Ronald J Kirkley. Archival Considerations: A Practical Approach from a Small Business Perspective. A Master's Paper for the M.S. in I.S degree. April, 2006. 29 pages. Advisor: Deborah Barreau

Procedures were developed to create a digital archive for a moderate sized engineering firm. The two part process involved data recovery from past backup tapes and from various magnetic and optical media from earlier attempts of archiving digital information. It also involved creating procedures for preserving and archiving present and future information. The archival procedures were developed utilizing the major recommendations found in the literature that generally focused on libraries, corporations and governmental organizations.

Headings:

Digital archive

Electronic archive

Digital preservation

Tape drive

Data recovery

ARCHIVAL CONSIDERATIONS: A PRACTICAL APPROACH FROM A SMALL BUSINESS PERSPECTIVE

by
Ronald J Kirkley

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2006

Approved by

_____

Deborah Barreau

Table of Contents

I. Introduction

The lifeblood of professional firms is the information they collect and produce. Managing this information in such a manner that it will be preserved in a useful, easily accessible format for future years is vital to all such firms. There are actually two major parts to the problem: (1) storing the information, and (2) insuring it is in a usable form. Often businesses, like individuals, can be good at storing, the first part of the problem. Very little is discarded. More likely items are placed in boxes and carefully stored in closets or warehouses. However, just as individuals can forget what is hidden away in their attics, the organizational memories of businesses can lose track of all the information that has been carefully saved. Therefore, the second part of the problem is more problematic. In order for the stored information to be useful, there must be a plan and ongoing review to insure that the usefulness of the information is maintained. There must be a way for people to know what is stored and how to retrieve it. The same is true for electronic information. All efforts of copying data to media for storage are wasted if the recorded media is placed in a closet and forgotten. Even if it is found or remembered years later, the initial efforts to preserve the data can be wasted if the means to read the saved data no longer exist.

The problem can be especially difficult for smaller firms that do not have the internal resources to dedicate towards this task. Without action to insure the preservation and restoration of historically stored electronic information or a plan to archive and

preserve presently generated information, firms stand a good chance of losing one of their most valuable resources.

II. Background

A. The difficulty of preserving digital information

The problem of preserving digital information is nothing new. Governmental and corporate archivists have long been aware of the special challenges of insuring the continual accessibility of digital information. Since the late 1980's as library holdings have moved increasingly from paper to digital, librarians and archivists have become increasingly involved with the problem (Waters & Garret, 1996). Despite the relative short history of digital information there are already famous cases of digital information being lost or nearly lost. There is the case of the 1960 census where the data was stored on a soon obsolete tape drive. When the National Archives requested the information in 1976 the Census Bureau had an extremely difficult time recovering the information from the obsolete tapes (Waters & Garret, 1996). Another example is the case of the British Broadcasting Corporation's (BBC) Doomsday Project. At a cost of $4.3 million the BBC created a snapshot of Great Britain that consisted of sound, video, photographs, maps and text. The information was placed on state of the art 12 inch video laser disks which less than 15 years later had become obsolete and nearly unavailable. It took some effort to reconstitute the data so that it could be made available with modern technology (O'Donnell, 2004). Not surprising, as more and more people conduct their personal and professional lives through digital means the problem has been receiving more widespread attention. In a *Scientific American* article Rothenberg (1995) describes the problems to be

encountered by his heirs in 2045 when trying to read a 20th century CD-ROM. *Forbes* ran an article that describes the problems with electronic information in terms of the digital photographs stored on our hard drives (Manes, 2006). Both articles clearly explain the problems of archiving digital information and offer some partial solutions. However both articles highlight that though the problem is clearly understood, there is no one simple solution.

B. Preserving the bits

The problem of the preservation of digital information is actually two problems. The first is the hardware problem which consists of the storage media and the hardware required to read the media. There are currently two choices for long term storage media: (1) optical disks, and (2) magnetic media, which includes floppy disks, hard drives and magnetic tape. For large organizations, traditionally only magnetic tape had the storage capacity required for their amount of data. However using hard drive space for archival purposes is becoming more popular. This is caused by lowering costs, the desire for live data and new configurations using low cost IDE disks. Smaller organizations also predominately use magnetic tape but the ease of use and perceived longevity of optical disks makes them attractive to smaller organizations. However in both the cases of magnetic tape and optical disks there is debate about longevity. Rothenberg (1995) conservatively estimated the physical longevity of tape to be one year. In contrast, the National Media Laboratory states that properly stored magnetic media should be usable for 10 to 20 years (Bogart,1997). There is similar confusion about optical media. In the *Cadalyst*, a trade publication for computer aided drafting, it is stated, “Optical media

have an almost unlimited storage life."(Green, 2005, p. 38). A study by The National Institute of Standards and Technology (NIST) seems to agree by concluding in the best case optical media my have a longevity of "several tens of years". However, the same study finds that lower quality media stored in less than ideal conditions may fail in as little as weeks. The NIST study finally concludes, "Unfortunately, it is very difficult for customers to identify these more stable media." (Slattery et al., 2004). For both the case of magnetic and optical media the debate about longevity is mostly irrelevant. The second and larger problem is that the hardware and software required to read the media is changing in as little as two to five years (Waters and Garret, 1996). This is especially true of less standard hardware as was the case of the 12 inch laser video disks used by the Doomsday Project. Therefore the practice of periodically migrating data to new media has become the accepted practice by archivists to ensure the survival of the bits (Mallinson, 1990).

C. Ensuring the bits will remain usable information

As shown, the preservation of the electronic data, while not trivial, can be accomplished by careful active maintenance. The second and larger problem is turning the stored data into usable information. If the hardware or software used to originally create the information is no longer available then all the effort to preserve the data could be wasted. There are several strategies used to avoid this problem. One is to migrate through translation files that are in danger of becoming obsolete to the current successor formats. A related strategy is to translate all files into formats that are more likely to survive for the long term (Caplin, 2004). For example word processing files could be

converted to plain text or PDF. Both these strategies have the disadvantage of losing some information through the translation process. Successive translation would only exaggerate the problem. Others suggest that the original file formats should be kept as is. This requires that either the hardware and software required to access the files natively be maintained or virtual machines that could emulate the original hardware and software be successively developed for modern technology (Rothenberg, 1995). An obvious problem is the difficulty of keeping older technology running when spare parts and expertise are vanishing. A problem with emulation is the continuing development of technology. Conceivably as technology advanced, emulators would be required to run within emulators in order to mimic the original technology. The resources required to continually develop these emulators would limit access to only the most important or common formats. A third, more intensive strategy is typified by the Universal Preservation Format (UPF). The premise of this strategy is that information should be archived in a "wrapper" that contains all the information required to reconstitute the stored information to its original usable form (Shepard & MacCarn, 2001). The difficulty of developing these wrappers would limit this strategy to larger institutions, governments or large partnerships.

D. The problem in small companies

The problem of preserving digital information is not unique to governmental organizations, large corporations and libraries. There are thousands if not millions of small to medium business and professional firms throughout the United States that have the same problem. An engineering firm is probably not atypical of other professional

service providers such as lawyers, accountants and architects. Their product is really the intellectual expertise of their employees. In other words they take a client's situation or need and provide the properly collected and formatted information needed to solve the problem. So in effect the product provided by an engineering firm is information. It is also typical that engineering firms do not begin as mature organizations with clearly defined procedures for managing their information. It is probably more common that firms start with one or a handful of engineers that set up an office and begin offering their services. Finding new clients and insuring that quality work is delivered in a timely manner are their top priorities. Because young firms begin relatively small, the internal managing of information is often done through informal processes. Everyone is familiar with what everyone else is doing. So, if some information is needed then you just ask the right person. Over time with new clients and an increasing number of active and completed projects, the information held by the company will grow. More people may be added to the company and some or all of the original employees and partners may leave. If the company has a long history then the technology used will have changed drastically. What was originally done with slide rules, typewriters, letters and drafting boards is now being done with spreadsheets, word processors, email and Computer Aided Drafting and Design software (CADD). Even companies with much younger histories have experienced changes in software, storage media, operating systems and hardware that if not properly managed will yield much of their historical information unavailable. Additionally, without a plan the information being produced today will probably be in the same danger of being lost.

E. One Case Study

The subject of this study is a moderate sized engineering firm of approximately 110 employees that was founded in 1936. Its history is probably not atypical of similar sized professional firms that were founded prior to the digital age. They developed with the technology. Unlike a firm that would be founded today, there were no established best practices or well developed information technology consulting services. Therefore it is not surprising that the technological history of the studied firm is somewhat haphazard. It began performing computer aided drafting (CAD) using a Hewlet Packard mini computer around 1987 and transitioned to Apple Macintosh computers using MicroStation around 1992. Then in 1998 the company forked, one department switched to Windows based machines running Bentley MicroStation for CAD, while the rest of the company continued running MicroStation on the Macintosh. Eventually Bentley dropped support for the Macintosh, forcing most of the users to switch to the Windows platform. One department continued to use Macintosh but with VectorWorks as their CAD software. During this time the technology for performing office productivity tasks such as word processing and spreadsheets roughly paralleled the technology used for CAD. Early users used DOS based machines. Eventually many of these users transitioned to Macintosh while some remained with DOS eventually transitioning with new technology to Windows. Throughout the 1990's and into the early 2000's both Windows and Macintosh platforms were used within the company. This alone created problems, but the problems were intensified through the use of multiple applications for the same task. For example, as late as 2001 Microsoft Word, WordPerfect for Windows, WordPerfect for Macintosh and ClarisWorks were all being used for word processing. The variety of applications and

operating systems would prove to greatly complicate the recovery of the historical digital information.

The approach to archiving and backups also paralleled the diversity of hardware and software platforms. Without a central file server until 2004, all data was stored on individual hard drives. Unlike paper based archives that consist of paper copies of correspondence and record sets of drawings there has not been a consistent effort to archive electronic data. However, there were many ad hoc efforts by individuals and small groups to backup and/or archive. These included copying important files to the media of the day such as 5.25-inch floppy disks, 3.5-inch floppy disks, ZIP disks, and eventually CD-Rs. Retrospect, a networked based backup software that backed up individual machines to a central server, was later used successfully with a succession of various formats of tape drives. Although there were some good attempts of periodically creating and documenting archives on media such as ZIP disks and CD's the existence of any real comprehensive electronic archive was non-existent. However, much if not most of the historical data still survive. It is on old tapes, most created using Retrospect, on floppy disks, both 5.25-inches and 3.5-inches, on various removable media such as Syquest disks and ZIP disks as well as on CD-Rs. Also, much of the old data survived through the process of copying files from a machine as it was being retired to its replacement. This data eventually made it to a file server, once a server based centralized network was established.

Presently the only archive consists of information that was previously recovered from some early archives created sporadically as projects were completed or at the institutive of individuals within various departments. The information was recovered

from the various media and translated into a format usable with the present Windows platform. It was then organized into individual projects and stored on a read only network accessible volume. Presently there are many completed projects that remain “live” within the project directory on the main file server. Therefore there is a need to develop a process to archive projects as they are completed and closed and there is a need to recover and make available the digital information contained on all the stored magnetic and optical media found throughout the company.

III. Objectives for the project

1. Insure all information produced presently and into the future remains easily available in a useful format, and
2. Make existing historical electronic information available in a useful format.

IV. Procedures

A. Create archival procedures for present and future projects

In the beginning the primary focus of this project was recovering digital information from all the found stored media. However the review of literature and the problems encountered recovering useful information made it apparent that a systematic plan for preservation and archiving was needed. Obviously a small company does not have the resources and expertise of large organizations to develop elaborate preservation and archiving systems. Using something like the Universal Preservation Format would be far too complicated. Also a small organization would not have the expertise to develop, nor could it expect there to be emulators available to natively view and manipulate files

into the future. However there are useful concepts found in the archival strategies of the large organizations.

An archival plan should include the following. It should provide for both the migration of bits and information. The bits should be copied from existing media to newer media whenever the existing media is in danger of becoming obsolete or it was approaching the end of its expected lifespan. The archived information should be migrated through translation whenever the format of the information is in danger of not being accessible with the present applications. Information should also be maintained in its original form because of the loss information incurred during translation and because of relational information that can be lost when files are removed from their original directory structure. Whenever it is reasonably possible, old hardware and software should be retained so that the original information can be accessed natively. Finally and most importantly the archive should be kept close to the original users. If the archives are easily accessible and kept as close to their original structure as possible then they will be used more often. This will help insure that they are maintained and that they will receive the resources needed to keep them fresh.

1. Selection of hardware and media for the storage of the archive

I chose to place the archive on a partitioned volume of our main file server and back it up to magnetic tape whenever additions were made. As stated by Caplin (2004), "…only use itself can ensure that archived copies remain usable" (p. 39). Storing the archive on a network volume allows users to search the archives with already familiar tools and to insure the archive is used. Because the archive is located on the main file server it would naturally be maintained. Hard drives have a limited life expectancy. In a

server environment they are configured for redundancy and will be refreshed and replaced much more often then tapes that are written and stored. Therefore the bits of the archive would be migrated as a part of the ongoing maintenance of the server. If space becomes a factor there are many reasonably priced hard drive storage devices such as Network Attached Storage (NAS) that could be used without sacrificing the ease of availability to users.

2. Develop a directory structure that facilitates access by users.

As much as possible I wanted to create a directory structure that would mirror the file structure on our main server as well as accommodate files recovered from the stored media. The accounting system provides a structure based on client and project. Since the 1980's each client is assigned a four digit number. Each project is identified by adding a sequential alphabetical suffix to the client number. For example the fourth project contracted for client 1212 would be 1212-D. This structure is the logical choice we use for our active projects on our file server. It follows that this structure should also be used for the archives.

3. Develop a procedure for archiving projects as they are closed

The most important part of any archival plan is insuring information is properly formatted and moved into the archives at the appropriate time. Therefore a Standard Operating Procedure (Appendix B) was written. It designates the personnel involved and their responsibilities during the archiving process. The bulk of the responsibility rests with the project manager and the Information Technology Manager. Because projects belong to and are the responsibility of project managers, it is their responsibility to initiate archiving of their completed projects. They inform all the personnel assigned to

the project that it is closing. Each person then insures that all the information associated with the project is within the project directory. Generally all the information should already be within the directory, with the possible exception of related emails and perhaps items that exist only in hardcopy that should be scanned to PDF format and added to the directory. If the project included computer aided drafting files then there are additional requirements that are handled by the CAD operators assigned to the project. Copies of any files located outside of the project directory such as base maps and reference files should be moved into the project directory. Because of the nature of references embedded in CAD files, the directory locations of all associated files must be carefully noted so that the original structure can be recreated. The CAD operator should also insure that plot files in a common format such as HPGL2, DWF or PDF are placed inside the directory to insure the design sheets can be replotted even if the original CAD program is no longer available. Finally, the CAD operator should insure there are scanned copies of the record drawing in a common TIFF format (Green, 2005). Once the project manager has determined that all associated information is within the project directory, the Information Technology Manager reviews the project directory for completeness, generates a Metadata.xml file (Appendix C) to place within the directory, and generates a copy of all the word processing and spreadsheet files in a PDF format also within the directory. The entire project directory is then moved as a read only copy into the archive directory.

4. Develop a procedure to periodically review archive

In order to insure the continual availability of the information contained within the archive it should be reviewed periodically. If there are files in formats that are in danger of being superseded, they need to be migrated to the newer formats. To facilitate

the review process the archival procedures require the creation of a metadata file (Appendix C) that is placed within the archived directory. Part of this metadata file is a listing of file formats for the enclosed files. Therefore a simple query could be used to produce a list of file formats that may need migrating. The Standard Operating Procedure for Archives (Appendix B) requires that the review be performed annually by the Information Technology Manager.

B. Make historical electronic information easily available

1. Find, investigate and catalog all the stored digital media.

A thorough search throughout the company produced many boxes of various media types as well as various hardware devices in varying states of function. Information on the media labels, some minimal documentation found along with the media, interviews with long term employees and web searches about the different media types yielded an overview of the extent and nature of all the media holdings. The results of these investigations about the found media are summarized in the table in Appendix A.

2. Recovering Digital Information

There were two major phases involved with making the historical electronic information available for easy use. First the information had to be recovered from the media. Secondly the information had to be formatted and stored in manner to ensure its preservation and to ensure its ongoing availability to users. Each of these phases involved multiple steps and decisions which are described in detail below.

For the purposes of this paper and also as a proof of concept, I decided to recover the information on the DDS and DDS-2 tapes. These were chosen for several reasons. First, preliminary investigations indicated they may provide the most valuable information. The later DDS-3 tapes mostly contained information that had been migrated directly to the central file server when it was installed. The DDS and DDS-2 tapes appeared to contain information from the mid 1990's. They were almost certain to contain some of the only electronic copies of projects that ended in this time frame. Therefore these files were recent enough that electronic versions of drawings and spreadsheets would contain more information than the corresponding paper copies. Also the limited documentation and longer term employees both indicated that these tapes contained the backups of individual Macintosh computers. Therefore they should contain the design work done during this time period. The second reason these tapes were chosen is that the hardware was available to read them and they had been made using Retrospect which was also still available within the company.

The first step was ensuring the tapes could be read. It was learned that only a DDS3 or earlier tape drive could read the DDS tapes (DAT Roadmap). It was also learned that, although Retrospect runs on both the Macintosh and Windows platforms, the tapes created with one platform are not readable with the other platform (kbadmin, 2004). An existing DDS3 tape drive was connected to an Apple G4 using a SCSI card purchased from eBay and an existing copy of Retrospect 5.0 was installed with all the updates from the Retrospect website. With this setup, Retrospect was able to read the tape labels, which was a good indication that the information could be recovered.

Because information on tapes is stored sequentially and not in a file structure the tapes had to be cataloged by Retrospect before any information could be recovered. Once Retrospect completed the cataloging the information was dumped from the tapes to a folder by retrieving all the information created prior to the present date. The cataloging and dumping of information was repeated for each of the tape sets. This created several folders containing the contents of each of the tape sets. Three of the sets were originally created specifically as an archive and somewhat mirrored the directory structure discussed earlier of "client number / project suffix". This will make it easy to move the files into the archive structure on the network volume. However the other four tape sets were backups of individual computers. Therefore the directory structures were much more complicated. Among the valuable project data there was a lot of "noise", consisting of a wide variety of information. This includes administration files, proposal efforts that did not result in a contract, personal files and redundant copies of files. The files were saved and named differently according to the user, the department, the type of information and the time the information was saved. Also these four sets of backup tapes had been created concurrently in order to produce redundant tapes in case of failure. Therefore much of the information contained on one set was also contained on the other sets as well. However, exploration of the contents revealed that the overlap of information was not complete. Some computers were backed up on one set of tapes and not on others. Therefore the information from all four sets had to become part of the archive. Also, because of the difficulty of trying to eliminate redundant data and because there was information in the structure of the directories and the computer names found on the tapes, it was decided to retain the directory structure of these tapes in the archive.

Because the information recovered from the tape sets was created on Macintoshes the file and folder names contained characters that were incompatible with the archive which is stored on a Windows volume. Also as discussed earlier in the background there were several formats of word processing files and at least two formats of spreadsheets. Therefore before the recovered information could be moved into the archives the names had to be made Windows compatible and the files had to be translated into formats compatible with the present applications within the company. This was accomplished by running the files through NameCleaner to make the names Windows compatible and then through MacLink Plus to translate the files into Word and Excel where appropriate. Once this was completed the files were moved into the archive. Because of the inevitable loss in information from the name changes and translations a copy of the recovered files was retained in their original form on the Macintosh G4.

V. Summary

Information stored on old media should be recovered as soon as possible. The difficulty for recovery increases with age. There may be some degradation of the media that causes partial or total loss of the data. The hardware and software needed to access the data becomes increasingly difficult to obtain. The people who have historical knowledge about what may be contained on the media and about what types of hardware and software were originally used to save the data may leave the company.

The recovery of data seems to have a diminishing return. As older and older media is accessed the effort to find and use the software and hardware required increases while the perceived value of the data decreases. For example a CAD file from the late

1980's does not contain much more information than an archived paper sheet, while a CAD file from the mid 1990's contains much information not shown on the paper sheet that may be valuable for a new project.

An understanding of the nature of the data is invaluable. For example, when archiving a word processing document all that needs to be saved is the file and perhaps some information about the software required to read the file. In contrast, a CAD file can be very different. For anything beyond the simplest of designs most CAD files actually consist of a series of files. There will be the main file and perhaps a common base map that lies outside the main file directory as well as some references to common detail drawings that may be in even another directory. If this is not taken into consideration when archiving CAD files, much of the information can be lost.

Decisions and compromises must be made between preservation of just the information and the information intact with its formatting. While it is true that there are some excellent translation utilities, nearly all translations result in some loss of information. The only way to ensure complete preservation of formatting is to use original software which sometimes means using original hardware as well.

If possible keep copies of old software and the hardware needed to run it. Insure that documentation such as product keys is kept with the software so that it does not become unusable. Also take care that the media is stored so that it will not degrade. It would be a good idea to make backup copies of the software.

Many professional firms operate on the concept of billable time, time that can be billed to clients. Archiving information and maintaining the archives would generally not be billable and would be considered overhead. Therefore the support of upper level

management is imperative to insure that project managers cooperate with the archiving process. Also it is important that the archiving process and the ongoing maintenance of archives is made the responsibility of one person.

As storage space has increased and become less expensive and because modern operating systems allow almost unlimited name lengths, users have become less disciplined. Many unnecessary files such as redundant copies of photographs remain in the project directories. Also, users are more likely to name files according to their whims. This makes future searches more difficult. Clear, strictly enforced naming conventions, and mandates for purging unnecessary files prior to archiving will result in smaller more easily searched archives.

The only certainty in trying to preserve digital information is that things will change and the change can be very rapid. The most important part of any preservation and archival plan is continuous review. One way to ensure this is to keep the archives available and immediately useful. The energy and expense to recover information and create an archival system will be wasted if once again the information is stored out of sight in some back room to once again be forgotten.

Bibliography

Bogart, J. W. C. V. (1997). What can go wrong with magnetic media? *Publishing Research Quarterly*, (4) Winter 96/97, 65-77.

Caplan, P. (2004). Building a digital preservation archive: Tales from the front. *Vine, 34*(1), 38-42.

Green, R. (October 2005). CAD manager – Document Archiving Strategies. *Cadalyst, 22*(10), 36-38.

DAT Roadmap. Retrieved March 17, 2006, from DAT Manufactures Group Web site: http://www.datmgm.com/datroadmap.htm

Mallinson, J. C. (1990). Magnetic tape recording: archival considerations. *Mass Storage Systems, 1990. 'Crisis in Mass Storage'. Digest of Papers., Tenth IEEE Symposium on*, 58-61.

Manes, S. (February 2006). Keeping Our Bits about Us. *Forbes, 177*(4), 60-62.

O'Donnel, D. (2004). The Doomsday machine, or If you build it, will they still come ten years from now?" *Electronic Medievalia,* 7. Retrieved March 4, 2006, from Electronic Medievalia Web site: http://www.mun.ca/mst/heroicage/issues/7/ecolumn.html

kbadmin. (2004) *EMC Retrospect Knowledgebase.* Retrieved March 17, 2006, from EMC^2 / Dantz Web site: http://kb.dantz.com. Article # 6054

Rothenberg, J. (1995). Ensuring the Longevity of Digital Documents. *Scientific American, 272*(1), 42-48.

Shepard, T and MacCarn, D. (2001). *The Universal Preservation Format: A Recommended Practice for Archiving Media and Electronic Records*. Boston, MA: WGBH Educational Foundation. Retrieved March 1, 2006, from WGBH Educational Foundation Web site: http://info.wgbh.org/upf/pdfs/20010818-UPF_RP.pdf

Slattery, O., Lu, R., Zheng, J., Byers, B. and Tang, X. (2004). Stability Comparison of Recordable Optical Discs – A Study of Error Rates in harsh Conditions. *Journal of Research of the National Institute of Standards and Technology 109*(5), 517-524. Retrieved February 16, 2006, from National Institute of Standard and Technology Web site: http://nvl.nist.gov/pub/nistpubs/jres/109/5/j95sla.pdf

Waters, D. and Garrett, J. (1996). The Challenge of Archiving Digital Information. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* (pp. 5-10) Washington, DC: Commission on Preservation and Access.

Appendix A – Summary of Found Media

Summary of Found Media

Media Found	Approximate amount of data stored	Approximate Range of Dates in Use	Nature of information	Recovery likelihood
5 ¼ Floppies	254 mb	1989 - 1996	Mostly word processing and some spread sheets	Very High The data stored on these media can be easily read using the appropriate drive installed on a modern machine. Most files can be opened using a translator or by using the modern equivalent application.
3 ½ Floppies	88 mb	1991 - 2001		
Iomega Zip disks (100mb)	200 mb	1995 - 2003	Mostly archived projects or data sets such as base maps.	
CDs	2000 mb	1995 - 2003		
¼ inch audio tape	13 gb (This number would be high because much of this is redundant full backups.)	1992 - 1995	Most likely backups of individual hard drives for Windows computers	Low to Moderate The hardware is on hand but there has been no success yet identifying the needed software
DC 2120 ¼ in 120 mb mini data cartridge	8 gb (This number would be high because much of this is redundant full backups.)	1994-1997	Most likely backups of the accounting system	Low There is no hardware on hand and the required software has not been identified. Because they are backups of the accounting system they ore of little value. The nature of the system retains all the historical accounting information within the internal database.
Syquest SyJet 1.5 GB cartridge	30 gb (This number would be high because much of this is redundant full backups.)	Late 1990's	Unknown the cartridges are poorly labeled and the contents are a mystery	Moderate The hardware is on hand but the required software has not been identified.
DAT DDS	10 gb (This number would be high because much of this is redundant full backups.)	1996	Backups of Macintosh computers done by Retrospect	Very high. These tapes seem contain the most potentially valuable information and the hardware and software are on hand.
DAT DDS-2	44gb	1996-1997		
DAT DDS-3	660gb	1997-2003		

Appendix B – Project Archiving Standard Operating Procedure

Standard Operating Procedure

Title	Project Archiving
Description	Prescribes the procedure for creating and maintaining electronic archives of completed projects.
Responsible Parties	IT Manager, Project manager, Marketing Director, all project personnel , CADD operator
Related SOP	
Maintained By	IT Manager
Review Schedule	Annually
Last Reviewed	March 18, 2006

Procedures

Project Manager

Ongoing

- Reviews assigned projects at least monthly to determine which projects are closing

Project Closure

- Informs the IT Manager and all the personnel assigned to the project that the project is closing
- Insures that all assigned personnel complete their responsibilities as related archiving the project as prescribed by this SOP

IT Manager

Ongoing

- Maintains the hardware and software required to hold and serve the archive directory on a network volume
- Performs a backup according to the backup SOP anytime there is a change in the archive

Annually

- Reviews this SOP annual and updates as needed
- Informs all personnel involved of any changes to this SOP
- Queries all Metadata.xml files in archive directory to insure that applications and hardware required to read and print all data is still available
- Performs needed translations or maintenance required to insure that all data within directory is
- Insures the hardware holding the archive directory is maintained to include updating and replacing as needed.

Project Closure

- Insures that all associated files are in the project directory

- Converts all email PST files to individual text files
- Reviews the files to create a list of all the applications and versions used to create files.
- Creates and saves within the project directory a Metadata.xml file (see attached) to fully describe data stored in project directory
- Moves project directory to the archive directory on a read only network volume
- Creates current backup of archive directory

CADD Operator

Project Closure

- Ensures that a copy of all referenced files is placed within the project directory and that references are updated within the drawings
- Note paths and drive letters so that entire directory structure can be replicated
- Ensure plot files are saved within the project directory in HPGL2 or PDF format
- Ensure scanned record drawings are stored within the project directory in common tiff format
- Ensure a copy of survey data is stored within the directory when applicable.

All personnel associated with the project

Ongoing

- Insure scanned copies of all items that exist only in hardcopy are scanned as PDF documents and placed within the project directory

Project Closure

- Export PST containing all email correspondence file into t:\Client\Project\Corresspondence\email

Marketing Director

Project Closure

- Updates or creates a description of the project and delivers it to the IT Manager to be included in the Metadata.xml file

Appendix C – Sample Document Metadata File

```
<Project>
      <Proj_ID></Proj_ID>
      <Client_ID></Client_ID>
      <Client_Name></Client_Name>
      <Start_Date>
            <Year></Year>
            <Month></Month>
            <Day></Day>
      </Start_Date>
      <End_Date>
            <Year></Year>
            <Month></Month>
            <Day></Day>
      </End_Date>
      <Proj_Descript></Proj_Descript>
      <Personnel >
            <Proj_Manager>
                  <First_Name></First_Name>
                  <Last_Name></Last_Name>
            </Proj_Manager>
            <Engineer>
                  <First_Name></First_Name>
                  <Last_Name></Last_Name>
            </Engineer>
            <CADD_Operator>
                  <First_Name></First_Name>
                  <Last_Name></Last_Name>
            </CADD_Operator>
      </Personnel >
      <Application>
            <WordProcess></WordProcess>
            <SpreadSheet></SpreadSheet>
            <CADD></CADD>
            <Design></Design>
      </Application>
</Project>
```

Appendix D – Resources and Applications

Retrospect

Backup application that was originally used to create the backup sets on the DDS and DDS-2 tapes. Version 5.0 was used to recatolog the tapes and recover the information. Fortunately, Retrospect remains backwards compatible fro reading tapes created with earlier versions. However even though Retrospect is available in both Windows and Macintosh versions, an installation on one platform will not read tapes created on the other.

The Retrospect web site at http://www.emcinsignia.com/supportupdates/technical/retrospect/ was helpful in providing updates for the application and for the tape drive. The site also provided information on compatible SCSI cards.

MacLink Plus

MacLinkPlus from DataViz was used to translate word processing s and spreadsheet files to Wood 2002 and Excel 2002 formats.

Information about MacLinkPlus is found at http://www.dataviz.com/products/maclinkplus/

Ebay

In order to install an existing tape drive I needed a SCSI card which was obtained through eBay. A SyQuest SyJet drive was also obtained to attempt to read the SyJet cartridges that were found.

Website http://www.ebay.com/

Atto

The Atto SCSI cards were one of the cards recommended by the Retrospect web site. The Atto website provided drivers and utilities to configure the card to operated properly with the Macintosh

Website: http://www.attotech.com/scsihostadaptertechnology.html