TITLE: Archiving Costs for Engineering Records

AUTHOR(S): Roger A. Stutz, NIS-9
Bruce C. Lamartine, MST-7

SUBMITTED TO: 41st Annual Meeting AACE
Dallas, Texas, July 13-16, 1997

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes.

The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
Estimating Archiving Costs for Engineering Records

ABSTRACT

Information technology has completely changed our concept of record keeping for engineering projects -- the advent of digital records was a momentous discovery, as significant as the invention of the printing press. Digital records allowed huge amounts of information to be stored in a very small space and be to examined quickly. However, digital documents are much more vulnerable to the passage of time than printed documents because the media on which they are stored are easily affected by physical phenomena, such as magnetic fields, oxidation, material decay, and by various environmental factors that may erase the information. Even more important, digital information becomes obsolete because, even if future generations may be able to read it, they may not necessarily be able to interpret it.

Engineering projects of all sizes are becoming more dependent on digital records. These records are created on computers used in design, estimating, construction management, and construction. The necessity for the accurate and accessible storage of these documents, generated by computer software systems, is increasing for a number of reasons including legal and environment issues. This paper will discuss media life considerations and life cycle costs associated with several methods of storing engineering records.
Introduction

Information technology has completely changed our concept of record keeping -- the advent of digital records was a momentous discovery, as significant as the invention of the printing press, because it allowed huge amounts of information to be stored in a very small space and be to examined quickly. However, digital documents are much more vulnerable to the passage of time than printed documents because the media on which they are stored are easily affected by physical phenomena, such as magnetic fields, oxidation, material decay, and by various environmental factors that may erase the information. Even more important, digital information becomes obsolete because, even if future generations may be able to read it, they may not necessarily be able to interpret it.

There are many factors that effect the cost of archiving records for engineering projects. The older methods of storing and indexing records using paper and microfilm has been changing over the last few years. More and more projects are being developed using computers systems that depend on magnetic and optical storage systems that use less then robust materials for the archiving of design, cost, engineering, and construction information. These records are often required in court cases during and after construction. In addition, there often is a need for these records tens to hundreds of years from the time of construction. The nature of the project will effect the length of time records need to be kept about a project. The nature of the media used to store information will affect the costs associated with maintaining them in a usable state.

When one considers the large numbers and wide variety of construction projects, there are a number of these projects that require not only the standard consideration for the retention of records but require extremely long retention of these records. Current examples include nuclear waste facilities where records may be needed in hundreds to thousands of years after completion of construction. Another example is the recently completed channel tunnel where the owners expect to continue operations for many years. Construction projections from the 1930s that also have long records retention requirements include Hoover dam, the Golden Gate bridge and many of the WPA projects. Going back to the late 1800s and early 1900s there are projects like the London underground system, railroads systems, and major sea locks such as the Panama Canal that are still in use.

Not only must the data be recoverable over long periods of time. it must also be the unchanged original information with added information about changes so the history of modifications and enhancements can be traced over long periods of time - centuries. For data storage over hundreds to thousands of years, there is reasonable concern about effects of man-made or natural disasters. Fires and floods have destroyed many major data bases. The great library at Alexandria burned in about 642 AD. Exactly who was responsible is debated, but an irreplaceable storehouse of knowledge was almost totally destroyed.
A new technology, HD ROM (ion micromilling), discussed below would survive most such disasters while tape, optical, and CD technologies would fail. The technology is not dependent on magnetic fields and can withstand high EMP and RF environments. This technology can use media like stainless steel. The melting point of stainless steel is approximately 2500 degrees F (1370 degrees C). Most building fires burn at about 1300 degrees F (700 degrees C), thus the probability of data survival is quite high. There are circumstances in which sustained fires can reach higher temperatures. Again, choice of materials for storage medium can be designed to resist the most aggressive fires.

**Comparisons of technology options**

Given the increasing need for long term digital and analog storage of information about construction projects and the overriding concerns about the cost of such record keeping new methods will be required. With these concerns in mind, one can look for methods where information has been preserved for very long periods of time. A primary example is paper. Paper has been used for several thousands of years and has proved, on the whole, to be a reasonably stable media for the storage of information. Paper provides a means of storing information in a native language format that can be understood by large numbers of people. The printing press expanded the role and dependence on paper as a means for storing information. Photocopying again expanded the role of paper. Nevertheless, paper has limitations that reduce its usefulness as a long term medium. European archivists are now beginning to see deterioration in 500-1000 year old documents produced on low acid paper. Many documents produced within the last hundred years on common paper (high acid content) are so deteriorated that they are even hard to micro-film. These papers are the common papers used to store information about construction projects from before 1930. Paper is an expensive method of storing information in both direct costs and life cycle costs. Working copies must be made to help insure the survival of the originals. Paper is a very low density storage medium that is slow in writing and reading speed and costly to index.

**Limitations of paper**
- Fire
- Mold
- Environmental reactivity
- Slow information search and read rates
- Fading of inks
- Media life of about 1000 years depending on storage conditions
- Need to make working copies to preserve the originals
- Low data density

**Advantages of paper**
- Native language
- Easy to copy in recent times
- Does not require special equipment to read or write
The next example, using more compact media, is the use of microfilm as a storage media. Microfilm provides a native language capability. It also provides a method of reducing the volume required to hold the information; however it still does not have a very high density compared with magnetic or optical systems. Copies are easy and reasonably cheap to make but still harder than making copies of paper documents. Microfilm is accepted in courts and generally is considered a replacement for paper documents. It is relatively hard to tamper with the information that is copied to microfilm. However, microfilm generally has the same disadvantages as paper with the addition that it requires an enlargement device to read the information. Each copy of the master microfilm information results in the loss of resolution.

Limitations of micro film
- Fire
- Mold
- Environmental reactivity
- Slow information search and read rates
- Loss of resolution
- Requires an enlargement device to read
- Requires chemicals for processing the film
- Media life of about 50 - 100 years depending on storage conditions
- Need to make copies to preserve the originals
- Need to make working copies or transfer to computer format

Advantages of micro film
- Reduced space requirements
- Hard to modify information
- Accepted by legal system
- Native language

The development of magnetic storage was next in the technology chain that provided a means to store ever increasing amounts of information on compact media. Magnetic storage can be divided into two main groups - tape and disk. Both can provide high data densities in the order of 1 Gbit per square inch. Magnetic technologies are reaching the limits for storing information. Furthermore, large improvements in areal density are not anticipated because of the minimum size required for a stable magnetic domains. The main advantages to magnetic media are the ability of machines to quickly read and write information, to index information, to store large amount of information, to update and append information depending on formats, and to correct some errors with built-in error correction information. However, there are a number of disadvantages to the digital storage of documents over printed documents. First, magnetic tape and disk media that are used to store digital documents are easily affected by physical phenomena, such as magnetic fields, oxidation, material decay, and by various environmental factors that may erase the information. Even more important, digital information becomes
obsolete because, even if future generations may be able to read it, they may not necessarily be able to interpret it. This is the result of requiring a bit stream interpreter to convert the information from a sequence of one and zero to numbers, text and images. Another concern for the archivist and future engineers is that information stored on magnetic medium can be changed without leaving any indications of a change. This last problem will effect the acceptability of digital information in the court systems.

Limitations of magnetic storage
- Fire
- Local RF and EMP fields
- Environmental reactivity
- Overwrite capability (advantage in some applications)
- Magnetic fade
- Requirement for bit stream interpreters
- Medium life of about 2-10 years depending on storage conditions (need to rewrite information every few years)
- Not native language
- Not uniformly accepted by courts
- Mold in some climates

Advantages of magnetic storage
- Rapid reading and writing
- Rapid indexing and searching
- Relatively high data densities
- Error correct capabilities

CD-ROM storage systems are a recent addition to storage methods. They provide a non-magnetic method for storing information. They can support reasonably high data densities and can be divided into three general classes. The first are WORM devices that are write once - read many times. The second class is read only. The third class is erasable CD-ROM. All device classes provide a reasonable method to store information at relatively high data densities in digital formats. There are advances being made in this area and densities will continue to increase. Some classes of CD-ROM are accepted in some court systems but there is no uniform acceptance of digitally stored information. The main draw-back to these devices is still the medium. The medium is subject to many limitations including low melting points, destruction by fire and common solvents. The erasable types pose, in addition, many of the same problems as magnetic methods related to their acceptability to the court systems.

Limitations of CD-ROM storage
- Fire
- Environmental reactivity
- Requirement for bit stream interpreters
Optical storage systems are a recent addition to storage methods. They provide a non-magnetic method for storing information much like CD-ROMs but at high data densities. These densities are approaching 2-5 Gbytes on a 5.25 inch disk. They can support high data densities and can be divided into three general classes. The first are WORM devices that are write once - read many times. The second class is read only which has the highest data density. The third class is erasable optical disks which have the lowest data density. All device classes provide a reasonable method to store information at relativity high data densities in digital formats. There are advances being made in this area and densities will continue to increase a little over current densities. Some classes of optical disks are accepted in some court systems but there is no uniform acceptance of digitally stored information. The main drawback to these devices is still the medium. The medium is much like a current CD-ROM and is subject to many of the same limitations; for example, optical media will melt at relatively low temperatures. The erasable types pose many of the same problems as magnetic methods related to their acceptability to the court systems.

Limitations of optical storage
- Fire
- Environmental reactivity
- Requirement for bit stream interpreters
- Not native language
- Media life of about 10 years depending on storage conditions (need to rewrite information every few years)
- Slow data writing
- Not uniformly accepted by the courts

Advantages of optical storage
- Relativity high data densities
- Relativity rapid reading
- Error correction capabilities
- Not affected by EMP or RF

Advantages of CD-ROM storage
- Relativity high data densities
- Relativity rapid reading
- Error correction capabilities
- Not affected by EMP or RF
The last method to be compared, HD-ROM, is new and was developed at Los Alamos National Laboratory. It became commercially available in 1997. The process can be compared with methods that were developed before the time of the pharaohs for long term storage of information. In those times men chiseled messages in stone as a means of creating enduring records. To be sure, these glyphs have imparted the information of their sculptors for readers millennia later. Some of the important factors that allow the information to be understood millennia later relate to the fact that the information is written in a native language and are written on durable medium. The Rosetta stone provided the key that allowed translation of one native language to another. Greek and Latin text can still be read without the need for a “rosetta stone” because they were written in a still active native language. HD-ROM uses an ion beam to chisel information into durable media. In fact, the durability of this high-density technique is so great that one observer suggested that “long-term” should be replaced by “geologic,” when describing the longevity of this data storage method.

This method allows data to be written in native languages like text on paper, direct human-viewable images like engineering drawings and photos, and in binary formats. The information types can be mixed on the same media. Therefore, it is possible to include in a human-viewable format, an engineering drawing just as printed along with a bit-map image of the drawing. Instruction on the bit-stream interpreter required to read the binary information can also be included if desired. The method allows the best of both the magnet and paper worlds to exist on a durable media that will provide the longevity needed to meet the needs of current and future engineers at a reasonable life cycle cost.

Limitations of HD-ROM

- Highest data densities currently requires Scanning Electron Microscope for reading -- SEM are large devices
- Large size of writer (writing can outsourced)

Advantages of HD-ROM

- Very high data densities 23 Gbits - 11,000 Gbits per square inch, higher densities are possible
- Media life of thousands of years
- Not effected by EMP or RF
- Not environmentally reactive depending on material used
- Several reading methods are available
- Native language formats are possible but not required
- Can have mixed densities on the same media
- Can have mixed data formats on the same media
- Rapid reading and writing
- Rapid indexing and searching
- Error correction capabilities
- WORM device (good for archiving information)
Cost Comparisons

The following tables compares the main cost factors that control the cost of storing large amount of data as could be required for major constructions projects. The model assumed that the data would be retained for a 200 year period and that the data would be rewritten at the end of the media life. The life of the tape and optical systems were set at 10. Many large users of tape and optical system rewrite the data at 1 or 2 years cycles. This model did not take into consideration the need to wind and rewind tapes each year. High quality micro file can have a life approaching 500 years if stored in expensive climate controlled areas. It should be pointed out that some of the technologies may be replaced over time but there are costs associated with the conversion and refiling of information from one storage technology to another. While the HD-ROM has a high initial cost it has the lowest life cycle and has the best chance for providing access to the data in the future. These are direct costs and do not include profits, escalation, taxes, or the cost of climate controlled storage.

<table>
<thead>
<tr>
<th>Item</th>
<th>HD-ROM</th>
<th>Removable disk</th>
<th>Tape</th>
<th>Micro film</th>
<th>CD-ROM/Optical</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years before first write</td>
<td>1000</td>
<td>10</td>
<td>10</td>
<td>50</td>
<td>1000</td>
<td>100</td>
</tr>
<tr>
<td>Cost of media for master</td>
<td>10</td>
<td>100,000</td>
<td>10,000</td>
<td>1,000,000</td>
<td>10,000</td>
<td>300,000</td>
</tr>
<tr>
<td>Cost of media for backup</td>
<td>10</td>
<td>100,000</td>
<td>10,000</td>
<td>1,000,000</td>
<td>10,000</td>
<td>300,000</td>
</tr>
<tr>
<td>Cost to write first copy</td>
<td>1,912,500</td>
<td>21,080</td>
<td>781,000</td>
<td>726,000</td>
<td>571,000</td>
<td>1,080,000</td>
</tr>
<tr>
<td>Cost to make backup</td>
<td>1,000</td>
<td>210,800</td>
<td>781,000</td>
<td>726,000</td>
<td>571,000</td>
<td>1,080,000</td>
</tr>
<tr>
<td>Cost to rewrite master based on number to times required</td>
<td>0</td>
<td>4,005,200</td>
<td>14,839,000</td>
<td>217,800</td>
<td>10,849,000</td>
<td>1,080,000</td>
</tr>
<tr>
<td>Cost of replacement media</td>
<td>0</td>
<td>1,900,000</td>
<td>190,000</td>
<td>3,000,000</td>
<td>190,000</td>
<td>300,000</td>
</tr>
<tr>
<td>Cost of storage space for 200 years</td>
<td>2,000</td>
<td>100,000</td>
<td>200,000</td>
<td>2,200,000</td>
<td>100,000</td>
<td>32,000,000</td>
</tr>
<tr>
<td>Labor to file master</td>
<td>25</td>
<td>250</td>
<td>25</td>
<td>5,000</td>
<td>250</td>
<td>250,000</td>
</tr>
<tr>
<td>Labor to file backup</td>
<td>25</td>
<td>250</td>
<td>25</td>
<td>5,000</td>
<td>250</td>
<td>250,000</td>
</tr>
<tr>
<td>Labor to retrieve &amp; refile for rewrite</td>
<td>0</td>
<td>4,750</td>
<td>475</td>
<td>15,000</td>
<td>4750</td>
<td>250,000</td>
</tr>
<tr>
<td></td>
<td>--------</td>
<td>----------------</td>
<td>------</td>
<td>------------</td>
<td>----------------</td>
<td>-------</td>
</tr>
<tr>
<td>Total</td>
<td>1,916,570</td>
<td>6,632,060</td>
<td>16,816,535</td>
<td>10,855,050</td>
<td>12,306,260</td>
<td>36,890,100</td>
</tr>
<tr>
<td>Cost per Mbyte per year</td>
<td>.01</td>
<td>.03</td>
<td>.08</td>
<td>.05</td>
<td>.06</td>
<td>.18</td>
</tr>
<tr>
<td>Cost per byte per year</td>
<td>9.58E-09</td>
<td>33.1E-09</td>
<td>84.1E-09</td>
<td>54.3E-09</td>
<td>61.5E-09</td>
<td>184.9E-09</td>
</tr>
</tbody>
</table>
The following table is another estimate of the cost to store one Terabyte of data for 200 years using services provider to write the data, assuming 100,000,000 8.5 by 11 inch pages at 6,000 bytes per page and 50,000 E size drawings at 8,000,000 bytes per drawing. The assumed life for the medium are as follows:
- paper 100 years
- micro film 50 years
- magnetic disk 10 years
- magnetic tape 10 years
- optical disk 10 years
- CD-ROM 10 years
- HD-ROM 1000 years.

These costs do not include escalation, taxes, or the cost of climate controlled storage rooms.

<table>
<thead>
<tr>
<th></th>
<th>HD-ROM</th>
<th>Disk</th>
<th>Tape</th>
<th>Microfilm</th>
<th>Paper</th>
<th>CD-ROM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost/page</td>
<td>0.05</td>
<td>0.02</td>
<td>0.03</td>
<td>0.08</td>
<td>0.1</td>
<td>0.05</td>
</tr>
<tr>
<td>Cost/E-drawing</td>
<td>0.4</td>
<td>0.2</td>
<td>0.3</td>
<td>0.8</td>
<td>4</td>
<td>1.00</td>
</tr>
<tr>
<td>Cost for 100,000,000 pages &amp; 50,000</td>
<td>5,020,000</td>
<td>2,010,000</td>
<td>3,015,000</td>
<td>8,040,000</td>
<td>10,200,000</td>
<td>5,050,000</td>
</tr>
<tr>
<td>Cost of backup</td>
<td>1,000</td>
<td>2,010,000</td>
<td>3,015,000</td>
<td>8,040,000</td>
<td>10,200,000</td>
<td>5,050,000</td>
</tr>
<tr>
<td>Rewrite data</td>
<td>0</td>
<td>38,190,000</td>
<td>57,285,000</td>
<td>24,120,000</td>
<td>10,200,000</td>
<td>95,950,000</td>
</tr>
<tr>
<td>Storage costs</td>
<td>2,000</td>
<td>100,000</td>
<td>200,000</td>
<td>2,200,000</td>
<td>32,000,000</td>
<td>100,000</td>
</tr>
<tr>
<td>Total life cycle costs</td>
<td>5,023,000</td>
<td>42,310,000</td>
<td>63,515,000</td>
<td>42,400,000</td>
<td>62,600,000</td>
<td>106,150,000</td>
</tr>
<tr>
<td>Cost/byte/year</td>
<td>2.51E-08</td>
<td>2.12E-07</td>
<td>3.18E-07</td>
<td>2.12E-07</td>
<td>3.13E-07</td>
<td>5.31E-07</td>
</tr>
<tr>
<td>Cost/MB/yr</td>
<td>0.0251</td>
<td>0.212</td>
<td>0.318</td>
<td>0.212</td>
<td>0.313</td>
<td>0.531</td>
</tr>
</tbody>
</table>

**Summary**

Technology has provided ideas that will solve many of the cost and life cycle concerns related to the long term storage of engineering records. Considerations should be given to the length of storage required, the medium used to record the information, the amount of information, and the life cycle cost of storing the information for each project. It makes no sense to use expensive storage methods and medium for a construction project when the life of the project is only a few years and the information storage space requirement are small. On the other hand, it is not cost efficient to use non-durable storage methods for a long term project.
The new HD-ROM technology provides data densities that are approaching the atomic limits for the storage of information. Thus, this system is not likely to be replaced with higher density or more durable storage systems or media. The fact that the HD-ROM system supports several methods reading data provides extra insurance that the information can be retrieved even as improved reading methods are developed without the need for a conversion from one system to another.

**Suggested readings**


Terry Cook, It’s 10 o’clock: Do you know where your data are?, *Technology Review*, vol. 98 number 1, pg. 48-53