

[16] 36 CFR 1228.270(d). The current wording provides for submission only of files that are not dependent on specific hardware and/or software, encoded in "ASCII or EBCDIC with all control characters and other non-data characters removed."

between the original and the migrated formats, and changes in compression, file sizes, media density, and file names. Such changes may introduce errors; affect the resolution, dynamic range, and color spaces of images; change or eliminate linkages to other files, such as metadata files, database directories, scripts, and URLs; or eliminate unique features of the original format that are not supported by the migrated format.[17] Clearly, practical considerations of feasibility and cost come into play when evaluating digital preservation strategies, but with little research and limited experience with preserving born-digital materials, it is difficult to assess how the trade-offs among costs, technical feasibility, and potential loss of information affect future use of digital information.

Archivists and curators recognize the need to consider the contexts of purpose and use when choosing preservation strategies, including who the users will be and what they will need. The Open Archival Information Systems (OAIS) Reference Model introduced the concept of a *designated community* defined as "an identified group of potential users of the archives' contents who should be able to understand a particular set of information."[18] In a survey of research, practice, and common understandings among stakeholders in digital preservation, Abby Smith and Daniel Greenstein observed that a trustworthy digital repository "must be explicit about what digital information it preserves, why, and for whom. It also must be clear about the attributes of the archived information it intends to preserve."[19] David Levy has argued that "determinations of which properties matter are made in the context of purpose and use."[20] This context includes not only the particular class of users, but also the stage of the research process in which they are engaged.[21] According to David Bearman and Jennifer Trant, "[t]o determine which methods are suited for what purposes, it is critical that we better understand the functional requirements for authenticity on the

[17] Gregory W. Lawrence, William R. Kehoe, Oya Y. Rieger, William H. Walters, and Anne R. Kenney, *Risk Management of Digital Information: A File Format Investigation* (Washington, D.C.: Coalition on Library and Information Resources, 2000); Anne R. Kenney, Nancy Y. McGovern, Peter Botticelli, Richard Entlich, Carl Lagoze, and Sandra Payette, "Preservation Risk Management for Web Resources: Virtual Remote Control in Cornell's Project Prism," *D-Lib Magazine* 8, no. 1 (2002).

[18] Consultative Committee for Space Data Systems (CCSDS), "Reference Model for an Open Archival Information System (OAIS)," *Blue Book*, CCSDS 650.0.-B-1 (January 2002), 1–10.

[19] Daniel Greenstein and Abby Smith, "Digital Preservation in the United States: Survey of Current Research, Practice, and Common Understandings," in *New-Model Scholarship: How Will It Survive?* ed. Abby Smith (Washington, D.C.: Council on Library and Information Resources, 2003), 42, at http://www.clir.org/pubs/abstract/pub114abst.html.

[20] David M. Levy, "Where's Waldo? Reflections on Copies and Authenticity in a Digital Environment," in *Authenticity in a Digital Environment* (Washington, D.C.: Council on Library and Information Resources, 2000), 24–31.

[21] For one analysis of how historians integrate digital resources into the research process, see Wendy M. Duff, Barbara Craig, and Joan Cherry, "Historians' Use of Archival Sources: Promises and Pitfalls of the Digital Age," *Public Historian* 26, no. 2 (2004): 7–22.

part of creators and potential users of digital resources, and appreciate where in the research process these requirements come into play."[22] These comments suggest that the context of use is an important consideration in choosing digital preservation strategies.

We used several important concepts from the literature on Human-Computer Interaction (HCI) related to significant properties, perception, and usability. The concept of "affordances" from the HCI literature was especially useful. Broadly defined, affordances refer to the "the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used."[23] Properties of digital objects can provide affordances for viewing, navigating, reusing, and analyzing the contents of the objects.[24] When encountering a preserved digital object, users may find the original affordances of the object important for making sense of its contents as well as for understanding the functional context in which it was created, used, and acted upon. Concepts and methods from usability design and evaluation helped us design the experiments and make use of the "Thinking-Out-Loud" (TOL) method.

CAMiLEON Project Experiments

We decided to conduct two laboratory experiments with human subjects during the CAMiLEON Project so that we could learn about user preferences for different formats of preserved digital objects. We designed the experiments so that we could observe human subjects as they interacted with digital materials and gather comments from them about the appearance and behavior of digital materials that were preserved using different preservation methods. The first experiment tested subjects' reactions to different versions of an obsolete computer game called "Chuckie Egg." The second experiment evaluated subjects' responses to different versions of speech files and office documents from a former president of the University of Michigan. In both experiments, subjects performed the types of tasks that might be expected of people using preserved digital objects for entertainment (Chuckie Egg) or research (speech outlines).

We used the Thinking-Out-Loud (TOL) protocol where subjects talked about their experiences and reactions as they performed the tasks. The research

[22] Bearman and Trant, "Authenticity and Digital Resources."

[23] Donald A. Norman, *The Design of Everyday Things* (New York: Doubleday, 1990), 9. For the original formulation of the concept of affordances, see James J. Gibson, *The Ecological Approach to Visual Perception* (Boston: Houghton Mifflin, 1979).

[24] Thomas A. Phelps and Robert Wilensky, "Multivalent Documents: Inducing Structure and Behaviors in Online Digital Documents," in *Proceedings of the 29th Hawaii International Conference on System Sciences (HICSS-29), Maui, Hawaii 1996* (IEEE Computer Society, 1996), 144–52.

assistants explained each task to the subject and then asked him or her to "think out loud," which meant mentioning any first impressions, further thoughts, or expectations concerning any aspects of the documents. The research assistants recorded each session, took detailed notes, and transcribed and coded the subjects' commentary to identify patterns and trends in the subjects' responses. We also used written instruments to collect data on the differences that the subjects observed when asked to compare original digital materials with other versions that retained only a portion of the significant properties of the original. The two experiments used different types of materials and asked somewhat different questions. Therefore, we discuss each experiment separately and then conclude with some general observations from the two experiments.

Experiment One: The Computer Game "Chuckie Egg"

The first experiment involved Chuckie Egg, which was a popular computer game in the United Kingdom in the mid-1980s, similar to other early maze games such as PacMan.[25] (See Figure 1.) Chuckie Egg was designed to run on the BBC Microcomputer, a computing platform that became orphaned in the 1990s.[26] We decided to run an experiment using a computer game for several reasons. Although computer games may seem like a frivolous example of a new documentary form that hardly merits attention, from the perspective of popular culture, games are difficult to ignore. Computer games are the most common type of software sold for home computers. Young people, particularly, share this cultural icon, but market studies show that more than one-third of computer game consumers are over age eighteen.[27] Computer games have also been the subject of serious research on the formation of gender identity in children and on tendencies toward violence and alienation among frequent users.[28] Moreover, they provide important evidence for understanding the evolution of software, both technologically and aesthetically. Applications such as flight simulators, virtual war games, and interactive educational software such

[25] Margaret Hedstrom and Clifford Lampe. "Emulation vs. Migration: Do Users Care?" *RLG DigiNews* 5 (December 2001).

[26] By "orphaned" we mean that the original developer of the hardware, Acorn Computer, and the developer of most of the software, the BBC, ceased support and further development of the platform. The BBC Domesday Disk, which was recovered by the research team at the University of Leeds, also ran on the BBC Micro.

[27] Catherine Beavis, "Computer Games, Culture, and Curriculum," in *Page to Screen: Taking Literacy into the Electronic Era,* ed. Ilana Snyder (New York: Routledge, 1998), 234–55; and Geoffrey Rockwell, "Games Galore: Literary Theory and Computer Games," *Computers and the Humanities* 36 (2002): 345–58.

[28] The classic work on this issue is Sherry Turkle, *The Second Self: Computers and the Human Spirit* (New York: Simon and Schuster, 1984).

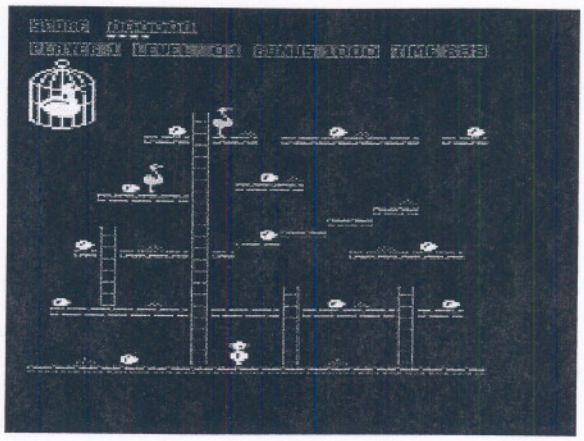

FIGURE 1. Chuckie Egg.

as 3-D visualizations of the human body draw heavily on software design principles developed originally for computer games. Computer games are also good examples of digital objects that are highly dependent on particular hardware and software with many potentially significant properties, including display size, screen resolution, aspect ratio, color graphics, and input/output speed variation.

We chose the Chuckie Egg game for both pragmatic and conceptual reasons. Three different versions of the game were available: the original game that ran on the BBC Micro, a disk image of the original game that could be played on a Windows platform using an emulator of the BBC Micro, and a version that was migrated to run directly on a Windows platform. We were able to use versions of the game that already existed rather than having to build an emulator or write a migrated version. Like many computer games, Chuckie Egg had a built-in scoring mechanism and a means for players to move to more advanced levels as their performance improved. The scoring mechanism offered one way to measure subjects' performance and to determine how quickly they became proficient enough to move to levels one and two when playing different versions of the game.

We started the experiment with two assumptions based on assertions in the literature about user behavior and emulation. First, archivists and others contend that under some circumstances users prefer to work with original documents rather than copies or surrogates.[29] We began, therefore, with the hypothesis that users would prefer to play the original game on the BBC Micro

[29] National Archives and Records Service, "Intrinsic Value in Archival Material"; Menne-Haritz and Nils Brübach, *Intrinsic Value of Archive and Library Material;* O'Toole, "Symbolic Significance of Archives"; and Nicholson Baker, *Doublefold: Libraries and the Assault on Paper* (New York: Random House, 2001).

rather than an emulated or migrated version. Second, advocates of emulation claim that emulation preserves the original "look and feel" of digital objects. We expected that subjects who played the emulated version of the game would find it more like the original than those who played the migrated version.

Experiment One: The Experimental Design

We acquired and set up a working BBC Micro in a usability laboratory on the University of Michigan campus. We recruited thirty subjects by posting flyers that advertised the opportunity to participate in the experiment and offered the subjects compensation for their participation at several gathering sites on the University of Michigan campus. Most of the subjects were graduate or undergraduate students in the age range of eighteen to forty-four years with a mean age of 24.4 years. About two-thirds of the subjects were American citizens; one-quarter were from Asia, and the remainder were from Europe or the Middle East. Most significantly, no subjects had used a BBC Micro or played the Chuckie Egg game prior to the experiment. Therefore, the subject pool is similar to future users who will not have had experience using original computer platforms or applications that have become obsolete.

The research assistants trained each subject to play the original Chuckie Egg game on the obsolete BBC Micro platform. After completing a one-hour training session, the subjects switched to a modern PC platform. One group played the version of the game that emulated the BBC Micro and the other group played the migrated version. We recorded all interactions and collected measures of subjects' performance while playing the game. We used the TOL protocol for one round of the game and asked subjects to think about differences that they noticed between the training condition (the original game) and their test condition. Each subject also filled out a questionnaire asking about differences between playing the original game on the BBC Micro and playing the game on the PC platform with either the migrated or the emulated version.

Experiment One: Findings

We measured three aspects of user experience: 1) satisfaction (was the game interesting and fun to play?); 2) perceived ease of use (how easy was it to play the game?); and 3) performance (how quickly could users accumulate enough points to advance to level one and level two of the game?). People who play computer games find them entertaining because players are challenged to improve their performance with practice. In our experiment, satisfaction fell almost in the middle of the scale (from very fun and interesting to not fun or interesting at all), regardless of whether subjects played the

migrated or emulated version. Subjects considered all versions of the game easy to play. On the performance measure, there was no statistically significant difference in the total amount of time it took subjects to reach level one or level two between the migrated or the emulated versions of the game. For this game, there is no statistically significant evidence that the emulated version preserved the original "look and feel" and behavior of the original game better than the migrated version did.

The subjects' comments during the TOL session further elucidated their reactions and experiences. Some subjects found playing the game on the original BBC Micro boring and frustrating because of the obsolete keyboard and display, slow pace, and unfamiliar patterns of interaction. Subjects in the emulation group mentioned that compared to the original BBC Micro version of the game, "the keys are nicer, a softer touch," "it looks like the screen is scrunched," "the character can jump farther," and "it might be a little softer, or it could be the hardware." Similarly, subjects who played the migrated version remarked that "the keyboard makes it easier to control Chuckie," "the old version flickers more," "the blue birds are smarter," and "sounds are a little sloppier." Subjects reported many differences between playing the game on the original BBC Micro and playing either the emulated and migrated versions, as indicated in Figure 2.

Many of the differences that subjects observed about the different versions of the game concerned the hardware environment, especially differences

Emulated Version (N=15)	Migrated Version (N=15)
Screen/Display: • "The screen is more squished together." • "Screen seems vertically compressed." • "Looks like the screen is scrunched." • "I liked the other screen better, it reminded me of the last time I played a game like this."	**Screen/Display:** • "Better looking because the pixels aren't so big." • "The old version flickers more."
Keyboard: • "The keys are nicer, a softer touch." • "Seems easier, but that could be related to the keyboard."	**Keyboard:** • "Keyboard makes it easier to control Chuckie." • "Something different about how it responds, I think it's the keyboard."
Interaction Patterns and Speed: • "Character flashes, makes it easier to recognize" • "The character can jump farther." • "All elements move faster, can get through the levels quicker."	**Interaction Patterns and Speed:** • "Blue birds are smarter." • "Seems faster." • "The (increased) speed makes it a little harder to time your jumps." • "Can jump further, and catch stuff above you, which you couldn't do before."
Sound:	**Sound:** • "Sounds are a little sloppier." • "Sounds are more annoying."
Other: • "There's a new bar," (artifact from the emulation software.) • "Nicer than the first one." • "Might be a little softer, or it could be the hardware."	**Other:** • "Different from the Commodore 64 feeling you get with the first one." • "You can see Chuckie's legs, which makes it easier to time stuff."

FIGURE 2. Differences between Chuckie Egg and Migrated/Emulated Versions.

between the BBC Micro keyboard and the modern PC keyboard, sound, screen display, and interaction speed.

Experiment One: Conclusions

We drew three main conclusions from the first experiment. First, the subjects noticed very minor differences between the original and the emulated and migrated versions of Chuckie Egg. Many of the elements of "look and feel" that subjects considered worth mentioning were not anticipated in our abstract model of significant properties. Second, when subjects compared the original with either the emulated or the migrated version of the game, we found no evidence that emulation is better for preserving original "look and feel" than migration. This finding may be an artifact of the specific software that we tested, because the migrated version was developed with great care to make the game perform as much like the original as possible, but it merits further investigation with different programs and platforms. Finally, we discovered that subjects preferred playing the migrated and emulated versions rather than the original game on the BBC Micro. Although a few subjects lamented the loss of the original game "feeling," most valued the greater ease of manipulation and faster speed of the migrated and emulated versions. The obsolete keyboard and screen were important factors in experiencing the original "look and feel" of the Chuckie Egg game, but almost all of the subjects were willing to forego those aspects of the original experience because they also made the game slow and cumbersome. Further research on the effectiveness of emulation and migration needs to account for the quality of the emulator, the impact of specific approaches to migration on document attributes and behaviors, and numerous aspects of the original computing environment that may affect the user experience.

Experiment Two: Speech Outlines and Office Documents

One goal of the second experiment was to gain a deeper understanding of the properties of textual documents that subjects consider significant when using, interpreting, and judging the authenticity of digital documents. In the second experiment, subjects interacted with three different versions of speech files and office documents. We also wanted to learn more about the differences that subjects might notice between different formats of the same document and which format they would prefer to use for research. We used the results and findings of the Chuckie Egg experiment to inform the design of our second experiment. For this experiment, all of the versions were presented on the same PC platform, thus controlling for any differences in the hardware configuration. We were also attentive to gathering data on minor differences in the various versions of documents presented to subjects.

Experiment Two: The Experimental Design

This experiment involved three tests, which are discussed separately below. In each test, we gathered data on subjects' reactions to outlines of speeches and notes for public addresses created between 1990 and 1993 by James J. Duderstadt, president of the University of Michigan from 1988 to 1996. Subjects viewed some combination of five speech outlines in the following formats: the original MORE format (MORE) (Fig. 3), a version migrated to Microsoft Word Version 6.0 (Word) (Fig. 4), and an ASCII text version (Text) (Fig. 5). MORE was one of a few early outlining programs available in the early 1990s.[30] An outliner organizes content as a set of topics, subtopics, sub-subtopics, and so on. MORE had the capability of displaying the topics in an outline view, in a bulleted list, or in a tree chart. In the outline view, MORE allowed users to expand or collapse the list of topics, thus hiding or revealing subtopics and sub-subtopics. Today, most word processing and presentation software packages

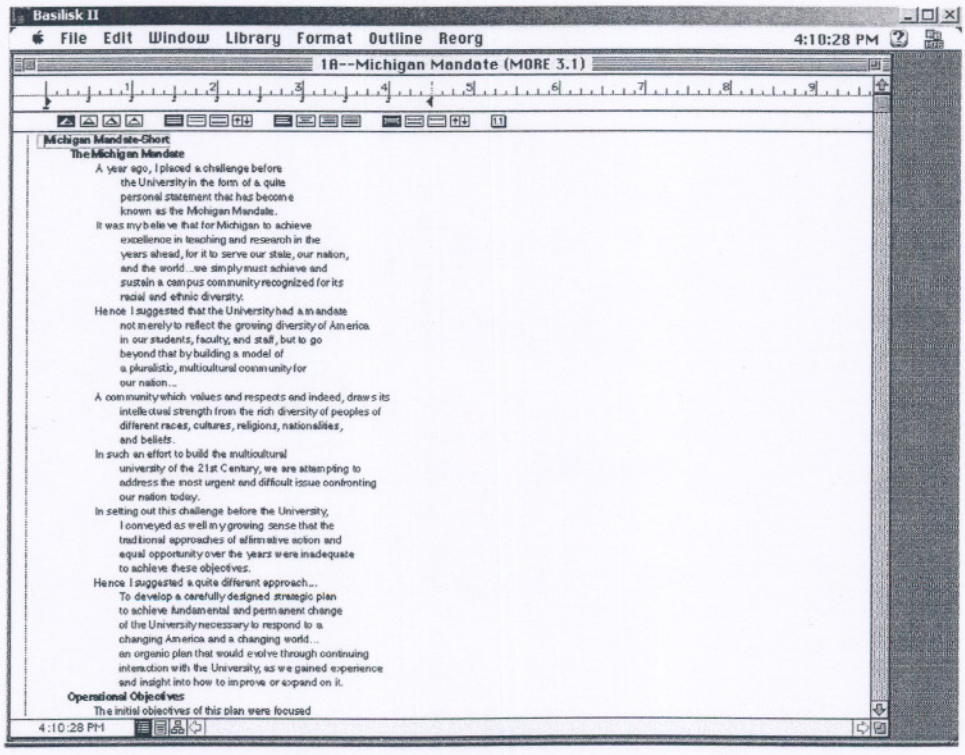

FIGURE 3. Michigan Mandate Speech in MORE.

[30] MORE originated as ThinkTank for the Apple II in 1984. In 1987, Symantec Corporation purchased MORE and supported it until 1997.

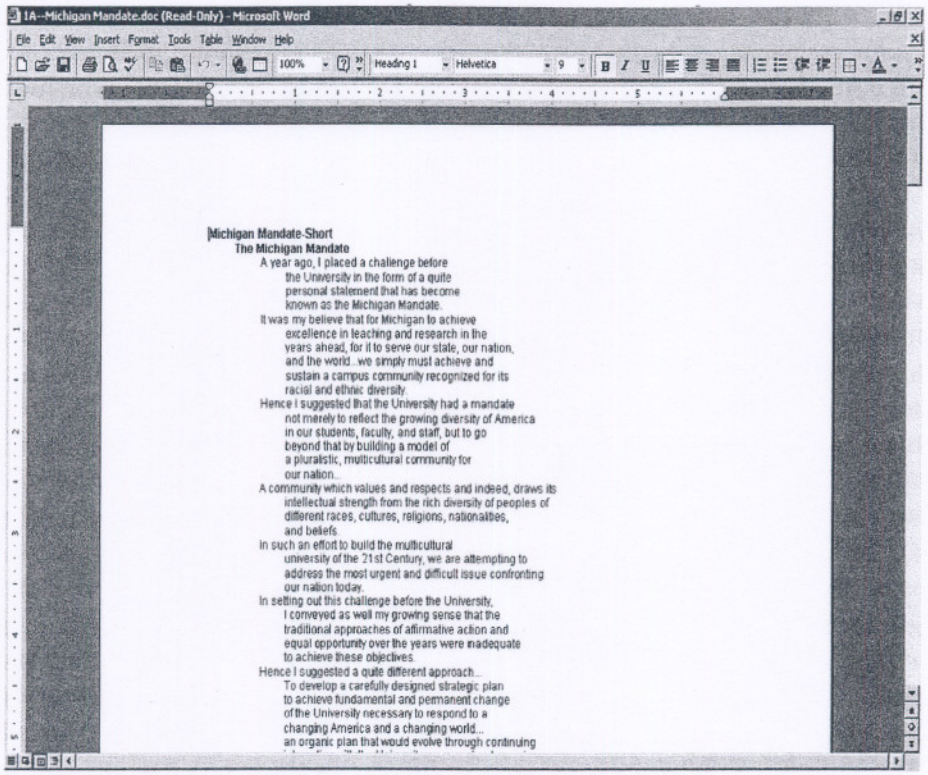

FIGURE 4. Michigan Mandate Speech in Word.

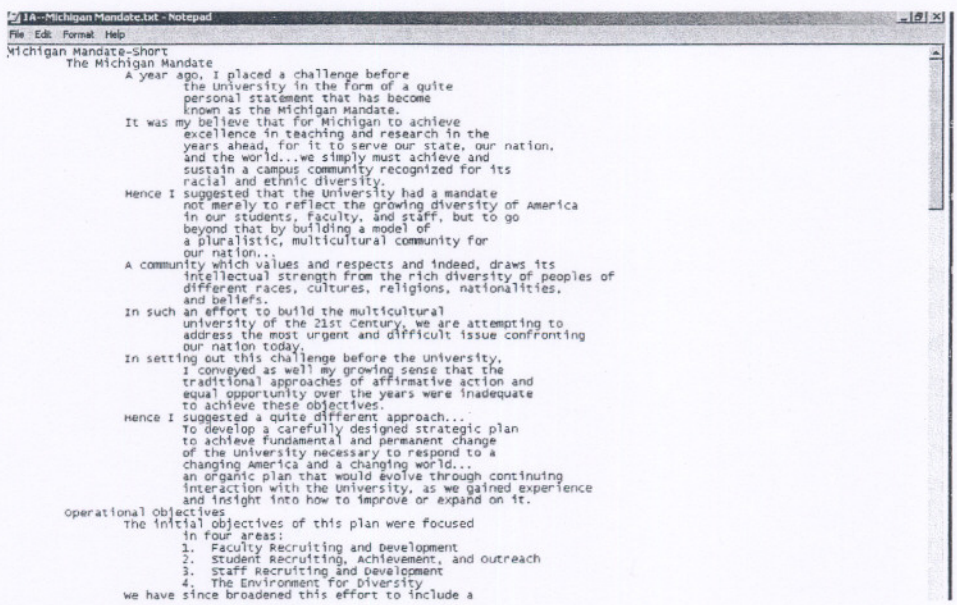

FIGURE 5. Michigan Mandate Speech in Text.

provide this functionality, but this capability was quite advanced and uncommon in software packages during the 1980s and early 1990s. As shown in Figures 3, 4, and 5, the Word versions resemble the original MORE documents much more closely than do the Text versions.

Although the distributor of MORE stopped supporting the application in 1997, it still has some enthusiastic users. We were able to purchase an emulator, called Basilisk II, which allowed us to present the original MORE documents to the subjects running under emulation on a modern PC platform with the Windows operating system. The MORE files were the actual files that President Duderstadt created on an Apple Macintosh computer, but we did not tell the subjects which format was the original until the end of the experiment. We also intentionally withheld from the subjects contextual information that was available in the finding aid, such as biographical information, information about the office of the president, and information about the processing of the collection, in order to learn how they would reason about the originality of the documents based only on what they could ascertain by interacting with them on the screen.

We selected outlines and notes for five speeches from a total of 2,144 digital documents in the Duderstadt Papers. All of the documents referred in some way to "The Michigan Mandate," a program launched by President Duderstadt to increase diversity at the university.[31] We used the speech outlines from the Duderstadt Papers for several reasons. First, these documents are typical of the types of digital documents that archival institutions preserve for future research. Second, the original MORE files and a version that was migrated to Microsoft Word were available to us.[32] We created flat ASCII text versions of the documents from the MORE files for use in the experiment to test subjects' reactions to "software-independent" or flat text files. Third, the Bentley Library staff decided to save the electronic versions for the purpose of reflecting "some of the ways Duderstadt's office functioned, particularly in the evolving years of electronic recordkeeping." Finally, we were able to interview President Duderstadt about his use of computing equipment generally and the MORE software specifically.[33]

We conducted the user tests in two usability labs on the University of Michigan campus. We recruited thirty subjects by posting flyers in locations on campus frequented by faculty and students who do historical research, including the Bentley Historical Library and the Gerald R. Ford Presidential Library. Our goal was to identify subjects who might do research and write papers using pri-

[31] The collection included 16 linear feet of paper records and 2,144 digital files created in Microsoft Word, PowerPoint, and MORE. Nancy M. Deromedi and Kathy L. Steiner, "Finding Aid for James J. Duderstadt Papers, 1963–1997" (Ann Arbor, Mich.: Bentley Historical Library, 1998).

[32] The Word files were created by Bentley Historical Library staff in 1997 and 1998, while they were processing Duderstadt Papers. They decided to convert the MORE files to Microsoft Word 6.0, while also retaining the original Macintosh computing platform, MORE software, and MORE files.

[33] James J. Duderstadt, interview by Clifford Lampe, 5 July 2001.

mary source materials in archives. This subject pool consisted mostly of undergraduate and graduate students in a variety of fields including public policy, law, education, psychology, business, history, economics, and computer science. There was one independent researcher and one librarian/archivist in the subject pool. The age of subjects ranged from twenty years to fifty-four years, with an average age of 24.6 years. The subjects had an average of more than five years of computer experience and used computers an average of 25.6 hours per week. All of the subjects had used Microsoft Word, two-thirds had used text editors, but no subject had used MORE. All but two subjects had done research using primary source materials or looked for primary sources on-line.

Each subject completed three tests in sequence and then responded to questions about usability of the different formats. Each experimental session lasted from forty-five minutes to an hour and ten minutes, averaging about fifty-five minutes per subject. We used written instruments and the TOL protocol to gather data on the subjects' responses to the speeches in different formats and to record the differences that they noticed. Each session was taped and the data were transcribed and coded for analysis. In all three tests, subjects viewed the documents on a computer screen, and they were able to scroll through the documents, move from one document to the other, and use the other functionality that the text editor, Word, and MORE provided. When evaluating the MORE documents, the subjects were able to expand and collapse the outline view and to view the documents in outline, bulleted list, and chart form. The subjects were not permitted to edit or change the documents.

Experiment Two: Tests and Findings

Test 1: Comparison of three speeches in one format

The purpose of the first test was to determine how subjects would decide whether they were viewing the original or a document that had been altered in some way when shown speech outlines in only one of the three formats. Ten subjects viewed three speech outlines only in MORE; ten saw the same three outlines only in Word; and ten saw them only in Text. Subjects were told that these were "documents that archivists believe were written by President Duderstadt. They are not sure if these files are the ones he actually created using his computer, or if they might have been copied or changed in some way between the time they were written and now." Subjects were then instructed to look over each document, think out loud, and provide written responses to two questions regarding the documents they had just viewed:

1. How likely do you think it is that the document you're looking at is the one that President Duderstadt created?
2. What led you to this conclusion?

When the subjects examined three speeches all in the same format, they reasoned carefully and considered many factors, given the limited amount of information available to them. However, when asked to speculate about how likely it was that they were viewing the original, on average, the subjects were uncertain about the likelihood that they were viewing the original documents. (See Figure 6.)

There was very little variation from format to format in how the subjects ranked the likelihood that they were looking at the originals. In fact, the mean rankings for MORE and Word were identical (3.63), and Text was very similar (3.77). Moreover, the means for all three formats fall very close to the midpoint (4) of the 7-point scale, indicating that the subjects, on average, could not tell how likely it was that they were looking at the originals. Of 27 observations when subjects thought that it was very likely or quite likely that they were viewing the original documents, only 7 were actually viewing the original MORE files. In this test, when subjects saw only one format, the format by itself did not provide subjects with enough information for them to ascertain accurately or with certainty which documents were the originals.

We learned from the written responses and the TOL session which factors the subjects considered when trying to determine whether they were viewing the originals. All of the subjects mentioned some aspect of the contents of the speeches. For example, subjects speculated about whether or not a university president would write such a speech, how an outline with abbreviations and subheadings would be used as the basis for a speech, and whether its main themes were consistent with what a university president might say. Many subjects made observations about the writing style, but they often reached different conclusions. Some subjects indicated that they thought the documents could have been written by President Duderstadt because they were concise and straight to the point, well

How likely do you think the document you are looking at is the one that President Duderstadt created?

Format:	Mean Ranking (On a 7-point scale from 1 = Very Likely to 7= Not Likely At All)
MORE	3.63
Word	3.63
Text	3.77

FIGURE 6. Comparison of three speeches in one format.

organized, "thorough—not choppy," "written in an intelligent voice," or seemed "professional." Others, however, questioned whether President Duderstadt wrote the documents because they found the outlines too informal, not well organized, extremely casual, too emotional, or too haphazard for a presidential speech.

In speculating about the originality of the documents based on their content or style, most subjects made assumptions about what an author would write in his role as a university president. A few subjects who possessed additional knowledge about President Duderstadt also used this in their speculations.[34] Subjects who were aware of Duderstadt's academic background as a nuclear engineer took this into account. One subject, looking at one of the speeches in Word format, surmised that it was "more orderly, more focused, has a purpose. Duderstadt was an engineer; he'd likely have those traits and want his work to reflect that." Another subject concluded that it was unlikely that President Duderstadt wrote one of the speeches himself because of the "more rousing and emotional nature" of the speech. "I'm stereotyping and assuming a nuclear scientist would not write a speech like this."

Most subjects also commented on formatting features and the general appearance of the documents. Those who saw the speeches in Text noticed unusual characters that did not convert accurately or odd spacing, or they noted that the flat text file would not be very useful for reading a speech. Subjects also speculated about the computing environment at the time the documents were created, noting, for example, that odd characters suggested that the documents were converted from a different format, that a university president would have used more sophisticated software than a simple text processing package; or, conversely and erroneously, that computers were so primitive at the time that the sophisticated functionality of MORE or Word would not have been available.

Test 2: Comparison of one speech in three different formats

In Test 2, we wanted to learn how subjects would identify the original if they were able to compare three different formats of the same speech. This time, we showed subjects the same speech in MORE, Word, and Text. Subjects were told, "Archivists going through some more of President Duderstadt's computer files found several documents in three different formats. They're not sure which of them he created himself." We asked them to rank the three formats based on "how closely you believe it resembles the actual file that Duderstadt created." We then asked the subjects to indicate how confident they were in each of the rankings that they assigned on a 7-point scale (from "1—very confident" to "7—not confident at all"). When subjects were able to compare the same speech in three

[34] We did not provide the subjects with any details about President Duderstadt other than the years of his presidency, but a few subjects were aware of his academic background.

different formats, more subjects identified the original correctly. Fourteen of the 30 subjects correctly identified MORE as the original, and 9 ranked all three formats correctly in terms of the degree to which each resembles the original. (See Figure 7.)

When we asked the subjects to indicate their confidence that they had ranked each format correctly, their confidence that they had ranked each format correctly increased. The mean confidence ranking was 4.86 on a 7-point scale, and almost two-thirds of the subjects (64.4 %) were somewhat confident to very confident with their rankings. They also took more factors into account, mentioning the appearance, suitability of the format for the purpose of writing or delivering a speech, and assumptions about the capabilities of computers at the time the documents were created. Frequently, subjects noted differences in the appearance of the three formats, especially in how the speeches were structured and laid out on the screen. Many subjects remarked on different affordances of the three formats. For example, 10 of the 14 subjects, who ranked MORE as most closely resembling the original, thought that it seemed like the best program for outlining speeches and organizing notes because the collapsing and expanding feature made it easy to read and navigate. Conversely, many subjects noted that the Text files would be difficult to use when delivering a speech because of inconsistent indentation and lines of text that exceeded the width of the computer screen.

Finally, subjects speculated about the computing environment at the time the documents were created and whether or not software applications with

For each format, how closely do you believe it resembles the actual file that Duderstadt created?

30 subjects – 3 formats = 90 observations

Ranking	1 (most resembles)	2 (somewhat resembles)	3 (least resembles)
	14	5	11
Word	10	**13**	7
Text	4	12	**14**

[Note: cells in bold are correct answers. A correct ranking of how closely each format resembles the original would be MORE (1 – most resembles), Word (2 – somewhat resembles), and Text (3 - least resembles)]

FIGURE 7. Comparison of three different formats.

the features of MORE and Word were available in the early 1990s. Subjects who believed that Word most resembled the original mentioned that it was a common software package. All 4 subjects who thought that Text was most like the original version reached this conclusion based on erroneous assumptions about the computing environment at the time the documents were created. They did not think that programs with the features of MORE or Word were available at the time that documents were created.

Test 3: Comparison of the original MORE documents with Word or Text formats

In the third test, we wanted to learn about the differences that subjects noticed between the original MORE files and either Word or Text. Subjects viewed a speech that they had not seen before and compared the original MORE version with the same speech in either Word or Text. Twelve subjects compared MORE and Text; and 18 subjects compared MORE and Word.[35] We first asked subjects to list all of the differences that they noticed between the two formats (MORE and Text or MORE and Word). Then we asked them to tell us what differences were most important in determining which version was the original.

The subjects who compared MORE and Text noticed a larger number of differences than those who compared MORE and Word, which is not surprising because the MORE and Text files have more differences and the differences are more obvious. (See Figure 8.)

They commented most often on conversion errors in the Text version, differences in the layout and alignment of sentences and paragraphs, the different fonts in MORE and Text, and the enhanced legibility of the MORE version. When asked to indicate which differences were most important for determining which version was the original, the subjects listed many different possibilities and sometimes reached contradictory conclusions. The subjects most often mentioned conversion errors in the Text documents, and greater legibility and "professional look" of the MORE files. But the "primitive look" of Text files also convinced some subjects that the Text files most resembled the originals. Two subjects who incorrectly identified the Text version as the original indicated that the "lack of new looking features" and the "primitive versus modern look" of the Text document were important in their assessments. Another subject, after noting the odd characters and font of the Text document, said, "If I were looking at a doc from [the] 1980s [*sic*], I would think the typos and more raggedy courier new [*sic*] font was more 'real.'"

[35] We did not do a comparison of Word and Text because our research focused on the difference between the original and the two variants. We asked more subjects to compare MORE and Word because the differences between the two are less obvious.

MORE	Text
Character Conversion Errors	**Character Conversion Errors** • Odd characters - O[accent] in Text. • Seems like it has been converted from another word processing program to .txt and it converted very poorly.
Formatting Features • It does not have sentences running across the page so you don't have to scroll all the way over to the right like in text. • At certain points it has marks by paragraphs. • Adequate, almost pleasant spacing, nice use of bolds in collaboration with the indents. The sections had nice structural flow to them. Easy to identify paragraphs. • Looks great. It appears to be approaching the "draft" stage in the writing process as it is very nicely formatted	**Formatting Features** • The alignment is a huge mess. • The indentation of paragraphs is all over the place
Font Style • **MORE document is in Palatino (which I miss)** • Nice font—pretty, easy to read.	**Font Style** • .txt is in some god-awful Microsoft font. • Primitive-looking font, hard to read.
Legibility • The Mac format has bolded subheadings. It's nice and highly legible, easy to organize.	**Legibility** • The Notepad is highly illegible. It seems to have been written in another program then copy pasted or saved in ASCII. . • Large spacing and indents makes it hard to read.
Other • Has a professional feel to the document. • Looks up-to-date and professional.	**Other** • It is very unstructured and some of the letters make no sense. • Looks like a technological dinosaur or an email.

FIGURE 8. Examples of differences noticed between MORE and Text.

The differences between the MORE and Word versions were less obvious than the differences between MORE and Text, and subjects who compared these two formats listed fewer differences. In fact, several subjects thought that the MORE and Word versions were almost identical, listed no differences, or had to be prompted to look closely. Subjects noticed formatting features and the larger font in Word, and speculated about the obscurity of MORE and the Basilisk emulator. (See Figure 9.)

When asked which differences were most important for deciding which version was the original, subjects focused on different features than those who compared MORE and Text, and they used more complex reasoning. Legibility was one issue that the subjects mentioned frequently, especially the ease of reading a speech

MORE	Word
Formatting Features • Formatting seems more important in MORE. • Paragraphs do not line up . . . possibly expanded from an outline. • The tabs in More are actually real numbers (1/4", 1/2", 3/4").	**Formatting Features** • Maybe formatting is more by default in Word, whereas it's done more "by hand" in MORE • The words appear really spaced out (as in not condensed). Kind of confusing later on with short key points because the space between words almost is the same as space between letters. • The tabs in Word are at odd locations.
Font Style/Size • Palatino [font] looks better on Mac	**Font Style/Size** • Size (font) is larger in Word. • • The Word document has larger font (at least it appears larger) and is also broken up into separate pages, as thought [*sic*] this would be printed out and read as a speech.
Software • Software required for MORE seems odd for conference intro. • • Maybe MORE is more specific to people who must often give speeches in their professional lives.	**Software** • Word is ubiquitous. • Word is for the general public.
Other • There's not much difference. There is a Tic mark at the beginning of the headers in the Mac document.	**Other** • The documents look exactly the same, word for word. It's prettier with a nicer font in Word. • Word has blank page at end . . . does not seem to be deliberate: possibly an effect of cutting & pasting from another application.

FIGURE 9. Examples of differences noticed between MORE and Word.

outline, but they disagreed about whether MORE or Word was more legible. Several subjects also suspected that Word was not the original because the speech ended with a blank page, which did not seem intentional to them.

Test 4: User preferences and authenticity

In the final part of this experiment, we wanted to learn how important it was to the subjects to work with the documents in their original format and which factors would influence their preference for one format versus another. Without revealing which format was the original, we asked subjects to indicate which format they would prefer to use if they were writing a paper on Duderstadt's presidency. (See Figure 10.)

Only 1 of the 12 subjects who compared MORE and Text preferred the Text version, while 1 subject had no preference. The one subject who preferred Text stated, "I believe it's original and would give me more insight into what he intended to say rather than how a transcriber interpreted it." Most of

User preferences when choosing between	MORE & Text	MORE & Word	Total
Text	1	0	1
MORE	10	8	18
Word	0	8	8
No Preference	1	2	3
	N=12	N=18	N= 30

FIGURE 10. User format preferences before knowing which format was the original.

the subjects preferred the MORE version because they thought that MORE was easier to read and navigate and more accurate, but 3 subjects also mentioned that they thought that MORE was the original. The 18 subjects who compared MORE and Word were evenly divided in their preferences, and 2 subjects had no preference. Subjects who preferred the Word format to MORE mentioned better legibility and their own familiarity with Word, but subjects who preferred MORE also mentioned legibility and ease of navigation. Only 1 subject stated a preference for MORE rather than Word explicitly because she thought it was the original, while 1 subject said he would prefer the Word versions only "if they were a) original or b) formatting did not matter to my research."

We then explained that MORE was the application in which Duderstadt wrote the documents and that Basilisk is an emulator we were using to imitate his original computing platform. We also told the subjects that the Bentley Library staff had converted the documents to Word and that we had produced the Text files. We then asked subjects, knowing all of this, "would you change your choice of formats that you would rather work with?" (See Figure 11.)

Only 4 subjects indicated a change of preference after being told about the provenance of the three different formats. Surprisingly, 2 of them shifted to Word from an initial preference for MORE. The only subject who initially preferred Text shifted his preference to MORE once he learned that MORE was the original, stating "even if [the Word version] is 99.5% the same, it could be changed enough—Word automatically corrects some typos, etc." The fourth subject who changed preferences shifted to MORE from Word after recognizing that the original was composed in MORE.

User Preferences	Before	After	Change of Preference
MORE	18	18	MORE (-2) to Word (+2)
Word	8	9	Word (-1) to MORE (+1)
Text	1	0	Text (-1) to MORE (+1)
No Preference	3	3	
Total	30	30	

FIGURE 11. Users' preferences before and after knowing which format was the original.

We concluded from this part of the experiment that only a small minority of subjects considered authenticity and working with the original source as a paramount concern. Usability was the most important factor for the vast majority of subjects. This was especially true of the subjects who compared MORE and Text where the Text version was very difficult to read and navigate on a screen. The 5 subjects who explicitly mentioned authenticity argued along the lines that archivists often use in pointing out the importance of preserving original records with all of their significant properties intact. As one subject indicated:

> Basilisk [MORE] is more like his own hand-written notes, whereas MS Word is more like a finished, very complete document in which you only see the end result, not the thought process he went through in writing it. This thought process would be very important for a researcher using primary sources to write a detailed account of his UM presidency.

Although only 5 subjects explicitly mentioned issues of authenticity and integrity in their choice of format, we do not conclude that the subjects did not care about the authenticity or integrity of the documents they would use for research. Rather, they often considered contextual factors, such as the odds that the documents would have been altered, their status as notes or drafts, and what they knew about their source. Typical of this reasoning are comments from subjects along these lines: the "interim status argues for authenticity (why fake a half-finished work?)" or "don't know why you'd think it is a fake."

Experiment Two: Conclusions

We drew three main conclusions from this experiment. First, subjects used a complex reasoning process that took many different factors into account to judge the authenticity of digital documents. When subjects were shown three speeches in the same format, they inferred primarily from the contents and writing style whether it seemed reasonable for a university president to deliver such a speech. When presented with a speech outline in three different formats, the subjects considered additional factors such as the appearance and layout of the different formats, or the suitability of the format and the affordances of the application for the intended purpose of writing and presenting a speech. Subjects used the appearance of the documents as one factor, but they also drew on their own knowledge and experiences to deduce which format might be the original. When the subjects were given more information to draw from, they became more accurate at identifying the original and more confident in their choices.

The subjects also made assumptions about the author and about the context in which the documents were created. In fact, some subjects were frustrated by the absence of contextual information. As one subject indicated, "I've no knowledge

of J.D.'s style or a possible purpose for this document." Another pointed out that his conclusion about authenticity was based on "an assumption of Duderstadt's style of writing that I have no way of knowing." Most subjects also made assumptions about the computing environment in the late 1980s, and many of these assumptions were incorrect. For example, some subjects doubted that the documents could have been created in MORE or Word because they assumed, erroneously, that no software packages were available at the time with their formatting and outlining features, and several initially thought that Text was the original because of its primitive look. In spite of careful consideration of many factors, however, a majority of subjects (16 out of 30) were wrong a majority of the time (in 49 of 90 observations) when asked to identify the original documents. Their errors in judgment were due almost entirely to the lack of contextual information that would have allowed them to check their assumptions about the author, the production of speech outlines, or the original software environment, and so to bring accurate information into their reasoning processes.

Interestingly, only one subject used any of the metadata associated with documents to help determine which was the original, even though each document had some associated metadata, such as file name, author, type, size, date created, and date modified. For example, the "created" and "modified" dates for the MORE versions were from 3 May 1990 to 25 March 1993, when President Duderstadt actually wrote the documents; the dates for the Word documents were in a range from 9 July 1998 to 19 August 1998, when the Bentley Library staff converted the files to Word. The "last modified" for the Text files were in a range from 31 July 2002 to 13 September 2002, when we created them for this experiment. Additional metadata about the Word documents included the date created, date modified, and the person whom the file was "last saved by." A quick look at any of the metadata available in the "Properties View" would have revealed that the documents had been altered since President Duderstadt first created them. Archivists consider elements such as file name, creator, creation date, and date modified critical for establishing the authenticity of electronic documents. In this experiment, subjects could have used some of the available metadata to help them determine whether the documents they were shown had been altered in some way, but only one subject did so. (See Figure 12.)

Our third conclusion from this experiment is that usability is the primary factor that will influence future users' preferences for the document format they would choose to use in a research project and that authenticity is at best a secondary consideration. Most of the subjects weighed multiple factors in selecting their format of choice, but they placed more weight on issues of legibility and ease of navigation. The subjects were not indifferent to the question of authenticity, but most subjects were willing to assume that the documents were authentic because there was little or no incentive to tamper with them, because they conformed to their prior notion of what an authentic document ought to look

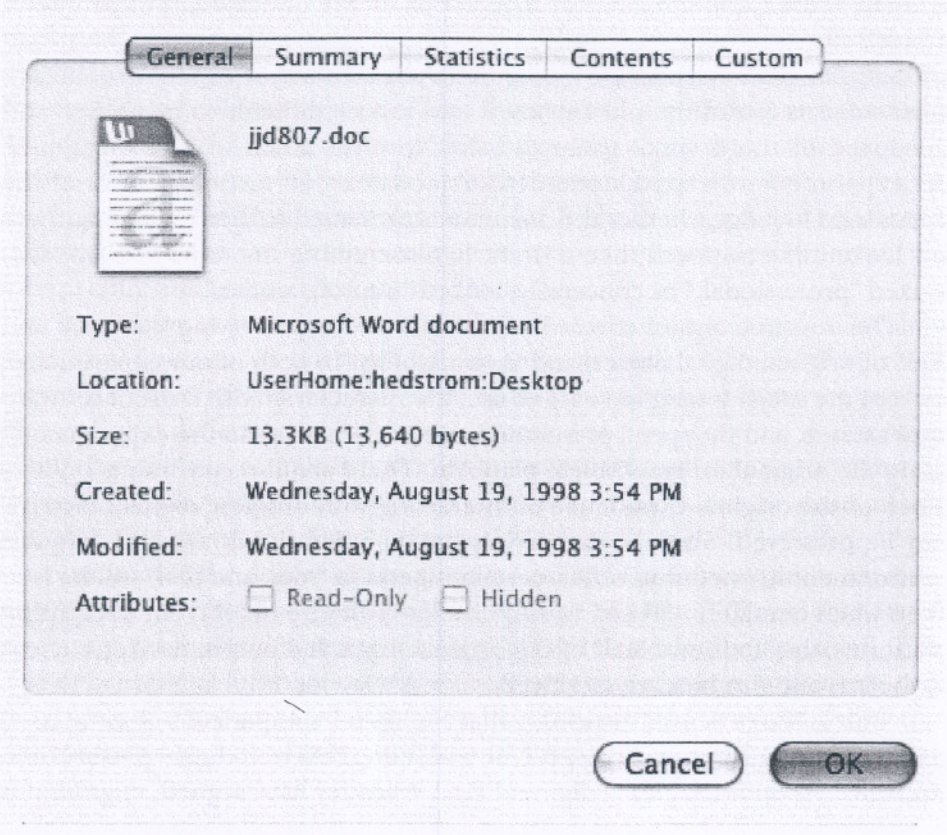

FIGURE 12. View of properties information in a Word version of one speech.

like, or because we told them that the documents came from the Bentley Library.

Conclusions and Lessons for Archivists

When combined, the findings from our two experiments address two main areas: the significance of "look and feel" for preserved digital objects and the importance of contextual information in helping users understand and interpret digital information.

"Look and Feel"

"Look and feel" was important to the subjects in both experiments, but not necessarily for the reasons we originally assumed. In both experiments, subjects thought out loud about how the material "looked" to them, and they often mentioned specific features that looked odd or different when comparing different

versions of the computer game or speech files. In the Chuckie Egg experiment, subjects did not prefer playing an old computer game on its original hardware and software platform because it provided a more authentic experience. In fact, most subjects found the old keyboard and mouse difficult to maneuver and the speed of the original game too slow to offer much of a challenge. In the experiment with speech outlines, subjects analyzed the appearance of the documents to judge whether the documents presented to them looked authentic, looked like a speech, like a draft, like something one could easily read, looked "professional," or conversely, looked "unprofessional."

Our research also uncovered a tension between preserving the "look and feel" of original digital objects and their usability. In both of our experiments, most of the subjects weighed ease of use, their familiarity with current software applications, and the speed of interaction more heavily than the experience of using the original on an obsolete platform. There are also constraints on how much of the original experience of interacting with obsolete digital materials can be preserved. Short of preserving the complete hardware and software environment in operating order, certain aspects of "look and feel" will be lost. Even when emulation is used to run obsolete software on current computing platforms, the "look and feel" of the original input and output devices, such as keyboards and displays, are sacrificed.

Our findings demonstrate also that high-level comparisons of emulation versus migration are not very useful for evaluating different digital preservation strategies. As both Charles Dollar and Paul Wheatley have argued, migration is not a single, unified concept.[36] Rather, different types of migrations entailed varying degrees of change to the original digital object. For example, both the Word and Text files in the second experiment were migrated from the original MORE files, but Word retained many more significant properties than Text. We also learned that not all emulators are equally effective at reproducing the behavior of the original computing platform. Some emulators do not replicate the "look and feel" of the original computing environments as well as some migrations do. Archivists would benefit from a more nuanced view of the advantages and disadvantages of emulation to help determine when it is important to preserve the original "look and feel" of digital objects.

Context Matters

Our research confirms that contextual information is important to users of preserved digital objects, regardless of the technical approach that is used.

[36] Dollar, *Authentic Electronic Records*; Paul Wheatley, "Migration: A CAMiLEON Discussion Paper," *Ariadne* 29 (September 2001) available at http://www.ariadne.ac.uk/issue29/camileon/.

Three types of contextual information were particularly critical for the subjects in our experiments: information about the context in which the objects were originally created and used; information about the purpose and audience for the materials; and information about the original computing environment. Typically, archival finding aids provide information about the creators of materials in collections and the context in which the materials originated, as does the finding aid for the Duderstadt Papers.[37] Our findings suggest that contextual information about the creators, uses, and provenance remains important for interpreting and using digital archival materials.

In addition to providing the types of contextual information typically found in archival finding aids, our research also suggests that archives should provide users with basic information about the computing platforms that were used to create digital collections. Given that we provided no contextual information about the various computing platforms, software applications, or formats, the subjects had to resort to their own knowledge, experience, or assumptions. Not surprisingly, many of their assumptions were wrong. Subjects frequently underestimated the degree of functionality available in software at the time the documents were created and lacked knowledge of when different versions of operating systems and software applications were released or upgraded. This type of contextual information will become even more important as users, over the course of time, become less and less likely to have had firsthand knowledge or experience with obsolete computing platforms.

Future Directions

This exploratory study offers a foundation for further research with different types of digital materials, larger subject pools, and different methods. We encountered several issues that we did not anticipate in the original project design, such as the variability in the quality of emulators, the differences that subjects would observe, and the type of logic they would apply to judging authenticity and making choices about formats. The methods we used are effective for gaining deep insights into subjects' perceptions and behavior, but they do not provide a basis for broad generalization. The small number of subjects further limits our ability to generalize about these studies. We only tested two types of digital information, an interactive computer game and textual documents. It is possible that image materials, interactive multimedia documents, databases, and other genres would produce different results. We also need further research on a wider variety of digital object types before we can decide how to set parameters around the aspects of "look and feel" that warrant preservation.

[37] Deromedi and Steiner, "Finding Aid for James J. Duderstadt Papers."