

Novice to Know-How Module Text

Course 2: Introduction to Bitstream Preservation

Module 3: Introduction to Bitstream Preservation

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.

In this module we will introduce a key form of digital preservation: Bitstream preservation.

We will look at what it is and why it is important, what risks to digital content it addresses and what processes and procedures are involved.

2. What is Bitstream Preservation?

At a fundamental level, all digital content is stored as a series of 0s and 1s. These are "binary digits" or "bits".

There are many challenges that digital preservation seeks to solve, not least of which is preserving our ability to access and understand the information contained within digital content.

But at the most basic and perhaps critical level, we need to make sure that the digital content, our files or streams of bits, are not lost or damaged.

Bitstream preservation covers the approaches to make sure that we can simply *keep the bits*.

3. Risks and Bitstream Preservation.

Our data faces a variety of risks. If left alone, it is not likely to survive intact into the future. So, what are the risks? Click the images below to find out more...

Media Obsolescence.

Media obsolescence is when storage media, such as tape, floppy disks or CDs become obsolete.

This obsolescence may be due to no longer possessing the hardware needed to read the media or due to manufacturers discontinuing production of the needed hardware. For

example, many new laptops do not have a disc drive, making CDs and DVDs obsolete on these machines.

The likelihood of this risk occurring can be reduced by monitoring changes in storage technologies and avoiding purchasing all of your storage from one supplier.

Media Degradation or Failure.

Storage media are commercial products and tend to have a reasonably short lifespan. Most hard disks tend to be reliable for around 5 years before you might expect them to start to degrade.

A common media failure is 'bit rot'. Though all forms of storage media are subject to different forms of decay, bit rot refers to the loss of data due to the small electronic charge of a bit being 'flipped' (changed from 1 to 0 or vice versa). This can be caused by cosmic rays or other high energy particles.

These risks can be avoided by using different forms of storage media and refreshing data onto new storage media over time. We can also check for bit rot using a process called integrity checking which is covered in a future module.

Manmade or Natural Disaster.

Digital content is as at risk from man-made or natural disasters, such as fire, flood and earthquakes, as analogue content. A fire or flood in your server room may easily result in the loss of the digital content you are preserving.

As well as all of the standard physical precautions you can take to prevent or reduce the impact of a disaster, there are processes you can put in place specifically for digital content.

The most important step to take is to ensure that you save at least one copy of your digital content at a different geographical location, preferably also with a different disaster threat profile. This might be with a cloud storage provider or as part of a reciprocal agreement with another organization.

Human Error or Malicious Damage.

Unfortunately, human beings are one of the biggest risks to digital content. This might be accidental damage due to human error, or through actions with more malicious intent, such as hacking.

To combat these risks, it is important to implement robust security measures. This may include:

- A security policy
- Strict control of access permissions to digital content
- Staff training

There are international standards for information security that can be implemented. Colleagues in your IT department will likely be familiar with them.

But what can happen if these risks do occur?

4. If We Do Not Protect the Bitstream?

The outcome is often unpredictable. Media degradation may lead to a complete failure of the storage device. So, you cannot read back any of the data stored on it.

In some cases, damage might be more subtle, like if individual bits in a file become lost or damaged. This might lead to an obvious result, as in the case of this before and after screenshot of a digitized newspaper page.

Alternatively, damage might be less difficult to recognize visually. Some of the newspaper pages that were also damaged looked fine until you zoomed in, and they became fuzzy.

5. Basics of Bitstream Preservation.

Ultimately bitstream preservation is about addressing the risks to the bitstreams of the digital content. Steps to take include:

- Keep several copies of all your files
- Keep one copy at a different geographical location with a different disaster threat profile
- Don't use the same storage technologies for each copy and don't rely on a single vendor to store all your data
- Perform periodic integrity checks to identify any changes/errors
- Avoid having any one person with write access to all copies of the digital content

6. Bitstream Preservation into Practice.

There is no one correct approach to bitstream preservation. Implementation decisions will be dependent on the types of digital content you need to preserve, the resources you have available, and the risks specific to your organizational context.

You may build bitstream preservation workflows from a collection of smaller tools or you may look to procure a repository system with inbuilt workflows.

In the next module we will take a look at some exemplar workflows that can help you plan, but first let us do a quick knowledge check on what we have learned about bitstream preservation.