

Novice to Know-How Module Text

Course 2: Introduction to Bitstream Preservation

Module 4: Introduction to Workflows

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. What is a Workflow?

In this course we will be defining a workflow as:

"a number of connected steps that need to be followed from start to finish in order to complete a process."

Developing clear workflows ensures consistency of approach over time, allowing us to document the actions taken, and provide evidence of adherence to good practice.

In this module we will be introducing some basic workflows for digital preservation. You can learn the skills to undertake these workflows in other modules and courses available on this platform.

2. What Workflows to Use.

Ultimately, the most successful workflows are the ones designed with your specific context in mind. Taking into consideration issues such as the skills of your staff, resources available, risks faced, and the types of digital content you want to preserve. We all, however, need a starting point and that is what we are going to look at here.

As part of their "Plugged In, Powered Up" digital capacity building strategy, the UK National Archives commissioned the creation of a set of exemplar digital preservation workflows for those looking to start work in the area. Over the next few slides we will introduce these workflows and examine the steps from each in a little more detail.

3. UK National Archives Exemplar Workflows.

The exemplar workflows set out four sections describing digital preservation processes: Select and Transfer, Ingest, Preserve, and Access.

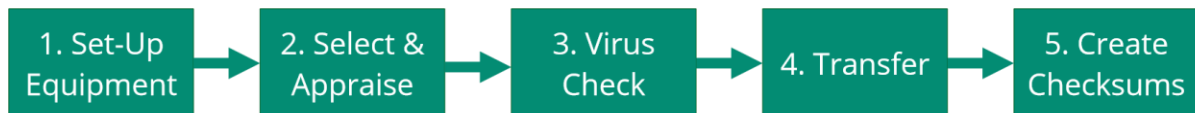
In this and future modules we will be focusing on the first three sections as they cover the activities of bitstream preservation, as well as selecting the digital content to be preserved.

Each section contains a number of stages which describe the activities within each, tools that can be used, and notes if the stage is essential or not. The various activities cover beginner,

intermediate or advanced skill levels. We will be focusing primarily on the beginner level skills and the relevant activities.

4. Select and Transfer.

The workflow below shows the steps in the Select and Transfer workflow, following on is a description of each step. If the step is considered essential to the preservation process, this is noted.



1. Set-Up Equipment – Essential.

- Set up a dedicated PC to connect to any storage media holding digital content. Ideally, only connect it to your organization's systems/internet to perform essential updates to systems.
- Add equipment to read various types of media you will work with. For example: readers for DVDs/CDs and floppy disks.
- Use software called write blockers to prevent changes to the content.
- Use encryption software if you work with sensitive content.

2. Select and Appraise – Essential.

- Confirm the digital content fits your organization's Collecting Policy.
- Capture information from the depositor about Intellectual Property Rights and access restrictions
- Ask the depositor to create a list of the content that is being transferred
- Carry-out basic appraisal (optional)
- Create an accession number or similar identifier
- (e.g. an identifier relating to a Digital Asset Register)
- Create a folder on the ingest PC to hold the digital content and related documentation.

3. Virus Check – Essential.

- Scan the content for viruses using anti-virus software.
- Remove any infected content and decide on action e.g. repair or contact depositor for clean copies.
- Ideally leave (quarantine) the content on your PC for 30 days and then re-scan them for viruses before proceeding to Workflow 2 (Ingest).
- Keep a record of what virus checks you have undertaken
 - (e.g. save any report the software generates.)

4. Transfer– Essential.

- Connect the transfer media to the ingest PC. If possible, scan the content for viruses using anti-virus software on the media before transferring them.
- Transfer the digital content from the media to the relevant folder on the PC using copying software, such as Teracopy.
- Alternatively, if receiving content internally or by email/the internet you may ask the depositor to use software such as Bagger or Exactly.
- Disk imaging is an alternative to copying the content. Software, such as FTK Imager Lite, can create an exact copy of the contents of the media, including original metadata.

5. Checksums – Essential.

- If checksums were created before transfer (by yourself or the depositor) they should now be checked.
- If not, use software to create checksums and if possible save them with the content (e.g. in the “metadata” folder for the digital content).
- At this point you may want to create a copy of the content that will be used for the steps outlined in section 2 Ingest (sometimes called a “working copy”). This will reduce the risk of loss if content is inadvertently changed.

5. Ingest.

The workflow below shows the steps in the Ingest workflow, following on is description of each step. If the step is considered essential to the preservation process, this is noted.



1. Understand What You Have – Essential.

- Use software, such as DROID, to identify what you have and create a list of the content. This should include file names, file paths, sizes, file format, last modified date etc.
- Save the list in an open format (e.g. CSV or XML) and store in the relevant “metadata” folder.

2. Validate Content.

- Validation software checks whether the content conforms to its file format specification. In some cases, the software can also fix issues.
- It is not always seen as an essential step but can help flag issues. For example, if the content does not conform to its specification then it may be more difficult to read or manage in the future.
- It can also be useful for checking the quality of digitized content.

3. Analyze and Investigate.

- Analyze metadata generated in stages 1 and 2 to flag any issues for investigation.
 - This includes looking out for corrupt files, compressed files, encrypted or password protected files. You may need to go back to the depositor to resolve these.
- It can also flag unidentified formats which could require further research.
 - Some archives also convert file formats to a preferred file format for preservation (see step 5 in Preserve).

4. Describe – Essential.

- As a minimum create a high-level description of the content, perhaps in a Digital Asset Register.
- You may decide to do more detailed cataloguing in accordance with your organization's cataloguing standards (either now or at a later date).
- You can add the descriptions to the metadata you created in step 1 and 2 or create the descriptions in a CSV or XML file.
- If you use a collection management system you may wish to record the descriptions there (e.g. the accession record or catalogue).

5. Appraise.

- You may have already carried out appraisal as part of the Select and Transfer workflow. At this stage you may wish to carry out further appraisal.
- As a minimum you could consider identifying and removing duplicates by comparing the checksums of the content.
- There is software that can help you do this. However, you may decide to keep duplicates if they have useful contextual information (e.g. file name).

6. Apply Access Restrictions.

- Some of the content may contain personal, sensitive or confidential information.
- The depositor should help you identify this during the transfer step. The cataloguing process may also identify sensitive content.
- There is forensic software that can help you identify personal information. Some of it is commercial and expensive, but there are free alternatives.
- Access restrictions or any risks should be recorded in the metadata.

6. Preserve.

The workflow below shows the steps in the Preserve workflow, following on is description of each step. If the step is considered essential to the preservation process, this is noted.



1. Set-Up Storage – Essential.

- If you have no dedicated storage, think about some practical solutions. For example, an interim approach could be to use your organization's storage network.
- The NDSA Levels of Digital Preservation are useful for planning storage. In particular, the sections on 'Storage' and 'Control'.
- Think about creating multiple copies, in different locations and using different technologies.
- Think carefully about who in your organization is allowed to access the digital content and the type of access that they have (e.g. read, write, move, delete). Keep a record of who has access.

2. Move to Storage – Essential.

- Before moving the content to secure storage check that any relevant documentation is stored alongside the digital content.
- Some archives will package the content and metadata in a 'bag' using software such as Bagger.
- Move the content to the storage. You could use copying software to do this to ensure date information and other file attributes are preserved.
- Perform an integrity check after the move to ensure no content has changed.
- Record the location of the content in your Digital Asset Register or Collection Management System.

3. Integrity Checking – Essential.

- Use checksum software to carry out regular integrity checks of the content in the storage.
- If checksums of content do change then investigate.
- For example, if the content is corrupt or has been accidentally changed, it may need to be replaced from another copy.
- Ideally, you should also keep logs of actions performed on content and carry out periodic reviews of these logs.

4. Monitor Storage – Essential.

- The lifetime of storage can be short. It can fail or cause corruption of content.
- You will need to review your storage every 3-5 years, moving content onto new storage.

5. Monitor Content.

- It is important to monitor your content to understand if any of the file formats you hold, or the software/technology required to access them, are at risk of becoming obsolete.
- One solution is file format migration where a file format is converted into a new file format.
- An alternative, Emulation, looks to recreate the functionality of the original software/technology.

- These activities are called content preservation and are beyond the scope of this course.

7. Wrap-Up.

We have now completed our walkthrough of the first three stages of the UK National Archives exemplar workflows for digital preservation. From identifying content for preservation, through processing it for preservation, to moving it into secure storage for long-term care. If you would like more information on these workflows there is a link in the resources related to this course.

For the remainder of this course we will be looking at a particular activity that forms part of bitstream preservation and is mentioned at several points in the workflows: Integrity Checking.