

Novice to Know-How Module Text

Course 4: Select and Transfer Digital Content

Module 1: Selecting Digital Content

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 the module "Introduction to Workflows" we looked at four simple exemplar digital preservation workflows published by the UK National Archives. The first workflow was "Select and Transfer" and in this module, and the others in this course, we are going to delve a little deeper into the issues to be considered when developing the processes for this workflow.

First up are considerations around selecting digital content for preservation. Many find it tempting when first approaching digital preservation to "just keep everything", as we can always buy more storage to put it on. But, just as in the analogue world, this can lead to high costs and an overwhelming amount of content, none of which is being properly preserved.

It is, therefore, important to take a robust approach to selection, so that we can ensure continued access to important information. If your organization already preserves analogue content, it is almost certain that you will have a Collecting Policy and/or a Retention Schedule detailing the content you look to acquire. Much of the policy or schedule will remain relevant for digital content, particularly statements relating to the subject matter of the content to be acquired.

There will, however, need to be some updates made to ensure the policy is fully relevant to digital content. The issues that you will need to consider will be the main focus of this module. If you do not have a Collecting Policy at the moment, it will also help you to establish what issues to cover in one.

But, before we start to consider the key issues around the selection of digital content, we will take a brief look at common routes taken by digital content into a repository. Knowing which are relevant to you will help guide the reworking or creation of your Collecting Policy.

2. Routes for Acquiring Digital Content.

There are many subtle differences in the potential routes for acquiring digital content that are context dependent and might impact the processes and procedures you will put into

place. But, for the sake of this module, we are going to consider four main routes. These relate to acquiring digital content:

- From an external depositor
- Via legal deposit
- In accordance with retention schedules
- And, as the outputs of digitization projects

Knowing which of these four main routes are relevant to your organization is essential to your approach to selection.

- If you are accepting digital content from an external depositor, ownership and legal issues such as intellectual property rights are key considerations.
- For material coming to you via legal deposit you need to understand the opportunities and restrictions afforded by the relevant legislation.
- If you are expecting digital content from internal records creators in accordance with established retention schedules, you will want to establish efficient transfer processes but may also need to consider training for staff to ensure the correct records are sent at the correct time.
- Finally, with digitization projects, you should aim to be included in development of the project if possible. This will mean you can shape from the start the content coming to you as well as making sure preservation is included as a core part of the process.

Once you have identified the routes through which you might expect to receive digital content, you can move on to consider in more detail the issues that will impact selection.

3. What are the Key Issues for Selection?

The key issues for the selection of digital content fall roughly into two main categories: legal issues and technical issues.

Legal issues are considerations around:

- Ownership
- Intellectual Property Rights
- Privacy and Sensitive Data
- And, Content Liability

Technical issues we may need to consider for selection include:

- Viruses and malware
- Storage media
- Formats
- Volume of content
- And, Documentation and Metadata.

In the next two slides we will consider each of these areas in turn.

4. Legal Issues.

So, let's look at some of the legal issues we might need to consider with regards to selection.

First up, if we are acquiring digital content from external sources, clarifying the ownership is essential. As it is relatively easy to make exact copies of digital content, we need to make sure that the potential depositor does, in fact, own the content offered. Transfer of ownership to the repository reduces legal liabilities. There are many circumstances, however, where ownership may not be transferred, e.g. where the content will remain in use by the depositor or when deposit is prompted by legislation covering legal deposit or public records.

Linked to this are considerations around intellectual property rights. Knowing who holds intellectual property rights is essential no matter the route through which the content comes to us. An inability to procure sufficient rights to the content will impact on our ability to preserve it and provide access. Ideally, we would hope for the transfer of intellectual property rights alongside ownership. If this is not possible then a perpetual license would be the next best thing. The license should grant permission for dissemination so we can provide access, as well as allowing us to change the content as needed for active preservation. Without the transfer of rights or a license we would again need to consider carefully whether or not to acquire the content. Also, with digital content there is an even greater chance that included with the content there is material whose intellectual property rights are owned by a 3rd party, particularly with content such as websites. This may require its own risk assessment before considering acquisition.

It is also important to know if the digital content may contain private or sensitive data. For selection we then need to balance the importance of the content, with the potential liabilities that might be derived from accepting the content for preservation. This, however, may be out of our control if the digital content is covered by a legal deposit responsibility or organizational retention schedules.

Finally, we also need to consider any liabilities in relation to the information within the digital content. Does it include anything defamatory or obscene? Does it promote illegal activity or infringe someone's human rights? If so, we may not wish to acquire the content, particularly keeping in mind how easy it is to copy and distribute digital content.

5. Technical Issues.

The other issues to be addressed in relation to selection of digital content fall loosely under the heading of technical issues.

This begins with how certain we are that the digital content doesn't contain any viruses or malware which might pose a risk to our other digital collections and systems. If we know there is or believe there could be malware or a virus present, we need to balance the importance of the content against our capabilities for removing or treating the problem. Do we have the skills and technology necessary? How much will it cost, financially and in time? Is the content important enough to accept the potential risk?

Next, we need to consider the storage media the digital content is currently held on. Do we have compatible technology for retrieving the digital content from the storage media? If we don't, do we have the resources to pay for a specialist to retrieve the content? It is also important to check if the media is damaged in any way. Even if it is damaged it might still be possible to retrieve the digital content, although this can often be prohibitively expensive.

Third, we need to think about the file formats present within the digital content. We talked in the "File Formats and Digital Preservation" module about establishing preferred file formats for preservation, that we have a high level of confidence in our ability to preserve. So, selection decisions may be guided by our list of preferred or accepted file formats. If we take a more open approach to file formats, we may still need to undertake an assessment of the file formats if a large number are rare or proprietary.

It is also important to understand the volume of digital content being offered, as a large deposit may strain or exceed your current storage capabilities. If this is the case, you may need to decide between procuring more storage and refusing the content. If have to turn down the deposit for this reason you may wish to suggest another larger repository who could preserve the content instead.

Finally, having sufficient documentation and metadata for digital content is essential. When working through the selection process you will need to assess what documentation and metadata will likely accompany the content and decide if this meets your minimum requirements. If it does not meet your requirements, is it possible to create the documentation and metadata and can you invest sufficient resources to do so? It is, however, essential to be realistic about how much documentation and metadata to expect and to remember that some metadata, especially technical, can be generated by characterization tools.

6. Digital Preservation Handbook Decision Tree.

Developing your own selection process for digital content can be a complicated undertaking, but there is help available! The Digital Preservation Handbook, a free online resource, offers an interactive decision tree mapping out steps for the selection of digital content. It covers all of the issues addressed in this module and provides guidance notes at each stage. Working through the decision tree for a potential acquisition of digital content can help you bring structure and accountability to the acquisition process.

7. Selecting Digital Content – Wrap-Up.

In this module we've established that selecting digital content for preservation requires us to consider new issues, in addition to those for analogue content.

Developing a robust selection process begins with understanding the routes through which you will receive digital content. You must then consider both legal and technical issues relating to the digital content, that may affect your decision to acquire it or not.

In the next two modules, we'll look a little more closely at working with depositors through the acquisition process. First, we'll consider what we need to do when working with

depositors external to our organization. Next, we'll have a look at internal depositors and the issues faced there.

But before we do that let's do a quick knowledge check based on what we've covered in this module.