Ten Thousand Years to Ten

Chris Krause

Preservation Management

Dr. Victoria McCargar

May 7 2009

In the ancient past in Mesopotamia the Sumerians recorded royal mandates and records of commerce on baked clay tablets the size of a fist. Using a wooden pen these men pressed into soft material to form a series of wedge shaped symbols. To this day the records remain as they were when originally created. And in this day we have thrown away such archaic media, celebrating the virtues of perpetually evolving digital technology. Overcome by the novelty of unique modes of expression and interaction, we have taken for granted the systems which constitute their application. With haste personal computers are now at the center of common life: replacing analog means of communication, record creation and storage. While these new means of information retrieval are celebrated by the public they are not without flaw: digital records become corrupt or reach obsolescence within a decade, and are accompanied by a great host of silent dangers. Information in the digital age is in constant state of flux and has become fundamentally impermanent. The illusion of permanent preservation of digital records comes through an expensive series of tricks and techniques; the media is at fault, although our diligence as librarians may prove fruitful in a preservation scheme. This paper aims to examine the story of ten thousand to ten, the number of years ancient and contemporary creators might respectively expect their records to survive against the ravages of time. Ultimately this is a report on the issues and problems of digital preservation and the solutions and alternatives available. As we will see, the ideal solution for preserving information at high densities in our time may be accomplished through a marriage of advanced digital technologies sculpting analog materials.

What exactly is digital preservation? What does it mean to make something digital (digitization)? Simply put: digitizing a record has nothing to do with preserving it.

Digital preservation speaks to the processes and procedures which ensure the long term access of a digital record. It is the science of avoiding the mechanical breakdown, as well as the technical obsolescence of digital records over time. While digitizing implies the transfer of an analog record to the digital, it does not necessarily imply a plan to ensure the renderability and understandability of that record by future generations. Digital preservation takes the wisdom of the archivist and applies it to the realm of computers and the internet: it is not enough simply to "hope for the best" when storing digital files, as Conway (1996) so succinctly demonstrated.[1] Specifically, digital files if left to their own devices, and without any special efforts taken to maintain them, typically become inoperable within a decade. This is the shortest lifespan of any medium to date, even lower than highly acidic paper.[2] And while highly acidic paper becomes brittle and browns with age, alerting owners of impending deterioration, digital records become corrupt silently and often en masse. In a world in which "born digital" records are becoming the norm, in which analog means of information retrieval and storage are replaced by the electronic, serious attention must be paid to preserving digital data lest the new host of documents detailing our cultural heritage be lost to neglect.[3] To summarize: digitization is a process of migration from the analog to the digital, digital preservation is the process, science and philosophy of ensuring digital records are not lost in time.

A major aspect of digital preservation is addressing the concern of digital obsolescence. This refers to incompatibility of older records in computer systems sporting newer hardware and software. This often results in the record becoming difficult or impossible to access. During the 1990s the de facto archival medium for digital files

was tape. In contemporary times there are few tape drives still remaining, to say nothing of the state of the tapes themselves. It follows then that it is difficult or impossible to access the information which was stored on such tapes. This is one example of digital obsolescence, wherein technological innovation quickly surpasses the rate at which information is stored. This situation is complicated by a lack of standard protocols regarding digital preservation, although OAIS and PREMIS have attempted to address these issues in recent years.[4] Wise digital preservation includes a serious evaluation of medium, both hardware and software, to ensure the permanent access of records.

The issues of physical deterioration and digital obsolescence can be answered by a suite of tools: metadata, refreshing, migration, copying, and emulation. Metadata refers to the attachment of contextual information detailing the object record itself, including data regarding its provenance, the technologies which created it, the hardware which it is stored on and more. The hope is that the attachment of such information would minimize the likelihood of obsolescence and benign neglect. PREMIS offers excellent standardized spreadsheets for this purpose, which are slowly but surely becoming standard in the digital archives universe. Refreshing refers the transferring of data between two storage mediums to ensure that the bits are constantly renewed and so do not degrade (or experience "bit rot"). Refreshing digital records is a short term solution to the long term and fundamental problem of digital medium disintegration.[5] Migration is the conversion of the record to newer system environments. In practice migration refers to the practice of moving a record from one file format to another, often for purposes of renderability and understandability, or from one operating system to a newer one so as to avoid obsolescence.[6] The overall goal of migration is to maintain functionality of the record and

to avoid the scenario in which the record cannot be accessed due to software or hardware constraints. Copying a record to multiple sources ensures that localized catastrophe does not spell the end of the record. Emulation refers to software which is capable of recreating the functionality of an obsolete microprocessor. Accordingly obsolete operating systems and the applications dependent upon them can be emulated and then be used to access records which would otherwise be inaccessible. In this fashion obsolete records can be accessed or migrated to newer media. Emulation is still in a theoretical stage, although it has some prominent supporters and is gaining impetus in the literature. Jeff Rothenberg recently launched the visionary, modular emulator called Dioscuri, designed to emulate a wide range of early computing applications and operating systems.[7]

Digitization for preservation is sound if the original analog record has deteriorated to such a degree that it is no longer feasible to maintain the record physically. The order is to preserve records with archival or otherwise cultural significance. Normally this entails the production of high resolution scans, "dark archive" master copies (often in .TIFF), and the production of derivatives (JPEGs, GIFs etc) for more general access. Great care must be taken to ensure that the digitization process is not an editorial or creative process; the digital record should be engineered in such a fashion so that it is as similar to the original as possible. Sharpening, light masks and other modifications should only be used to best replicate the composition of the record.[8]

Now for a review of the seminal works of digital preservation theory and the implications and concerns those studies expounded. During the 1980s and early 1990s concern began to rise within the literature about the viability of popular analog media.

The high acid content of wood pulp paper, a reality since the advent of mechanized production following the industrial revolution, was spotlighted as an imminent and serious threat in librarianship. While all media has some degree of inherent vice, i.e. a capacity to disintegrate and become corrupt over time, acidic paper degrades so quickly and had become so prolific as a means of storing information, that some within the library science community feared an impending mass destruction of stored cultural heritage.

Slowly increasing awareness of these issues culminated with the "slow fires" and apocalypticism of the late 1980s, as it soon became apparent to professionals that treating the vast majority of archives with techniques such as mass deacidification would be a monumental and effectively impossible task.[9] Few spotted the acid fires, started in quest to produce paper quicker and in greater volume, before it as too late. Indeed, one might say that the "conservation consciousness" movement came a quarter of a century too late to be of any great significance.

As O'Toole reminds us, the problem of the slow fires and of permanence have not been addressed clearly by the current generation of archivists; the community, faced with such a monumental task as mass deacidification, chose a variety of reactions: inaction, selective conservation, copying and treatment.[10] Some institutions, perhaps compelled by looming economic threats, opt to neglect the idea of permanent collections and in the vein of their 19th century brethren, have taken interest in copying as a means of ensuring survival.[11] These institutions often opt for digital copying. Yet electronic media have their own problems and "vice" of sorts, notorious for even shorter longevity than acidic

paper, albeit with higher density.[12] The hope is that the vast density of electronic

information systems will outweigh the individual failure of the parts.

   In some aspect the seminal works of digital preservation theory represent an

extension of the aforementioned concerns regarding analog media. The first generation of

digital preservation theorists observed not a *slow fire* but a raging conflagration in digital

media. Media was becoming obsolete within a couple years time, lost in a dizzying

lineage of transient microprocessors, operating systems and software. The notion of

storing information in any meaningful way by means of a digital system appeared

questionable. Of course, these concerns were not echoed by the public, who enraptured

by the novelty of creation, adopted and began to standardize the use of these technologies

with little consideration for responsible information storage and retrieval. Today we deal

with the consequences: costly, tedious curation projects, mass corruptions, missing files

and staggering restoration projects.

   Jeff Rothenberg's crucial 1995 paper "Ensuring the Longevity of Digital

Documents" contributed to the maturation of discourse concerning digital preservation.

Previous works were abortive and highly theoretical, focusing principally on the

philosophy of information retrieval in a digital age.[13] Rothenberg introduced technical

expertise for consideration: analyzing the physical nature of the devices and media of

digital systems with the knowledge of a computer scientist and the wisdom of an

archivist. Yet Rothenberg's findings were not stabilizing. Indeed, the paper begins with
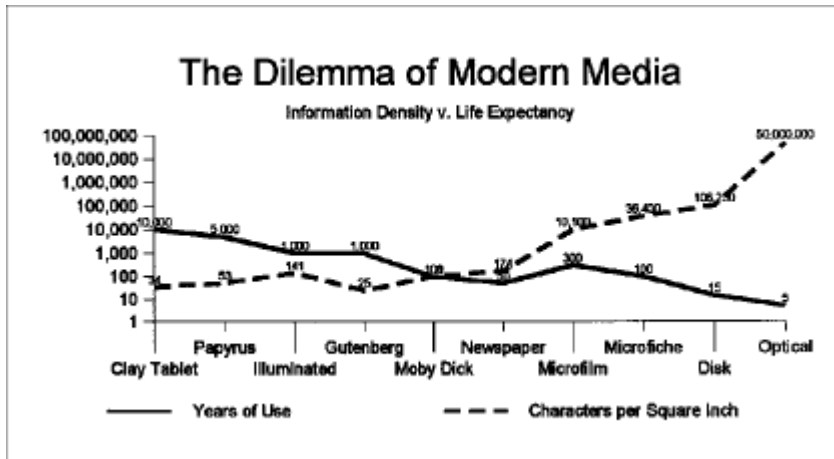
this audacious query, framing the whole of the work:

"Digital documents are replacing paper in the most dramatic record-keeping revolution since the invention of printing. Is the current generation of these documents doomed to be lost forever?"

Rothenberg proposed a number of challenges to overcome if digital media were to become a respectable canvas by which to store information: physical decay of, loss of information about the format, encoding, or compression of files, obsolescence of hardware, and unavailability of software. Unlike analog media, digital media is useless if not carefully governed by a rigorous scheme of metadata and constantly managed to ensure fidelity. To state at the risk of simplification: digital objects do not exist in independence of the systems which read them, they are fundamentally linked to the context and provenance of their creation. Without the proper microprocessor, operating system, disk controller, appropriate size and speed drives, file structure, and software and hardware devices, without a compatible computing environment, digital records lack intelligibility. These themes would come to dominate, and continue to dominate, digital preservation theory. Vast frameworks and models (OAIS, PREMIS) have been constructed to address the themes which Rothenberg first elucidated in his critical work. Alfred Whitehead once claimed with some hyperbole that all of western philosophy is merely a series of footnotes to Plato; the same can be said of contemporary digital preservation theory and Rothenberg.

Aside from proposing these core concerns of digital preservation, Rothenberg also spoke on the nature of the media itself, the implications of which are critical in understanding the discipline. Rothenberg argued that magnetic data was fundamentally volatile in nature: if not perpetually refreshed the bits and bytes constituting digital files

would become corrupt within a decade's time. And while the lifetime of the physical media was shockingly low, no more than sixty years for compact disc, thirty years for tape and ten years for magnetic disks, the rate of physical degradation of media was only exacerbated by a five year average obsolescence of dependant systems. While Rothenberg concluded his paper with a pessimistic note, urging for the immediate attention and action of those in the field in order to avoid a new type of information apocalypse, he nevertheless served a helpful role as whistleblower and advisor, setting the stage and tone of discourse for subsequent works to include a serious technical consideration of the technologies at hand.

Another seminal work, and perhaps of greater relevance for terms of this paper and arguably the library science community at large, is Paul Conway's 1996 work "Preservation in the Digital World," a report of the Commission on Preservation and Access. Conway's major contribution to the discipline was placing digital preservation within a grand historical context, as well as speaking to the relation of information density and media longevity. Specifically: Conway discovered an inverse relationship between the two characteristics; the shorter the lifespan of the medium, the greater the density. The most durable media, clay tablets from Sumer created six thousand years ago, still intact and intelligible to this day, contained text at thirty four characters per square inch. Optical media was examined to contain fifty million characters per square inch yet only had a mean lifespan of five years:

The Dilemma of Modern Media
Information Density v. Life Expectancy

Source: Conway 1996.

This chart is an illuminating portrait of modern information retrieval and storage. The data contained therein has implications foremost for archives and record preservation at large, posing the question: is it really feasible to contain culturally relevant or otherwise archival grade information on a medium which requires constant human intervention in order to maintain? Models such as OAIS are a reaction to the popular inclination toward utilizing digital records, but does the system tender a viable way of managing information in an archival or permanent sense? Is there clearly an advantage to such staggering densities of information when the records in question will become obsolete or corrupt within a decade's time? Conway does not answer these fundamental questions, instead focusing his work on suggesting a suite of tools and methodologies to minimize them as technical hurdles to be overcome with institutional strategies. Accordingly Conway and Rothenberg set the stage for further developments in the field of digital preservation.

We face a real threat, anticipated by Rothenberg, of a "digital dark age," in spite of our best efforts to organize institutionally to preserve digital data. The term refers to a future in which historical documents are no longer accessible due to the problems of

digital information retrieval including obsolescence and physical deterioration.[14] We could face a future in which the vast majority if not all government and public records are exclusively digital, and without the means to access archaic file formats, we might reduce our historical awareness to less than a decade, as older records will be impossible to read.[15] In 2007 it was discovered by the National Archives of the United Kingdom that millions of archival records, stored in obsolete Microsoft Office formats, no longer could be accessed by contemporary computer systems. These files, a mere ten years old, were by that year impossible to access and the information contained within the important government records (covering such diverse topics as nuclear waste storage and census data) effectively lost forever. If not for an expensive restoration campaign spearheaded by Microsoft Corporation, which designed and implemented an emulator known as Virtual PC 2007 to retrieve the obsolete files, little other recourse would have presented.[16]

Supposing that the programmers involved in designing the archaic Microsoft Office software were deceased or unavailable for assistance in designing such an emulator, or if the Microsoft Corporation had declared bankruptcy or otherwise ceased to function prior to the discovery of such a circumstance it is likely that such records would remain inaccessible to this day. A similar event occurred in the 1990s when the magnetic data tapes from the Viking landings were deemed to be unreadable by NASA. The Ampex reading device, designed in the 1970s and necessary to access the tapes was no longer available, and the computer scientists who designed the data structure on the records had long since deceased or parted ways with the organization. Accordingly the incredibly valuable Viking data concerning the planets and objects constituting the solar

system was inaccessible until an expensive reverse engineering and restoration project spanning several years was undertaken.[17] These major public catastrophes overshadow the vast host of anecdotal cases in which digital records of importance are lost to benign neglect and obsolescence. The events also beg the question: is it possible to maintain such a system, even with aid of exhaustive models such as OAIS, if the majority of records in society are digital?

As we may recall, digital preservation implies a permanent storage and enduring access of information. Another means by which longevity of digital records can be addressed is by creating a system catered to "digital sustainability." Rather than address the particular technical challenges of the media digital sustainability aims to create a framework in which interdependent continuous development alleviates the central problems of obsolescence and medium deterioration. Kevin Bradley writes:

"What distinguishes the contemporary sustainability approach from earlier aspirations to a "permanent" solution is the concentration on systems architectures and schemas that will aid in future management of digital information, rather than on the solution itself."[18]

OAIS is a model which follows the aforementioned maxim, aiming to identify the various components of a sound digital information system (ingest into and storage in a preservation infrastructure, data management, accessibility, and distribution) without specifically providing a blueprint for one. In a sense digital sustainability is a sort of floating philosophy and frame of reference which should color the application of computer science and archival work in a digital setting. All well and good, if such a high minded approach to digital preservation was adopted in practice, which it is not. A 2007

survey of organizations within the United Kingdom concluded that less than twenty percent had in place a strategy to accommodate the degradation and loss of digital objects.[19] A similar share of institutions in the United States have taken serious measures to accommodate information in a digital environment.[20] Furthermore, "digital sustainability" is similar to string theory of theoretical physics: it presupposes that supposed future developments will remedy the fundamental problems underlying the discipline. Perhaps this may be the case, but perhaps not.

I claim that it is unethical to continue to use digital records for archival purposes. While digital systems excel at providing access to records (and by no means should be curtailed in that function), they do not excel at preserving records. Digital media is fundamentally volatile and does not afford credit as a sound material for information storage of cultural heritage records, even despite our best efforts to come up with elegant solutions. For every digital record of importance, let there be an analog cognate. If we continue to exclusively utilize born digital records over analog, catastrophes resulting from obsolescence and media disintegration will become more common and have more immediate consequences for the public at large. As librarians we are duty bound to work to avert this outcome: a world in which knowledge and history is transient and in flux.

Yet this proposed outcome does not necessarily entail a return to clay tablets or even paper, both of which have information density potentials several orders of magnitude lesser than digital media. We as a people may still continue to operate in a mostly digital world of experience and record creation. The ideal solution to the issues of digital preservation may indeed be a marriage of digital technologies with the analog: the former used to sculpt the latter.

The High Density Rosetta (HD-Rosetta) is one such technology of interest. Designed by Norsam Technologies and utilizing the technologies of the Los Alamos National Laboratory, the HD-Rosetta is a metal disc of nickel some 2.2 inches in diameter. Data, ranging from graphic to plain text information, is etched onto the surface using a high precision focused ion beam. The data source may include digital files of any sort. In this fashion data is physically engraved onto the medium, not unlike the clay tablets of old, but at a minute scale invisible to the naked eye. Information densities are only limited by available magnification hardware: up to 196,000 pages of data may be contained upon the disc if retrieved using an electron microscope, up to 18,000 pages if using an optical microscope. Accordingly no special computer software or hardware is needed to access the information contained upon a HD-Rosetta disc, only a simple magnifying lens. Norsam claims the HD-Rosetta has an effective longevity of over one thousand years.[21]

And while the HD-Rosetta achieves the high information density requirements of a contemporary medium, it also has the durability and stability (and thus archival grading) which surpasses clay tablets. Specifically the HD-Rosetta is immune to water damage, extremes of temperature and environmental fluctuations, technological obsolescence and electromagnetic radiation (a key culprit behind bit rot in digital systems). While digital systems must operate in a fragile environment and are prone to malfunction given the slightest variance in controls, the HD-Rosetta is marked by resilience. A study by the Los Alamos National Labs found that:

> The HD-Rosetta, an analog, data storage disc, was resistant to degradation at moderate temperatures in stagnant air. At times up to 65 h, little or no

degradation occurred for temperatures up to 300 oC (570 oF). At 450 oC (840 oF)

and higher oxidation of the nickel surface occurred, which rendered parts of the

text unreadable. If the discs were stored in an inert gas atmosphere, it is expected

that they would resist degradation at temperatures above 300 oC (570 oF).

Results of exposure and electrochemical tests indicate that the HD-Rosetta

disc had a high resistance to corrosion in saltwater, tap water, and marine air.

After 15 weeks of exposure no pitting was observed for any of the environments,

and the text appeared almost pristine. The corrosion rate n 3.5% sodium chloride,

as calculated by linear polarization resistance during a 7-day test, was on the order

of 1 mm per year. Extrapolation of the corrosion results to very long times is not

prudent because chemistry changes in the environment may cause localized

corrosion.[22]

In sum the HD-Rosetta is a storage medium which offers both high information

density and the durability of yesteryear. The discs themselves are produced on a common

metal and the information contained thereon may be extracted with common tools.

Perhaps the only thing in danger of long term obsolescence is the ion beam technology

which produces the records, but that potentiality would bear no effect on the renderability

of the medium itself, as the HD-Rosetta is a self-contained record, free from the matrix of

dependencies common among digital records. The HD-Rosetta is also economically

sustainable: it does not require energy (to operate, migrate and access), new hardware or

software, constant management or refreshment.

The same cannot be said of digital systems which consume rare materials like

rubber, petroleum based plastics, and gold and copious other raw materials including

lead, mercury, tin, silicon, aluminum, iron and copper.[23] Digital systems require constant replacement of degrading and obsolete parts, producing an endless stream of waste. And computer waste contains powerful toxins and carcinogens including dioxins, polychlorinated biphenyls, cadmium, chromium, radioactive isotopes, and mercury.[24] Are digital systems really sustainable in perpetuity? Even if we did adopt rigorous systems which certainly refresh corroded bits or developed a universal emulator that would make the issue of format obsolescence moot, would these resource intensive machines continue to be producible at manageable expense thousands or even hundreds of years into the future? As librarians and archivists we must pose these questions in order to ensure the survival of our records, not for us alone but for our civilization's future. I find that the answer is no: we must discover a more stable, dependable, independent medium to store records of prime cultural significance.

While the public at large might continue to use digital systems in order to accomplish the operations of the day, archivists must not be so careless in storing and maintaining historical records. History shows us that once a new technological paradigm passes, it rarely if ever recedes. Computers and digital systems will not cease to be publically desirable, the demand will not cease by natural circumstance until it is too late and the consequences manifest as severe. Accordingly we in the librarian science profession must have the foresight and prudence to take action now to properly preserve our cultural heritage records: not on acidic paper, not via ephemeral digital files, but on a sound, permanent medium. The HD-Rosetta is one inspiring alternative; we must act with conviction and courage to make such virtues known to our institutions.

Bibliography

Beagrie, Neil et al. *Trusted Digital Repositories: Attributes and Responsibilities An RLG-OCLC Report*. Mountain View: RLG Inc, 2002.

Bradley, Kevin. "Defining Digital Sustainability," *Library Trends* Volume 56, Number 1 (2007): 148-163.

Conway, Paul. "Preservation in the Digital World." *The Commission on Preservation and Access* Newsletter Number 88 (1996).

Deegan, Marilyn and Simon Tanner. *Digital preservation*. London: Facet, 2006.

Gladney, Henry. *Preserving digital information*. New York: Springer, 2007.

Hedstrom, Margaret. "Digital Preservation: A Time Bomb for Digital Libraries," *Computers and the Humanities* Volume 31, Number 3 (1997): 189-202.

Liu, Ziming. *Paper to digital: documents in the information age*. Goleta: ABC-CLIO, 2008.

Rothenberg, Jeff. "Ensuring the Longevity of Digital Documents." *Scientific American* Volume 272 Number 1 (1995): 42-47.

Watry, Paul. "Digital Preservation Theory and Application: Transcontinental Persistent Archives Testbed Activity." *International Journal of Digital Curation*, Volume 2, Number 2 (2007).

---

[1] Paul Conway, "Preservation in the Digital World," *The Commission on Preservation and Access Newsletter* Number 88 (1996).

[2] Rene Teygeler, *Managing preservation for libraries and archives: current practice and future developments* (Aldershot: Ashgate Publishing, 2004), 83-86.

[3] Margaret Hedstrom, "Digital Preservation: A Time Bomb for Digital Libraries," *Computers and the Humanities* Volume 31, Number 3 (1997): 189-202.

[4] David M. Levy and Catherine C. Marshall, "Going digital: a look at assumptions underlying digital libraries," *Communications of the ACM* Volume 38 , Issue 4 (1995): 77-84.

[5] Claire Tristram, "Data Extinction," *Technology Review* 105 (2002): 37-42.

[6] Donald Waters and John Garret, *Preserving Digital Information. Report of the Task Force on Archiving of Digital Information* (District of Columbia: The Commission on Preservation and Access, 1996).

7 Jeffrey van der Hoeven, "Dioscuri: emulator for digital preservation," *D-Lib Magazine* Volume 13 Number 11/12 (2007), http://www.dlib.org/dlib/november07/11inbrief.html

8 Cornell University Library/ Research Department, "Moving Theory into Practice Digital Imaging Tutorial," Cornell University, http://www.library.cornell.edu/preservation/tutorial/conversion/conversion-02.html

9 National Materials Advisory Board, *Preservation of Historical Records* (Washington, D.C.: National Academy Press, 1986)

10 James O'Toole, "On the Idea of Permanence," *American Archivist* 52 (Winter 1989).

11 Samantha Willner, "To Cut Costs, Library Unloads 95,000 Volume Duplicative Collection," Cornell Sun Times, http://cornellsun.com/section/news/content/2009/11/04/cut-costs-library-unloads-95000-volume-duplicative-collection.

12 Conway.

13 Jay David Bolter, "Text and Technology: Reading and Writing in the Electronic Age," *Library Resources and Technical Services* 31 (January/March 1987): 12-23.

14 Margaret MacLean and Ben H. Davis, *Time and Bits Managing Digital Continuity* (Los Angeles: J. Paul Getty Trust, 2000).

15 Stewart Brandt, "Escaping the Digital Dark Age," Library Journal, Volume 124 Number 2 (1999): 46-48.

16 Maev Kennedy, "National Archive project to avert digital dark age," guardian.co.uk, http://www.guardian.co.uk/technology/2007/jul/04/news.uknews

17 Sandra Blakeslee, "Lost on Earth: Wealth of Data Found in Space," New York Times, http://www.nytimes.com/1990/03/20/science/lost-on-earth-wealth-of-data-found-in-space.html?sec=&spon=&pagewanted=all

18 Kevin Bradley, "Defining Digital Sustainability," *Library Trends* Volume 56, Number 1 (2007): 148-163.

19 Danny Kingsley, "Fading away: the problem of digital sustainability," On Line opinion Australia's e-jounral of social and political debate, http://www.onlineopinion.com.au/view.asp?article=6324

20 NEDCC, "Surveying Digital Preservation Readiness: Toolkit for Cultural Organizations," Northeast Document Conservation Center, http://www.nedcc.org/resources/digtools.php

21 Norsam, "HD-Rosetta Archival Preservation Services," Norsam Technologies, http://www.norsam.com/hdrosetta.htm

22 Jennifer A. Lillard, "A Survey of the Environmental Degradation Resistance of the HD-Rosetta Data Storage Disc," Norsam Technologies, http://www.norsam.com/report.html

23 Russell Morgan, "Tips and Tricks for Recycling Old Computers," SmartBiz, http://www.smartbiz.com/article/articleprint/1525/-1/58

24 Ibid.