Selecting a Proper Storage Device

Anne Roach, AG©
Family History Library, Information Commons

1. Purpose

Educate genealogical developers, instructors, and professional genealogists about choosing appropriate storage devices, addressing pros and cons of various storage mediums, and addressing solutions for data disaster-recovery.

2. Introduction

Creating a long-term digital storage medium continues to be a challenge for genealogists at every level. Digital Preservation is contingent upon a combination of the following co-dependent elements:

- The correct physical computer equipment
- The correct operating system for that equipment
- The correct computer program to read the data in a coherent format, compatible with the computer equipment
- The correct version of that computer program that will correctly read and organize the data compatible with the operating system
- The correct drive and drivers with which to read a data storage device compatible with the computer program
- An uncompromised version of the data storage device, or the container designed to hold the data compatible with the drive and drivers
- An uncompromised or uncorrupted copy of the data

Unlike paper copies, that, when partially damaged may be re-created and are still useable, digital data may only require a minute flaw to render all of the data unusable. The longevity of digital data, due to its short lifespan from conception to the present, is uncertain. This paper will discuss possible solutions to using and choosing digital storage devices, including information about creating a system that will best satisfy the need to increase the lifetime of digital media through regular maintenance and data disaster planning.

3. Early Computer Storage Devices

1947	Williams-Kilburn Tube: a modified cathode-ray tube with the storage capacity of 1024 bits.
1961	IBM 1301 Disk Storage Unit: disks spun at a rate of 1800 RPM with a storage capacity of 28 million characters. This unit is similar to units used today for storage, including a read-write arm for each disk.
1962	The birth of Virtual Memory, allowing computers to use storage to run outside computer programs
1962	IBM 1311 Disk Storage Drive containing the first removeable disk pack at 10 lbs per pack. Disks were spun at a rate of 1500 RPM with a storage capacity of 2 million characters. This unit used a hydrolic actuator to read and write from the disks with one head per disk.
1967	IBM 1360 Photo-Digital Storage System: Stored up to a trillion bits of information, data was stored on small strips of film developed within the machine.

Although some of these devices were soon outdated, some have been developed and have survived in advanced forms. Like earlier versions of storage devices, one major threat and benefactor to older data storage devices is newer formats. While these may create an end to old formats, improved storage methods, if updated within a reasonable amount of time may prove helpful to old computer data and provide a greater lifetime to computer data.

4. Genealogical Applications for Storage Devices

- **The Consumer:** The consumer genealogist uses storage devices to store genealogical database files, including names, dates, places, photographs, historical documents, source details, family biological information, e-mail archiving, correspondance records, family contact information, and living family information. Correct choices about storage devices impact the consumer genealogist on every level as the consumer seeks for the best way to preserve the information gathered on past, present, and future family members, and a format that will be useable to future generations and will not make the data obsolete.

- **The Producer:** The producer genealogist uses storage devices to invite consumers to use a product. This often involves creating trust with the consumer by providing useable, non-deteriorating hard copies of software, or viable downloadable versions of software. This may also involve creating an online storage system from which a consumer may view images of documents on a regular basis. The reliability and integrity of such online databases is contingent upon proper selection of data storage systems and data disaster recovery plans. Without reliable service, the producer loses consumer confidence and patronage. The producer may provide a virtual storage system to the consumer that allows the consumer to store genealogical data in an online database for collaborative purposes. Once again, the producer's business integrity rests solely on its ability to provide a reliable, non-failing product with a sound digital storage system.

  Virtual Storage allows the producer the opportunity to create a constant flow and backup of information from remote business offices and sites. This is especially important for database-producers who employ record gatherers, digital camera operators, international indexing labor forces, and on-site researchers to contribute to the consumer product. The data gathered through these remote sites may be instantly added to the data pool of the main office and processed, stored, and secured without the lag time and risk of transferring data physically. Remote offices may also use proper storage procedures to avoid the expense of leasing large storage devices on site, and avoid the time, expense, and down-time of data when leased equipment is no longer under contract.

5. Dynamic Storage versus Static Storage

a. **Dynamic Storage:** This type of storage device allows the user to work within the storage device, making changes while working within the file, and easily saving changes to a file or database. This type of storage device has many advantages for the consumer. This allows the consumer to work within genealogical databases and frequently alter the data file, and add, remove, and change information.

b. **Static Storage:** This type of storage device allows the user to save data at a specific point in time and create a non-vulnerable version of the data that may not be easily altered or deleted, as with dynamic storage devices.

6. Available Digital Storage

- **'Dynamic' Memory Cards:** Memory cards are not just used by consumers for digital cameras. These storage devices may be used in PDAs containing genealogical databases, digital images, e-mail archives, and other information. Music devices such as I-Pods and MP3 players containing memory card storage may be used as digital recorders genealogically to record oral histories and interview ancestors. Cellular phones or Blackberry devices may also contain many PDA features, and will increasingly allow for storage, online viewer capabilities, and storage for computer programs. Consumer computers may use memory cards directly in portable computers, much like 3¼ floppy disks.

  Producers may use memory card storage while gathering information onsite or in creating digital sound files of conferences that may be sent to remote business offices. This type of memory is not susceptible to magnetic deterioration, but may be damaged through digital malfunction. Memory cards are not a good choice for a permanent type of storage, but should be used as a temporary transfer of information.

- **'Static' Optical Disks:** CDs, DVDs, DVD-Rs, DVD+Rs, CD+Rs, and BlueRay Optical disks are often useful for consumer and occasionally producer storage devices. Consumers use optical disks to store data such as photographs, movies, and non-changing data. Optical disks are most frequently used by the producer for distribution of databases, or product creation and distribution. Producers often prefer consumer purchased downloads or online-viewable data versus the expense of packaging and marketing optical disk versions of a product.

  Both consumers and producers often view optical disk storage as a permanent or semi-permanent storage solution. As with most digital storage, optical disks may be corrupted in a variety of ways. First, if information was incorrectly written to the disk, the entire disk may become unreadable. Second, oxidation may create data damage. Digital data is highly volatile and with a minor flaw, all data may become unreadable and unrecoverable.

  Also, while optical disks are often seen as a long-lasting solution to digital storage, the lifetime of optical disks has not been proven. According to the National Institute of Standards and Technology, or NIST, there is no clear timeline at present by which to measure digital deterioration. In regard to digital optical media, NIST states, "The lifetime of this storage media is limited because of chemical and physical deterioration as well as the obsolescence of the format and playback technology." Many examples of digital deterioration are apparent, even among prominent repositories such as the Library of Congress. 1-10% of their 150,000 CD collection has experienced some type of significant digital failure.

Re-writable CDs and DVDs have shown deterioration in quality over time through re-writing material to them. Consumers often misunderstand the concept of re-writable optical disks and consider them a replacement for the dynamic floppy disk; however, re-writable disks are only erasable entirely, and not partially.

- **'Dynamic' Flash Memory:** Flash memory is a consumer form of memory that is often used by producers in the workplace. Flash memory is a non-volatile form of memory that allows the user to write/read without changing the integrity of the memory. While this type of storage has a high rate of usability, flash memory devices are easily damaged and often non-recoverable. This should not be used as a permanent form of storage, rather a temporary storage transfer device.

- **'Dynamic' Hard Drives**: Hard drive memory is used by producers and consumers at many levels. This may include computers, servers, and external hard drives. Hard drives also make up an important part of a data disaster recovery system and SAN networks. Hard drives consist of several dual side platters, with read/write arms that are all attached and move together across the platters at the same time. Platters are made of light aluminum alloy and coated with magnetizable material like ferrite compound applied in liquid form and spun across the plate evenly or electroplated onto the platters like chrome. Newer technology uses glass/ceramic platters because they are thinner, and more efficient at resisting heat. Disk spacers hold the platters apart.

  Unlike their early predecessors from the 1860s, hard drive disks spin at around 5400-7200 RPMs. Hard drive storage is based on magnetic particle placement, and is therefore susceptible to magnetic surge damage. Also, because the space between the heads and the platters is so small, a small fleck of dust could prevent the disk from spinning and they must be contained in a clean environment. Hard drives are also susceptible to power surges, and breaking parts. Often hard drive information is recoverable if data is lost. Despite the failure factors of hard drive storage, this continues to be a reliable form of digital storage. Hard drives in server arrays may be configured to work together to provide immediate recovery for a failed drive or server. These configurations allow a consumer-transparent recovery, and immediately compensate for device failure. In these configurations, no data is lost and a replacement server or drive will not use any down time to load information because the information is spread evenly across several servers or drives. The capability of many hard drives to work as one hard drive compensates for the occasional hard drive failure and makes hard drive storage a superior choice for dynamic storage. For producers, this also equates to consumer confidence in online image and pedigree databases.

- **'Static' Tape:** Tape storage is often only used by producers, and is not a consumer-typical storage medium. Most large companies, repositories, and storage facilities utilize a linear tape device to back up data from a LAN or SAN. Data from tapes often stored at an off-site location where this information may be accessed in case of a data disaster. Data from tapes is not rapidly restored like RAM memory, however, because retrieving data from tape is sequential, rather than random. Tapes are a strip of plastic magnetically coated to allow encoding, not unlike music cassettes. The tapes are not as easily damaged as music cassettes due to the archival thickness and size of the tape strips, but these, like hard drives, are susceptible to magnetic damage, and are also susceptible to heat damage. The portability and cost of these devices make them a frequently used data recovery option. Many tapes now contain WORM technology, Write Once Read Many, that makes these tapes suitable and acceptable in a court of law as evidence. Tapes are not a suitable option for regular file access, however, and are only used for long-term storage.

7. Data Disaster Planning

A disaster may come in many forms, including device failure, outdated format for a storage medium, natural disasters, electrical surge damage, electromagnetic damage, vendor extinction, and loss of key leaders in an organization.

Keeping data current is one of the most vital steps to data disaster planning. At the consumer level this may have greater consequences than the producer level due to the volatile nature of the consumer market and the need to release new and better product to the consumer. Consumers hold the responsibility to keep their own data in a readable, useable format. For example, this may require transferring 3¼ floppy disks to a hard drive, CD, or other storage device before the format is outdated. The producer is also impacted by the consumer format change when producing product for the consumer. Format is an essential part of producing packaged product and software, and the target genealogy market, often operating on a lower technological scale than the average computer user, must have a means to read and use the software or product.

The producer faces this issue on a different level. Replacing servers and building SANs require upgrades and changes, but not at the same frequency as consumers. Servers and SAN components are often built to work together over a longer period of time, without having to replace SAN components entirely when upgraded components become available.

In addition to upgrading technology and transferring data to newer formats, a secondary source of storage should always be used. For the consumer, this may consist of a storage system not unlike a SAN. The consumer may choose to store small amounts of data on a flash memory device to work with on a regular basis, backing up the information to a larger storage pool such as an external or computer hard drive. The consumer may then create a regular static storage system of optical disks.

Likewise, the producer may use information regularly from a SAN through application servers and file servers, and keep static backup copies using inexpensive linear tape technology. The important element in this system is the ability to utilize more than one storage medium. If one storage medium fails, for example, the SAN, or perhaps optical disks oxidize and become unusable, another medium is available for use. For producers, this ability to recover rapidly, or even without the consumer being aware of the failure, creates confidence between the consumer and the producer.

If such a failure occurs, the first priority must be replacing the failed storage device. This allows the user to continue having a secondary or mirrored copy of the data for quick and easy data disaster recovery.

Finally, for genealogical purposes, both the consumer and the producer must create as part of the disaster recovery plan, an individual who may replace the primary leader or curator of information. This plan allows an easy transfer of passwords, data, or other important information without bringing an organization to a halt. As recorded in the Center for History and New Media, “The Ivar Aasen Centre of Language and Culture, a literary museum in Norway, lost the ability to use its large, expensive electronic catalog of holdings after the death of the one administrator who knew the two sequential passwords into the system. The catalog, an invaluable research tool stored in an encrypted database format, had taken four years to create and contained information about 11,000 titles.” Fortunately a 25 year old hacker was able to access the catalog; however, this type of disaster is often overlooked as a possible loss of information.

The above example refers to the producer-level genealogist, however, consumer genealogists must also consider locating an individual who will know where to locate valuable files, and who will have access to locked genealogical files. This back-up individual should be technologically savvy and able to easily locate and take ownership of important genealogical data in case of a death or serious accident of the primary information curator. This individual should be familiar with present research, and understand how and where to begin research on incomplete family lines.

8. Summary:

All storage devices have benefits and specific uses. Understanding how to use these devices will help consumers and producers keep information safe and current. Preparing for data disaster will prevent consumer and producer productivity halts and allow information flow to continue without interference.