Unlocking Our Sound Heritage

BRITISH LIBRARY

# Long-term storage of digital sound files

#SaveOurSounds

HERITAGE FUND

# Long-term storage of digital sound files

Taking care of any type of digital file, whether audio or otherwise, requires some careful planning and action, to ensure they are stored safely, and can be found or used easily. This is sometimes referred to as **digital file management**, and this document is intended to help you manage your files more effectively.

## What is digital file management and why do I need it?

Any digital file management solution should follow a policy often called **multiple copy resilience**. At a very basic level, this requires that at least two copies of every file be held on completely separate pieces of I.T. equipment.

If these separate pieces of equipment are simply two external hard drives plugged into the same PC, then the risk of total disaster (i.e. the loss of both digital copies) is relatively high. A more complex setup will result in significantly lower risk, but at a higher overall cost. Be aware that the cost will be an ongoing rather than one-off commitment, and so should be planned for appropriately.

Any digital file management system, however simple or complex, will need to be reviewed periodically to ensure it remains fit for purpose, and to ensure timely upgrades in both hardware and software.

### Redundancy vs. Resilience

These two terms are often confused, and both are critical to an effective digital file management solution.

> **Redundancy…**
> is the provisioning of at least one duplicate device which is designed to take over active operation if the primary device should fail.
>
> **Resilience…**
> is the ability of a device to recover or self-heal in order to restore normal operations after a disruptive event. In order for resilience to exist, we must also have a redundant device.

In order to provide an adequate level of redundancy and ensure resilience, the preservation system should ideally consist of on-site, near-site and off-site storage. For the same reasons, mirror versions of on-site systems might also be provided should regular access be required.

### Storage and Security

Adequate storage capacity for all files should be maintained and regularly reviewed to cover any required future expansion; this is known as Capacity Management. Provision for secure networking, adequate connectivity and the ability to restrict access to certain users might also be considered.

All I.T. equipment within the digital file storage system should be protected by power surge protection systems. Disaster recovery procedures should also be in place.

## What are the options for a digital file management solution?

Below are descriptions of digital file management solutions which progressively provide increasing levels of security, resilience and redundancy. As you might imagine, these benefits also come at an increased overall monetary cost.

### Hard Disk Drives (HDD)

The basic element of almost all storage solutions and the building blocks on which the more resilient systems listed below are based. How that storage is managed is key to long-term preservation.

As mentioned previously, the absolute minimum for any file management solution would be to have two hard disk drives connected to the same computer, with a mirror or backup process designed to keep copies of all files. Both hard disk drives would contain the exact same files as each other. This would provide bare minimal redundancy.

### Network-Attached Storage (NAS)

One or more hard disk drives contained within an external enclosure and accessed as a shared drive over your local network (LAN). This has the benefit of being a relatively cheap method for adding an increased level of redundancy, the storage being geographically separate from the accessing computer.

## Redundant Array of Independent Disks (RAID)

A natural progression from the last option providing an additional method for increasing the resilience and redundancy of your Network-Attached Storage (NAS). Whilst there are a number of different RAID options or levels, by far the most common are…

- **RAID 1 or mirroring mode** is used to make an exact copy of data on each hard disk drive which is contained within your NAS.
  - In this setup, your NAS would have two hard drives and all your digital files would be on both drives at the same time; one holds a complete copy of the other.
  - This mirroring relationship between the two hard drives is automatic and provides a reasonable level of redundancy, one hard drive having a complete backup of the other at all times and the entire NAS being geographically separated from the accessing computers.
  - The downside is the increased cost of requiring double the number of hard disk drives to provide the same total storage capacity of one (e.g. 1TB total storage requires 2x 1TB hard drives when running under RAID 1).

- **RAID 5 or parity mode** distributes your files across all the hard disk drives contained within your NAS along with enough recovery information to allow the temporary loss of one of the drives.
  - Your NAS must contain three or more hard drives for this method to work, thereby increasing the cost.
  - However, a high level of redundancy and resilience is gained due to the fact that, if one hard disk drive fails, the remaining disks can keep file access going until a replacement hard disk drive arrives and is installed.
  - Once the replacement hard disk drive is installed, the entire NAS is automatically rebuilt and full file security is then restored.

## Locally managed network storage

A file management solution which is typically provided and managed by an organisations' I.T. Technology team. Such a team would grant you storage space for your digital files, whilst at the same time also providing a high level of resilience and redundancy for safeguarding the long-term preservation of your audio.

This will include provision for:

- the regular renewal of old equipment
- any security patching required
- the supply of electrical power
- ongoing maintenance
- the ability to scale/upgrade your storage as and when required etc.

This type of service is usually cost effective for medium to large organisations and would be billed internally on a regular basis.

### Cloud managed network storage

This can effectively be thought of as if you are renting another company's network storage solution, and paying for them to manage it. This would include all the benefits detailed in the last section. Some things to bear in mind though are:

- Cloud-based solutions are not cost effective when it comes to the large storage capacity usually required for long-term preservation (see Further information section below).
- The speed is dependent upon your internet connection with upload/ download performance likely to be nowhere near that of a Locally managed storage solution.
- Options for customisation vary between suppliers but tend to be quite limiting, even on the more expensive solutions.
- Whilst the resilience of Cloud-based solutions is high, redundancy is not as access to your files relies on a single point of failure; namely, the internet. For this reason it is common practice to employ both Local storage (NAS or similar) for business-as-usual access, with the addition of some Cloud-based storage for the longer-term.

### Digital Asset Management

- In addition to the actual storage of your audio files, you may also consider how best to access, discover and generally manage them on a day-to-day basis. For this, a Digital Asset Management (DAM) application may help.

## Further information

**The National Archives, Preserving Digital Collections**
https://bit.ly/2XWpwL9

**Enterprise Storage Forum, NAS Vs RAID: How they Differ and Overlap**
https://bit.ly/3sHmYyy

**NAS vs. Cloud: What to Choose?**
https://bit.ly/39NbnVW

**Enterprise Storage Forum, RAID Levels Explained**
https://bit.ly/3bXorLe