Digital transfer initiation - readiness assessment

1 Introduction

Archives New Zealand (Archives) has developed tools and methods for organisations to use for the transfer initiation stage. The following steps can help organisations understand their current digital health and assess their readiness for transfer. A key component is the preparation of an initial test extract or copy of digital information and records (files) that are eligible (i.e. that have acceptable transfer characteristics), along with their accompanying metadata. Creation of a test extract is necessary for Archives to determine the feasibility of a full transfer.

A list of requirements for this initial extract will be developed by Archives from information provided by the transferring organisation about their information management environment and capabilities. Organisations considering the use of these steps should contact Archives.

2 Step 1: Understand current digital health

Organisations can use automated tools such as DROID (Digital Record Object Identification)[1] to identify:

- Which file formats they hold, particularly any old or obsolete formats and unusual format modifications
- Duplicates and versions, by generating and comparing checksum values for each file
- Layers of content, such as embedded objects and
- Any system files, missing files and empty folders

This can assist organisations in reducing data storage and retrieval costs.

3 Step 2: Assess digital transfer readiness

Organisations can use appropriate tools such as SQLint[2] to:

- Quality check the accuracy and consistency of file sentencing and
- Locate obvious sensitive, non-business related and/or draft material.

This will assist organisations in identifying and managing any access risks.

4 Step 3: Identify what metadata is needed

At a minimum, Archives expects organisations to provide the mandatory metadata elements required by the *Minimum requirements for metadata* (16/G7). We have set no fixed requirements for the format or structure of this metadata. Although we prefer CSV, TXT or XML file formats, Archives' only requirement is that the metadata is **structured** consistently, and that there is someone in the organisation who understands the metadata enough to assist us in understanding it and to facilitate mapping to our systems.

---

[1] DROID is a file format identification freeware created by The National Archives in the United Kingdom, and can be downloaded from their website (http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/).

[2] SQLint is a simple command-line linter which reads SQL files and reports any syntax errors or warnings it finds. A linter or lint refers to tools that analyse source code to flag programming errors, bugs, stylistic errors, and suspicious constructs (https://github.com/purcell/sqlint).

Organisations will need to identify and provide any extra metadata necessary to add value and enable discovery of the files. Completing this step accurately will ensure the capture of metadata necessary to add contextual value to the files.

Organisations can use the export format options of the system(s) in which their files are stored to export or copy the files and their associated metadata. As some systems do not have metadata export functionality, organisations may need technical knowledge and/or IT support to do this. Archives can provide some advice and support, but organisations may also need to consult the system designers or vendors.

Generating a list of metadata is in essence the same as populating a list template in the paper environment. The tool DROID 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:

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

A checksum value is an essential metadata element that is required to ensure the integrity of files. Checksums must be generated by organisations before transfer to Archives, either within the original storage system or straight after the files are exported from it. The checksum values must be provided to Archives as part of the transfer so they can be validated to make sure all the files have been transferred successfully and no changes or errors were introduced during the transfer.

Checksums can be generated by DROID as well as free online tools such as Free Commander (Windows), and SHA1SUM or MD5SUM (Linux).

5 Step 4: Create an initial test extract or copy

In order for Archives to determine whether a full digital transfer is feasible, the transferring organisation must identify and assess an initial test set of eligible files and their metadata. The organisation must then extract or copy these onto a removable hard drive that can be secured with encryption (which Archives can provide if required), or arrange an alternative method for secure transport to Archives. Archives recommend using the tool 'rsync'[3] (Remote Sync) for the copying and synchronising of files. This tool preserves the integrity of the files and the metadata, and Archives is happy to advise transferring organisations on its use.

NOTE: Organisations must not delete any files or metadata – at this stage, Archives only needs a reliable copy.

6 What happens then?

Once the initial test extract is received by Archives, several analytical processes (both automated and human) are run over the files and metadata to identify any content, technical, metadata and accessibility issues that may affect a full transfer to Archives and ingest into the Government Digital Archive. Archives then consolidates the analysis results in a report for discussion with the transferring organisation, which outlines their transfer readiness (i.e. the quality and consistency of sentencing decisions) and their current digital health

[3] **rsync** is a utility for efficiently transferring and synchronising files across computer systems, by checking the timestamp and size of files (https://en.wikipedia.org/wiki/Rsync).

(i.e. the identification of unique file formats and potential digital preservation issues). The report concludes with a recommendation for the organisation to undertake further preparation, to postpone, or to proceed.