ELECTRONIC SUBMITTALS

CSI-HOUSTON

TECHNICAL COMMITTEE

AUTHOR: Brett M. Wilbur CSI, CCS, AIA

CONTRIBUTORS: Tom Atwell, Amy Brom, Ken English, Di Ann Hassloch, Scott Lestus, Lance McHaney, John Purcell, Greg Roberts

4/09

PART II

In Part I, we addressed electronic file transfer and review of electronically issued submittals. Data longevity is not without its own set of issues which need to be reviewed and addressed in order to make educated decisions about storage requirements, storage media type, storage location, file size, and accessibility. These decisions will effect how, when and why you might chose to use electronic file transfer protocols, electronic construction documents and electronic submittals. It must be understood that data longevity can be limited both by media deterioration and by technological obsolescence.

FILE STORAGE

An understanding of the characteristics of the currently available common storage media will assist in decisions on data storage. There are two main technologies in use for storing digital data: **volatile** and **non-volatile**. Volatile storage loses its data when the power is switched off, (for example, the random access memory (RAM) inside a computer), whereas non-volatile storage will retain the data for a period of time without the need for a power supply (for example, a CD-R).

Non-volatile storage consists of magnetic media (tape, floppy and zip discs, hard discs), Solid State Media (USB memory sticks), and Optic media (CD-ROM, DVD).

Magnetic media are the most common types available and consist of a surface, usually a plastic, covered with a magnetically sensitive coating.

- **Floppy** discs and **Zip** discs have a typical shelf life of 10 years but can only be relied on to store data for 3 to 5 years before environmental factors start corrupting the data (for example, leaving a floppy disc close to a main cable or electric motor).
- **Hard discs** have more mechanical parts to go wrong, so their working life is limited to 3 to 5 years. Although many drives last longer, the risk of a hard disc "crash" increases considerably after 5 years. With removable hard discs, if the drive mechanism goes wrong the data can at least be read from the disc using another drive.
- **Tapes** are designed for data backup and made to a high standard, so, provided they are not used too much (the surface coating wears thin) and they are stored in a magnetically shielded

environment (usually a steel cabinet) they can be expected to retain data reliably for up to 10 years.

Solid State media memory devices retain data even when power is removed. The technology has developed rapidly due to the increased use of digital cameras, but in the context of data storage and backup, it is the USB memory stick that is commonly used as a higher-capacity replacement for the floppy disc. Although these devices are now relatively cheap and reliable, their performance and reliability does vary, and it doesn't always pay to get the cheapest memory stick available. A hardware failure can make the entire contents of the 'stick' inaccessible. It is also possible for environmental conditions to slowly corrupt the data, but under good conditions, a memory stick should retain data at least as long as a floppy disc, typically about 5 years.

Optic media, such as CD-ROM and DVD discs are written to and read optically using a laser to reflect off minute pits in the surface of the disc. Drives capable of reading CD-ROMs are sold on almost all computers so the CD-ROM is a good way of making data available to most computer users. CD writer drives are becoming a standard feature on most computers and the price of the blank discs is decreasing, making CD-ROMs a cheap way of backing up as well as exchanging data.

CD-ROM discs normally contain four layers consisting of a pre-grooved substrate, a dye layer, a metal layer, and a protective coating. All but the metal layer contain organic compounds that can degrade as a result of changes in their chemical structures or because of unstable defects that grow in size. Depending on susceptibility, wear, and environment, various layers of optical discs may undergo oxidation, hydrolysis or mechanical stress, leading to damage (variously described as "rot" or "pin holes" or "mirroring", etc.). Exposure to radiation, inks, other chemicals, water, or pollutants can adversely affect CD quality. In addition, CD-ROM ageing processes can be accelerated by high temperature and humidity environments. Although environmental degradation should be avoided, a major cause of deterioration is improper user handling based upon overconfidence in the robust construction and error correction of CD-ROM media. Mechanical stress induced by rapid environmental changes may result in excessive differential expansion of the various layers and delamination or excessive decomposition. These and other ageing modes limit media longevity. Handling or storage conditions may degrade good discs, especially if the thin, vulnerable protective coating on the label surface is flawed or damaged.

Though CDs are unaffected by magnetic fields they are susceptible to bad handling such as scratches. In addition, high quality media that is recorded in poor writers may fail interchange standards. It is generally accepted that CD-ROMs are a better medium than a magnetic device for long-term storage of data but it remains to be seen whether the data will remain retrievable and for how long.

Since the late 1990's, a common misrepresentation of findings from a National Media Laboratory study has created a myth that the life expectancy of most CDs is about 5 years. [1] However, natural and accelerated ageing studies by the Library of Congress showed relatively little change in disc quality over the course of ten years for the specific population of discs, though some discs did fail with regard to successful playback (4 %). [2] Other tests have come to different conclusions, stating that with proper care, CD's should be readable one thousand times or more. [3] It is believed that at the actual life

expectancy of CD to maintain reliable, readable data may well be over 50 to 200 years. [4] However, physical lifetime is rarely the limiting factor, since at any given point in time, a particular format of a given medium can be expected to become obsolete within no more than 5 years.

Regardless of your means of storage, there are also issues with backward compatibility of the operating system and software. In other words, software programs used to store the data may not be available for access of the data years in the future.

Technological developments are rapid and it may well be the case that, even though a disc or tape has stored data reliably, no device is available to read the data in 5 or 10 years time, let alone 75 or 100 years! If archival storage is needed, it is likely that the data will have to be read off the existing medium and written to a new one periodically; hence, making copies of copies. Maintaining, managing, and locating these files becomes an issue in itself.

There are a number of factors that need to be considered when deciding where to store data: who will be accessing the data and how? Will the data be accessed from a number of different locations or by a number of different people? How critical is the data to your job? The following notes and examples are intended to give guidance:

Local storage, for example on a hard disc of a desktop computer, is good for personal work which is not needed by others or is not project related. As an example, most email falls into this category. Data kept on a local drive is normally less secure than on a network drive, though the latter is only as secure as the user's password, which should be known only to the user.

Central storage, that is, on a server run by the IT Department, is secure and is typically backed up to tape on a regular basis. It is therefore the place to store information if you work from more than one computer, if you share data with others, or if the data has value to you or your project. Central storage is finite and may be used by many users, so most directories have a maximum amount of data that can be stored in them; however, data retrieval is easy and instantaneous. With central storage, the security and backup service is provided by the IT Department.

The size of storage needed for data files depends partly on the file's content and partly on the file format used.

File content is usually dictated by the application being used. Plain text is compact; formatted text such as Word files uses more space; vector graphics files, for example, AutoCAD files are reasonably compact; but bitmap graphics files from scanned documents or programs like Photoshop can be large, as can digitized video files.

File format can make a big difference to the space used to store data. Files that contain formatting information can often become quite bloated. Word files can often be compressed to a quarter of their original size which gives an indication of the wasted space they contain. There are a number of compression programs available for PC and Mac that can reduce file size for longer term storage and transmission.

How Much Storage Space Do You Need?

Number of Bytes	What That Relates To
1 Byte	One character (letter or number)
1KB (Kilobyte) 1000 bytes	3 or 4 typed manuscript style pages
1MB (Megabyte) 1,000,000 bytes	Average size of a novel (300-400 pgs) or 1 diskette
1GB (Gigabyte) 1,000,000,000 bytes	Approximately 20 sets of encyclopedias
1TB (Terabyte) 1,000,000,000,000 bytes	A small library (approx. 5,000 books)

How quickly will the data need to be accessed? If data is needed quickly and frequently then it will need to be stored on a hard disc. This could be on an individual computer if other considerations make this appropriate, or on a server. Data that is needed less frequently or less quickly could go on to removable disc such as a CD-ROM, though it would be more difficult to up-date the information regularly. However, this disc may be networked or used individually. For infrequently accessed data that is not needed immediately, such as data used for facility management like record documents, tape storage might be appropriate, especially if large quantities of data are involved.

How freely available does the information need to be? If the information needs to be accessed from more than one computer, by more than one person, or needs to have some sort of selective access applied, then the data is best kept on a server, whether this is a hard disc or a networked CD-ROM.

If access to the data needs to be restricted to individuals or groups, the data will need to be on a central server with arrangements made with the IT department.

RECOMMENDATIONS

Some good practices for storing and archiving electronic data:

- *Determine owners storage media format and location to determine best transfer protocol for facility management;*
- *Store data on a central server for rapid and frequent access to shared data.*
- *Store data on a local hard disc only if the data is to be accessed by one individual at one computer.*
- *Store data on CD-ROMs if less frequent access to the data is needed.*
- *Store data on tape if it might possibly be needed after a short delay.*
- *Use HTML email for information that needs to be available regardless of computer and location. Deliver this information using a Web server.*

- *Compress your data before archiving it.*
- *Choose a file format that gives adequate quality without wasting space.*
- *Use well-established, open standards for file formats in preference to proprietary formats that may not be around in 5 to 10 years.*

REFERENCES

1. http://www.slais.ubc.ca/PEOPLE/students/student-projects/C_Hill/hill_libr516/longevity.htm
2. http://www.loc.gov/preserv/rt/projects/cd_longevity.html
3. CD-R. (2008, December 8). In *Wikipedia, The Free Encyclopedia*. Retrieved 18:58, December 9, 2008, from http://en.wikipedia.org/w/index.php?title=CD-R&oldid=256585258
4. http://www.slais.ubc.ca/PEOPLE/students/student-projects/C_Hill/hill_libr516/longevity.htm