Digital transfer preparation - Extract creation

1. Introduction

In order for us to determine whether a digital transfer is feasible, your public sector organisation must first create an extract or copy of a test or sample set of eligible digital records and their metadata for us to analyse. This process can also be used to prepare a full extract of all eligible digital records when we have reached agreement to proceed with a full transfer.

This is the most technical part of a digital transfer however there are some free online tools available as outlined below that you can use to assist in extracting your digital records and their metadata. We encourage you to contact us for specific guidance and advice if you are considering the use of any of these tools.

2. Identify extract requirements

We will meet with you and create a list of requirements for the test extract of eligible digital records and their metadata. These specific requirements will be based on information gathered about your organisation's information environment and technical capabilities.

To guide this information gathering, a *Digital transfer information gathering checklist template (23/Fm10)* may be used which is available on our website.

3. Export extract

It is recommended that you use the export format options of the system(s) in which your digital records are stored to export copies of both the selected records and their associated metadata. The easiest option is to export all the metadata fields available in the system and then analyse those in collaboration with us to decide which fields provide context and assist with discovery.

However, as some systems do not have metadata export functionality, you may need technical knowledge and/or IT support to do this. We can provide some advice and support, but you may also need to consult the system designers or vendors.

4. Create a transfer metadata file

A transfer metadata file (TMF) or list that includes metadata for each digital record must accompany the extract. If you are transferring digital records from a shared drive or similar where there is no separate descriptive metadata available, we still require a TMF or list to be created or generated. In essence, generating a TMF is the same as populating a list template in the paper environment.

From our experience with previous transfers from shared drives, the automated tool DROID (Digital Record Object Identification)[1] can be used to generate and copy file format identification metadata into a .csv file and thus create a list for transfer. This metadata is limited but includes:

[1] DROID is a file format identification freeware created by The National Archives in the United Kingdom and can be downloaded from their website (File profiling tool (DROID) - The National Archives).

- File path (identifier)
- File name (title)
- File size
- Date last modified
- File/Folder and
- Checksum values (see 4.1).

If you have more metadata than this available, some manual input may be necessary, or we can work with you to merge this into a more complete metadata set. Completing this step accurately will ensure the capture of metadata necessary to add value to the digital records.

4.1 Checksum values

A checksum value is an essential metadata element that is required to ensure the integrity of the digital records. Checksums must be generated by you before transfer to us, either from within the original storage system or immediately after the records are exported from it. The checksum values must be provided to us as part of the metadata describing each digital record, or in a separate file (ideally both). Providing the checksum for each transferred record allows us to validate each record and make sure that all the digital records have been transferred successfully and that no changes or errors were introduced during the transfer.

Checksums can be generated by DROID as well as other free online tools such as Free Commander (Windows), and SHA1SUM or MD5SUM (Linux). For more information on checksums, see our factsheet *Checksums overview (17/F25)* available on our website.

5. Transport extract to Archives

You will need to copy the extract and the TMF onto a removable hard drive that can be secured with encryption (which we can provide if required) or arrange an alternative method for secure transport to us such the download option from a secure system/environment or Government Cloud services.

Before transporting to us you must check that all the digital records are synchronised or have been copied correctly. We recommend using the tool 'rsync'[2] (Remote Sync) for copying and synchronising of digital records. This tool preserves the integrity of the records and their accompanying metadata, and we are happy to assist you in its use.

IMPORTANT NOTE: You must not delete any digital records or metadata – at this stage, we only need a reliable copy.

[2] rsync is a utility for efficiently transferring and synchronising records across computer systems, by checking their timestamp and size (https://en.wikipedia.org/wiki/Rsync).