

Novice to Know-How Module Text

Course 6: Preserving Digital Content

Module 2: Key Types of Storage

The development of this course was funded by The National Archives (UK) as part of the "Plugged In, Powered Up" digital capacity building strategy.

1. Introduction.

This module will consider the different storage options available for digital preservation, their respective pros and cons, and how a mix of types of storage might offer the optimum storage solution. Making the right storage choice is not straightforward.

The best approach is likely to be a compromise between competing requirements. It is important to remember that most storage used for digital preservation is low cost, commodity storage designed to have a relatively short lifespan (often only 3-5 years). This makes it cost effective!

Well understood processes of effective bitstream preservation (keeping multiple copies, monitoring and replacing defective storage units, and implementing frequent integrity checks) will ensure that this cheap but possibly fragile storage still meets our digital preservation requirements. So, what types of storage could be considered?

2. Hard Drive (or "Spinning Disk").

A Hard Drive is an electro-mechanical storage device that utilizes a moving head to read and write data to a spinning disk (or series of disks called a platter) that is coated with magnetic material. This arrangement allows rapid access to any of the data stored on the disk by moving the read/write head to the required position.

Hard drives are typically used as the main storage device in a desktop PC, but can also be found in mass storage devices that hold large numbers of drives in racks that are designed to be managed in a climate-controlled server room by dedicated IT professionals.

Hard drives offer fast read and write access but can be expensive if storing large volumes of data.

3. Magnetic Tape.

Magnetic tape consists of a reel of plastic tape coated with a magnetic material. The tape is fed through a drive to enable reading or writing of data. A tape library or "jukebox" provides the hardware to automatically manage a large number of magnetic tapes incorporating a

mechanism (sometimes referred to as a "robot") to manage and fetch different tapes for reading and writing.

Magnetic tapes typically offer a lower cost of storage compared to hard drives, but access speeds are significantly lower as tapes are moved from the jukebox to be read, and tape is wound from reel to reel to reach specific data to be read.

4. Cloud Storage (Outsourcing).

Outsourcing your storage to the "cloud" is rapidly becoming a popular choice amongst digital preservationists. Cloud storage is most likely to rely on hard disk or tape storage, but this is usually completely opaque to the customer who simply pays (often different rates) to write data, store data, or access data that is stored in a remote location.

All necessary management operations (such as replacing failed storage units, refreshing storage hardware over time, data integrity checking and so on) are (at least in theory) performed by the cloud storage provider.

Verifying that the provider is actually performing these management operations may be a matter of trust - not something that digital preservationists like to rely on. Keeping all of your data in the cloud (or at least, with the same cloud provider) might be viewed as a risky approach. Having a clear "Exit Plan" is essential when using cloud storage.

5. What to Avoid...

Outsourcing your storage to the "cloud" is rapidly becoming a popular choice amongst digital preservationists. Cloud storage is most likely to rely on hard disk or tape storage, but this is usually completely opaque to the customer who simply pays (often different rates) to write data, store data, or access data that is stored in a remote location.

All necessary management operations (such as replacing failed storage units, refreshing storage hardware over time, data integrity checking and so on) are (at least in theory) performed by the cloud storage provider.

Verifying that the provider is actually performing these management operations may be a matter of trust - not something that digital preservationists like to rely on. Keeping all of your data in the cloud (or at least, with the same cloud provider) might be viewed as a risky approach. Having a clear "Exit Plan" is essential when using cloud storage.

6. The Future?

Whilst the preservation community remains sceptical of most "long lived" storage technologies, DNA or genetic storage appears to show greater potential and may dramatically change how we store digital data at some point in the future.

At the time of writing in 2020, genetic storage is still a number of years away from being a practical and cost-effective solution to digital storage. Although it's not a realistic choice at the current time, it's worthwhile keeping an eye on developments as research and productization brings costs down and makes these technologies more practical.

7. An Optimum Choice?

Ultimately there is no single and optimum choice of preservation storage that will meet all requirements. Digital preservation requirements will vary considerably from organization to organization.

Keeping 3 copies of your data on 3 different types of storage could take advantage of the benefits of each of these storage types whilst mitigating the respective shortcomings of each. This is just one possible approach:

1. A copy held locally on hard disk delivers convenience and rapid speed of access.
2. A second copy on magnetic tape (perhaps held at a different geographical location to mitigate the potential for natural disaster) provides resilience at a low cost.
3. A third copy in the cloud provides further resilience, again at low cost. Some cloud providers charge low rates for writing data and high rates for reading it. This third copy might be storage of last resort, to be accessed and read back only if the other copies become lost or damaged.