

Preserving Information Forever and a Call for Emulators

Presented by SteveGilheany@ArchiveBuilders.com at Digital Libraries Asia 98: The Digital Era: Implications, Challenges & Issues, 17-20 March, 1998, Singapore
<http://www.DLA98.DigiLib.org.SG>

Abstract

Astronomers gather information from the start of the universe. Paleontologists gather fossils that are millions of years old. Information has lasted for very long periods of time. The information plate on the Voyager spacecraft will last for a similar period into the future now that the craft has left the solar system. This paper reviews some of the issues and perspectives of planning to store information and make it accessible on this time scale. The extremes of this review may help to clarify similar issues for shorter preservation intervals. Long term preservation must include a plan for preserving meta data as well as data. This paper also discusses the need for emulators to permanently preserve the essence of the machines that execute the algorithms that convert abstract data into viewable images. We will need emulators to reproduce chronistically accurate images printed from common word processing programs. While these files stay the same over time, each new release of software interprets them in a slightly different way, creating different page images for viewing. Beyond the mechanics, we must also address the need to move up the scale: detectable differences-bits-data-information-knowledge-wisdom at any given point in the future.

Introduction

We now have the promise of preserving information forever(1)(2). Twenty years ago the Voyager spacecraft was launched with a gold plaque that the engineers estimated would last over 1 billion years (3). We now have many digital files that have been created with the intention of permanent preservation. We have also accumulated a history of managing those files(4). Recent revelations about the need for meta-data and data migration for these files are mileposts on the road to permanent preservation, not roadblocks. In almost all cases, digital emulation will provide the final key for assured reproduction forever.

Even if we can preserve information forever, to succeed in our quest to pass our record on to the future, we must stand simultaneously in the paradigm of technology, and the paradigm of creation, the creation of that which we seek to preserve. We must do so because our ability to preserve, and the ability of future interpreters to make use of that which we have preserved, depends on creation. Creation is an ability we have, but may never be able to understand or reproduce, in contrast to our ability to understand and reproduce our technology.

What we do, in our preserving, will change what we seek to preserve, that is, our society, our civilization(5). Making our civilization aware of these possible changes is the professional responsibility of archivists, librarians, and museum curators, just as preservation is our professional responsibility.

This paper is divided into four parts. The first reviews the goals of archivists, librarians, and museum curators and presents those goals as the basis for discussions on how to use digital technology to achieve those goals. The second part describes the technology of permanent preservation. The third part attempts to describe creation, as that which is embodied in the materials we seek to preserve, as the creative act that preservation is, and as the creation that will be done in the future to make use of preserved materials. The fourth part describes what might be our responsibility for the consequences of perfect, permanent, all-encompassing preservation.

1. How We Preserve

1.1 Introduction

The function of libraries, archives, and museums are well defined and are easily extended to make use of newly available technology. Many of the problems attributed to technology are problems that have previously been worked out in libraries, archives, and museums and merely need to be applied in the new context. Choosing what to preserve, preserving as much as can be afforded, using the most economical means, and operating in a passive, low profile manner appear to be important techniques when applying technology to existing preservation problems.

Knowing why we preserve information, and how the task has been carried out thus far, can help us to make our choices of technology and procedures in a future world that is both limitless and increasingly defined and recorded.

All current library, archives, and museum policies and procedures make sense, because they evolved to fit their environment -- our civilization and society. Not many new ones are needed.

Straying from the policies and procedures that have evolved over time is expensive, often prohibitively expensive, to the point of reducing the number of items that can be accessioned (adding to an archives) (2). An example of a departure from the underlying passive model of archives is the proposed practice of periodic migration of files(4) to make them compatible with current computer software. Another example is an over emphasis on links with the goal of matching the richness found on the internet(4), more than is required in paper archives.

Preservation is not a solution to a technical problem, it is the practice of archives administration, museum curatorship, and librarianship. Preserving is not merely copying. Preservation is an act of creation, of creating an authentic, accessible record by accessioning the materials sent to archives. These materials are the record of that which we seek to preserve, our civilization.

1.2 Libraries, Museums, and Archives

Archives and libraries are very similar. Archives place preservation over access and libraries place access over preservation. If preservation and access are both available universally and at a very low cost, the goals of archives and libraries begin to merge. Museums also seek to preserve a record of the past and make it accessible. Access to museum materials can be improved by providing a digital record of the materials. Like archives and libraries, museums will, over the long run, have to deal with the fact that eventually the physical artifacts will disappear, leaving only the digital copies.

1.3 Libraries

In libraries we preserve interpretations. Interpretations are creations that describe what has been done and creations that form the framework for what will be done in the future.

The transition from an oral tradition to a written tradition, and the transition from manuscript to book is similar to transition from book to computer. Librarianship was around for parts of the first transition, and all of the second one. Librarianship can contribute many solutions to the third, the transition to the use of the computer.

Libraries have had to pick which books to keep and make available. When libraries can store unlimited volumes of many types of materials digitally, libraries will have a responsibility to help society decide what not to keep, and what to make it illegal to keep, if anything.

1.4 Museums

In museums we preserve physical objects, as a means for forming a connection with the creators of the objects, their ideas, their creations. Plants, animals, and other naturally occurring objects may also be selected. In addition to illustrating an evolutionary or geologic process, these naturally occurring objects also reflect the creativity of the person who found and selected the objects. Access to the physical objects provides the most primary source for checking the fidelity of data and notes, stored in archives, that are derived from the physical objects and the interpretations, stored in libraries, that are about the physical objects.

Digital scans of the objects, both surface scans and internal composition scans done with procedures such as Nuclear Magnetic Resonance Imaging (NMR or MRI)(6) will provide future access to the objects when the objects themselves have deteriorated. While no analog of an object, manuscript, original document, or book is as good as access to the original, the analogs can protect the originals from unnecessary access. The digital analogs of objects can provide better access to the originals objects, than the original objects themselves can provide, when the original objects have deteriorated or been otherwise compromised by the vulnerability of their physical nature. In archives and libraries, this point is often illustrated by microfilm or xerographic copies that provide the only access to original records that have been destroyed.

1.5 Archives

Archives preserve the information artifacts, the raw information, the bits, the data, the information, the knowledge, the wisdom that was created by, available to, and used by the people who created. Archivists create an archives from what would otherwise be the mere collecting and preserving of documents. Archivists serve the goals of their archives by their decisions on what is chosen, what is organized, what finding aids and links are provided, what is explained and interpreted, and what is presented to later generations.

Archives provide the raw materials for validation of interpretations preserved in libraries. Archives also support research being done to create new interpretation of events documented by archival materials. Using the resources of archives, researchers can reduce their dependence on previous

interpretations and can perhaps be more accurate in their new interpretations.

Archival preservation does not mean 'keep-a-long-time', or forever, it means everything archives are today: the preservation of original materials and the maintaining of original materials in their original form, where the creator of the materials picks the format of the materials, not the archivist. When a record is accessioned, added to an archives, all possible provenance (the order of the records as defined by the creator of the records and the subsequent history and circumstances of the record)(2) is preserved and linked. Even if the provenance is not complete, the record can be preserved. Provenance easily encompasses the concepts of meta-data and emulation described in this paper.

The professional act of accessioning, recording, and maintaining the chain of custody of an accessioned fonds (record series)(2) covers digital signatures, seals, and encryption. You can not know what you preserved unless a person, an archivist, declares that you have what you think you have. The archivist provides their professional guarantee that no illicit alteration was made, no sham was substituted, at any point in the process of accessioning and custody. An archivist generally cannot inspect individual items in accessioned fonds, but individual items are protected by archivists during and after accessioning. Archivists can sign their work, accessioning, with their digital seals. This care, and this professional guarantee, means that the act of preservation is the creative act of a person. Preservation is not mere copying.

1.6 Preserving Accountability

In addition to an archivist's accountability for carrying out their professional tasks, archivists are also charged with preserving a record of the accountability of the organization of which the archives are part. This further distinguishes the archival profession from a collection of technologies and techniques.

1.7 A Proven Archives Technique Applied in the Digital Age

Most commercial issues disappear when records are sealed for 100 years.

Archivists have the practice of sealing records for a period of time. This would work well for preserving the digital content of copyrighted materials. Submission for copyright could include the requirement to submit both a digital raster image and a digital source document and their digital provenance. The digital information could then be sealed for the length of the copyright. From a purely preservation standpoint this would completely solve the problem.

1.8 Development of Digital Media to Meet Traditional Archives Requirements

Archival integrity suggests that digital media, with perhaps more appropriate properties than currently available media, might be created for use in archives. Because all document management systems first write their documents on magnetic disk, archives have no need for the high data rates or short sectors now provided by optical disk vendors. This is because the a magnetic disk buffer can free the archival disk system from the requirement to keep up with high speed input scanners or frequent researcher access. Disk media storage systems can be developed for archives that are significantly larger in capacity,

lower in accesses per unit time, and much less vulnerable to environmental variables, than the media used in commercial document management systems. An example of this type of media is presented in section 2 of this paper.(7)

1.9 Professional Responsibility

Given the role of the their professions, librarians, museum curators, and archivists have a professional responsibility to know the technology and how to employ it as needed to carry out their professional duties.

1.10 Summary

Tools and techniques that are well known to librarians, archivists, and museum curators are very effective in solving digital preservation problems. This is particularly true in the areas of making accessioning decision, allocating scarce resources, and solving organizational problems.

2. The Technology

2.1 Introduction

It appears that the technology will soon be ready to do everything we might ask for. We have identified fossils that are 3 billion years old, providing a precedent for long term preservation of information(8). We routinely build components and record information at the resolving limit of visible light. Nanotechnology (9) promises molecular machines that operate under computer control on the same molecular scale we personally employ to consume food and move about. Physical and molecular structures can be designed for any length of statistical longevity(9). In short, we now have solid theories from which we can begin to build mechanisms and structures to preserve information forever.

2.2 We First Preserved Information Forever Twenty Years Ago

We now have the ability to preserve information forever on a limited scale. Twenty years ago the Voyager spacecraft was launched with a gold plaque that the engineers estimated would last over 1 billion years (3). The life of the plaque is limited by the diffusion of gold into space.(3) The plaque was designed to be read by other, non-human civilizations and was intended to give an overview of our world. The plaque is part of our civilization's journey to become independent of our planet and its vulnerability to asteroids and other threats. The 1 billion year estimate is conservative. The plaque will probably survive the end of our sun in 5 billion years because the Voyager spacecraft has now left our solar system. The plaque is beyond our current means to alter it, and beyond court orders. By the time we have the means to alter the plaque, we will have only interest, not a desire to change it.

2.3 How Long is Forever?

The people who gave us the big bang theory also provided the possibility of the big crunch, when the universe falls back on itself. The whole process, bang-to-crunch, a pulse, is projected to take 100 billion years, and we are about 15 billion years into the pulse that our universe might be (10). This provides a convenient definition of forever. Just like 'beyond' is part of 'inside' for the curved space-time

continuum, 'forever' is part of 'during' in the space-time continuum. These assumptions give us a fixed size universe to record, index, to provide information to and to store information in(10).

The big bang-crunch provides a convenient definition. If the crunch does not happen, then it will take a billion trillion trillion (a decillion) (10^{33}) years for the 100 trillion decillion decillion (10^{80}) particles in the universe to decay into energy.(10) This possibility would be inconvenient. It might be necessary to go back to defining forever as a-long-time, but a-long-time would be fine for the purposes of this paper.

2.4 Engineering Bits to Last Forever

The longevity of individual bits on the plaque can be calculated using statistical chemistry. The longevity of the information represented by the bits can be engineered to be of any desired length through the use of error correcting codes (ECC's)(11). There are already ECC's in wide use(4). For example, the ECC that raises the raw error rate on CD's from one error in ten to the fifth power, or one-in-one-hundred-thousand, to one error in ten to the twelfth power, or one-in-one-trillion. Given easily achievable rates of bit fading over the life of the universe, any desired level of certainty of bit survival can be achieved.

2.5 Current and Near Term Preservation Technology

Going beyond current CDs, and beyond DVDs, using ion milling(12) with a 50 nanometer spot size, 165 gigabytes can be stored, this year, for a conservatively projected 1 thousand year life, on a CD sized nickel disk from Norsam Technologies.(7) The nickel disk is corrosion resistant and is stable over a wide range of environmental condition.

An example of technology available in the near term is an ion milled, 8 to 12 inch iridium disk. The iridium disk uses a 7 nanometer spot size to store 50 terabytes (50 trillion bytes) for a possible, but not yet engineered, 100 billion year life. To put this in perspective and to show what is coming, using future technology, the content of this disk, 50 terabytes, could be sent over a single optical fiber, with the 1300 nanometer optical carrier modulated at one bit per baud, in under two seconds.(13)

2.6 Is It Economically Feasible to Protect the Bits from All Dangers?

The cost of magnetic disk storage has dropped by a factor of about 1-million-to-1 over the last twenty-five years (14). Even easier to calculate, over the last twenty-five years, the cost of RAM (Random Access Memory) had dropped from about one US dollar per byte to about one US dollar per megabyte, also a factor of 1-million-to-1(14). By the time we can retrieve the Voyager plaque, we will also be able to scatter an effectively unlimited number of Voyager plaque-equivalents throughout our universe, for essentially no cost, in a pattern that will make them irretrievable in their entirety, and thus unalterable, while simultaneously guaranteeing retrievability for at least one copy.

The rate of cost reduction has actually been exceeded by the reduction in physical size. In 1972, the IBM 2314 magnetic disk system required 10 square meters (120 square feet) of floor space, including access panel swing and operator walk space, a volume of 25 cubic meters (1 thousand cubic feet), to store 8 magnetic disk drives with a capacity of 25 megabytes each, or 200 megabytes. Today a 9 gigabyte, half-

height, 3 1/2 x 1 1/2 x 5 inch, .5 liter (.015 cubic foot), magnetic disk drive stores 45 times the information in 1/50,000 times the volume, a reduction of 2-million-to-1 in twenty-five years.

In 1972 a megabyte of RAM required 4 square meters (40 square feet) of floor space or 10 cubic meters (350 cubic feet). In 1998 a chip containing 8 megabytes of RAM requires 2 milliliters (1/8 th cubic inch), a reduction in volume of 40-million-to-1.

These reductions are a million-to-1 in 25 years, a trillion-to-1 in 50 years and a trillion-trillion-(a sextillion)-to-1 in 100 years. And, we tend to underestimate technological predictions in the long run(15).

We are well on our way to Eric Drexler's world of nanotechnology(9), where the 6.022 x 10 to the 23 rd power (more than a half-trillion-trillion) atoms in a mole of iridium weigh only 192 grams (4 ounces) (16) and we do not need many iridium atoms to store a bit forever.

Sending information over the internet is free. This is inconceivable, even today.

Historically, information technology has maintained steep price reduction trends over long periods of time. There is a story that a few centuries ago England was happy to spend a royal fortune to build a chain of signal towers from Lands End to London to send the single bit of information that the Spanish Armada was coming in 1588.

As storage declines in price, space limitations disappear as a constraint for old record formats. For older records this will break the sequence of weeding out duplicates and then losing the one remaining original through misfiles, disaster, theft, or accidental destruction.

Declining prices and shrinking volume do not solve the problem entirely, however. New types of documents will continue to need ever increasing numbers of bits per item. This will maintain the need for professional judgement to limit accessions to accommodate space constraints.

2.7 Can our Civilizations be Recorded as Ones and Zeros? Exclusively?

All of this 'promise of technology' assumes that all of our records can be recorded as ones and zeros, in binary. Until recently this was held in question by many people. With the advent of CD's for music, DVD's for movies and video, the use of computers in publishing and the office environment, digital cameras, and digital fax machines, recording information as ones and zeros has become synonymous with recording information.

Digital copiers and fax machines render older analog documents into a digital form on a regular basis. Few people listen to music with the realization that it is stored as ones and zeros, but everyone readily accepts the fact when it is presented. Today, the basis for most of our world is digital, with an analog veneer(17).

Modern manufacturing starts with a digital model, which is rendered into physical form by computer controlled machines (18). Chemistry and genetics now start with a digital model of elements and bond properties that combine to form molecules with computer model predicted properties.

Not only can existing library, archives, and museum items be digitized, most new items are now first created in digital form(19). The best image of a book is the digital file produced by the program used to typeset the book. Librarians will have to come to

terms with the practice of preserving the degraded paper version of the original digital image.

2.8 Can the Perfectly Preserved Bits Be Correctly Interpreted?

The Need for Emulators

One of the harbingers of the coming digital age is the suspicion that we may lack the ability to make sense of our recorded ones and zeros in the future. At first meta-data was hailed as the savior. Then we discovered that meta-data deals with declarative records and not with procedural information (20).

For procedural information, such as the actual interpretation of a word processor file by a specific version of a word processor, one needs an emulator. Further, the emulator must provide for running a specific version of a specific operating system on specific computer hardware. The output of the first emulator must then be sent to an emulator running a specific version of display software, on an emulator of a specific graphics display card, or a specific version of a page description language (PDL) (e.g. PostScript Level 3 (21)), running an emulator of a specific version of a printer.

An emulator, which is the key to the puzzle, the digital library conundrum, is the binary embodiment of a specific configuration of hardware in binary form that can be preserved along with the meta-data and data. With emulators, a complete set of support software, meta-data, and the data, the entire document environment can be recreated from stored bits at any point in the future.

2.9 The Theory of Emulators

Emulators depend on the theoretical computer science concept of a Turing Machine(22). A Turing Machine is a mechanism that implements the minimum number of instructions needed to be a stored program computer. Because any stored program computer can be emulated by a Turing Machine, and because any computer can emulate a Turing Machine, there is a transitive relationship that says that any computer can be emulated on any other computer throughout all time.

Initially emulators are implemented in hardware or microcode because these implementations are much faster. Over time, as hardware continues to become faster, speed becomes a moot point, and software emulators provide all the speed required.

The goal for emulators is a sealed module that correctly accepts binary file inputs and provides binary file outputs. Unfortunately emulators are never perfect, although a very close approximation is available to archivists. Over time archivists will occasionally discover, and perhaps correct additional bugs, making the emulators about as perfect as the various hardware versions of a given computer ever become.

2.10 Displaying Digital Records that Lack a Display Form

Beyond word processor records, there are many digital records that do not have an obvious display form. They rely on the execution of a program to produce a displayable form of the record. These programs can only be executed on the computers they were designed for. When the computers are no longer available, the programs cannot be run and the documents can no longer be displayed.

Documents that digitally represent the chemicals, Computer Aided Design (CAD)(18) files,

manufactured items, databases, and Geographic Information System (GIS)(23) records are even more dependent on emulators for accurate renditions of digital records into viewable and otherwise useable forms.

2.11 Current Emulation in PCs

Because versions of an operating system emulated previous versions of the same operating system, low fidelity emulation of some operating system features may change the image of a document from one version of the operating system to another. All of these elements constitute the provenance of a digital document.

Part of the accession process is identifying these elements and preserving contemporary print and display bit-map images of the document for future verification of the correct operation of the provenancial emulator and software. Automation of this check is easily defined because, in most cases, the emulator should produce a perfect match for the preserved contemporary bitmap. (Archivists also have a professional responsibility to understand that because to the Nyquist sampling theorem(24), a 600 dpi computer generated bitmap from a PDL is equivalent in quality to a 1200 dpi raster scanned bitmap.)

2.12 A History of Emulators: IBM

Emulators are another example of finding a solution in history. They are built into IBM mainframes so they can run all of the legacy applications in the mainframe world, some of which are over half a century old. IBM wrote the Report Program Generator (RPG) language to emulate Electronic Accounting Machine (EAM) equipment from the 1940s and 1950s on the 1401 and 1410 computers of the early 1960s (25). The 1401 and 1410 were then emulated in microcode on the 360 computers of the late 1960s and 1970s (14). One of the key benefits of IBM's Multiple Virtual System (MVS) operating system of the 1970s was that it allowed customers to run multiple versions of obsolete operating systems in order to support legacy applications that depended on the obsolete versions of the operating system on emulators of obsolete computers. These legacy features are themselves emulated on the IBM 390 computers of the 1990s.

2.13 A History of Emulators: Intel

of the Intel chip generations emulates(26) all previous generations, as does Intel's latest Merced chip which uses microcode, a binary representation of the complex instructions formerly implemented in the hardware of earlier processors.

Software emulators are actually built for all of the most popular microprocessors, before they are actually produced, so that the operating systems for the new generation of microprocessors can be written and tested while the microprocessors are still being developed.

Bill Gates' legendary basic program for the Altair was written on a one hundred percent software emulator of an Intel 8080 microprocessor that ran on a Digital Equipment Corporation PDP 10 computer.(27)

2.14 A Call for Emulators

Archivists have a responsibility to preserve these emulators in their archives as part of the provenance of archival materials. The emulators must be

preserved in a form that is useable throughout time (binary).

A series of emulator layers, applied sequentially (executed in series) is part of the digital provenance of records. To avoid migration of emulator code, when the computer that emulator A is designed to run on becomes obsolete, the emulator A is run on Emulator B, where emulator B emulates the computer emulator A was designed to run on. This problem, of requiring a series of emulators, can be solved in a more general way by writing emulators to run an emulation of a Turing Machine (or any standard model of a computer instruction set). Then there would never be more than two layers of emulation: the first layer would be the original emulator that runs on a Turing Machine and the second would be a Turing Machine emulator that ran on a future contemporary computer. The UCSD (University of California at San Diego) Pascal compiler generates an intermediate P-code that is the equivalent of the Turing Machine emulator in this case. The P-code makes the compiler independent of the target machine instruction set. The Java virtual machine is a more rigorously defined example of a universal intermediate platform(28).

2.15 Where Emulators May Fail Themselves

Like all systems, there are some aspects of emulation that emulators will probably not provide. Emulators emulate the instruction set of a computer, not the microcode instruction set that implements the instructions. Therefore, bugs in the microcode, such as the one in the early Intel Pentium processors (33) will not be emulated. It is possible that archivists may also preserve microcode emulators. This would certainly begin to show the depth, diversity, and richness of the archival profession.

2.16 Dynamic, Linked Databases Stored in Snapshots

Modern databases are changing. In the past, when new data was entered in a database field, the data previously in the field was overwritten. Today, the newer databases provide the option of implementing data fields as pushdown stacks. In a pushdown stack, the new data is placed on top of the old data, the old data is pushed down one level, and meta data about the change is recorded as an attachment to the new data. Meta data can include the date of the change, the person making the change, and the authorization for the change. Thus, the database extends back in time.

Emulators do not solve the problem of preserving a database that is effectively infinitely extensible both physically and in time, such as the internet. Emulators do, however make snapshots a much more viable solution for a partial, temporary solution to the problem. Emulators handle records that are published with aplomb. To the extent that these databases can be published as snapshots, emulators can permanently preserve the dynamic linked databases while more complete solutions are being developed.

By definition, hypertext is dynamic and infinitely extensible(59). That does not mean that it is not publishable. A snapshot from time to time may provide a stability that is not otherwise available. Eventually the dynamic hypertext and permanent preservation may grow together, but that solution does not seem to be at hand. Perhaps the free and nanoscopic world of molecular machines will do it(9).

2.17 Short Term Preservation

Preservation within the lifetime of a civilization concentrates on preserving the digital record, the bits and not the medium the bits are stored on. This is because the civilization can support copying. Beyond a civilization, only multiple, very long lived, copies can be used. Within a civilization, the life of digital media is not important, only the digital record, the bits, need be preserved. In this case, removing the bits from the incoming media is part of accessioning the record. An example of this is the United States National Archives and Records Administration which has transferred all binary records to a standard magnetic tape format which it plans to refresh to the latest format about every decade(17). For trans-civilization preservation, the permanence of the digital media is synonymous with the preservation of the digital record.

In digital copying, a perfect digital record is recreated using additional redundant information recorded along with the original digital record. This redundant information is needed to implement an error correcting code (ECC)(11). The newly corrected, perfect digital copy is then recorded on new media. Thus in digital archives, the newest, most copied documents are the most accurate, with the fewest raw bit errors. Conversely, in traditional archives, the oldest, least copied, documents are the most accurate.

This is not to say that there are not serious short-term problems in managing digital information. A long term plan may provide some guidance for short term plans, and long term, permanent storage may insure that materials are not lost as the short term use of dynamic storage, networks, and computing platforms makes permanence elusive.

2.18 Chain of Custody, Integrity

For shorter-term preservation, a civilization provides the ability to encrypt a digital seal on the record. An encryption key can only last the life of a civilization because the integrity of the separation of the record from the key requires a civilization.

In the short term, records can be protected absolutely with digital seals. A digital seal is something like an error correcting code that has been encrypted. While the content of the record may be non-encrypted, any change to the content of the record would be identified as an error when the error correcting code was decrypted and applied to the record. Records found to have been altered can be discarded and replaced with unaltered copies of the original records that have been stored at an alternate location and that have had their digital seals checked for alteration as well.

At the start of each day, an archivist can check the integrity of an entire archives, using a hierarchy of digital seals, and make certain that not even a single bit has been lost or changed. By managing access to the digital seal over time, an archivist can establish and maintain the chain of custody of items stored in an archives. With the chain of custody, the integrity of the archives can be assured.

Digital seals can also protect records intended to provide trans-civilization preservation through the use of very long lived media. This trans-civilization use can be assured by only encrypting the digital seal and not encrypting the preserved information. More problematical is the alternative of encrypting the preserved records and assuming that future advanced civilizations will be able to easily break the encryption and decrypt the preserved information.

2.19 Creating a Vision as a Given and Growing into It

Planning as Though the Vision was Going to Happen

Rather than collecting information about technology, we can start with the vision that everything will be preserved throughout all time. We can start to plan, build, prepare, making policy decisions and incorporating technology and accessioning records in accordance with our vision before we know the exact technology that will be used. Eric Drexler suggests the need for careful discussion planning well in advance of the availability of advanced technology(9).

While the technology is at hand to preserve materials forever, it is still true that vision, precedes the creation of the tools to support the vision. The whole is not the sum of the parts, the whole determines the parts and comes first. Freedom is not the result of our work, our work is the expression of our freedom.

This is not so very unusual. All planning assumes the success of the plan. It is when the possibility of the vision is unbelievable that difficulties arise. The United States Declaration of Independence and Constitution laid out a framework, a vision, a creation, which the United States is still growing into.

Action is also required. George Santayana the philosopher said "Those who cannot remember the past are condemned to repeat it." However, those that remember the past, but do plan far enough ahead to apply the lessons of the past to the future, are condemned to relive the mistakes of the past in their future.

A large percentage of what is preserved today will be preserved forever.

2.20 Summary

Emulators solve many of the problems not solved by storing meta-data and not solved by migration of digital media. Media is being developed to meet the needs of archives. The archival media is physically robust, and does not require periodic copying to compensate for rapid deterioration over centuries or millennia. Costs are dropping rapidly and storage capacities, per unit volume, are increasing rapidly. Many archival preservation problems can be solved by making numerous, low cost, complete copies of entire digital archives. The promise of permanent preservation is at hand. We can, and should, begin to plan for it now.

3. The Act of Creation

3.1 Introduction

Creation, the act of making something from nothing, is the essence of what is preserved in libraries, archives, and museums. That which is merely a transformation, a mechanical copying, is not of interest, except as part of the provenance of an item.

The act of publishing, making something ready to fit to the linked, larger world is a creative act. Bringing together two creations, an item and the linked world, to form one creation, makes publishing a creative act. Preservation is a creative act because it is an extension of publishing.

A future researcher's exhuming of a preserved item and linking the item into their future world is a creative act because it brings together two creations,

the preserved item, and the creation that is their future world, forming one creation.

3.2 The Meaning of Publishing, the Meaning of Document

One could say that a published document is a presentation of information in a form that can be integrated into the sum of recorded information. Publishing could be said to be the extra effort that is taken to create a document from stored information.

Historically, the effort of publication has been imposed by the limitation of previous technologies. When the technologies were swept away by computing, many thought that all the effort in publishing could be dispensed with. Unfortunately this eliminates the production of published documents and leaves just raw information. Publishing a document indicates some level of checking, some level of group agreement and approval, some promise that it is likely that the document can be correctly interpreted, and that a good citizen effort has been made to limit the number of targets to which multiple reference might be made. Raw information carries none of these traits.

In publication, reference targets or versions should be limited in number to balance the demand for a readers time and understanding between the one published item at hand and the universe of items the published item has been linked to. In the case of versions, more is definitely not better.

3.3 Did We Get It?

Preserving starts with what is preserved. The connection between the information, knowledge, and wisdom and the digital record is the person certifying that the digital record is accurate in both quality and content. In accessioning a record to a digital system, the original record must be identified and the fidelity of digitizing process must be certified. Only a person can do this. For this reason, all digital records must be sealed by a person operating under the authority of their civilization. At all times in the conversion and preservation process, the chain of custody of the record must be secure and the provenance preserved.(4)

3.4 The Archival Basis for Interpretation

Archives are a good model for evaluating all the versions of 'refresh' and 'migration' that digital libraries propose. (Refresh is defined as the exact copying the bits in a binary file, with the application of an error correcting code, to new media. Migration is defined as updating preserved document or record files to be compatible with the current version of the applications that created the files.) The archivist's responsibility to preserve the actual original documents can be projected to preserving: the original data or document files, the original bitmaps of printed and displayed documents, and the provenance of the software version that created the files, the operating system version that supported the software version, the emulator for the computer that the operating system ran on, and the emulators for the software and hardware used to print and display the bitmap images of the documents.

3.5 Interpretation

Any form of improved access or interpretation, such as migration, is beneficial, as long as the original documents, in their original form, are still accessible. Historically, the survival of original documents could

not be guaranteed, due to deterioration, theft, fire, or willful distraction. This is no longer the case. We can guarantee survival of original documents through the production and dispersion of multiple identical, low cost, long lived, digital copies.

It is important to remember that all forms of migration are interpretations: image processing on raster bit maps, raster based OCR (Optical Character Recognition)(60), raster to vector conversion, 2D (dimensional) to 3D conversion and 3D to solid model conversion in engineering drawings, the 16 to 32 bit conversion, the coming 32 to 64 bit conversion(26), and the Y2K (year two thousand) conversion in software, auto-refining of catalog databases, and source file version updates for documents. Each of these forms requires the creative act of proofing, review, republishing, and in some cases re-engineering.

Preserving the original assures that no damage can be done in interpretation, as happened in the past when the interpretation supplanted the original. The digital equivalent of replacing the original with a copy is expending funds on interpretation (digital migration) while original documents are being lost. During interpretation it is important to remember that provenance of the document is a fundamental part of the original document and all interpretations should themselves be linked by their own provenance to original documents.

3.6 An Example of Mechanical Migration Without Interpretation

To illustrate the problems of mechanical migration of records, without the creativity inherent in the professions of librarianship, archival administration, and museum curatorship, we need merely call on the famous author who wrote under the pen name of Twelve Feet (Two Fathoms or Mark Twain without the mechanical migration).

3.7 Interpretation, an Act of Creation

Because interpretation is republishing, which is an act of creation, it requires the act of a person. If someone does not check and sign the interpretation, the act of creation is on the part of the person who is the viewer of the interpretation to judge if usefulness. It is the responsibility of the archivist to make the person using an archives aware of the provenance of what is being viewed. If the item being viewed is an interpretation, the provenance of the interpretation must be made available and explained.

An example of interpretation is OCR. The popular program, Adobe Acrobat (34) interprets a bitmap image of a document and replaces it with the raster registered vector based character glyphs that most exactly match the scanned bitmap characters, their font and their size. Here the few glyphs that are not recognized are left as in the original raster scanned form. This process produces beautiful, typeset quality documents because the few glyphs that are left in their original raster form are invisible. The raster glyphs are invisible because the eye fixes on the overall quality of the vector glyph typeset document and effectively averages out the ragged edges of the few raster scanned glyphs.

Patrons assume that typeset quality documents have been through a publishing process in which errors were removed. With the replace bitmap form of Acrobat, this assumption is not correct. In fact, because a dictionary is included in the OCR process, words can be changed to better match the OCR model of the document, with the result disguised as

beautifully typeset text. In this case, archivists also have a special responsibility to understand the mechanisms of vision and perception(29).

Archivists' professional responsibility extends to accommodating disabilities. When necessary, the provenance of a displayed image should include a transformation of color spaces to match the color gamut available to color blind patrons(29). That is, archivists must make available systems that produce individualized color shifts for each colorblind patron so that each person can see images with the best fidelity available.

Archivists also have a professional responsibility to know about the Acrobat option to preserve the original bitmap and to use the OCR text output exclusively as a finding aid.

3.8 Paradigms

A paradigm is an act of creation. In creation, from nothing comes everything. Paradigms, like our universe, are all encompassing. In our universe, 'beyond' exists only 'inside' our curved space-time and 'after' exists only 'during'. Because a paradigm is all encompassing, within a paradigm, there is no outside, and no beyond. The paradigm of technology admits to no other paradigm. Within a paradigm, the paradigm is invisible, like good typography. Things in the paradigm 'are', as in 'are reality'. Things outside the paradigm 'are not', as in non-existent.

Speaking to someone in another paradigm is difficult, if not impossible. Words do not have the same meaning. Obvious, essential basic concepts have no meaning or are attacked. The complete lack of common ground brings mistrust. Seeing something from the other person's point of view means giving up everything you believe in. You may learn how to predict the other person's actions, but you can never relate to them.

We can conceive of more than one paradigm, we may be able to stand in more than one paradigm, but beyond that, it is extremely difficult to operate or reason in more than one paradigm without shrinking our multiple paradigms back to a single paradigm.

Because all paradigms include, and are created complete with, their own history, once we have personally entered a paradigm, we also discover that there has always been a history of the paradigm that showed that there were good reasons for the paradigm's existence. For example, when we see that emulators are useful in preservation, we enter the paradigm of emulators. Being in the emulator paradigm, we can see that there is a history of emulators that makes it obvious that emulators are the best choice for permanent preservation. Outside of the emulator paradigm, it is not at all obvious that emulators are in any way useful for permanent preservation.

3.9 Interpreting, Creating a Paradigm

One does not truly enter a national park until they have created a context for the park in their paradigm. Because Horace Albright realized this, he wrote the first instructions for national park rangers to interpret national parks for visitors. The interpretation is intended to assist park visitors in creating their own context for a park(30).

3.10 Will the Information Be Interpretable Across Civilizations?

What is the Meaning of Meaning?

There are many answers to this question.

Given the ability to guarantee the reproduction of bits until the end of the universe, there is still the issue of meaning. The concept of the sociology of knowledge(31) says that information and knowledge is specific to a civilization and that meaning dies with a civilization.

The most extensively funded work on long term (over ten thousand years) information preservation has focused on preserving civilization along with the recorded information. Monasteries have been proposed to maintain the interpretation of keep-out signs on nuclear waste dumps.(35)

In the face of this possible need to preserve our civilization in order to preserve our records, we can assume that future interpreters will be able to reconstruct our civilization as we have reconstructed past civilizations lost to us. To assist future interpreters, we can include what we hope is a Rosetta Stone (36) with our preserved information. This is what was done with the Voyager spacecraft plaque.(3)

We have many examples of interpreting lost civilizations. Egyptian and Mayan temple carvings(32), separated from us by millennia, can be interpreted in many ways, none of which are certain. But, we are much closer to these civilizations than the future interpreters will be in a billion, or even a million years.

However, in some cases, contemporary events are difficult to interpret for contemporary observers. Even in our personal lives, we are out of sync with the times of our surroundings. This suggests great problems when a decade, much less a century or millennium separates the creation of a digital record from its interpretation. Today, the six million transistors of a Pentium II micro-processor cycle 300 million times per second, executing about 300 million instructions per second to do our word processing(26). This fact is beyond interpretation in most of our personal contexts.

An alternative to the need for future interpreters to reconstruct our lost civilization is to assume that in preserving the record of our civilization forever, we may cause our civilization to last forever. This possibility is outlined in the last section of this paper.

3.11 First People, Then Technology

We can, and have, engineered bits to last forever. This prospect of bits that will last until the end of time has nothing to do with our ability to know, to learn, to create, or to be wise. All that we speak of, all that we record, the essence of our civilization, is outside our recording mechanism, which is a part of our engineering. It is possible that we will never know how we know, learn, create, or act wisely.

The ability to describe, understand, and engineer our technology must be dealt with, lived with, 'groked' (37), simultaneously with the realization that the essence of creating the material we are preserving is outside of the paradigm of technology and possibly outside our ability to describe or understand. Creating may be something we are only able to do, not something we can describe or understand. For this reason, technology must always play a secondary role to the creation we wish to embody in our preserved records and our preserved documents.

It would be better to lose everything we have recorded, and all of our technology, than to suffer even the slightest diminution in our ability to create.

3.12 Moving Up the Scale

Discernable differences (e.g. pits in the smooth surface of a CD) can be used to represent bits. No machine can make a discernable difference into a bit, only a person can declare that a discernable difference is a bit. No matter how many discernable differences are seen in the record of our civilization that might be preserved forever, only an interpreter can see them as bits to be interpreted.

There is no number of bits, large or small, that can make the bits into data. The beginning and end of a data element or measurement must be defined by a person. And, the data representing elements or measurements must be assigned meaning by a person.

Moving from data to information again requires interpretation, creation of meaning. A person must create an idea in which the data can be understood as information.

Knowledge could be said to be an analog of reality. Knowledge can only be created by a person. No amount of information can produce knowledge.

Wisdom carries with it the concept of 'should' or morality. Knowledge does not produce wisdom. Only a person is able to become wise. Only a person can assist another person in becoming wise.

When we record information forever, we are hoping that our permanent discernable differences will foster wisdom forever. We have no proof that it will happen, or even if it is possible. We may not be able to know if it is possible, but is a noble cause.

Lord Kelvin said that (paraphrased) : "If it cannot be measured, it is unimportant(38)." A possible corollary, with respect to moving up each step on this scale is "If it can be measured, it is unimportant." The ability to move up from step to step, an action on which our digital records rely for usefulness, is immeasurable, perhaps unknowable. It is, however, something we can do, and something it is possible that future interpreters may be able to do as well.

We know we can preserve the bits. We must guess that we, or some other interpreters, will be able to make use of them.

3.13 Technologists Create

We must dispense with the technologist's question: "How can you say you created it from nothing if so much existed before you created it?" To exist, every bit requires an act of creation by a person. This is true even though the discernable difference existed before the bit was created.

Technologists also live in this gap. All digital records and signals are analog in nature, by definition. The finest digital distinctions, found in quantum mechanics, are probabilistically not really there (39). All of our records preserved as bits are digital analogs of the documents they preserve, a seeming contradiction in terms, even an oxymoron. We therefore use analog digital signals and media to preserve the digital analogs of our documents.

Using logic, it is possible to say that it is logical to say that it is not logical to be logical. That is, to paraphrase Goedel, you can't prove things about a system from within the system (40). There are many aspects of ourselves that are not amenable to being figured out, among them is the basis (creation) for things we seek to preserve forever in binary form.

Beyond analog and digital, what is preserved in a bit is a pure mathematical concept, a one or a zero, not a physical embodiment. Collectively, with error correcting codes, the mathematical concept of bits can

last until the end of the universe when it will not be possible to distinguish a one from a zero.

There is nothing wrong with good engineering. The Library of Congress building was completed ahead of schedule, under budget by the United States Army Corps of Engineers in 1898 and is an incredible work of art (41). It is merely helpful to see it as an act of creation.

3.14 Creation in a Moment of Insight

In the same way that our universe suddenly was, with no 'before' before the big bang, creations spring forth instantly, with no past, created from nothing.

An archives is not a room full of records, it is an idea, a commitment, an agreement, a creation. An archives is created in the instant a person with the authority says: "We will have an archives. Lock the records room. Write procedures. Restrict and manage access." An archives ends in an instant as well, when a person with the authority says: "We do not need an archives. Unlock the records room. Throw away the procedures. Let anyone who needs the records have them."

Archivists are devoted to continuity. Archivists and their archives represent their parent organizations' devotion to continuity. Continuity is very helpful, very important. Continuity and creation are a strange combination, but archivists must create and support creation on the part of others and on the part of the archives parent organization.

3.15 Seven at a Time: And Other Limitations

One of the limitations we seem to have as people is that we can only hold or consider no more than about seven concepts at the same time. For this reason when we catalog, outline, or organize our writings, objects, or any of our creations, with more than seven divisions in one category, we group some of the items to form a sub category. (This section of this paper, Section 3, has an informal subgroup of subsection that have the word 'create' in their title.) That is why there are seven layers in the ISO OSI (International Standards Organization Open System Interconnect) protocol stack. That is also the source of management's span of control tending toward seven people working for one person, and forming formal or informal subgroups if there more than seven people working for one person. The best description of this limitation is in a paper entitled the "Magical Number Seven Plus or Minus Two published in 1956.(42)".

The limit on the number of concepts we can hold simultaneously is linked to the functioning of our short term memory(42). It is also linked to our vision(29). We merge image that are presented to us at more than about 60 per second. This produces the illusion of motion. (43) Conversely, we can perceive individual images present at less than 60 image per second. This is used in the craft of book design (44) to make it possible to flip through the pages of a book at 60 pages per second. This can be seen easily by observing how many pages are flipped to find a given page in a book and then reflecting on the number of seconds the flipping required. The most common example of this is in using a dictionary.

The downside of the speed with which we can perceive images is that if we do not get stimulus to hold our attention on a topic, our short term memory will fade or be replace in under one second. More than one second will then be required to refresh our short term memory, to recontextualize our actions. For tasks such as word processing, database interaction, and most recently and most importantly

internet access, a task that requires a fraction of a second to complete will have its execution time extended by more than a second if the system does not provide the user with a response in a fraction of a second. Addressing this perceptual parameter can increase by a factor of one, two, or more(45).

Librarians, Archivist, Museum Curators, and Records Managers have a professional responsibility to be aware of these limitations when designing access mechanisms. These limitations are also some of the differences we may have with future intelligences, who might receive our preserved record.

3.16 Summary

The concept of creation is essential to preservation. Without it, one may come to believe that copying is preservation and that future reproduction is the goal of preservation.

Most of what is needed in preservation constitutes creative acts. Preservation and preservation systems are domains of creative people.

What is preserved is not conveyed to the future until an interpreter recreates the item or concept, the creation, for themselves.

We do not understand creation. We do not know if our attempts in support of future creation will succeed. But, if success is not certain, neither is failure. We can try.

4. The Effect of Our Effort

4.1 Introduction

Heisenberg said that to measure something is to change it (5). When we preserve and catalog the record of our civilization, we change our civilization. We increase understanding and reduce the diversity of our civilizations. Diversity moves from unawareness to known differences. As we share in the celebration of our differences, we become more alike. As we remember better who we have been, who we have been is more a part of our future.

In the face of this, creation is independent of all that exists. That is why creation is said to be an act of producing from nothing. Leaving the world of mechanistic logic, we have the possibility of drawing on our preserved record, or not drawing on our preserved record, as we create our future. Future users are similarly free to use or not use our preserved record to create their futures.

In addition to changing our own civilization by accessioning our record into the fully linked structure of the future, we will add to our civilization's knowledge of our universe.

4.2 How Will Preserving Our Record Change Us

To measure something is to change it, an axiom described by Heisenberg (5). Certainly to record something is to change it, even if the record is inaccurate or completely fictitious. In Southern California, where the record of what is built is fictitious, existing only in the popular media of entertainment and advertising, buildings are built to look the way they are supposed to in the movies or real estate brochures. Building designs are influenced by fictional descriptions both to be appealing and to earn income as on-location movie sets. An example of this is the seaside Hyperion waste treatment plant on the south side the Los Angeles International Airport which is built to resemble something out of a space opera (46).

4.3 Merging Civilizations

Creating a perfectly preserved record of our civilization may, in turn, preserve our civilization in the same way that it is suggested that the English language has changed very slowly since the time of Shakespeare(47) because his writings were so wondrous that no one wanted to give up the ability to understand his writings in their daily language. Shakespeare's arrival appears as a seminal event. English had changed more in the century from the Middle English of Chaucer to the Early Modern English of Shakespeare than it has in the five centuries since then.

The philosophy of Confucius is said to have been the underpinnings of the Chinese civil service system and the Chinese civil service tests, which slowed the change of Chinese civilization for two and one half millennia. The last of Confucius' descendents, Kong Decheng of the 77 th generation, moved out of the family home, where Confucius had lived in 479 BC, in 1940 AD(48).

As with Shakespeare, whose version of English has absorbed all future changes rather than being deposited or Balkanized by them, recorded information will tend towards creating one civilization where there were many. Today, television announcer English captures the mechanism of homogenization and hegemony in the term. Narrowing the gulf of interpretation, understanding, and communication between civilizations reduces the separateness between them, creating a single civilization. It is difficult to describe the merging of paradigms because by definition paradigms are fully self contained.

4.4 The Future, A Context

Our information can be cataloged by referencing it to a fixed framework of a known size. In our case, we are 15 billion years into the 100 Billion year lifetime of our universe(10). In physical dimension, our universe is a sphere with a radius of 15 billion light years that is expanding at the speed of light. Libraries, archives, and museums seek to preserve or provide a group memory for humankind, to provide a context for our thoughts and activities. Cosmologists define their domain as all of space and time. We can create syndetic structure to hold and organize our records of what is known of space and time and our thoughts and activities therein.

4.5 Linear Measure

For linear measure the following units provide different contexts or paradigms that are not easily related to one-another: angstroms, microns, points, inches, cubits, yards, meters, fathoms, chains, furlongs, li, miles, leagues, astronomical units, light years, and parsecs. The metric system brings the possibility of seeing all distances as being related, a new step in knowing our universe. Distance and time become dimensions along which information can be organized. We can see that a rainbow is an octave when the rainbow and octaves are cataloged by the linear dimension of wavelength.

We can see our place in the universe by locating ourselves on our planet and following our planet through time and space in a three dimensional astronomical catalog. The movie and book "Powers of Ten"(49) shows the relationship of size along a scale from atomic particles to the entire universe.

4.6 Time

We do not often see time as merely another dimension. Our calendar is taken as reality. We begin to see that our calendar is merely a cataloging tool when we learn that the Catholic Church was off by six years when our calendar was established, that Jesus Christ was born in 6 BC (a contradiction in terms), and that the third millennium started in 1994(50).

The Jewish calendar(51), which makes 1994 the year 5755, brings order to a longer period but has been in common usage a few hundred years less than the Catholic calendar. Events during the earlier periods covered of the Jewish calendar are even less certain because time as a numeric dimension convenient for cataloging is a relatively recent invention.

Many of the event series and time relationships in the past, that we relate to easily, would have been recorded quite differently if every action was timestamped and recorded exactly and forever, as it could be in the future. Will the future be different if everyone counts microseconds, as they do when using their PCs, and ignores the Earth, the Sun, the Moon, and the stars, thusly relegating them to the status of cataloged items?

Arthur C. Clark combined the calendar and a three dimensional model in his 'simulacrum' to catalog and display changes in the history of a city in 1956(52). Visitors could use a 4D projection of a simulacrum of the city to move a cursor through the city in any direction or through time. Any scale or perspective could be selected, and by moving smoothly one could fly over, move through, and watch the growth and decline of parts of the city over time. Hypertext will allow us to catalog using an effectively infinite number of dimensions. Every object, every item, will have the possibility of a location in time. How well cataloged should the future be?

4.7 GIS

Cartography has become the production of Geographic Information System (GIS) databases(23). In GIS databases map features are made up of digital points and vectors on a Digital Terrain Model (DTM). In a GIS, the image of the land's surface can be draped over the DTM using digital orthophotogrammetry(53). The cameras used in this process are flown in a plane and us the GPS (Global Positioning Satellite)(54) information to record very exact digital position, the camera's exposure, camera angle, and time, and the photograph which can also be recorded digitally. These features are spreading to all types of digital cameras and spoken annotations can be added easily in digital form. Soon no photograph will be without comprehensive digital annotations.

In southern California surveys are done often because the land frequently stretches, slides, thrusts, subsides, accretes, compresses, liquefies, shrinks, tears, rotates, drifts, expands, bounces, and washes away. There is no fixed property size, elevation, or location, only a property configuration on a given day. Some subdivisions, such as Portuguese Bend, are built on permanent landslides. Older technologies could not keep up. Now GIS is an important component of emergency planning and response.

4.8 CAD (Computer Aided Design)

CAD(18) links the common components of everything manufactured and all structures built. In combination with GIS(23), we can locate structures in space. We will be able to find the plans for our house by locating our house on a map. This will support

increased detail in zoning regulations and increase compliance through access to rules, regulations, and a history of variances granted. With CAD and GIS, we can see where things are and we can discuss where we would like new things to be. Building shadows and tree growth can be modeled and debated.

4.9 Chemistry to Nanotechnology

Reactions, molecular structures, and physical properties are modeled in the computers and then rendered by arranging atoms. The merging of manufacturing and molecular construction has been mathematically modeled by Eric Drexler in his *Molecular Machines*(9). In Eric's world of free everything, only our thoughts and records will have any value.

4.10 Phylogenetics and Genealogy

Eventually all individuals and species will be linked in phylogenetics(55), with differences shown as a sequence of mutations in the genetic code. The human genome project(56) will make it possible to link mutations and genealogy.(57)

Phylogenetics will grow up with genealogy and in the fully linked future world it will seem normal that someone with the name Kung would know they are a 74 th generation descendent of Confucius(39).

Today, seed banks preserve the biodiversity of many species to protect against the dangers of monoculture(58). Phylogenetics can completely defines biodiversity. In many cases biodiversity will be defined down to the individual.

4.11 The Foundation for Hypertext(59)

In addition to explicit document links, hypertext will be founded on the physical syndetic structures of space and time found in CAD, GIS, and 3D models of the universe, and the chemical structures of nanotechnology and phylogenetics. Every means of organization will become a dimension of hypertext. The current possibility of recorded-but-misfiled-and-lost will essentially disappear.

4.12 Hypertext in Support of Document Creation

Hypertext will also reduce the amount of information stored. Today we store a complete copy of a file every time we make a back-up. Hypertext treats every keystroke and mouse click as a linkable object. Commands are just objects made up of keystrokes and mouse clicks. These objects are collected in a string as a person records information. Hypertext is an efficient means of using any part of this string to create a new object called a document. Every keystroke creates a new version, with publication at the discretion of the author.

In hypertext, a delete command is just another object that creates a new form of a document. Similarly, an un-delete command is just another object to be added to the string of stored objects. While this makes it impossible to make an unrecoverable mistake, it also makes it impossible to delete anything. The actual function of a word processor or a hypertext system is an implementation decision. In many contexts it is also a matter of rules and regulations.

4.13 Hypertext Links

Adding links to the hypertext database will become synonymous with research, taking notes, with the simple act of viewing or reading. The act of choosing something to view or read is an act of creation and

can be saved in a log of reading history. Nothing need be lost. It is the archivists' professional responsibility to encourage debate on what should be lost and what should be saved. Hypertext is designed so that all notes, all viewing and reading histories for all individuals can be made part of the hypertext database without getting in the way. Adding links will be a way of life. Those research notes and document drafts that were previously lost, thrown away, or filed and later discarded, or accessioned into an archives, will become permanently available links. Even the least formal links will be used because almost any lead is of value when starting from scratch on a research topic. Information will, by definition, be fully hyperlinked, unless we decide not to.

4.14 Hypertext Provenance

The original scanned document image should be held inviolate, with all enhancements and modifications maintained as a series of algorithmic appliques.

Whenever historic artifacts or documents are accessioned, preparation almost always includes a strong desire to complete the documentation of the record, or rearrange the material in order to return the documents to their original order so as to assist the researcher. How often do photographs contain dates, let alone a summary of the event in which it was taken? Digital archives provides the first opportunity to both preserve the item exactly as it was received, while allowing the researcher to apply any number of interpretive aids without modifying the original.

Because a raster digital image has no intrinsic scale, both the original physical size and the intended viewing size and distance must be explicitly recorded for preservation as part of the provenance of the image. This is because a pixel can cover part of an atom, or an entire galaxy, depending on the subject of a digital image. Billboards are intended to be viewed from a distance, and stereopticons from the end of one's nose. A digital image should make note of this meta-data.

4.15 Applying Traditional Archival Techniques to Digital Imaging

When scanning multiple images on a single piece of film, the entire microform should be scanned as a single image. Individual images on the microform can be identified by added digital tags, hyperlinks. Separation of images on the microform should be logical, not physical.

This is the digital equivalence of preserving provenance. If a film is scanned section by section, there is a chance that some portion will be missed. More importantly, the relative position of scanned segments are hard to document. By scanning the film in one piece, provenance is preserved as part of the scan, just as it is incorporated during the arrangement of paper records, because of the importance of maintaining the integrity of the order of the film (paper) document.

Scanning the microform as one image also automatically picks up the resolution test chart(60) at the start and end of microfilm rolls and any identifying strips or tags filmed along with the documents. The microform meta data which was developed over a period of a century is seamlessly and transparently transferred to the digital world, preserving the provenance of the microform record.

By treating each microform as a single image, the policy of refreshing is favored over the more expensive policy of migration of the images on the microform, and scarce cataloging funds can be

conserved. All of the film based finding aids will continue to work on the new digital analog of the film record. No new indexing, beyond indexing the single new digital image, is required. Additional indexing can be done in hypertext, over the internet, at any time, by anybody, to any pixel or group of pixels (sub-image) on the scanned image of the microform. If, in the future, funding becomes abundant, detailed, sub-image by sub-image (frame by frame on fiche and roll film), indexing can be done.

Scanning microforms as one image is more passive, less expensive, preserves provenance, and reduces the archivist's role in republishing (migration).

4.16 Beyond Links: Searches

One example of a search technique is a simulated neural network, the same method we use when we look at something and review everything we have ever seen to figure out what to do. For us, one result might be: "Yes, that is my toothbrush." When applied to a comprehensive collection of the OCR produced text of all documents in a fonds, this search could be used to discover plagiarism of scripts by comparing every new script to every script every submitted or registered with the Screen Writer's Guild. It could also be used to make hyperlinks from newspaper clipping books to the original papers, reconnecting the theme of the scrapbook to the other events of the time.

4.17 Statements to the Universe

To spread the recognition of their creativity, we can provide every person, every family, every group, every group of groups with the opportunity to publish a document from time to time. This document would be their statement to the universe for all time. This could be a way to apply ISO 9000(61) to the quality of life.

Because these personal statements are the very things that are most likely to be lost over time, it may be that most people benefit from being able to forget what they say and write in their youth or in moments of stress or confusion. In our rush to preserve permanently, we need to create tools and mechanisms to provide the benefits (for example, the ability to loose records) of the systems that existed in the past, while also providing the benefits of the new systems we might be proposing.

4.18 Should We Do It?

Not long ago we were always within ten minutes of destroying our planet. This brinkmanship lasted almost one-half century and was made possible by a technological advance. Soon we will be able to preserve a record of our civilization forever, with no possibility of it being lost. The converse is that we will never be able to destroy the record, and the record will include everything: how you drove to work each day of your career, everything you bought, what you read, what you wrote, and what you erased. And not just for you, for everyone. What should we forget as individuals? What should we forget as a civilization? Perhaps nothing, but the question can never be dismissed.

In the book 1984, people watched the video from cameras above street corners. In the real future, we only check the recorded video when an event occurs. We check the video to see what lead up to the event, what happened during the event, and what occurred after the event (usually a robbery). Eventually it could be possible to review someone's entire life on

video as soon as the person does something of interest(60).

What correlations will be possible when every tax deduction requires a traceable transaction number identifying the payer and payee? Archivists have avoided this problem by saying that it relates to business records and not to archival quality records. Business records also carry with them a professional responsibility for judgement. Records managers(61), who turn their records over to archivists as one of the alternatives when a record reaches the end of the record's retention schedule, share in the professional responsibility of librarians, archivists, and museum curators to preserve the records of our civilization with judgement.

4.19 The Existence of a Permanent Preservation Plan Will Sustain the Program

No matter what problems exist today with preserving information forever, it will soon be possible. A permanent preservation plan establishes a goal, a matrix, a set of expectations, and a place to link ideas for review and addition. Even if permanent preservation is not immediately available, a plan says that it is reasonable to assume it will exist in the near future. If it is said to be likely to exist in the near future, then it is also likely that we will begin planning for it.

What will it look like to us when we are done? This question, created by a plan, will lead to designs and to growth in every dimension.

4.20 A Reward

Perhaps one of the benefits of being able to record, and have access to, all information is that we will see that we are more than that, more than all the information that it is possible to store. We may see that more information does not solve our fundamental problems and does not deal with the fundamental issues of civilization. It may deliver the insight that it does not mean anything. We will still have to create our civilization, moment by moment.

In the paradigm of creation, the world could end tomorrow. It is only our current scientific context that says that the world was not created a moment ago along with all the existing scientific history of our world's existence.

To survive, maybe we will have to evolve, change, forget our civilization and start an entirely new one.

4.21 Summary

When we show an order and structure in our catalog for what we have stored (not 'the' order, 'an' order), we create the possibility of order, of a fitting together, of all-encompassing, of understanding. We have been able to do all these things before, just not so perfectly and pervasively. We have a professional responsibility to everyone affected by our possible future success.

We may also cause our civilization to persist as the record resists change. A new civilization can no longer be created without the knowledge of previous civilizations because the record of the previous civilization is continuously at hand.

Two civilizations that are intimately linked in a common record must grow together, forming a composite civilization. They must follow the path of two paradigms that move from co-existence to common understanding.

Preservation may slow the evolution of civilization and reduces diversity between coexisting civilizations. This may be good or it may be bad, but it could be the effect. The effect will probably become stronger as the preservation becomes more comprehensive, and more permanent, and more available.

Conclusion

Preserving everything forever will be technologically feasible in the near term. Emulators stored as bits will make traditional static preservation feasible in the permanent digital context. It is reasonable to begin planning as though preserving everything forever is possible today. Planning for permanent preservation will provide insights in solving short term problems. It will also preserve materials that might otherwise be lost when funds are expended on procedural matters to the exclusion of simple static preservation.

Our planning should be founded on the professions of librarianship, archival administration, and museum curatorship. We must see our efforts in the context of creativity, the creativity embodied in what is preserved, the creativity of preservation, and the creativity inherent in the use of preserved records. Our preservation efforts are not merely the solution to a technical problem. We must know, and make it known, that preservation is not merely copying, not a mechanical task. We have the professional responsibility to use the newly available technology wisely, in service to our society, to our civilization, and to future recipients of our preserved record.

Appendix

Science Fiction, A Short Story:

Passing the Baton Over the Wall, The Electron Story

Imagine you are a scientist in a billion year old civilization at the end of the last big bang - big crunch cycle of the universe. You know you cannot stop the cycle, but you want to pass your archives on to civilizations in the next cycle. You also know that an electron can be described as a particle or as a wave. As a wave it can be modulated like a radio wave, except it has a much higher frequency, and therefore a much higher information carrying capacity. You use your science to change the big bang ever so slightly, so that when electrons are created in the next pulse, the next universe, they are defined as being waves that are modulated with the information from the archives of your civilization. To see if this happened, all we have to do is demodulate the information, if any, from any electron(63).

This short story acknowledges the oral tradition as an important way to preserve information and understand our world. This oral tradition was one of the earliest examples of the need to refresh files (64). The Bishop Paiute Tribe has not yet made the transition to a written form for their language. Their language may go directly from an oral tradition to the internet and permanent preservation until the end of time. Some of the discussion of the oral tradition was done with the Cynthia Andrade of the Bishop Paiute Tribe Telework Center that can be found on the web at www.Paiute.com and with Dolly Manuelito of the Bishop Paiute Even Start program. One of the aspects of the American Indian oral tradition is that all actions are taken in the context of their effect on the next seven generations.

Acknowledgements

Some of the research for this paper was done by, and many of the sections of this paper were discussed with: Alan Bain, Director of the Archives Division of the Office of the Smithsonian Institution Archives, and Archivist of the Smithsonian Institution, Charles Dollar, Associate Professor, School of Library, Archives, and Information Studies, University of British Columbia, Canada, Susan N. Newcomer, President, Crescent Meadow Systems (providing consulting in records, library, and information management), and Guy Shaw, Sr. Software Engineer, Sun Microsystems. Many others also helped in the discussions and research for this paper. Thank you all. All of the errors and omissions are mine, Steve Gilheany.

All trademarks are the property of their respective holders.

Notes and References

1. Additional resources for this paper and updated versions of papers can be found at www.ArchiveBuilders.com.
 2. Archives: Many of the terms used in this paper are from the archival profession and can be found in A Glossary for Archivists, Manuscript Curators, and Records Managers, Bellardo, Lewis J. and Lynn Lady Bellardo, Chicago, The Society of American Archivists, 1992. www.Archivists.org.
 3. Voyager plaque: p15, Ferris, Timothy, The Whole Shebang, Simon & Schuster, 1997. p226, Sagan, Carl, Billions and Billions, Random House, 1997. Launch dates: Voyager 2 August 20, 1977; Voyager 1, September 5, 1977.
 4. Digital libraries: Garrett, John and Donald Waters, Preserving Digital Information, the Report of the Task Force on Archiving of Digital Information, commissioned by the Commission on Preservation and Access and the Research Libraries Group, Research Libraries Group, May 1, 1996.
www.SLA.org (Special Libraries Association), www.ALA.org (American Library Association), www.ASIS.org (American Society for Information Science) www.RLG.org (Research Libraries Group).
Digital libraries listservs: send: subscribe DigLib to listserv@infoserv.NLC-BNC.ca and send: sub DigLibNS to listproc@sunsite.Berkeley.edu (news service).
- From 4: p 12. Checksums: There are various well-established techniques, such as checksums and digests, for tracking the bit-level equivalence of digital objects and ensuring that a preserved object is identical to the original.
- P21. Appraisal: Archives cannot save all information objects; they must appraise and select for retention the most valuable items. Selection processes for archives of all kinds -- paper and digital -- are matters of intellectual judgment about what to include and save and what to exclude. Criteria for such judgments are largely tied to the intrinsic qualities of the material and many of the criteria that have proven useful in the paper world will no doubt translate to and prove equally effective in the digital environment.
- p43, Clifford Lynch (1996:142) observes that in linking names and locations in systems of citation for digital objects, we seem to be demanding a higher standard than for information objects in the print environment.

- Lynch, Clifford (1996) Integrity Issues in Electronic Publishing. In Robin P. Peek and Gregory B. Newby, eds., *Scholarly Publishing: The Electronic Frontier*, Cambridge: The MIT Press, pp. 133-145.
- p14. Provenance has become one of the central organizing concepts of modern archival science (Dollar 1992: 48-51). The assumption underlying the principle of provenance is that the integrity of an information object is partly embodied in tracing from where it came. To preserve the integrity of an information object, digital archives must preserve a record of its origin and chain of custody.
- Dollar, Charles M. (1992) *Archival Theory and Information Technologies: The Impact of Information Technologies on Archival Principles and Methods*. Macerata: University of Macerata Press.
- Alan L. Bain, Smithsonian Institution Archives: Its History and Activities on Digital Imaging, Proceedings of the Second Documenting Japan International Executive Seminar, January 21, 1998, Tokyo, Japan.
5. Heisenberg: measuring affects the thing that is measured: p37, 12-13, Feynman, Richard P., *The Feynman Lectures on Physics*, Reading, MA, Addison Wesley, 1963. The Heisenberg uncertainty principal is used to protect the theory of quantum mechanics from possible attacks based on reasoning based on classical mechanics.
6. 3D (three dimensional) scanning: NMR (Nuclear Magnetic Resonance) changed to MRI (Magnetic Resonance Imaging) to eliminate emphasis on the word nuclear. www.RSNA.org. The Radiological Society of North America.
7. www.Norsam.com. Norsam Technologies scientists have taken a look at the Voyager calculations which are based on diffusion of the metal itself over time. The environmental considerations were very limited since space offers few variables. Under similar diffusion limited circumstances, any material or metal with a higher melting temperature than gold (such as the nickel Norsam Technologies is currently studying), will last longer than the gold disc on the Voyager. We have not considered damage from cosmic rays however.
8. Fossils: Mark Stoneking - Editorial: Ancient DNA: How do you know when you have it and what you can do with it? *Amer. J. Human Genetics* (1995) 57:1259-1262.
- Soltis, P.S. and Soltis, D.E. 1993. Ancient DNA: Prospects and limitations. *New Zealand Journal of Botany* 31 (3) 203-209.
- Lister, A.M. 1994. Ancient DNA: Not quite Jurassic Park. *Trends in Ecology and Evolution* 9(3): 82-84.
- Berand, Hermann, *Ancient DNA*, Springer Verlag, 1994.
- Eldredge, Niles, *Fossils: The Evolution and Extinction of Species*, Princeton, NJ: Princeton U. Press, 1991. p 180, The earliest fossils are 3.5 billion years old. p 78, A DNA study has been done of a 125 thousand year old fossil.
9. Nanotechnology: p 159: Where indefinitely prolonged system life is required, the standard (molecular machine) engineering answer is a combination of redundancy and replacement of damaged components. This is feasible in nanoscale systems, but at the cost of substantial increases in system complexity. Drexler, Kim Eric, *Nanosystems, Molecular Machinery, Manufacturing and Computation*, John Wiley & Sons, 1992, Drexler, Kim Eric, *Engines of Creation - The coming Era of Nanotechnology*, Anchor Books, 1986, Kim Eric Drexler, *Unbounding the Future - the Nanotechnology Revolution - the Path to Molecular Manufacturing and How It Will Change Our World*, Quill, 1991.
- The static ion milled surfaces proposed by Norsam Technologies use larger design features and have to contend with only diffusion in space, and chemical reactions on earth.
- The unbroken (as far as we know) chain of RNA and DNA reproduction over the last 3.5 billion years shows that the molecular machine of life on earth has proven Eric's point, using the same techniques but different design requirements.
10. Big Bang: p 171, Mkrauss, Lawrence, *Beyond Star Trek, Physics from Alien Invasions to the End of Time*, Harper Collins, 1997. The big bang to the big crunch will span 100 billion years. If there is no big crunch all the particles in the universe will decay into energy in 10^{33} years (1 billion trillion trillion years). p129, Hawkings, Stephen, *A Brief History of Time*, Bantam Books, 1988. There are 1×10^{80} particles in the universe.
11. ECC (Error Correcting Codes): Pohlmann, Ken C., *The Compact Disk Handbook*, 2 nd ed, AR Editions, Madison, WI, 1992. A Reed Solomon Error Correcting Code (ECC) reduces the raw error rate from 10^{-5} to 10^{-6} (p 67) to a corrected error rate of 10^{-10} to 10^{-11} for audio (p 73) Sherman, Chris, *CD ROM Handbook*, 2 nd ed, McGraw Hill, 1994, 101, An ECC further reduces the corrected error rate to 10^{-12} for CD ROMs.
12. Ion Milling: p. 153-195, Rai-Choudhury, ed, *Microolithography, Micromachining, and Microfabrication*, v.2 *Micromachining and Microfabrication*, Bellingham, WA, SPIE (Society of Photo-Optical Instrumentation Engineers) Optical Engineering Press, 1997. Chapter 4, Focused Ion Beams for Micromachining and Microchemistry by Diane K. Stewart and J. David Casey, Jr.
13. Speed of document transmission on future fiber optic links. A 1300 nanometer wave length is about a 230 terahertz carrier frequency. In most cases, the speed of transmission of documents in binary form on fiber optic links can be increased by replacing the transmitter and receiver and retaining the fiber that has been previously installed.
14. Increase in data density over time: In 1994, IBM said that: IBM's magnetoresistive head technology, which underlies magnetic disk design and prices, had been increasing the areal bit density of magnetic disks at a rate of 60 percent per year since 1989. IBM projected that the 60 percent rate would continue for the foreseeable future. IBM's laboratory results confirmed this rate until at least the year 2000: The Era of Magnetoresistive Heads Grochowski, Ed, IBM Research Division, Almaden Research Center, San Jose, CA., 1994.
- IBM recently said that it will be able to continue its increase in disk density of 60 percent per year that it began in 1989: IBM press release, *New Technology Stores 11.6 Billion bits of Data on One Square Inch of Disk*, December 30, 1997, www.IBM.com.
- The calculation of the historic rate of decline in magnetic disk storage cost is based on the price of the 5 MegaByte RAMAC disk drive introduced by IBM in June, 1957, at a monthly rental of \$3,200.00*(in 1957 dollars), and on 1994 disk prices. The decline from the 1957 RAMAC cost of \$100,000 per MegaByte (Adjusted to 1994 dollars, and adjusted for IBM's historic desire to rent rather than sell), to the cost of \$ 1 per MByte for drives available in 1994, represents a decline of 27 percent per year for 37 years: Bashe, Charles J., *IBM's Early Computers*, Cambridge, MA: MIT Press, 1986.
15. Technology Forecasting: When viewing these figures, its is important to remember that projections of technology beyond ten years tend to be conservative because they do not adequately account for new technology. H. A. Linstone and M. Turoff, eds., *Delphic Method: Techniques and Application* (Addison Wesley: [where], 1975). Also see J. P. Martin, *Technological Forecasting for Decision Making* (New York: American Elsevier, 1972).
16. Physical and chemical basis for future data storage densities: By definition, one Mole of molecules is Avogadro's number, 6.022×10^{23} , of molecules and the weight of 1 mole, in grams, is equal to the atomic weight of the molecule.
17. (A) The present day world has become very digital. (B) NARA, the United States National Archives and Records Administration) policy for digital record preservation is to immediately copy all incoming digital records to a standard form of magnetic tape. The digital records on the standard digital tape are copied to the new NARA standard for digital storage media about every ten years.
- From Charles Dollar: (A) Future historians are likely to view the last three decades of the twentieth century as a watershed where the convergence of digital technologies reshaped the information landscape and thereby fundamentally altered how people create, retrieve, use, and view information. This convergence is particularly evident in the telecommunication industries where audio, traditional print, still pictures, motion pictures, and telephone signals increasingly are being stored and retrieved in a common digital base. The traditional distinction between information objects such as letters, books, audio recordings, maps, photographs, movies, video, and telephony based upon the means of transmission or carrier of the information that has supported separate technologies, disciplines, professions, and industries is being eroded. The magnitude of this transformation and its long-term implications for society are barely recognized, much less understood, although many contemporary observers believe that the transformation is similar to what happened with the introduction of writing three millennia ago.
- Although few contemporary observers fully understand the magnitude of this transformation, there are several generalizations that can be offered. First, every indication is that reliance on digital information will increase in virtually every segment of society, ranging from the home, to government, and to the workplace.
- From Charles Dollar: (B) National Archives of the United States Archival Preservation System: This is based on the concept of transferring or copying electronic records in bit form from one media and associated recording format to the same or different media with the same or different recording format without the loss of authenticity or integrity.
- Ensuring Access Over Time to Authentic Electronic Records. Strategy, Alternatives, and Best Practices presented to the National Association of Government Archives and Records Administrators, July 17, 1997, Charles Dollar, University of British Columbia, Canada.
18. CAD (Computer Aided Design), standard. ANSI X14.26M. *Digital Representation for Communication of Product Definition Data*, ANSI, Published by ASME, New York. www.ASME.org (American Society of Mechanical Engineers), Vendor:

Preserving Information Forever and a Call for Emulators

- www.Autodesk.com, product: AutoCAD Vendor: www.Intergraph.com.
19. Computer Typesetting: Bigelow, Charles and Donald Day, *Digital Typography*, Scientific American, vol. 249, N. 2 p. 106-119, August, 1983. www.GATF.org (Graphic Arts and Technical Foundation).
 20. Procedural and declarative (meta data) files: Wilson, Leslie B., and Robert G. Clark, *Comparative Programming Languages*, Addison Wesley, 1988. p250 Procedural Languages, p325: Declarative Languages.
 21. PDL (Page Description Language) example: Postscript Language, Reference Manual 3rd ed., Palo Alto, CA.: Adobe Systems www.Adobe.com.
 22. Turing Machines: p 80-101, Ullman, Jeffrey D., *Formal Languages and Their Relation to Automata*, Reading, MA, Addison Wesley, 1969.
 23. GIS (Geographic Information Systems) vendor: www.ESRI.com, Environmental Systems Research Institute, Redlands, CA. www.URISA.org Urban and Regional Information Systems Association.
 24. Computer printed images are twice the quality of scanned images of the same resolution: Nyquist Sampling Theorem: H. Nyquist, "Certain Topics in Telegraph Transmission Theory," (April 1928; reprint, New Jersey: Bell Telephone Laboratories, August 1928): 1-3. A 1200 dpi (dots per inch) scanned image is equivalent in quality to a 600 dpi PDL (Page Description Language) computer generated print image.
 25. Emulation: p 93, Hasson, Samir S., *Microprogramming: Principals and Practices*, Englewood Cliffs, NJ, Prentice-Hall, 1970.
 26. Emulator and 64 bit Merced chip, www.Intel.com.
 27. Bill Gates emulator: Manes, Stephen and Paul Andrews, *Gates: How Microsoft's Mogul Reinvented an Industry - and Made Himself the Richest Man in America*, Simon & Schuster, 1993.
 28. P-code can be a metaphor for a Turing Machine: The UCSD (University of California at San Diego) version of the Pascal programming language uses an intermediate P-code to make it independent of the target machine instruction set. Aho, Alfred V., Ravi Sethi, Jeffrey D. Ullman, *Compilers, Principals, Techniques, and Tools*, Reading, MA, Addison Wesley, 1988. P734 P-code. The Java Virtual Machine is a more robust platform independent, intermediate machine. www.Sun.com.
 29. How vision and perception work: Vision is processed in three separate ways: resolution (black and white line art), color, and motion. Manipulating the presentation of these three elements independently or in unusual combinations produces optical illusions. For colorblindness see p134-144, *The Science of Color: The Committee on Colorimetry of the Optical Society of America*. www.OSA.org (Optical Society of America) and www.SPIE.org (Society of Photo-optical Instrumentation Engineers).
 30. Interpreting the United States National Parks: Horace Albright and the creation of the National Park Service: The Founding Years 1913-1933, Salt Lake City, Howe Brothers, 1985.
 31. Sociology of Knowledge: Stark, Werner, *The Sociology of Knowledge: Toward a Deeper Understanding of the History of Ideas*, 1958, Reprinted: New Brunswick, Transaction Publishers, 1991.
 32. Egyptian and Mayan temples: Budge, E. A. Wallis, *The Mummy: A Handbook of Egyptian Funerary Archaeology*, 2nd ed Cambridge U Press, UK, 1925, Reprinted Dover, 1989. Sharer, Robert J, *The Ancient Maya*, 5th ed, Stanford, CA, Stanford U Press, 1994. Coe, Michael D., *The Maya*, New York, Thames & Hudson, 1993.
 33. Pentium bug: Halfhill, Tom R., *The Truth Behind the Pentium Bug*, Byte Magazine, March, 1995.
 34. Adobe Acrobat and Postscript: www.Adobe.com.
 35. Ten thousand year old 'keep out' sign: Applicants seeking approval for high-level radioactive waste disposal sites must include a "conceptual design for monuments which would be used to identify the controlled area after permanent closure." U.S., Code of Federal Regulations, title 10, part 60 (Washington, DC: Government Printing Office, 1985). It is interesting to note, that while the importance of the Rosetta Stone was instantly recognized, its existence did not make the solution automatic. Furthermore, it was probably never the intent of the Egyptians to communicate with future generations. By providing a third language, Greek, which survived longer than their own, the Egyptians provided a code breaker for their hieroglyphics. For this and other information about communicating information to future generations see Thomas A. Sebeok, *Communication Measures to Bridge Ten Millennia*, BMI/ONWI-532, prepared by Research Center for Language and Semiotic Studies, Indiana University, for Office of Nuclear Waste Isolation, Battelle Memorial Institute, Columbus, OH., 1984, 20; Human Interference Task Force, *Reducing the Likelihood of Future Human Activities That Could Affect Geologic High-Level Waste Repositories*, BMI/ONWI-537, Office of Nuclear Waste Isolation, Battelle Memorial Institute, Columbus, OH, 1984, 37; and Percy H. Tannenbaum, *Communication Across 300 Generations: Deterring Human Interference With Waste Deposit Sites*, BMI/ONWI-535, prepared by Survey Research Center, University of California, Berkeley for Office of Nuclear Waste Isolation, Battelle Memorial Institute, Columbus, OH., 1984. All reports prepared for Battelle may be purchased from the National Technical Information Service, Springfield, VA.
 36. The Rosetta Stone is part of the collection of the British Museum, London, collection number AE 24.
 37. 'grok': Raymond, Eric S., compiler, *The New Hacker's Dictionary* 3rd ed, MIT Press: Grok: from Heinlein, Robert A., *Stranger in a Strange Land*, where it is a Martian word meaning literally 'to drink' and metaphorically 'to be one with' vt. 1. To understand, usually in a global sense. Connotes intimate and exhaustive knowledge.
 38. "If you cannot measure it, it is unimportant." Paraphrase of Lord Kelvin, who was William Thompson: p793 Beck, Emily, *Bartlett's Familiar Quotations*, 14th ed, John Bartlett, Boston, Little, Brown, 1968, *Popular Lectures and Addresses*, 1891-1894.
 39. Probabilistically not really there: Capra, Fritzof, *The Tao of Physics: An Explanation of the Parallels Between Modern Physics and Eastern Mysticism*, Shambala Publications, 1991.
 40. Goedel: You cannot prove something from within itself: Nagel, Ernest and James R. Newman, *New York, NYU Press*, 1958 p85. *The Heart of Goedel's Argument: On formally undecidable propositions of Principia Mathematica and related systems*.
 41. Goodrum, Charles A., *Treasures of the Library of Congress*, Abrams Books, 1980.
 42. Holding seven concepts at once: Stark, Werner, *The Magical Number Seven Plus or Minus Two, Some Limits on Our Capacity for Processing Information*, Journal: *Psychology Review*, v63, p 81-97, 1956, An original study of the limits of human short term memory.
 43. It is likely that the seven layers of the ISO/OSI protocol stack standard (e.g. TCP/IP - Transmission Control Protocol / Internet Protocol) proved to be the most efficient number of layers to manage the programming and coordination effort for the reason given above. The standard, ISO/OSI (International Standards Organization/Open Systems Interconnect), Voelcker, John, *Helping Computers Communicate*, IEEE Spectrum, 23, no. 3 (March 1986): 61-70. Also see, *Proceedings of IEEE* (December 1983).
 44. Blurring into perceived motion at more than 60 images per second: www.SMPTE.org (Society of Motion Picture and Television Engineers).
 45. Book design, the art of designing a book so that the pages can be flipped and perceived at up to 60 pages (images) per second: McLean, Ruari, *The Thames and Hudson Manual of Typography*, London, Thames and Hudson, 1980, Chapter 5: *Book Design* p120-147.
 46. Reloading our short term memory when our contextualization for interaction times out (We lose interest and our attention drifts): Doherty, Walter J. and Arvind J. Thadhani, *The Economic Value of Rapid Response Time*, White Plains, NY, IBM (GE20-0752-0, 11-82), 1982 www.IBM.com International Business Machines, Inc.
 47. The things that are being recorded change to match the record, even if the record is fictitious: *Hyperion Treatment Plant serving 4 million people in 10 cities at 12000 Vista Del mar, Playa Del Ray, CA 90291*.
 48. The relatively high rate of change from Middle English to Early Modern English and the relatively low rate of change from Early Modern English to Modern English, Chaucer to Shakespeare, and Shakespeare to today, the Folger Shakespeare Library: Baugh, Albert C., *A History of the English Language*, Appleton Century Croft, 1957 and McCrum, Robert, William Cran, and Robert McNeil, *A History of the English, The Story of English* (video), Viking, 1986.
 49. Confucius: p 298-306, Atiyah, Jeremy, *China, The Rough Guide*, Penguin Books, London, Rough Guides Ltd., 1997, Confucius' stress on education was a basis for the Chinese civil service exam that lasted until the 20th century. The Confucius family home, in Qu Fu, China, was occupied by family members continuously from 479 BC until 1940 AD. p280-285 Taylor, Chris, *China 6th ed.*, Hawthorne, Victoria, Australia: *The Lonely Planet Press*, 1996. The Confucius family archives survive in Qu Fu, China.
 50. The Powers of Ten: Morrison, Philip, and the Office of Charles and Ray Eames, *Powers of Ten*, San Francisco, Scientific American Press, 1982.
 51. Winkler, Jude, *The Good News About the Year 2000, Is it the End of the World?* Conventual Franciscan Friars, St. Anthony of Padua Province, Ellicott City, MD, 1995.
 52. Calendar, *Encyclopedia Judaica*, Jerusalem, Israel, Keter Publishing, 1971, v 5, p 45.
 53. 4D (four dimensional) simulacrum: p46-58, Clark, Arthur C., *The City and the Stars*, Harcourt, Brace & World, 1956.
 54. Digital Orthophotogrammetry (Aerial photographs pixel registered to a GIS vector database): www.ASPRS.org (American Society of Photogrammetry and Remote Sensing).

54. GPS (Global Positioning System) via satellite based triangulation: Brinker, Russell C. and Roy Minnick, *The Surveying Handbook*, 2 d ed, Chapman & Hall, 1995. Chapter 15 GPS, p334-382.

55. Wiley, E. O., *Phylogenetics: the Theory and practice of Phylogenetic Systematics*, John Wiley & Sons, 1981, p 1-20, Introduction. p 63-100, Araff, Rudolph and Thomas C. Kaufman, *The Developmental - Genetic Basis of Evolutionary Changes*, Bloomington, IN, Indiana U. Press, 1991. Chapter 3, *Morphological and Molecular Evolution in Embryos, Genes, and Evolution*. Chapter 3 shows Phylogenetic trees. The first books on phylogenetics described the theory and promise. Now, with genetic databases like the Human Genome Project, the molecular foundation for phylogenetic research is being established.

56. *The Human Genome Project*: Wingerson, Lois, *Mapping Our Genes, The Human Genome Project and the Future of Medicine*, Penguin books, 1990.

57. *Genealogy: The largest store of genealogical records in the world is held by the Church of Jesus Christ of Latter-day Saints*. www.LDS.org The LDS Church also has a world wide network of libraries of books on genealogy. The LDS Church is actively pursuing a program of genealogically linking all of the individuals in the genealogical records it holds. The LDS Church provides access to both the genealogical records and to books on genealogy.

58. *Monoculture/Biodiversity: Monoculture: farming large areas of land with genetically identical crops*. p77, Caerncross, *Costing the Earth*, Harvard Business School Press, 1991. *Biodiversity: the genetic diversity of life as it has evolved and as it is found when not disturbed*. p40-53, Blum, Elissa, *Making Biodiversity Conservation Profitable* in Orwen, Lewis and Tim Unwin eds, *Environmental Management*, Oxford, UK, Blackwell, 1997.

59. *Hypertext*: Nelson, Ted, *Literary Machines South Bend, IN.: The Distributors*, 1981. See also Stephen A. Weyer, *Searching for Information in a Dynamic Book Ph.D. diss.*, Stanford University, 1982, Nelson, Ted, *The Hypertext, Proceeding of the World Documentation Federation*, 1965, Nelson, Ted, *A file Structure for the Complex, the Changing and the Indeterminate*, *Proceedings of the Association for Computing Machinery*, 1965, Nelson, Ted, *A conceptual Framework for Man-Machine Everything*, *Proceedings of the National Computer Conference*, 1974, Nelson, Ted, *Computer Lib*, South Bend, IN, Nelson, Ted, *Selected Papers of Ted Nelson*, Ted Nelson Press, 1965-1977.

60. *Microfilm and document imaging: ANSI/ISO 3334-1979, Microcopying - ISO Test Chart No. 2-Description. The history of aperture cards*: MacKay, Neil, *The Hole in the Card, The Story of the Microfilm Aperture Card*, St. Paul, MN, 3M (Minnesota Mining & Manufacturing Co.), 1966. Rewritten as: Connors, Richard J. and William M. Amundson, *Microfilm: Active and Vital*, St. Paul, MN, 3M, 1975. Aperture cards were first used to marshal the vacation photos of the 1920's and 1930's to provide information and images in support of the Allied invasion of Europe in World War II. This collection represents one of the first comprehensive organizations of personal information on the part of a government agency. www.AIIM.org (Association for Information and Image Management) AIIM is a source of information on Document Management, Document Imaging, OCR (Optical Character Recognition) and related technologies. AIIM administers the CDIA program. (Certified Document Imaging Architect).

61. ISO 9000 (International Standards Organization quality standard) (ISO 14000 is the equivalent for the environment) Both ISO 9000 and ISO 14000 depend on record keeping for implementation.: Randall, Richard C., *Randall's Practical Guide to ISO 9000 Implementation, Registration, and Beyond*, Addison Wesley, 1995. Brumm, Eugenia K., *Managing Records for ISO 9000*, Milwaukee, WI, 1995. Canter, Rob, *ISO 9000 Answer Book*, Essay Junction, VT, Oliver Wight Publications, 1994. ISO 14000: Kuhre, W. Lee, *ISO 14010's Environmental Auditing*, Upper Saddle River, NJ, Prentice Hall, 1996. www.ASQC.org (American Society for Quality Control).

62. *Records Management*: Robek, Mary F., *Information and Records Management*, Glencoe - McGraw Hill, 1995. www.ARMA.org (Association for Records Management and Administration), the ICRM (Institute of Certified Records Managers) is an affiliate of ARMA.

63. *Passing the Baton Over the Wall, The Electron Story*, first spoken orally in 1981 in a job interview. Is important to our ability as people to build our individual worlds in which we each deal with the universe.

64. *Federal cylinder project: A loss of provenance has already occurred for an entire class of machine recorded information. The description of a large part of the Library of Congress' early cylinder field recording, documenting the traditional culture of America and the world, has been lost. ". . . the cylinder collections had not always been carefully organized and cataloged in the first place, and their subsequent orphan-like history, shuffled from boxes under beds to boxes in a variety of institutional basements, succeeded in dispersing much of the information about the cylinders to the winds. To the preservation challenge then, was added a major organizing and cataloging challenge". Also, the sound on the cylinders was at great risk of being lost. Many cylinders could be played only once before they disintegrated. Now that the sound had been transferred from the disks, there is a possibility of saving it forever.* Erika Brady, *The Federal Cylinder Project, Introduction and Inventory*, vol. 1, Library of Congress (Washington, DC: Government Printing Office, 1984): vii.

Updates and More Detailed Descriptions

When using the information in this article, please check the website www.ArchiveBuilders.com for updates. The version number of this article is just before the page number below. The website also has articles that provide more details on some of the terms and concepts in this article.

Comments

Please let us know how you like this paper, or if you had any questions. What would you like to see in the future? For more, and the most recent version of this article, please visit our web site at www.ArchiveBuilders.com. Please send your comments to SteveGilheany@ArchiveBuilders.com. +1 310-937-7000. Also, please let us know where you saw this article.

Acknowledgements

Reprinted from *Archive Planning*, Volume 2, number 7, 1998, *Archive Builders'* analysis newsletter for document management.

See www.ArchiveBuilders.com.

All trademarks are the property of their respective holders.

Note to Editors

Paper 22010v52

We will continue to update these articles as we get comments. Please contact us for the most current version before you publish. Also, please request permission to publish the article. Permission will be given freely for most purposes.

Steve Gilheany
Archive Builders
1209 Manhattan Ave., PMB C-14
Manhattan Beach, CA 90266
Tel: +1 310-937-7000 Fax: +1 310-937-7001
SteveGilheany@ArchiveBuilders.com

Bio

Steve Gilheany, BA in Computer Science, MBA, MLS Specialization in Information Science, CDIA (Certified Document Imaging System Architect), AIIM Maser, and AIIM Laureate, of Information Technologies, CRM (Certified Records Manager, ARMA) has seventeen years experience in document imaging and is a Sr. Systems Engineer at Archive Builders.

Author

Steve Gilheany is a Sr. Systems Engineer at Archive Builders. He has worked in digital document management and document imaging for seventeen years.

His experience in the application of document management and document imaging in industry includes: aerospace, banking, manufacturing, natural resources, petroleum refining, transportation, energy, federal, state, and local government, civil engineering, utilities, entertainment, commercial records centers, archives, non-profit development, education, and administrative, engineering, production, legal, and medical records management. At the same time, he has worked in product management for hypertext, for windows based user interface systems, for computer displays, for engineering drawing, letter size, microform, and color scanning, and for xerographic, photographic, newspaper, engineering drawing, and color printing.

In addition, he has nine years of experience in data center operations and database and computer communications systems design, programming, testing, and software configuration management. He has an MLS Specialization in Information Science and an MBA with a concentration in Computer and Information Systems from UCLA, a California Adult Education teaching credential, and a BA in Computer Science from the University of Wisconsin at Madison. His industry certifications include: the CDIA (Certified Document Imaging System Architect) and the AIIM Master, and AIIM Laureate, of Information Technologies (from AIIM International, the Association of Information and Image Management, www.AIIM.org), and the CRM (Certified Records Manager) (from the ICRM, the Institute of Certified Records Managers, an affiliate of ARMA International, the Association of Records Managers and Administrators, www.ARMA.org).

This Paper Presented at:

DLA98, Digital Libraries Asia 98, March 17 to 20, 1998 in Singapore.

Contact:

SteveGilheany@ArchiveBuilders.com
Tel: +1 310-937-7000 Fax: +1 310-937-7001

For more information, courses, and papers:

<http://www.ArchiveBuilders.com>