A Thesis Presented to

The Faculty of Alfred University

Technological Advances of Imaging As Applied to Library Archives

Bernadette D. Youells

In Partial Fulfillment of
the Requirements for
The Alfred University Honors Program
May 8, 1995

Under the Supervision of:

Committee:

Laurie McFadden

Toni Olshan

Rima O'Connor

Technological Advances of Imaging as Applies to Library Archives
by Bernadette D. Youells

Thesis Outline

Introduction

- I. Basic information
  - A. Definitions
  - B. Hardware
    1. General
    2. Technical scanner information
    3. General scanner information
    4. Hand-held vs. Page
    5. Storage
  - C. Software
    1. General
    2. OCR (Optical Character Recognition)
    3. Retrieval
  - D. History

- II. Applications
  - A. vs. microfilm and photocopy
  - B. Access
  - C. Future

- III. Popular Hardware and Software
  - A. General cost
  - B. Hardware
  - C. Software
  - D. Turn-key, Multi-industry systems
  - E. Networking
  - F. Copied pages from Cornell

- IV. Other programs

- V. Problems
  - A. Standards
  - B. Memory
  - C. OCR
  - D. Hardware
  - E. Fragile storage

Conclusion

An Executive Summary

Main Recommendation

This report recommends that digital imaging be utilized as a method of conservation, preservation, and storage for the items in Herrick Memorial Library's Special Collections. The suggested products are the HP ScanJet IIIp($600), used in conjunction with either, Calera's WordScan($1000), OmniPage($700), or Caere's PageKeeper($600). A collaborative effort with Scholes Library and their Excalibur turn-key system should be considered, as well.

The Problem

Currently Special Collections houses old and rare books and photographs that, over the years have begun to fall apart. Unless something is done, these items will either be completely destroyed, or their use and handling will have to be restricted or prohibited.

What this report Covers

In this report I discuss the advantages of digital imaging, its applications, and how much it costs. I also review the necessary hardware and software needed to implement a digital imaging program.

The advantages of using digital imaging as a means of storage and retrieval are conservation, preservation, and the increased accessibility of rare and disintegrating items.

The fast paced advancement of computer technologies such as imaging have changed the ways in which we communicate, work, and do leisure activities. These technologies have changed so fast that it has become nearly impossible to keep oneself updated on all that is new or improved. Scanning, or digital imaging is one of these technologies that has become available, yet has not been quickly integrated into everyday knowledge. Although digital imaging is being utilized by companies to cut down on their overwhelming amounts of paper, the advantages of it being used in a library setting has just begun to be studied. Other areas of computer science, such as on-line cataloging, have now successfully been put to work in libraries, opening the door for other projects, such as imaging. At Alfred University's Herrick Memorial Library, a growing curiosity has been circulating through the preservation, conservation, and archives personnel about digital imaging and what it could do for their library. Other libraries as well must be looking at the same things and asking the same questions that they have at Herrick Library: What is imaging? What could it do for us? What special staff and equipment is needed? And most importantly, what is it going to cost? During my study of digital imaging I have investigated all of these questions and attempted to take an educated, logical approach to them and the answers that I have found. I have explored what imaging is, its history, and how it works. I have looked at the technical as well as practical side of imaging, including the hardware and software involved and its applications. I also sought out the best and most popular systems in use today, and the libraries and companies using them. During this study I have learned, above all else that there is an infinite amount of information out there about technological advancements that is yet to be learned.

Discussing computer-related subjects can sometimes be difficult because of complex the vocabulary that is used, therefore I felt it was essential to give some key definitions:

Digital imaging has been described in many ways and words, such as, "a group of technologies which combine to allow images to be created, stored electronically, manipulated, transmitted, and accessed,"(Boss 644), "a process of converting printed material to electronic form"(Hawkins,1991 4), or "the method of converting images...to digital code for information processing or storage"(Boss 710). What digital imaging is, in layman's terms, is the transferring of a document, book page, memo, etc, from paper to computer file using appropriate computer hardware and software.

ASCII, or **American Standard Code for Information Interchange**, is the standard coding method used to change text to a digital form that all computers can read and understand (Boss 709).

Binary Scan is "a scan that assigns each pixel or dot an attribute of either black or white(New York State...21).

Bit is "the smallest unit of data a computer can understand" (New York State...21).

Bit map is the "representation of colors or graphics by individual pixels" (New York State...21).

Byte is a unit of storage that usually equals one character. It also equals eight bits.

DPI, or Dots Per Inch is "the common term used to describe the 5~number of pixels per linear inch which make up the digital image" (New York State...22)

Pixel is a dot. When a series of these dots are put together in a particular pattern, it becomes a letter, or picture.

These are words that come up frequently when speaking of scanning, or any computer applications. Unfortunately, they become so familiar to many people that they forget that not everyone is so well versed in the computer lingo and laypeople are sometimes left feeling a bit overwhelmed. Digital imaging is an interesting subject of study for everyone, not just those that deal with computers on complex levels.

Digital imaging is not something that a company or library can say "OK, let's get started" and just do. Digital imaging requires specific hardware that is compatible with the system it is going to be used for, as well as, the context in which it is going to be used. In general, "the hardware translates the relative brightness and location of each point into digital data..."(Greh 20) More specifically, a digital imaging set up will consist of a scanner, an IBM or compatible PC, or Macintosh computer, and storage hardware and medium.

A scanner is an input device that reads the data and sends it to the computer. Technically, a scanner is an intricate and complicated device.

> "The document being scanned is first illuminated by a fluorescent or incandescent light source. The reflected light goes through a lens and onto a photosensor called a CCD (charge-coupled device). The CCD...contains a row of tightly packed light sensors; the output voltage of each light sensor is proportional to the light striking it. A single scan line from the document is focused onto the CCD, and the output voltage of each light sensor is converted to a digital value (a binary bit pattern) by an ADC (A/D Converter). A bi-level scanner actually uses a simple voltage comparator instead of an ADC and outputs either a black (0) or white (1) level for each light sensor. A gray-scale scanner will output a value that depends on the number of gray levels supported."(Alford 349)

In other words,

> "it bounces a beam of light off the surface of the original image, reads the intensity of each light reflection and sends that information to the computer where software can interpret the data and convert it into visible form"(Greh 20).

Although this may sound like a complicated tangle of unnecessary information, it is sometimes important to know how something works, not just that it does.

Although scanners differ between manufacturers, there are some basic commonalities found in all scanners. Even though it is said that scanners "read" data, no scanner actually does. The information obtained by a scanner is a photo-like image of the material. Because of this, a scanned image alone cannot be searched. There are three different types scanners: color scanners, and two kinds of black and white scanners. Gray scale scanners use "the spectrum, or range, of shades or black" of an image, and can support sixteen, sixty-four, or 256 shades of grey, depending on whether they are four-, six-, or eight-bit scanners(New York State...22; Alford 347). On the other hand, halftone scanners produce "a graphic, usually created from a photograph, in which dots are used to represent continuous tones"(New York State...23). The difference between gray scale and halftone are that halftone uses more shades so that there is a gradual instead of defined change in shades. Scanners are categorized as one or the other on the basis of resolution. Eight-bit scanners, for instance, are halftone and work well with continuous-tone images because they are "able to discern roughly 256 levels of gray"(Heid 292). To compare scanner resolutions to something more tangible, all that is needed is knowledge of other types of output devices' resolutions. Scanners have resolutions comparable to fax machines(200 dpi), high-quality photocopiers(400 dpi), and laser printers(600 dpi)(Boss 645). Besides color or black and white, and halftone or grayscale, there are still other things to consider when looking at scanners.

There are two different types of outward appearances to regard when thinking about scanners. There are hand-held scanners and there are array, page, or desk-top, scanners. The desk-top scanners then break up into three mini-categories: flat-bed, over-head, and sheet-fed. Hand-held scanners are somewhat self-explanatory, and page scanners work similar to a copy machine.

> "To scan an image with a hand-held scanner, you drag it across the image you want to scan. Typical hand-held scanners use a four-inch scan width...page scanners...are capable of scanning an entire eight-and-one-half by eleven-inch or eight-and-one-half by fourteen-inch page" (Alford 348).

With page scanners,

> "the scanner mechanism is a line of solid-state sensors spanning the paper's width. Either the sensor or the page moves to scan the length of the image, a process that takes four to thirty seconds" (Lu 185).

The three mini-categories of page scanners correspond to familiar pieces of equipment. Flat-beds most resemble the typical office copier, overhead scanners, as the name suggests, look like overhead projectors, and sheet-fed scanners favor the typical fax machine (Alford 348). Choosing a scanner, however, may not be as difficult as it seems, if the particular situation is considered carefully. For a library setting it is a simple process of elimination. Hand-held scanners may be more affordable, but this fact is outweighed by the fact that the scanning width is only about four inches wide (Greh 20). Another disadvantage of the hand-held scanners is that they can snag pages and cause tears in the materials. Because they are run by the human hand, they are also prone to unsteadiness, which causes accuracy and speed to decrease. Sheet-fed scanners can also be quite easily eliminated if the project will be working with books that need to be preserved. To use a sheet-fed scanner, the material being scanned must be in sheets. Books would first have to be completely disassembled and essentially destroyed before being able to pass through the sheet-fed scanner. For use in a library, especially archival, these drawbacks are too much of a risk.That leaves flat-beds and over-heads, but from what I have seen, over-heads are rare, whereas, flat-beds seem to be the scanner of choice.

Another consideration that must be made when deciding on an imaging system is storage. The amount of storage space needed when using digital imaging is phenomenal. "Grey-scale images require more than one bit per dot..." (Hawkins,1991 4). "Twenty pages, twenty images, twenty megabytes" (Griffith 20). Data compression technologies are making images smaller and easier to store, "twenty pages or images per megabyte" (Griffith 20), but storage remains an important issue. Optical disc storage is the medium in which, currently, the most information can be stored in the smallest amount of physical space. The most common type of storage is the WORM, or write-once-read-many, optical disc. WORM is "the optical medium that accommodates the permanent and unalterable storage of data" by "burning permanent bubbles or pits onto the medium surface with a laser" (Boss 713; Kalstrom 139). WORM discs are so reliable because neither computer crashes, stray magnetic fields, nor environmental factors can affect the information (Kalstrom 139). Aside from being extremely fast, another helpful plus for optical storage is the optical jukebox. An optical jukebox can hold many optical discs which it can alternate as needed. The concept can be somewhat compared to a multi-CD music player. Again, there is not much deciding to be done where storage is concerned.

Software is the next important consideration in exploring digital imaging. "...text management is the key to the image data base, and most imaging systems would be incomplete without it" (Hawkins,1992 94). The main components of any imaging software program are OCR, or optical character recognition, and retrieval capabilities. OCR is not completely necessary, but it is extremely beneficial to have. What good would imaging do if there was no way to retrieve what has been scanned? "Image-based systems also enhance information security. Physical security is more easily assured for images than for paper. Establishing back-up files is easier..."(Moore 32). Having all of the sophisticated hardware available would be of no real use without the proper software to get the information into the system, organize it, and get it back when it is needed.

"...If one wishes to search the text of scanned documents, an additional step is necessary...optical character recognition [OCR]"(Alford 349). Because scanning is similar to taking a photograph of a document, the computer only sees the image, not the actual words, therefore, searching the text would have to be done by calling up the document and skim reading it with the human eye to find the data that is needed. "OCR utilizes sophisticated algorithms to analyze bit streams, making statistical judgements that certain patterns of pixels represent certain digits or characters" (Moore 30). One of the things to watch for when choosing an OCR package is whether or not it has omnifont capabilities. Omnifont is "the ability of an optical character reader to recognize any font without having to learn the type face in advance" (Boss 711). This is terribly important when dealing with a variety of different material types. If, however, the system will only be called upon to scan, for example, documents made by the same type of printer, typewriter, or other instrument with the similar letter appearance, then omnifont capabilities would be not be required. Most scanning jobs, however, will include the scanning of an array of materials and fonts. OCR and omnifont programs are not inexpensive, but without them, digital imaging will be more time consuming and less effective.

There are several different kinds of retrieval methods that can be included in software programs for data imaging. The type of search that is chosen depends completely on the retrieval needs and preferences of the establishment. There are four main types of searches: word, boolean, statistical, and concept. Word searching is fairly self-explanatory and simple to use. Boolean searches "involve the user performing AND, OR, or NOT logic on the results of word searches" (Hawkins,1992 93). "Statistical retrieval uses the frequencies and positions of words to derive a statistical picture of the document and to determine its subject" (Hawkins,1992 93). Concept retrieval is similar, but uses the subjects of the document, instead of specific words.

Now that we have gone over some of the basic information on data imaging, it would probably be helpful to have a look at its history. Although it seems that imaging and OCR are new and exciting discoveries, they have been around for much longer than one would think. Studies show that digital imaging products have been on the market for well over ten years, and that OCR technologies were first patented in 1912 (Conway 43; Mallen 9). Early in their history, scanners were less accurate and could only scan line art and black and white pictures (Alford, 347). Since then, knowledge and use of it has increased tremendously. This is mostly due to the fact that digital imaging systems have become more affordable in recent years as a result of the advancement of imaging technologies. Included in this is an increase in the storage capabilities with the perfection of the optical disc and greater data compression, faster high speed networks, and greater resolution monitors and scanners (Boss 644). Charge-coupled devices have boosted the resolution of scanned images, and very large scale integrated circuits have expanded processing power and memory capabilities in computers (Boss 644). This more recent history of data imaging has been quite extensive and impressive, despite its early slow start during the first seventy years of the twentieth century.

Many people hear about digital imaging and think it is interesting, but wonder what it can be used for. What are digital imaging's practical applications and why would it be worthwhile to implement such a system? Can it benefit us by being more efficient than our current practices? Businesses have been looking into digital imaging as an office tool to cut down on use and waste of paper. Libraries, on the other hand, are considering it as a preservation and conservation method. In both of these environments, digital imaging must be compared with the procedures already in use: microfilming and photocopying. "Digital technology combines the desirable qualities of photocopy reproduction with microfilm's ease of duplication and space saving compression" (New York State...2). Microfilm may equal imaging's storage, but imaging greatly surpasses it in ease of use; "the challenge of locating and using microfilm may offset its advantages for patrons" (Conway 42). Imaging also out-performs both microfilming and photocopying in multiple reproduction situations. "...a digital image can be reproduced over and over again with no resulting loss of quality" (New York State...2). Although quality and storage are important factors, so is cost. The cost of setting up and maintaining a digital imaging system may appear to be much higher than the alternatives, but in the long run it will be most comparable. Compared with microfilming, imaging is only slightly more expensive, and with compared photocopying it is estimated that the cost will be about twenty percent lower (Boss 646; Kenney 2). So, as an alternative to photocopying and microfilming, digital imaging seems to be an attractive application.

Another appealing application of digital imaging is the increased access to materials that it allows. This was explored briefly in regards to microfilming, but can be especially important when considering those material that are contained in library archives. As Nicholas Finke, assistant law professor at Cincinnati College of Law, said, "If it was important enough to put into an archival library in the first place, then it is important enough to digitize it" (Dunn 28). Imagine walking into an archival library looking for a book that you need to do a research paper. You ask for the book at the desk and while you are busy putting on gloves and a hair-net, the librarian disappears into the maze of old books for what seems like hours. You must practically sign your life away to get the book and still you are guarded carefully every second in which the book is in your possession. After an experience such as this, one may be discouraged from ever giving the archives another try. If digital imaging were implemented, access to rare books, even those previously forbidden, would expand exponentially. The ease and speed with which these books could be accessed would also be phenomenal. Patrons could search for items themselves on a computer workstation in Special Collections, or even from terminals across campus. He/She would need only to type in a few key words and the item would almost instantly appear on screen. Patrons would not only be getting the information, but instant gratification as well.

As with all new markets, one must continually look toward the future, and what it will do for this market, or, sometimes, what the market will do for the future. Digital imaging will not be a fad that will come and go in the blink of an eye. That it will, is of some concern to prospective buyers, but "...the imaging market is growing at a compound annual rate of thirty to fifty percent..."(Hawkins,1991 3). This growth is causing software and hardware costs to plummet, thereby, encouraging more people to buy, and causing costs to go down even further. The future of digital imaging looks so promising to writers and editors that several publications have already been created to accommodate the interest. Some include the journals *Advanced Imaging, Electronic Imaging, Imaging World, Computer Pictures, Imaging Magazine* and books, such as, Real World Scanning and Halftones, by Blatner and Roth. Digital imaging has also helped shape the future by making possible a networked digital library that will allow patrons to view books from workstations world-wide (Kenney 31). This "national electronic highway" would be available to schools, colleges and businesses, as well as, the general public (USC). Prospective buyers should not worry about imaging disappearing like a shooting star, instead, they should worry more about the imaging train pulling out of the station and leaving them behind.

It is difficult to pinpoint how much a general imaging system will cost, because of the great variety of manufacturers and products available. There are turn-key, or multi-industry systems, and then there are pieces of imaging systems that can be bought separately and pulled together to form a more customized system. In general, though, systems can cost less than $50,000 or over one million dollars (Conway 43). Broken down, the cost can be anywhere between five cents and one dollar per page (Griffith 21). In order to better assess the cost of an imaging system that suits specific needs, one needs to examine the assortment of available products, as well as their capabilities.

The first category of products that will be explored will be popular hardware. Since many systems begin with a standard PC, and the printer depends upon the particular needs of the user, the scanner is the most important piece of hardware to be chosen.

The ScanMan 256 is a popular and powerful scanner for small jobs. It works in a Windows 3.0 environment and comes with its own editing package, called Ansel, in addition to some other easy to use editing tools (Linderholm 133). Unfortunately, the ScanMan is a hand-held scanner and is subject to all the difficulties associated with that type of scanner.

The Fujitsu M3096G appears to be a promising scanner, but carries with it a fairly weighty list price. It is a high speed document scanner that can use 200, 300, or 400 dpi with speeds of 15-25 pages per minute. For a large, profitable company, this may be the perfect option, but I would not suggest it for any establishment on a budget.

The Canon IX-3010 is a more reasonably priced scanner with a great deal of perks, but it also requires a bit more time and effort to use, and to learn to use, than some similar products. This scanner affords great control and flexibility and comes with Caere's PageKeeper Portfolio document-management system and Image

Assistant image editor (Poor 39). The Canon can scan a five by seven photograph in thirty-six seconds at 300 dpi, 136 seconds at 600 dpi, and 424 seconds at 1200 dpi (Poor 39). This product sells for $500-600 (Poor 39). Compared to the HP ScanJet IIIp discussed later, this is a slow scanning time.

The ArtiScan, ScanMaker and XRS's Omni-Media are some other scanners that were reviewed, but none of them were highly recommended. The ArtScan was fine at normal resolution, but when it was slowed down to increase resolution it took more than twice as much time to scan the same sample (Hurlow 101). Neither the ScanMaker nor Omni-Media showed any greater or lower quality or speed when settings were changed (Hurlow 101). Because of the scarce, yet negative, information on these scanners, I would suggest not buying one of them unless more extensive research was done.

By far the most reviewed and revered image scanners were those made by Hewlett Packard. The HP ScanJet IIp, HP ScanJet IIcx, and HP ScanJet IIIp, were popular products with reviewers. The Iip

> "offers many ease-of-use features, while partner Caere adds OCR software, which accurately converts into word processing characters the text that HP's machine scans, making retyping unnecessary...costs $1,295(in 1992)..."(Mamis 128).

According to an HP pamphlet, the Iip has greyscale scanning, 1200 dpi enhanced resolution and 300 dpi optical resolution, while the Iicx can be used for color or greyscale scanning, has 1600 dpi enhanced resolution, and 400 dpi optical resolution. The IIcx cost $1179 in 1994. The IIIp is the newest HP model, as well as the most inexpensive. The IIIp "uses SCSI interfaces to produce high-quality 256-greyscale images" (Poor 38). It is an eight-bit scanner with OCR, image-editing software and includes PictureScan software, WordScan, and AccuPage (Poor 38). WordScan and AccuPage work together to produce the OCR quickly and accurately (Poor 38). The IIIp is "blazingly fast" scanning a five by seven photograph at 300 dpi in twelve seconds (Poor 38). It is easy to use, can enhance to 1200 dpi and has automatic settings that "not only produced fast scans, but excellent image quality as well" (Poor 39). Even with all of these extras, the IIIp still costs less than the IIp and the IIcx at a mere $600 (Poor 38).

The next category that will be examined will be the software needed to maintain a working digital imaging system. The available software for imaging systems is nearly overwhelming, but this canvass of the alternatives will give a clearer view of the best choices.

AccuScan is an OCR addition to Scanning gallery Plus 5.0 and can be used in a Windows 3.0 environment (Diehl 234). It has an accuracy rate of ninety-six percent, but it is not fast and it cannot scan images and text at the same time, on the same page (Diehl 234). As far as simplicity goes, "it is easy to run Accuscan without ever cracking open the manual," and it sells for about $600 (Diehl 234).

Caere Corporation's PageKeeper and Personal PageKeeper were not mentioned in any of the reviews, but it is worth mentioning because of its rating as "Product of the Year" by Imaging Magazine in 1993. The following information has been taken from an Imaging Magazine circular on the best products of the year, that was included in an informational brochure sent out by Caere. "PageKeeper and Personal PageKeeper might be the best-kept secrets" in document management. PageKeeper also uses a revolutionary new search method. The program reads the file that has been input and creates its own "profile". "It knows the difference between nouns and verbs. It counts how often a word appears. Which words appear together-phrases. It presupposes how a person will want to find that document later" Unlike most search methods, it does not assume that the searcher knows exactly what he/she is looking for. It also allows the searcher to use an article that is being looked at as an archetype to search for similar articles. Even though PageKeeper provides some interesting options and received rave reviews from Imaging Magazine, it sells for less than $600.

Calera's CDP 9000 is an OCR package designed for speed and those for whom money is no object. It is used with a Mac or Pc with Windows 2.x and can decipher eleven inch page in three seconds (Dunn 28). The price tag on the 9000, however, is out of reach of most consumers at $21,950 (Dunn 28).

The Discover 386 5.0 is an OCR program that has a much more reasonable price of $600, but it also has its problems (Diehl 236). It can be used on a PC, and can be used with the scanning equipment of your choice (Diehl 236). It runs "faster and more accurately than almost any other tool..," but is complex and difficult to use (Diehl 236).

ALOS Micrographics Corp's DocuWare Pressman is a package that works on a PC and comes with a hand-held scanner from Logotech Inc. It sells for around $400 (Verity 1266).

Image-Read 2.0 is an example of a poor OCR software package that cannot recognize any font or letter without the programmer giving it a learning session (Diehl 232).

Ocron's Perceive 1.0.5 is a Windows 3.0 software package that has powerful learning capabilities and sells for only $600 (Diehl 234).

ReadRight 2.01 has the attractive price of around $500 and comes with 640 kilobytes of memory, but it is very limited in the area of editing (Diehl 234). There is "no way you can interactively teach or even correct characters..." (Diehl 234). The inability to correct mistakes is a definite drawback for the program.

OmniPage and OmniPage Professional were mentioned more frequently than some of the others and the most information was available about these products. OmniPage 386 3.0 is an OCR program that runs under Windows 3.0 (Diehl 228). It had some problems with fax and nine pin dot matrix documents, but it had no problems with underlining and was easy to use (Deihl 228). OmniPage costs about $700 (Diehl 228). OmniPage Professional is a similar package that retains the original format of the document, but cannot handle handwriting, or color photographs (Diehl 228).

Calera's WordScan Plus came highly recommended. "It's fast accurate and dependable" (Diehl 221). "...WordScan software by Calera Recognition Systems combines scanning, OCR, verification, and conversion processes into one convenient package" (Hawkins,1991 4). WordScan Plus is a Windows 3.0 OCR package that out-performed all other packages in its price range. It sells for about $1000 (Diehl 232).

Although many of the above products come with all of the pieces of the digital imaging puzzle, they are not considered or listed as multi-industry, or turn-key systems. A turn-key system is one in which all the software, hardware, and expertise are specifically created, or adapted to work together as one complete, imaging system. The advantage of such a system would be the convenience of only having to deal with one vendor for all equipment needs. The disadvantage is that the manufacturer may not be highly specialized in any of the specific areas of digital imaging. There are, however quite a variety of turn-key systems available, including some that can be expressly geared to the library atmosphere.

Advanced Computer Concepts, Inc., has created a turn-key system that can be run in Windows, or by using a MUMPS operating system, on a digital VAX computer (Boss 684).

CARL Systems, Inc.'s Pictoral turn-key system is a 386-based PC system that runs on Windows and includes a 300 dpi scanner (Boss 684). It uses TIFF and stores one image per file. This particular system has already been purchased by the University of Hawaii, and Boulder and Atlanta Public Libraries (Boss 684). The system costs

$9600 plus $250 per station (Boss 684).

Voyager from Carlyle Systems Inc., comes equipped with the software needed to create a database by linking the scanned images with chosen key words (Boss 685). This database can be searched by the boolean searching method (Boss 685). The setup also includes the editing tools needed to enlarge, reduce, zoom, and pan the images once they are retrieved (Boss 685). It offers both color and greyscale scanners (Boss 685).

DEC imagr Express, a multi-industry system that can be used on a digital VAX system, comes equipped with a MD410 DEC image greyscale scanner, LN03Q laser printer, DEC image 1200 terminal, and a twelve-inch standalone WORM and jukebox (Boss 666). This product sold anywhere from $30,000-$200,000 in 1992 (Boss 666).

One system that I have personally seen in action is Excalibur. Excalibur with PixTex EFS is a unique product in that it conducts what are called "fuzzy searches". It searches and recognizes binary patterns instead of words. When a search is begun, the system breaks the word or phrase down to a binary level before beginning the search. In this way, it can hit on those things in the database that most resemble what is being searched for. "PixTex EFS can accurately search raw OCR-processed text with no need to correct or rekey the data, which can run one to five dollars per page" (Griffith 47). It also "adapts and grows smarter with use. As the database and its associated indexes grow, the software is able to recognize variations of patterns it has seen before" (Griffith 47). The cost quoted by the reviewer in this particular article was

$2500-$12,000 per user, but when I interviewed Mark Smith and Carla Freeman, who have worked with the system at Alfred University's Scholes Library, they informed me that, including all of the equipment, installation, and training costs, their system came to about $50,000. Even after the training they found it difficult to use for their purposes and have not been completely satisfied with the results.

EyeCom 100 Electric Filing Cabinet from Eye Communications System is a "one-desktop document imaging and storage system" that comes equipped with a flat-bed scanner, laser printer, and 940MB WORM drive (Dunn 29).

Centel Turnkey System from Federal Services Corporation includes image systems with hardware, software, instillation, training and even ongoing support (Boss 685). It runs DataLib for digital VAX, Data general, and Unix-based systems (Boss 685).

Feith Systems and software offers a turnkey system called Single Station Solution that

> "includes a PC with a nineteen-inch monitor, mass storage (up to fifty gigabytes using an optical jukebox), a modified laser printer and a scanner. Software features the *Fieth Document Database*, which supports OCR, voice and video annotations, graphic annotations, and other operations" (Dunn 29).

Image works offers a system that includes a database with browser, icon-based filing system, and storage on either WORM optical, or magnetic media Boss 665). In 1992 a four user system sold for $135,000 (Boss 665).

Kodak has even gotten into the act with its Image Management System 4000 (Boss 667). This runs in an IBM PS/2 based environment and comes with 5.25 inch optical disk storage, desktop scanner, and laser printer, for $55,000 to $100,000 (Boss 667).

In one study, Micrografx Picture Publisher came in as the best multi-industry product for its cost (Marshall 51). This product had "superior documentation, abundant filters, [and] strong painting tools..."(Marshall 51). For desktop publishing, or photographic manipulation this system would be a very strong candidate, but it may not be the best for simple data entry and retrieval, even for under $500.

Omnidesk contains WORM optical disks, an OCR subsystem and is compatible with MS-DOS, Windows, or IBM OS/2 (Boss 677). $35,000 plus $30,000 per additional user is the average cost (Boss 677).

Specifically for Macintosh, Photoshop is a product that is rated highly in terms of importing and exporting, as well as, editing speed (Marshall 51).

Although dealing with only one vendor has many advantages, I am not convinced that this would be the way to go. In my opinion one could be sacrificing quality for convenience. Bringing together different products, each from specialized vendors, to work as a system is cheaper than a turn-key system and may be more effective because of the specialized equipment. The only difficulty with this, would be that each piece of equipment would have its own vendor, warranty, and repair policy. If, however, one scanner and one software program were the only pieces of equipment, then there would also be only two vendors.

Having a database of digital images makes it very tempting to expand into networking. When getting into networking, there are some other things that need to be considered. "...you may need enterprise library services (ELS) to keep documents properly filed and accessible..."(Dunn 29). There are some ELS services available to help break down the barriers that networking can cause. Soft Solutions Image Manager (SS/IM) is a system that requires no special equipment, but does work on a 386 PC running on Windows (Dunn 29). SS/IM "provides enterprisewide management and control of image-based documents" (Dunn 29). Another option, Mezzanine, by Saros Corporation, works with the network to access up to date, current documents relative to the search (Dunn 49). Having a large database of images and getting those images on a network do not necessarily automatically lead one to another. If an organization chooses to get onto a network, it will need some extra, more specific services.

Any organization that is looking towards imaging for storage, preservation, or simply as an office tool, needs to consider which other businesses or establishments have used digital imaging, in what capacity they used it, and how it benefitted them. It is surprising the number of digital imaging systems that are in use and the familiar names that are using them. Companies such as AT&T, Kraft, and Federal Express, as well as numerous libraries and organizations, are getting involved with digital imaging projects.

The AT&T Information Services Network began their scanning project in 1989 and have been scanning an average of fifty new documents per day ever since (Boss 701). AT&T is using WORM storage disc to store images of inter-office memos scanned at 400 dpi (Boss 701).

Federal Express has been experimenting with digital imaging, and has decided that the best way to reduce employee errors when scanning, is to limit scanning times to three hours at a time with a break between two three hour shifts, to do other types of work (Ubois 39).

Kraft Foodservice Inc. has been using a software program from FileNet Corporation for their digital imaging needs (Ubois 39). They are also using an IBM computer and an optical storage medium that carries a capacity of 630 gigabytes (Ubois 39).

The Boulder Public Library has undertaken an extensive digital imaging product using a Microtek 300z scanner, with Picture Publisher software in a Microsoft Windows environment (Boss 703). The system uses a PC-compatible machine with eight bytes of memory, and an additional 200 Megabytes of disk storage is available (Boss 703). The library is scanning photographs to be added to the online catalog as a part of its CARL System Public Access Catalog (Boss 703). It has been estimated that the average photo requires 100 kilobytes of memory and cost approximately eighteen cents to store on optical disk (Boss 703). The estimated cost was $9600 plus $250 per station (Boss 703).

Vanderbilt University's Peabody College Education Library has taken on the scanning and storage of ten vertical file cabinets (Boss 707). The library is using the Macintosh IIx and Apple scanner with a Macintosh SE/30 directory server as an indexing tool (Boss 707). For storage it uses the Pioneer 5.25-inch double WORM drive (Boss 707). The library has also hooked up to MARS, multi-user archival and retrieval system, and an Ethernet LAN to link the system to the campus-wide network (Boss 707).

At the University of California, Berkeley's Bancroft Library has created a Pictoral Collection using a Macintosh Quadra 840, a VRAM expansion kit, and the Arcus Plus 1200 dpi scanner with divers for Macintosh Photoshop, and Exel Software (Brenemann 46).

L. Suzann Kellerman, Penn State's first preservation librarian, and Xerox Corporation have joined forces to create a "prototype document production system that could scan, store, and print on demand information while maintaining the integrity of the original work, including graphics and text" (Penn State). According to a fact sheet given out by the Digital Preservation Project, Penn State employs a Xerox WG-40 high resolution scanner and a Xerox 4030 draft printer, along with Xerox Document on Demand system software. The technicians at Penn State are achieving an average rate of ninety-five pages per hour at a resolution of 600 dpi. They are totaling an average of sixty hours per week. Penn State has estimated the average cost, including production scanning, image inspection, and optical disk storage, to be approximately fifty-four dollars per 300 page document.

Columbia University's law library is currently attempting to create an online library using Xerox Scanworks scanners (Dunn 28).

Future digital imaging projects are also being discussed by several other establishments, including the Library of Congress. Money has already been raised and appropriated for digital imaging projects at the Library of Congress (Ohnemus 64). The "...total amount of private money raised for digitizing LC collections [has risen] to $14.5 million," and a goal has been set to "digitize collaboratively five million items relating to American history by the year 2000" (Ohnemus 64).

Proficiency Testing Services in Brownsville, Texas has progressed to a system that scans about sixty documents per minute and OCRs it with about ninety-five percent accuracy (Dunn 28).. "...Four people working eight-hour shifts now handle a workload that previously required two eight-hour shifts with ten operators each" (Dunn 28).

The United States Patent Office uses six 386-based PC workstations, three Fujisu 3096 scanners and an Anamet OCR software program to scan 10,000 drawings per week at 400 dpi (Boss 704).

The Archivo General de Indias in Seville, Spain has begun an extraordinary digital imaging project to preserve and store documents relating to the voyages of Columbus and the colonization of the Americas (Kalstrom 136). "Since the project began, approximately nine million pages of letters, maps, and drawings have been scanned into optical discs and are accessible via a text database created for the computerized information system" (Kalstrom

136). The Archivo has made available a three query indexing system for use with these digital images (Kalstrom 138). Searches can be done by descriptor tags, the origin of the document, or direct use of reference numbers given by the archivist (Kalstrom 138).

Other prototype imaging projects are also in the works for the National Archives, the National Library of Medicine, the National Agricultural Library, and the Naval Research Laboratory (New York State...3).

By far, though, the most research and development of digital imaging projects has been done by Cornell University. Cornell has teamed up with the Xerox Corporation to lead in the research and development of digital imaging systems. Together they have developed CLASS, College Library Access and Storage System (Kenney 6). During the study, 950 books were scanned and stored at 600dpi (Kenney 9). This created 285,000 digital files, or one file per book page (Kenney 10). The files were compressed so that a six by nine inch page would require only fifty kilobytes of storage, rather than 2000 (Kenney 11). Cornell's two technicians were able to scan approximately 5000 pages per week, at a rate of about five pages per minute in a purely productional mode, but in order to achieve these numbers, they were required to completely disassemble the book before scanning using a sheet-fed scanner (Kenney 41). In terms of sheer numbers, the technicians scanned the average 300 page book in 1.75 hour, or 103 minutes, averaging 175 pages per hour (Kenney 41). Set-up time was an average of twenty-four minutes per book and the average production scan took about 67.8 minutes

(Kenney 42). In a survey Cornell determined that sixty-five percent of those surveyed preferred scanned over photocopied versions, seventy percent preferred scanned when there was only text, and eighty-three preferred scanned photograph over photocopied (New York State...16). Cornell and Xerox did a thorough investigation into digital imaging and its uses, and came to some very promising conclusions:

> "The Joint Study concludes that digital image technology represents a new method for the preservation reformatting of library materials that in the future will replace or compliment microfilming and photocopying. The use of digital technology is currently cost-effective as a reformatting option, and the quality is comparable to light lens processes. The technology offers a means for replacing paper with paper, while simultaneously providing new access opportunities. In the future, researchers will be able to access not only catalog records, but also the full text to which the records refer" (Kenney 35).

During this study Cornell also calculated some very important cost estimates. These figures can be found in Tables A,B,C, and D in Appendix I of this thesis.

Cornell University has also joined the Digital Preservation Consortium, a group that includes Harvard, Penn State, Princeton, Stanford, Tennessee, the University of Southern California, and Yale (Conway 44).

In an investigation such as this, regardless of how positive the author feels about the subject, the difficulties and challenges that can occur must also be explored. The most common problems I discovered during this study, were a lack of common standards, a lack of adequate memory, the difficulties associated with OCR, and the frailty of even the best storage methods.

Digital imaging is such a new medium, and hardware and software programs are being created by so many different sources all at once, that no one has taken the time to devise rules that every company must follow. One of the problems that this creates, is one of incompatibility (Conway 44). Scanning done in different systems cannot yet be transferred or shared. This is a problem that many organizations are starting to discover. "The most pressing issues for imaging as a whole lie in the area of standards" (Boss 648). There is a "crying lack of interoperability standards for imaging and document management systems...Vendors must agree on standards" (Booker,1994 62). "The library community has not yet specified standards for the conversion of microfilm to digital data. The absence of quality standards often leads to system design compromises that are not in the best interests of the long-term retention of the image files" (Conway 43). It seems that if digital imaging is going to be the wide-ranging data storage device of the future, someone has to formulate some rules and regulations for the producers, vendors, and users.

Another problem plaguing imagers is the great need for memory that images have. Technology is starting to catch up to this need, but there are still some lack of memory factors to be considered. "The average desktop PC, even with an eighty-megabyte hard disk and efficient compression algorithms, is not a good device for storing images" (Hawkins,1991 4). "...If the Mac can't receive the data as quickly as the scanner can send it, the scanner must halt and wait for the computer to digest the information before starting again, thus negating the advantage of a higher speed setting" (Hurlow 101). Hopefully, as storage mediums and techniques and compression technologies advance, this will become less and less of a problem.

OCR is such an intricate and technical operation, that it seems amazing that it would work at all, let alone at nearly 100 percent accuracy, but this is where the problems with OCR lie: its accuracy. Even ninety-five percent accuracy hurts when you are dealing with thousands of letters and words.

> "A really good package can achieve better than ninety-nine percent accuracy, but think about it: That means you'll have to fix one out of every 100 characters--that's twenty characters in a single-spaced typewritten page of 2000 characters" (Diehl 220).

Although this is a problem, it may not always be the OCR packages problem. "The biggest pitfall of degraded documents is broken characters. Many times, an OCR program will read a broken character as two distinct characters" (Diehl 226). Fonts and style can give an OCR program additional difficulties. "...Bold faced words that mysteriously appeared in the middle of plain text; bold italic text recognized as plain text;...underlines where none should have been;

[and] extra spaces between words or letters" are all difficulties associated with some OCR programs (Warner 95).

Storage is the most important facet of digital imaging. All of the other steps in imaging are futile if there is no way to effectively store the information once it is scanned. Optical disk storage is the medium of choice among those that have used digital imaging in any capacity, but there is a growing concern over even its long-term viability. Some people have expressed dissatisfaction with the permanence of optical disks. "Optical disks are frail, so you need to have a good backup and disaster recovery plan--we've had to reenter data several times..." said a spokesman from Federal Express (Ubois 39). But even more so than the physical frailty of optical disks, the longevity of the technology has also been questioned. There is a good point to be made that the written page, in all its simplicity, may actually outlive optical disks because of the rapid technological obsolescense that tends to occur in these changing times. "Most files contain information that is meaningful solely to the software that created them" (Rothenberg 42).

Another slight problem that has been encountered when dealing with digital imaging is the pulling ahead of one technology over another. Great advances have been made in the resolutions of scanners, monitors, and programs, but copiers have not quite kept up. "Images that look realistic on screen look course when laser printed" (Heid 306). I predict, however, that this problem will work itself out, eventually.

Digital imaging is a modern, developing technology that points directly to the future. Despite some of its drawbacks, it is a movement that merits extensive study and consideration. When considering an imaging project, one must weigh the scales, between risking something so new, or being completely left behind in the technological race. Yes, imaging has a few "bugs" left that need to be ironed out, but I believe that deciding against imaging as a viable preservation and storage medium, could hinder an organization in the future. I believe digital imaging is not a fad. It is not a medium that will quickly fade out of consciousness. Already corporations and organizations world-wide are turning to digital imaging and praising its applications and amenities.

Libraries are a specific area where digital imaging can be an amazing asset. Fragile books can be accessed conveniently, quickly, and without handling the book itself. Any library with an archives needs to consider the best way to store and protect rare and brittle books that also provides patrons with a way to view them. Digital imaging may be this way.

For a library or small business that is contemplating the move to digital imaging, there are many products, and methods that will need to be considered. Hopefully, this study will provide enough details to make informed decisions about digital imaging.

Researching this topic has enabled me to evaluate many of the available products, and come to some of my own conclusions about what products I feel are the best for the money. Hewlett Packard scanners were given high marks by reviewers and have impressive capabilities. The HP ScanJet IIIp, specifically, was also reasonably priced in the $600 range. The top reviewed OCR software packages were Calera's WordScan Plus and OmniPage/OmniPage Professional. Both products are priced a bit higher than average, but not quite out of the reasonable range. OmniPage is around $700, while WordScan Plus is about $1000. The deciding factor between the two will come down to a specific evaluation of needs and to personal preference. Another promising option, however is Caere's PageKeeper at $600. All of the information that I had on PageKeeper was positive and quite impressive. Unfortunately, all of it came directly from Caere circulars and could be misleading. Most of the turn-key systems were priced far out of the range of a typical library or small business, and the one's that were a little less extravagantly priced, lacked some basic imaging capabilities. If you are willing to pay the extra price, I would suggest a careful review of the available products, as well as, in depth research and discussion with the vendors. I believe if you are looking for complete cost effectiveness putting together your own system would be the best bet, but if you feel that it is necessary to have installment and training options, a turn-key system might be more appropriate. Some of the systems worth considering would be Pictoral, at $9600 plus $250 per extra station, as well as Voyager and Centel. If money is no object you might want to consider DEC Image Express for $30,000-$200,000, Excalibur at about $50,000, or Omnidesk for $35,000 plus $30,000 per additional user. Soon there will be literally thousands of digital imaging products on the market. Because of the costs and important objectives of digital imaging, choosing which products to purchase should not be something that is done solely on the advice of a salesperson. Getting the correct pieces to the puzzle and creating a system that accomplishes the objectives efficiently and cost effectively is the first and most important step in any major project, especially one in digital imaging.

At Alfred University's Herrick Memorial Library, the librarians are right to be curious about digital imaging. To use an old cliche, it is the wave of the future. They were also right to question how it could benefit them and at what cost. At Herrick Library, as well as, other libraries around the world interested in preservation, digital imaging should be looked at as a new and exciting, but also effective, means of conserving, storing, and viewing their rare and delicate volumes.

Appendix I

TABLE A. COMPONENT COSTS OF USING SCANNING TECHNOLOGY TO CREATE PAPER REPRODUCTION OF PREDOMINANTLY TEXT OR LIGHTLY ILLUSTRATED 300 PAGE BOOK AND TO MAINTAIN THE DIGITAL MASTER
(see Appendix III for Assumptions)

Unit Costs in 1992 Dollars (unit = 1 book)
5% rate of change represents normal inflation

	Rate of Change	Year 1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	10 Year Average Cost
Scanning Costs												
1. Labor	5%	$20.64	$21.67	$22.76	$23.89	$25.09	$26.34	$27.66	$29.04	$30.49	$32.02	$25.96
2. Equipment	-10%	$4.92	$4.43	$3.99	$3.59	$3.23	$2.91	$2.61	$2.35	$2.12	$1.91	$3.20
3. Total Scanning Cost		$25.56	$26.10	$26.74	$27.48	$28.32	$29.25	$30.27	$31.40	$32.61	$33.93	$29.17
Optical Storage Costs												
4. Optical Jukebox	-15%	$1.60	$1.36	$1.16	$0.98	$0.84	$0.71	$0.60	$0.51	$0.44	$0.37	$0.86
5. Optical Disk	-30%	$5.36	$3.75	$2.63	$1.84	$1.29	$0.90	$0.63	$0.44	$0.31	$0.22	$1.74
6. Technology Refreshing*						$2.12	$1.61	$1.23	$0.95	$0.75	$0.59	
Printing Costs												
7. Printing Equipment	-5%	$6.75	$6.41	$6.09	$5.78	$5.50	$5.22	$4.96	$4.71	$4.48	$4.25	$5.41
8. Acid Free Paper	5%	$1.50	$1.58	$1.65	$1.74	$1.82	$1.91	$2.01	$2.11	$2.22	$2.33	$1.89
9. Total Printing Cost		$8.25	$7.98	$7.74	$7.52	$7.32	$7.13	$6.97	$6.82	$6.69	$6.58	$7.30
Binding Costs												
10. Library Binding	5%	$7.00	$7.35	$7.72	$8.10	$8.51	$8.93	$9.38	$9.85	$10.34	$10.86	$8.80
11. In-line Finish	5%	$1.00	$1.05	$1.10	$1.16	$1.22	$1.28	$1.34	$1.41	$1.48	$1.55	$1.26
12. Unbound/Stapled	0%	$0.25	$0.25	$0.25	$0.25	$0.25	$0.25	$0.25	$0.25	$0.25	$0.25	$0.25
13. Weighted Binding Cost **		$1.90	$1.99	$2.08	$2.18	$2.29	$2.40	$2.51	$2.63	$2.76	$2.89	$2.36
Access Costs												
14. Access Cost /Book ***	-10%	$1.60	$1.44	$1.30	$1.17	$1.05	$0.94	$0.85	$0.77	$0.69	$0.62	$1.04

* On average a book will be refreshed twice in a decade. For instance, a book created in 1992 will be refreshed in 1996 and 2000.
** This weighted average binding cost assumes 20% library binding, 40% In-line, and 40% unbound/stapled.
*** This is highly dependent on the choice of technology.

Table B	1992 Cost	Average
Digital Technology		
Scanning: Labor and Equipment [3]	$25.56	$29.17
Printing [9]	$8.25	$7.30
Overhead - 30%	$10.14	$10.94
Library Binding [10]	$7.00	$8.80
Total Printed Copy from Digital	**$50.95**	**$56.21**
Photocopy		
Total Photocopy	**$65.00**	**$74.52**

TABLE B. PHYSICAL REPLACEMENT OF BOOKS (NO STORAGE OF FILES)[28]

Table C	1992 Capture and 10 Year Maintenance	
Digital Technology		
Scanning: Labor and Equipment [3]	$25.56	
Storing: Optical Jukebox & Media [4&5]	$6.96	
Refreshing - 10 years [6]	$2.87	
Overhead - 30%	$10.62	
Total Digital	**$46.01**	
Microfilm	**One-up**	**Two-up**
Creating Archival Master	$58.50	$29.25
Creating Print Master	$5.00	$2.50
Storing 2 Generations for 10 Years	$6.66	$3.33
Total Microfilm	**$70.16**	**$35.08**

TABLE C. CREATION, STORAGE, AND MAINTENANCE OF DUPLICATING MASTER FOR 10 YEARS[29]

Table D	1992 Cost	10 Year Average
Digital Technology		
Access [14]	$1.60	$1.04
Printing [9]	$8.25	$7.30
Overhead.- 30%	$2.96	$2.50
Binding [13]	$1.90	$2.36
Total Digital	**$14.71**	**$13.20**
Photocopy Technology		
Subsequent Copy	**$65.00**	**$74.52**

TABLE D. COST TO CREATE A SUBSEQUENT PRINTED COPY OF A BOOK[30]

28 Costs associated with digital technology are derived from Table A. The numbers in [brackets] refer to line numbers in Table A. Overhead reflects the general and administrative costs and profit margin that would be included by an outside vendor. The 1992 cost of photocopying is based on two quotes for photocopying and binding a 300 page book (Library Bindery Service and Ridley's Book Bindery). The average annual inflation rate is calculated at 5%.

29 The numbers in [brackets] for digital technology refer to line numbers in Table A. A book scanned in 1992 will be refreshed twice in the next decade, in 1992 and 2000. Overhead reflects the general and administrative costs and profit margin that would be included by an outside vendor. Microfilm figures are based on 1992 prices quoted by MicrogrAphics Preservation Service (MAPS). Cost of archival master is based on $.195/frame for one-up and two-up filming. Cost of print master is $15. For two-up filming, assume six books can be stored on each roll; for one-up filming, assume three books. The cost of one book on the print master will be $5.00 (one-up) or $2.50 (two-up). Storage costs are based on $1/year to store one roll of film. The cost of book storage/year will equal $1 divided by 3 (one-up) or by 6 (two-up). Since two generations are being stored, the cost equals $.66 (one-up) and $.33 (two-up) per year times 10 years, or $6.66 and $3.33 respectively.

30 The numbers in [brackets] for digital technology refer to line numbers in Table A. Overhead reflects the general and administrative costs and profit margin that would be included by an outside vendor. The binding cost included here assumes that 20% of all requests for subsequent copies will be bound with a full cloth library binding , 40% will be bound using Docutech in-line tape binding, and 40% will be unbound or stapled. If we assumed that all subsequent copies were bound in a full cloth binding, the total digital

The numbers below refer to the line numbers in Table A.

Scanning Costs:

1. Labor. It takes a scanning technician 1.72 hours to scan an average 300 page book. The 1992 hourly rate, including benefits, is $12.00/hour. ($12 X 1.72 = $20.64) The ten year outlook does not include an increased production rate, which almost certainly will occur. The costs of labor are projected to increase by 5%/year.

2. Equipment. Equipment includes scanner, personal computer, high resolution monitor, application software and network connections. Annual scanning equipment cost is $9,440 ($20,000 X .472). The equipment is assumed to be operational for 2 shifts. Assuming a standard shift equals 37.5 hour work week, the total number of hours/year that the equipment is in use is 3,300. The hourly cost for equipment ($2.86) is computed by dividing the total number of hours per year by the annual scanning equipment cost ($9,440/3300). Since the average scanning time per book is 1.72 hours, the scanning equipment cost per book is therefore $4.92. This figure declines by 10%/year.

Transfer to Optical Storage:

4. Optical Jukebox refers to the cost of the time it takes to write a disk on the jukebox (20 minutes for 300 page book). This figure is calculated by taking the annual cost of jukebox ($75,000 X .472) and dividing it by 21,900 (which equals 20 minutes of a year: 365 days/20 hours per day,[2] divided by 3 for the 20 minutes). This figure declines by 15%/year, which is very conservative, given declines in technology cost and time taken to write the disk.

5. Optical Disk Cost. This figure represents the cost of the 12" disk ($500) amortized over 4 years, and divided by the number of books that will fit on a disk.
$500 X .472 / 44 = $5.36. This figure declines by 30%/year.

6. Technological Refreshing. The cost of refreshing, which begins in year 4 and is repeated every four years, is calculated by adding the cost of equipment (#4 above), which is the cost of the time taken to transfer to a new disk, and the cost of that portion of a new disk that the volume occupies (#5 above). Certainly the time between refreshing will increase as data exchange standards are developed.

Printing and Binding:

7. Print Equipment. This figure includes the cost of the Docutech printer, maintenance, set-up for printing, and associated costs. Cornell has set a price of printing at $.0225/side (excluding paper), which is based on full costing of the Docutech, including allowances for space and overhead. A 300 page book would cost $6.75. This figure will decline by 10%/year.

8. Acid Free Paper. The cost of acid free paper is calculated at $.01/sheet (a 300 page book would be printed on 150 sheets). This figure will increase by 5%/year.

10. Library Binding. The binding figure includes the cost of full cloth binding (Ridley's Book Binder) and preparation for binding. It will increase by 5%/year.

11. In-line Finish. The Docutech offers a number of finishing options, including a heat-set tape binding, the price of which is listed here. This figure will increase by 5%/year.

12. Unbound / Stapled. A nominal figure is included here for stapling.

13. Weighted Binding Cost. Binding costs for subsequent copies are based on the assumption that not all requests will result in a library binding. It is assumed that a full cost library binding will be done for 20% of the print requests, an inexpensive in-line finishing will be used for 40%, and that the remaining 40% will be stapled or left unbound.

14. Access Cost / Book. The cost of accessing the book assumes that the optical disk containing the book will be mounted on the jukebox when needed. This cost is calculated as the cost of one tenth of an hour ($1.20) and 5 minutes of optical jukebox time ($.40) (See #4 above). This figure will decline by 10%/year.

Works Cited

Alford, Roger C. "How Scanners Work." Byte June 1992: 347-350.

Anthes, Gary, H. "How the Feds Find Data." ComputerWorld Aug. 30, 1993: 47-50.

Booker, Ellis. "Counting on Paper Cuts." ComputerWorld March 21, 1994: 62.

Booker, Ellis. "Low Cost Imaging Abounds at AIIM." Computer World April 12, 1993: 68.

Boss, Richard W. "Imaging for Libraries and Information Centers." Library Technology Reports Nov.-Dec. 1992: 637-726.

Breneman, William E. "Reference Implications of Digital Technology in a Library Photograph Collection." Reference Implications of Digital Technology Winter 1994: 45-49.

Cohen, David. "New Books From Old: A Proposal." The Serials Librarian v.23 no.3-4: 149-150.

Conway, Paul. "Digitizing Preservation." Library Journal v.119 iss.2 Feb. 1, 1994: 42-45.

Diehl, Stanford, and Howard Eglowstein. "Tame the Paper Tiger." Byte v.4 iss.16 April 1991: 220-238.

Document Solutions Profiles. University Park: Pennsylvania State University.

Dunn, Si, and Connie Dunn. "Image Building." Delaware Valley Computer User April 1994: 1, 28-30.

Greh, Debbie. "Scanning Your Imagination." School Arts v.91 May 1992: 20-23.

Griffith, Cary. "There's More to Imaging Than Image." Information

Today April 1994: 20-21.

Hawkins, Donald T. "Imaging--Another New Technology for the Information World: The Basics." Online iss.15 Nov. 1991: 3-5.

Hawkins, Donald T. "Text Retrieval Technologies for Image(and Other) Databases." Online v.16 iss.1 Jan. 1992: 92-94.

Heid, Jim. "Getting Started With Image Processing." MacWorld v.8 iss.11 Nov. 1991: 291-307.

Hurlow, Mark. "Pushing Scanners to the Limit." MacWorld Oct. 1992: 101-102.

Kalstrom, Dave. "Archive Project Preserves Columbus Documents Optically." CD-ROM Professional v.6 iss.4 July 1993: 135-139+.

Kinney, Anne R. and Lynn K. Personius. The Cornell/Xerox/Commission on Preservation and Access Joint Study in Digital Preservation. Washington: Commission on Preservation and Access, 1992.

LaPlante, Alice. "Los Alamos Laboratories Get a Grip on Research Data." InfoWorld v.15 Dec. 6, 1993: 76.

Linderholm, Owen. "Logitech Puts Photo-Realism in Windows." Byte v.15 Nov. 1990: 133.

Lu, Cary. "Optical Scanning Technologies." High Technology March 1986: 185.

Mallen, Ed. "Intelligent Character Recognition: It's Not Just Recognition Anymore." American Society for Information Science v.18 iss.5 June 1992: 9-11.

Mamis, Robert A." The Scanner: Next Peripheral?" Inc. v.14 June

1992: 128.

Marshall, Patrick, Doug Green, and Denise Green. "Image Creation and Editing." InfoWorld v.14 iss.32 Aug. 10, 1992: 51-66.

Moore, Frank A. "Spelling Out the Benefits of Imaging." Inform Feb. 1990: 29-30.

New York State Program for the Conservation and Preservation of Library Research Materials. Preserving Archival Material Through Digital Technology: Final Report. Ithaca: Cornell University Library Department of Preservation and Conservation, 1993.

Ohnemus, Edward. "Digital Library Gets a Boost in Funding." LC Information Bulletin Feb. 20, 1995: 64-65.

Patalon, William III. "Kodak's Digital Deals." Rochester Democrat and Chronicle March 28, 1995: 1(A) and 5(A).

Poor, Alfred. "Two of a Kind Scanners Move Words and Images From Page to PC." PC Magazine v.13 iss.22 Dec. 20, 1994: 38-39.

Rothenberg, Jeff. "Ensuring the Longevity of Digital Documents." Scientific American Jan. 1995: 42-47.

Ubois, Jeffrey. "Clearing the Paper Jungle." CFO Aug. 1994: 38-42.

USC and the Kodak Library Image Consortium. Los Angeles: University of Southern California James Irvine Foundation Center for Scholarly Technology, 1993.

Verity, John W. "Don't Clip Those Articles--Scan Em' In." Business Week Nov. 29, 1993: 1266.

Warner, Tim. "Lab Notes: What to Expect From OCR." MacWorld v.9 iss.7 July 1992: 95-96.