

Digital Archiving in the Context of Cultural Change

By Kimberly Douglas
Director of the Sherman Fairchild Library of Engineering and Applied Science
Caltech Library System

March 2000

Contribution to “Balance Point” in *Serials Review* v. 26(3) October 2000, pp. 55-59

We see much discussion about how works created in the digital medium will be kept permanently functional as technology changes, as if only the technology will change. These technical explorations focus on how the digital medium will achieve what the print environment has done so well. While mimicry is an accepted form of flattery it does not necessarily allow for unknown potentials to be fulfilled.

Let’s look at the words for a moment. “Digital archiving” is a phrase that uses a noun rooted in antiquity. It first appears in English literature shortly after the turn of the 16th century. This is about 150 years after the invention of the printing press. Next, we modify that deeply culturally laden word with a modern term of the 20th century, “digital,” which describes a dynamic new technology. With these two words, we have taken a subject that we well understand, within the framework of the information world born of the printing press, and push it into our future society by using a new descriptor. It may not seem particularly useful, but it is fair to ask, “Will this phrase survive? True, it’s a bit more useable than, “horseless carriage.” But by how much? It’s possible that “digital archiving” will eventually be added to the oxymoron lexicon. How can something that is fundamentally required to be of a fixed nature be stored in a medium well known for its volatility and mutability?

So, do we really know what we mean by those words, “digital archiving?” With this phrase we are doing our late 20th century best to describe some unknown activity that will exist by the end of the 21st century. Therefore, any discussion or essay about digital archiving must be about the future.

In 1986, in an introduction to his new book, *July 20, 2019*, Arthur C. Clarke quoted this phrase, “There are two futures, the future of desire and the future of fate, and man’s reason has never learnt to separate them.”¹ Seldom, Clarke says, does the real future coincide with human aspirations. Faced with a question about the future of print or how to maintain an historical archive he would assuredly challenge, “Are we sufficiently imaginative when contemplating the future?” Our best approach, it seems, is to take the process of archiving, i.e. selecting fixed works and preserving them in their contemporary manifestation in perpetuity,

¹ Clarke, Arthur C. *Arthur C. Clarke’s July 20, 2019: Life in the 21st Century*, New York: Macmillan, 1986. p. 5 quotes Bernal, J.D. *The World, the Flesh and the Devil*, Bloomington and London: Indiana University Press, 1929. p. 3.

and repeat it in the digital environment. We naively aspire to save it all digitally, reflecting our heritage of the printing press and more precisely the age of “typographical fixity.”² The printing press gave to western society the capacity to record and store its intellectual creations. This accumulation preserved over centuries, is the foundation of our current concepts of thinking and knowledge. So naturally, we believe that the models that have shaped our world and served us well for more than 400 years will continue.

But let’s be clear. The issue is not the computer versus the printing press. The agent of change is not even the computer. It is the network, perhaps an unintended result of the computer’s invention. The digital network links personal computers and creates a killer communication conduit for the whole world. All new mechanisms for human contact drive huge societal changes. The printing press made it possible for an individual author’s works to be quickly and reliably copied for distribution. When authors could reliably reach untold strangers within a fraction of the usual time, the world changed.

Traditional scholarly output, books and research journals, are the new targets now that the postal service, horse and buggy and the daily newspaper have undergone major evolution, in one case to near-extinction. The modern scientific age was born of Copernicus in 1543 with his publication of the *Revolutions of the Heavenly Spheres*. It took another 120 years for academies to form and publish research results in a printed journal, the logical outgrowth of diaries and letters that were shared widely and quickly among interested laypersons and scientists. These societies, originally under the patronage of royalty, added the “value” that warranted the saving and preserving of these publications. While we no longer require a royal presence to certify the value, we do still currently require a stamp of approval by a validating body such as an editorial board. With this validation comes the mandate to preserve the content, and it is perfectly natural to continue to think about preserving it with the new technology. But communications theory and futurist projections predict that it will not happen that way.

We have extensive preprint archives on the web. There are now new aggregation models called overlay Kenneth Pickar recently commented at Caltech that the Internet is the “mother of all disruptive technologies.”³ To 15th century society, the printing press was equally disruptive, or so no doubt the monks of the Middle Ages felt. For us it is crucial to understand that it is the network that fundamentally alters our cultural constructs, such as our method of carrying out daily business and how we place value. The question before us is how will those differences affect the concept of archiving as we know it? Ithiel de Sola Pool warned that in an environment of converging communication technologies (e.g. personal conversations, broadcast news services, and published research) projections into the future cannot be based

² Dewar, James A. *The Information Age and the Printing Press: Looking Backward to See Ahead*. 1998. Rand Publication P-8014. <http://www.rand.org/publications/P/P8014/> (25.08.2000).

³ Kao, Justin. “Creating Tomorrow’s Entrepeneuers [sic],” *The California Tech*. Vol. 101(19):1, March 3, 2000.

on the elasticity or known demand of just one of the services, the preservation of the scholarly journal in this discussion.⁴ To do so is myopic and the conclusions will certainly be flawed.

It's possible that the need to digitally preserve all the articles published in all the research journals may not survive the network. This, however, is not to say there won't be journals. Just as the daily newspaper has survived radio and TV, albeit far reduced in numbers and modified to compete, the printed scientific journal will undoubtedly continue, but it will change. The new technologies provide mechanisms that allow the service or purpose to be ever more narrowly differentiated to meet certain needs or markets. There is really nothing that dictates that the same facts cannot be packaged in different ways to meet different market purposes, and news services certainly know this. That's how all other information services work. There is no such thing as complete replacement, though rag paper did, in time, succeed parchment. Even stone carvings continue when we want to be especially sure that the message endures – witness the use of gravestones and monuments. The words on the Lincoln and Jefferson memorials may be all that is left of the United States in the year 3000.

As shown in this admittedly extreme case, archiving requires copying of words from the original. Generally, in the print world, the print artifact served as the archival piece. The same is not so, as we already know, in the digital environment. Digital storage media have very limited physical lifetimes. The digital creation is likely to last only a few years in its original storage medium. This means that works have to be continuously and iteratively manipulated if they are to be kept for future use in the same decade, let alone in future generations. In our current printing press-based constructs, rights are necessary for such further handling because copyright law forbids systematic copying without explicit permission. However, as we will see, this copyright compliant practice will not work on the global network.

Only the copyright holder gives permission for the creation of duplicate or derivative works. Thus, the copyright holder establishes all rights of online access because such access creates a derivative work. The combination of these controls puts all copyright holders in a completely new position i.e., they become the de facto “archivists in perpetuity.” This is very likely an unintended consequence of copyright applied to the digital medium and publishers have been searching for a business model or a way to make money for the effort ever since.

This brings us to the problem of profitability⁵ that manifests itself differently depending on one's point of view. Profitability for the copyright holder or licensor creates an altogether different set of actions than if

⁴ de Sola Pool, Ithiel. *Technologies without Boundaries*. Boston, MA: Harvard University Press, 1990. p. 228.

⁵ Rebecca Graham of the Digital Library Federation, Council on Library and Information Resources, included this issue of profitability in her presentation on digital archiving for the LITA Interest Group,

one considers profitability in the context of academic scientific research. Many copyright holders are for-profit business concerns. Archiving, including access at some level, has been accomplished in the print world by libraries and museums, with a publicly funded subsidy, because there was no strictly "for-profit" model. The wanderings of the Special Libraries Translations Center file in search of a self-supporting home (a sci/tech librarian's nightmare) is a case in point. Most public, school and research libraries are funded by non-profit organizations to provide services that would never be affordable in the commercial sector.

Archiving has not been a moneymaking venture nor was it every intended to be. Can the digital network now transform this effort into business opportunity for the for-profit sector? We know clearly that in the electronic world information that is easy to get to is the information that will be used. Perhaps there's a business opportunity, but there's also a built-in dilemma. Barriers are necessary to achieve a commercial pricing model but, we know that barriers stifle use. In addition, the iterative costs of continual migration and upgrading plus the cost of traditional intellectual property control is quite likely to outweigh any benefits. Moreover, it appears the users themselves will not put up with it. "The reality is there's no such thing as an invincible copy protection system. It's impractical to make it both invincible and usable,"⁶ said Len Kawell, president of Glassbook Inc., one of the e-book publishers distributing the recent Stephen King novella. He then confirmed that hackers had cracked the 40-bit encryption technology used to protect the story from copyright violations.

Profitability has a different look from the scholarly researcher's point of view. Financial benefit from their writings is indirect and derivative. Through peer recognition and acknowledgement, a measure of fame is earned, and employing institutions reward that recognition with promotion and remuneration. However, an author's works must reach a sufficiently large and appropriate audience for that to occur. Scholarly writers need their work to be distributed and they want it to be easily accessible by others. Mildly put, there's a growing discrepancy between the profitability objectives of the publishing sector and those of the scholars.

If the scientific ethos, as Robert K. Merton wrote in 1942,⁷ continues to embrace common ownership of research findings, then society will witness an evolution of intellectual property concepts. It must, because the network medium is so disruptive that human behavior will create different models. Let's examine one of Jeff Rothenberg's examples of software emulation as a model for archiving. He says, "[There is a] highly active retro-computing community, whose members delight in creating emulators for obsolete video game platforms and other old computers. There are numerous World Wide Web sites listing hundreds of

Electronic Publishing/Electronic Journals, at the American Library Association Midwinter Conference in January 2000.

⁶ Kane, Margaret "Scary! Stephen King e-book Pirated," *ZDNet News US*. Thu, 30 Mar 2000 09:18:09 GMT. <http://www.zdnet.co.uk/news/2000/12/ns-14465.html> (25.08.2000).

free emulators of this kind that have been written to allow old programs to be run on modern computers.”⁸ It is not so much the technical process but the human motivation and sharing that is fascinating. Niches of users with a real need and/or passion are acting to keep the access available. It’s possible that over the network, a geographically dispersed user group for any particular work can leverage their cumulative presence to successfully re-write, migrate, and reformat any work or application into the newer environments. Is it likely?

The network delivers communication and sharing capabilities that our society has only just begun to realize. Here are some emerging themes, by no means complete, that provoke us to seriously contemplate how the network will change society’s behaviors for retaining and passing on its cultural record.

- Small-world network⁹ - The web is not a controlled, centralized or planned organization. It is a small world network typical of social or biological systems. This type of network is neither regular nor completely random. It is a regular network with many random links that have been shown mathematically to be the most efficient in their self-regulatory nature. Recent articles prove that despite the enormous size of the web (800 million estimated documents¹⁰) any one document is only nineteen clicks away from any other, assuming the agent acts intelligently! Could it be that any effort to organize and plan a systematic approach to archival functions cannot achieve its goal, the preservation of important works? An understanding of small-world network connections may provide the inspiration for a different solution.
- Computational power – The available computing power and the incredibly rich behavioral and factual data on the network have generated many ideas about agents and robots. Could automatic harvesting robots upgrade files as they are used? If so, could there be alerting agents for those files that are *not* accessed at appropriate intervals. Since the network is so powerful a tool, humans will certainly devise new conventions for fixedness. Mechanisms for communicating with peers in real time versus communicating with posterity over decades and centuries may well be differentiated for the purpose.

⁷ Merton, Robert K. “Science and Technology in a Democratic Order,” *Journal of Legal and Political Sociology*. vol. 1 (1942):115-126.

⁸ Rothenberg, Jeff. *Avoiding Technological Quicksand: Finding a viable technical Foundation for Digital Preservation*. CLIR Publication 77. (Washington DC: Council on Library and Information Resources, January 1999), p. 25. <http://www.clir.org/pubs/abstract/pub77.html> (25.08.2000).

⁹The recent work of Duncan Watts in papers in *Nature* and his book, *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton NJ: Princeton University Press, 1999 is particularly illuminating.

¹⁰ Lawrence, Steve and C. Lee Giles. “Accessibility of Information on the Web.” *Nature*. July 8, 1999. vol. 400 (6740) p. 107.

These perspectives foreshadow functional computational solutions. The LOCKSS (Lots of Copies Keep Stuff Safe) project at Stanford is an encouraging example of a network-based solution. It has a long way to go but the good news is that it represents an embryonic attempt to use the network’s power to address the problem of perpetual access.

But, back to the theme. The technological solutions will only come into being and be deployed if the social mandate drives it; or as Arthur C. Clarke said, “Anything that is theoretically possible will be achieved in practice no matter what the technical difficulties, if it is desired enough.”¹¹ This will be only possible as long as intellectual property issues evolve away from the printing press model. The alternative, which appears to be the path we are currently pursuing, is that of tight traditional control. This is likely to put a stranglehold on creativity,¹² not to mention overwhelm society with the “expense of sorting out and clearing rights even for ephemera.”¹³

Pool’s 1990 posthumously published work¹⁴ comments vigorously on the futility of applying the intellectual property laws that evolved with, and complemented the technology of, the printing press to the global network. He argues, “The law could impose serious burdens on the development of electronic information handling. Unenforceable laws, like the Eighteenth Amendment, may not prevent individuals from doing what they want to do, under the cover of privacy and corruption, but they do prevent substantial responsible institutions such as corporations and universities from addressing themselves to meeting public demand effectively.”¹⁵ We’ve already seen the signs of this and know the frustration.

The National Research Council’s report, *Digital Dilemma, Intellectual Property in the Information Age*,¹⁶ discusses how librarians are starting the process of creating new models of licensing. They are explicitly negotiating the right to “archive”, i.e. maintain a digital version of the work by specific agreement with the copyright holder. To what extent this model will be economically or socially viable remains to be seen.

¹¹ Clarke, Arthur C. *Profiles of the Future*, New York: Harper & Row, 1973. p. 11.

¹² Dewar, op. cit. argues under his heading of “Making Policy,” that those countries that regulated the printing press the least (Europe) gained the most. The western world needs to apply that lesson in fostering creativity to its handling of the Internet.

¹³ National Research Council (U.S.). Committee on Intellectual Property Rights and the Emerging Information Infrastructure. *The Digital Dilemma: Intellectual Property in the Information Age*. Washington, DC: National Academy of Sciences, 2000. Chapter 3. p. 120. <http://www.nap.edu/books/0309064996/html/120.html> (25.08.2000).

¹⁴ de Sola Pool, Ithiel. op.cit.

¹⁵ Ibid. p. 256.

¹⁶ National Research Council (U.S.). Committee on Intellectual Property Rights and the Emerging Information Infrastructure. op. cit. <http://books.nap.edu/books/0309064996/html/R1.html#pagetop> (25.08.2000).

Although one can liken it to re-arranging the deck chairs on a sinking ship¹⁷ it is worthy and necessary step toward a new model on the continuum of cultural change.

As a self-organizing system the network’s ability to serve society may well depend on removing restrictive rules, such as intellectual property control as we know it, and focus instead on discovering what’s possible. Scholarly communication is well positioned both in terms of a value system and access to technology to break these bonds and chart the path.

In conclusion, our most important digital archiving issue is intellectual property rights, especially those controlled by copyright. Without drastically different treatment of a scholarly author's legacy we will never truly gain the full advantage that network technology makes possible. Research librarians are pushing the boundaries on copyright when it comes to handling the scholarly record by engaging in tough negotiations with publishers regarding perpetual access. These efforts have resulted in policy changes by some publishers in a new recognition of their societal obligations. However, such adjustments do not change the underlying fact that copyright restrictions have led the for-profit publishing industry down an untenable path.

Now is not the time to relax the pressure. The academy, university administrators, librarians and scholarly authors must step up the demands on publishers to forego the traditional restrictions on electronic access. Recognizing the historical base for our current constructs provides a strengthened position from which to recapture scholarly intellectual property for the academy.

¹⁷ Barlow, John Perry. The Economy of Ideas, *Wired* 2.03, March 1994. p. 85.
<http://www.wired.com/wired/archive/2.03/economy.ideas.html> (25.08.2000). See also
www.eff.org/pub/Intellectual_property/idea_economy_article (25.08.2000).