Preservation in the digital age

NANCY BIRK[1]

Abstract

In this paper, I will examine Preservation in the Digital Age from two different viewpoints, two sides of a coin if you will. The first side will be an examination of the expanded opportunities offered to the field of preservation/conservation utilizing the new technologies of digitization and imaging. The flip side of that coin and my second approach will be to look at the preservation technology itself and what issues it raises about its own preservation needs. My perspective on this topic is a personal one based primarily on my own experience in building a virtual archive.

There is no long term preservation standard for digital materials. What is needed is a change in thinking, in our perception of what is truly archival. Is it the physical disk, whether that physical disk is a hard disk, an optical disk, a CD-ROM, or is it the electronic byte of information that merely changes format? We have come to realize that it is the latter - an electronic byte of information that merely changes format. This means we must be diligent in our efforts to track and maintain current formats for all our archival digitized information, at least until a long term preservation standard emerges and I say that "until" without much confidence because the reality is constant evolution in this area.

Pertinent issues examined include:
- Technological obsolescence of software and hardware, of storage,
Retrieval and access tools
- Legal ramifications particularly copyright
- Lack of standards for digital preservation
-

[1]Associate Professor, Kent State University Libraries & Media Services, Kent, OH 44242, USA, nbirk@lms.kent.edu.

- Ease of Enhancement/Change in imaging
- Selection — how do we choose what to keep and what to discard from electronic records.
- Selection ~ how do we choose what we will digitize from our past, we can't do it all

We have a moral obligation to provide for the preservation of the new digital libraries that we are all so eager to create. We cannot focus solely on their creation and access. We must also provide for this information to endure into the generations to come. We cannot afford to ignore preservation in the digital age.

Keywords: digital preservation, digitization, imaging, electronic records, scrapbook preservation.

How many of you have ever lost a computer file?
Had a disk crash?
Erased something you wished you hadn't?
Or simply forgot where you had filed something electronically?

I expect that all of us have had a least one if not all of these things happen to us.

In this paper, I will examine Preservation in the Digital Age from two different viewpoints, two sides of a coin if you will. The first side will be an examination of the expanded opportunities offered to the field of preservation/conservation utilizing the new technologies of digitization and imaging. The flip side of that coin and my second approach will be to look at the preservation technology itself and what issues it raises about its own preservation needs. My perspective on this topic is a personal one based primarily on my own experience in building a virtual archive.

But first, I thought I would begin by telling you a little about myself and my connections to Greece and what brings me here today.

I first came to Greece in August of 1985 and remained until June of 1986, working in the English department library of Aristotle University as part of a new exchange program established between my university, Kent State University and Aristotle University. Since then, I have returned to Greece nine times, sometimes to continue work with Greek libraries but also on holiday to visit many friends and colleagues, as Greece has become a second home to me.

Concerning preservation and digitization, both of which I am going to talk about today, I became involved in preservation issues early in my career in the early 80s. As University Archivist and Associate Curator of Special Collections, my work with rare and unique documents is inter-twined with preservation.

My experience with digitization began in 1986 with work with an optical character reader ~ a Kurzweil scanner that was one of only several dozen produced at a then cost of $45,000 and which is the size of a copy machine. This particular scanner has artificial intelligence and can be trained to ready any typeface. It was critical to our work at Kent State at the Center for Conrad Studies. There we are producing a textual edition of the works of Joseph Conrad, being published by the Cambridge University Press. That was my entree into digitizing and I was hooked.

Then in 1994, we were preparing for the 25th annual commemoration of the shootings of protesting students on the Kent campus by National Guardsmen. Along with Mike Kreyche, we began our library's web site with an online exhibit' created by scanning original documents and images. This was the beginning of what is becoming a virtual archive of primary source materials on a web site dedicated to telling the story of this troubled time in the history of the United States.

Our web site has also expanded since then to include archival inventories to other unique collections, many of which include images of documents and photos, some **solving** preservation problems and some **creating** them.

Then there is the new dizzying world of electronic records.

Think about the varying storage formats that have evolved since the advent of the personal computer in the early 1980's.

- Cassette tape
- Floppy disks
- Hard disk drives
- 3.5 inch disks
- CD-ROM
■ Removable hard drives like zip and jazz drives

'http://www. library, kent.edu/exhibits/4may95/index. Html

Then there is the mainframe computer with:
• Punched cards
• Hard disk drives
• Tape in all of its various evolutions, from reels to cartridges
• Optical disks
Now think about all the programs we use to manipulate the data stored on these formats:
• Databases
• Spreadsheets
• Word processing programs - pick iust one, WordStar, WordPerfect or Microsoft's Word, and then think back 10 years or so to its inception. It's not just a concern with differing and outdated programs but software changes within programs.

Imagine an archive collecting an author's work who sent in a floppy disk instead of a manuscript. It would be impossible and simply ludicrous for an archive to try to save the information on that disk in its "original format". Not only would you have to have the software (and someone who understood how to use it) but you would need the hardware current at the time.

So what can an archive do?

Must they ultimately store everything produced electronically on paper?

What about records that only exist in electronic form? Particularly records in database form?

It's overwhelming, it's dizzying, and it is a crisis.

Alternatively, we can turn to digitization of documents and images and particularly brittle materials as a means of preservation itself.

But we must be cognizant that we are creating something that is transient by its very nature. We have come to the realization that reformatting is our future, but we mustn't lose anything along the way, because if we do, it may be impossible to go back and pick it up again. It truly is the best of times and the worst of times.

 So let's look at it first as the best of times and see what some of the exciting areas are out there on the horizon for preservation.

For many years, there was great resistance to digital formats in preservation. This stemmed from the very conservative view of preservation/conservation professionals, who felt that a preservation for-mat MUST meet the test of time and be able to last for centuries. Microfilm was (and still is for many) the long-term preservation medium in which to preserve original and unique materials.

There is no long term preservation standard for digital materials. What is needed is a change in thinking, in our perception of what is truly archival. Is it the physical disk, whether that physical disk is a hard disk, an optical disk, a CD-ROM, or is it the electronic byte of information that merely changes format? We have come to realize that it is the latter — an electronic byte of information that merely changes format. This means we must be diligent in our efforts to track and maintain current formats for all our archival digitized information, at least until a long term preservation standard emerges and I say that "until" without much confidence because the reality is constant evolution in this area.

**What do digital formats offer in terms of access?**

Obviously, the Internet has given access to digitized materials a global perspective, and while that is a very exciting reason for undertaking such projects there are other reasons to undertake electronic formats for archival materials.

Within our collection of materials at Kent State documenting campus unrest in the late 60s and early 70s is a vast collection of flyers in mimeographed format ~ you remember those from the days before photocopy machines with that dizzying smell when producing copies from a mimeograph master? If you have such papers in your libraries they are most likely similar to many of ours ~ the purple ink is fading fast to the place where the inky text is almost invisible. Similar to this are the early faxes which came through on that shiny paper but whose inking was faint and now is growing fainter. But after such papers have been scanned, they can very easily be manipulated with simple programs like Adobe Photoshop or Aldus Photos Tyler and then color enhanced so that the text reappears.

Such programs also have a zoom feature that allows one to enlarge a digitized image or a portion of a digitized image. In the images from the scrapbook that we will look at in a few minutes, the zoom feature

allowed a Romanian historian to identify the faces in photographs. Zooming in on images of manuscript allows textual scholars to examine more closely the handwriting of an author and that most difficult part of a manuscript to discern -- to read the text below a strikeover.

Additional access is gained through full-text searching. Optical character recognition software converts a scanned image of text to ASCII characters searchable and downloadable to word processors, giving scholars and researchers quick access to full text searching.
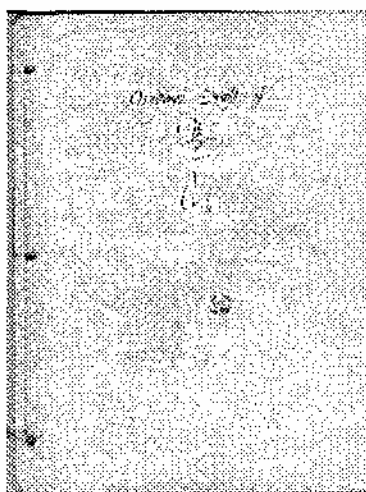
This Gary Snyder manuscript[2] shows an extension of this digital approach to other types of rare material. The manuscript was "tied together in the Chinese fashion". Anyone who tried to open the manuscript book would cause damage to the binding as it would crease the paper cover and also all the subsequent pages, eventually leading to their breaking at the folds. We decided that, for the preservation of the documents, we would need to untie and disbind the booklet. Even though the poet, Gary Snyder, had felt that the tying of this booklet was important to its content, we had no real choice. This of course would destroy the original format (something always avoided in archival work) but it had to be done. We decided to "preserve" this Chinese tying by digitizing it and so snapped a photo with a digital camera and captured the Chinese knots, which now only exist digitally.

Increased access brings its own problems. In our experience, it has increased demand for answers. While we have a great deal of content available concerning May 4, 1970 and the shootings at Kent State, including a chronology with hypertext links to images, an online exhibit, and all the finding aids to the collections, researchers demand more. They ask critical questions, which we cannot answer, questions like, "why did this happen at Kent State?" "Was justice ever served?" "Who was responsible, the National Guard or the students?" We try to lead them back to the documents, so that they can answer these questions for

http://www.library.kent.edu/speccoll/literature/poetry/snyder.html

themselves. We give them SOME images, they want them ALL. We recently put the archival inventory up on the Web for the SDS, the Students for a Democratic Society, papers[3]. One of the items listed was a membership list for 1969. When a patron requested a photocopy of that be sent to him, we went further -- we scanned the list and put it up on the Web, hopefully addressing the needs of other researchers. We hope to continue to add to our virtual archives in this way ~ prioritizing what will be scanned next on an as needed basis.

Digitization of Scrapbooks

One of the most difficult preservation issues facing archives and special collections concerns scrapbooks. The typical scrapbook is a unique collection of memorabilia that includes clippings, photos, pressed flowers, tickets, invitations, and even small objects. They are usually attached using the worst kinds of adhesives from rubber cement to various forms of "magic" tape. They are also usually attached to extremely acidic papers. The paper is disintegrating, the adhesive no longer adheres in many cases and the scrapbook as an object is usually in various states of disrepair. Almost everyone keeps scrapbooks, the famous and not so famous.

In many cases such scrapbooks must be disassembled to protect their valuable contents from highly acidic scrapbook housings. Other times, because of the brittle paper and bad adhesives used, they literally disassemble themselves and photos and clippings fall off and out of the pages when they are used.

Scrapbooks which must be disassembled for their own self-preservation can have their format preserved by means of digital imaging. The scrap-book then exists in its original format only in digital form, with the photographs and memorabilia removed and placed into an acid free environment. Patrons can view the digital version without causing wear and tear on the original.

We have, in our collection at Kent State, the papers of Queen Marie of Romania[4]. Within these papers are several scrapbooks. Let me begin by giving you some brief biographical information about Queen Marie and

---

[3]http://www.library.kent.edu/exhibits/4may95/boxl07/l 07.html [4]http://www.libra ry.kent.edu/speccoll/women/queen.html

how the scrapbook came into our possession and then I will return to the specific problems of the scrapbook.

Queen Marie was born Marie Alexandra Victoria on October 29, 1875 in Kent, England. She was the daughter of Alfred, Duke of Edinburgh and the former Grand Duchess Marie of Russia. She was the granddaughter of both Queen Victoria of England and Tsar Alexander II of Russia. The English Royal Family planned for her to marry her first cousin Prince George, who later became George V of England. Instead, in 1893, her mother decided that she, at the age of 17, would marry Crown Prince Ferdinand, the heir-apparent to the Romanian throne. Marie, who became Queen of the Romanians in 1914, was popular among the Romanian public and served as a nurse during World War I. Queen Marie had six children, of which Princess Ileana was fifth, being born in 1909. The Queen died in 1938.

Princess Ileana and her husband, Anton Habsburg, Archduke of Austria, also had six children with Stefan, being the oldest. Princess Ileana moved to the United States after her deportation from the communist state of Romania. Divorcing Anton, she moved her family to Newton, Massachusetts and began a writing career and made several public appearances to support the family. Much of her writing reflected on her life in Romania and her mother. In her later years Ileana entered an Orthodox convent in Pennsylvania and was known as Mother Alexandra. She died in 1991.

How did such a collection come to be at Kent State?

The collection was donated in several installments and at varying times by Queen Marie herself, her daughter Princess Ileana, Stefan Habsburg, and Roy Baker Harris, an American friend of Queen Marie's. Both the Queen and Mr. Harris were searching for a suitable home for her papers and did a search to see where most Romanian immigrants had settled within the U.S. They found the largest concentration of Romanians was in Northeast Ohio, and that Kent State was the only university in the area which taught Romanian studies. So the collection came to Kent.

The collection itself consists of 18 cubic ft., including correspondence, photographs, newspaper clippings and personal belongings. Other sorts of memorabilia include paintings by Queen Marie, Romanian coins, and even 2 large marble commemorative stones presented to the Royal

Family on   a visit to the U.S. in 1926. There are costumes donated by the family but these are part of Kent State's Fashion Museum.

There are several family scrapbooks and photo albums in the Queen Marie Collection. This Photoalbum[5] has a beautifully embroidered cover, embroidered by Queen Marie herself.

The pages of the scrapbook were composed of an acidic black paper, not the stiff paper usually found in scrapbooks but a rather flimsy, flexible paper. The paper had begun to deteriorate, flaking at the edges. The pages were tied together with a heavy thread and were beginning to break at the inner fold. The adhesive holding the photos to the scrap-book pages had lost its stickiness and the photos were beginning to become detached from their pages and the identifying remarks written by Queen Marie. This treasure could not be safely handled by any researcher without risking a loss of information.

We wanted to show the Habsburg family what the potential was both for the preservation of their materials and for access to it by scholars around the world. We scanned each page of this scrapbook to minimize use of the actual item. The electronic site provides visual access to the album in its entirety. In this case we did not disband the scrapbook, but many of the other photo albums in the Queen Marie Collection have been dis-bound and the digital images remain the only evidence of the placement of photos on the page.

What are some of the other advantages to digitization in preservation?

• Help for the brittle paper crisis. It is estimated that anywhere from 1/4 to 1/2 of the books and materials in U.S. research libraries are brittle. Long standing preservation projects to deal with this crisis have focused on major microfilming. Enhancement capabilities make digitization desirable for brittle papers.
• Microfilm can now be digitized as well
• Use of a digital camera enables one to easily document history and events as they happen, particularly important in an archival setting
• With this same digital camera one can preserve quickly disintegrating and transient formats

---

[5]http://www.libra ry.keiit.edu/speccoll/women/marie.html

For example, one of our collections is from the Great Lakes Theater Festival[6] based in Cleveland, Ohio. We are their archival repository and their collection greatly enhances our primary sources in theater. You probably have never heard of them but two actors that I am sure you know made their start with this theater group, John Lithgow and Tom Hanks. As part of the educational program of the Great Lakes Theater Festival, they create lobby posters for every production. These posters explain historical information about the production, the actors, and the playwright. They are oversized and mounted on foam-core. The foam-core is the problem. It is a rapidly disintegrating medium that self-destructs and within 10 to 15 years will be crumbs. Add to that the potential storage difficulty of such large pieces and the problem is even more cumbersome. We are now capturing these posters with a digital camera and then discarding the original.

This discarding of the original is not the normal procedure in an archive however.

I've often been asked what an archivist does with the original once it has been digitized. Because of the intrinsic value of the original, we don't discard it once digitized.

What are Some of the Issues to Consider in Beginning an Imaging Project:

• **Time** required

• **Minimal Equipment** Needs

    Hardware
• PC platform, Pentium processor w/ multi-gigabyte hard drive, network connection
• Flatbed scanner - 11 χ 17 surface area at least
• High resolution 17" monitor
• Color Printer
• HP LaserJet Printer
• Digital Camera

      Software
• Image capture
• Image processing and manipulation
• Web Interface--HTML editor

"http://www.library.kent.edu/speccoll/theater/gltfhist.html

• Space necessary to conduct such a project

- Final Queen Marie scrapbook project consisted of 72   mb and has 429 files/links/images
- But that 72 mg. is the final product utilizing GIF   images which are images compressed for the World Wide Web

  Initial scanning produces a high quality TIFF image. Uncompressed TIFF images can be 10 times larger than the final compressed GIF. We equivocated with the idea of saving these TIFF's as archival images but quickly had to abandon that idea - we simply did not have the storage space.
  The compression comes mostly with color images where you take them from millions of colors down to an indexed 256 bit color. 256 is more than most monitors can see anyway so to the naked eye when you compare the uncompressed TIFF with the compressed GIF, it is often difficult to see any change in the image anyway.

- Recommended resolution is 600 dpi, ours vary from 300 to 600, but 600 is about the highest achievable in a production mode
- We worked on either our local network drive or on the hard drive of the computer itself, depending upon where more space was
  available. When work on the image was completed, it was then moved to our Web server and linked to the HTML marked up page for the collection. That way we compressed each image as we went, saving space along the way.
  - **Archival backups**
  Our Web server is a Unix which is backed up weekly. We also have backed up special archival projects like this one onto an optical drive. We now have a jazz drive on which we do periodic backups of our entire web site.

There are Access Issues to Consider at the Point of Creation

- The larger the GIF image, the more time it takes users to download the image to their local web browser. Our images for this project are large, what is recommended for the Web are images of about 50 kb. Most of these are 350 kb. That means if you are using a modem for access to the Web, that it may take some time to download the images. I recently received an email message from a high school student who complained specifically about this and suggested that we get

rid of the black borders around the pictures and load them separately. I explained that this would normally be a better way to view images but that our goal was to duplicate the scrapbook page and that the placement on the page had value as well as Queen Marie's identifying text that accompanies many of the photos. In other words this was more than an access issue, it was a preservation issue and we wanted to meet both needs.

• **Design** issues

How do you arrange information on a page so that it is both attractive and informative? We usually choose content over glitz, but the issue of how to arrange the information in the scrapbook in an informative way is one of the most challenging.

• **Enhancement** issues

How far do you go in enhancing an image ~ when does enhancement go too far so that it enters the arena of change. We avoided "bettering" any image preferring to reproduce the image exactly.

One of the most famous May 4 photographs shows runaway Mary Vecchio leaning over the body of one of the slain students, Jeffrey Miller. The photograph was picked up over the wire and used in almost every major newspaper in the world. The student who took the photo, John Filo, won the Pulitzer Prize for his work. At the left is the photo as it appeared in 1970.

This photo was discovered to have had some "touching up" done at one time that interfered with its archival integrity.

As you can see in this second photo which was published in Life magazine in 1995, the fence post behind the head of Mary Vecchio has disappeared. This is when enhancement becomes questionable and even unethical.

• **Copyright** issues

What a complex issue this one is -- copyright on the Internet is a raging area of concern. As of yet our laws don't address the Internet. I don't want to get too deep into it here — that's fodder for another presentation, but let me say that we do have the permission of the Habsburg family to place this scrapbook on the WWW.

But be warned, placing unpublished archival material on the Web means that anyone can download it and publish it. I was very pleased when I saw an issue of the ARL newsletter[7] addressing the issue of digitizing technologies and that they had used a photo from our Queen Marie scrapbook

But I didn't know they were going to use our image of Queen Marie until I saw it in print. This photo had never been published before. Obviously I am thrilled to see it as the example they selected to advertise one of their SPEC kits, especially one on preservation imaging, and a kit I recommend highly to you if you are considering undertaking a digital imaging project for preservation. The SPEC kit highlights 29 preservation imaging projects[8].

The Romanian scholar that I mentioned earlier was quite concerned when he learned that we were planning to mount this project on the Web. He realized quite correctly that his unpublished information would be available to any scholar with access to the Web. That meant that he no longer had exclusive use of these photographs. But he was completely won over when last spring, Stefan Habsburg — the grandson of Queen Marie and the son of Princess Ileana — paid us a visit and we showed him the site and how we were preserving the scrapbook. He was particularly pleased when he realized he could have access to it from his own home computer in Michigan and could alert other family members in the U.S. and Europe to its new expanded Web life.

• **Standards** issues

There are as yet no firm standards for preserving digital formats. Quickly emerging are text based versions of SGML with its various DTD's or straight ASCII text. What about images? Will an uncompressed TIFF be

---

'"SPEC Addresses Today's Challenges", ARL: A Bimonthly Newsletter of Research Library Issues and Actions. August 1996: 10.

"SPEC Kit 214. Digitizing Technologies for Preservation. Washington D.C.: ARL, 1996.

best or is the JPG for the Web good enough, and what resolution will emerge as the standard? The Commission on Preservation and Access along with the Research Libraries Group have been examining this issue and their Task Force on Archiving of Digital Information has been charged with putting forward a standard. In their initial report published in 1996", they espouse migration as the essential function of digital archives, that is, a constant reformatting and "refreshing" to the newest stable formats.

I will turn now from those materials which began their lives in traditional formats but were subsequently converted to digital format to those documents which began their lives in digital form, electronic records. Both of course share the problems mentioned earlier:

• Quickly obsolete formats for storage, retrieval and access
• Continuous innovation in both software and hardware
• Absence of digital preservation standards

But electronic records add another area of concern, an area all too familiar to librarians, and that is in selection. Archivists, probably more than librarians, have seen the shift in the past decade to records they collect once produced on paper but now only in digital form. Do we save everything? Some might argue that it may, ultimately, be easier to do so, that the process of selection, or actually deselection might be too costly. But we must remember that as librarians, we are also faced with the problems of access, and we all know that the more information there is out there in cyberspace, the more clogged our search engines become. Somehow we must find a solution to this issue.

To sum up, what are all the pertinent issues?
• Technological obsolescence of software and hardware, of storage, retrieval and access tools
• Legal ramifications particularly copyright
• Lack of standards for digital preservation
• Ease of Enhancement/Change in imaging
• Selection — how do we choose what to keep and what to discard from electronic records

---

"Preserving Digital Information: Final Report and Recommendations. Commission on Preservation
and Access, Task Force on Archiving Digital Information, May 20, 1996. URL: http://www.rlg.org/ArchTF/index.html
PDF: http://www.ifla.Org//documents/libraries/net/tfadi-fr.pdf

• Selection — how do we choose what we will digitize from our past, we can't do it all

Let me conclude by saying that we have a moral obligation to provide for the preservation of these new digital libraries that we are all so eager to create. We cannot focus solely on their creation and access. We must also provide for this information to endure into the generations to come. We cannot afford to ignore preservation in the digital age.