Digital transfer initiation – Readiness

1. Introduction

We can provide advice and guidance on tools and methods which can be used by your public sector organisation to assess the readiness of eligible digital information and records (digital records) for transfer to us.

The following activities are useful for understanding the current digital health of your records and how well they are being managed for digital continuity. These activities underpin a successful transfer as they help you to discover and address in advance any issues that may affect ingest or transfer of your digital records into the Government Digital Archive.

2. Check current digital health

Before initiating a conversation with us about a digital transfer, you must understand the current digital health of your organisation's digital records, including identifying any unique file formats and potential digital preservation issues.

There may be some duplication in the records. You are encouraged to assess the risk of this and dispose of duplicates prior to transfer. Duplication can also occur if the records were created at a time when your organisation's information management policy was 'print-to-file'. This can create considerable work in identifying what constitutes the 'authoritative' version, i.e., the digital or paper record.

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

- which file formats you hold, particularly any old or obsolete formats and unusual format modifications
- duplicates and versions – this can be done by generating and comparing checksum values for each digital record
- layers of content, such as embedded objects, and
- any system files, missing files and empty folders.

This can also assist you in reducing data storage and retrieval costs. See our website for more information on *File formats for digital transfer*.

Other tools such as SQLint2 or Demystify3 can be used to discover more details about the digital records intended for transfer. They can also be used to:

- quality check the accuracy and consistency of file and content sentencing (for example, showing timelines based on last modification dates), and

[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).

[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).

[3] Demystify is a way to analyse DROID CSV and Seigfried export files. Demystify breaks the export into its components and stores them within a set of tables in a SQLite database; creates additional columns to augment the output where useful; and queries the SQLite database, outputting results in a readable form useful for analysis. Demystify provides an easily readable overview and statistics of the files in the transfer (https://github.com/exponential-decay/demystify).

- locate obvious sensitive, non-business related and/or draft material by listing potentially problematic words or characters in file and folder names. This will assist you in identifying and managing any access risks.

3. Identify what metadata is needed

A transfer of digital records consists of not only the records but their metadata as well. A file or list that includes metadata for the records must accompany the transfer. At a minimum, we expect organisations to provide the mandatory metadata elements required by our Information and records management standard (16/S1).

Although we currently have no mandated requirements for the structure of a transfer metadata file (TMF), we do recommend that the TMF:

- uses UTF-8 coding
- has file folder names free of non-standard characters (only ASCII), and
- most importantly, is understood by someone in the organisation who can assist us in mapping it to our systems.

For each digital record included in the transfer, the TMF must include the following metadata elements to enable us to process the transfer:

- checksum value (generated using one of the methods outlined below)
- file path which would be a complete pathway to the relevant record within the transfer set (not the pathway as it was in the original pre-transfer system).

Checksum values can be generated using free online tools such as Free Commander (Windows), and SHA1SUM or MD5SUM (Linux). The tool DROID can also be used to generate checksums (MD5 and SHA1). We will identify any other technical metadata that we need to enable the preservation of the digital records. You can also provide any extra metadata that you may need to add value and enable discovery of the records.

For more information on checksums, see our factsheet *Checksums overview (17/F25)* available on our website.

This metadata, when open, will appear on our Collections search tool, which provides access to digital public archives held in the Government Digital Archive, so that members of the public can search, browse and find relevant information and records.