

# Digital transfer preparation – Pre-ingest analysis

---

## 1. Introduction

When we receive a test extract or copy of eligible digital information and records (digital records) and their transfer metadata file (TMF) for potential transfer, several analytical processes (both automated and human) are undertaken to identify any issues that may affect ingest or transfer into the Government Digital Archive. This process is also used to analyse a full extract of all eligible records when we have reached agreement with your public sector organisation to proceed with a full transfer.

## 2. Integrity check

We use the checksums supplied by your organisation to ensure the digital records are not altered or corrupted from the time they are exported or copied from your system(s) until we receive them. We do this by generating new checksums and comparing them to those in the TMF. If they match, this reassures both of us that nothing has changed in the copying and transit processes and the integrity of the records has not been compromised.

For more information on checksums, see our factsheet [Checksums overview \(17/F25\)](#).

## 3. Content analysis

A number of automated tools such as DROID (Digital Record Object Identification)<sup>1</sup>, SQLint<sup>2</sup> or Demystify<sup>3</sup> are used by us to assess the capability and accuracy of the sentencing or selection of eligible digital records against your organisation's relevant disposal authority(ies) and/or our General Disposal Authorities, GDA6 and GDA7.

These tools are also used to check for any duplicate information and records and obvious sensitive, non-business related or draft material, as well as missing files, empty folders and layers of content, such as embedded objects like images or videos, that you may not have picked up during your readiness assessment – see our factsheet *Digital transfer initiation - Readiness (18/F27)* available on our website.

---

<sup>1</sup> 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](#)).

<sup>2</sup> 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>).

<sup>3</sup> 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>).

## 4. Technical analysis

File format identification and validation tools are run over the extract. This allows a closer manual analysis of any accessibility and preservation issues which will need to be addressed such as obsolete file formats, unknown file formats, broken files, invalid file formats etc. Other checks are done to identify duplicate files, and system files which might need to be excluded from the transfer after discussing with you.

## 5. Metadata analysis

We will manually map the metadata fields in the TMF to the descriptive metadata fields in our Collections search tool (which provides access to the digital public archives), and the technical metadata fields in Rosetta (which stores and preserves the Government Digital Archive). This manual mapping is checked and confirmed with you before importing the extract into the test environments of Collections and Rosetta as appropriate.

For more information on creating a TMF, see our factsheet *Digital transfer preparation – Extract creation (23/F33)* available on our website.

## 6. Analysis results

We will consolidate our analysis results in an extract analysis report or 'warrant of fitness' for discussion with you. This report assesses your digital records' transfer readiness (i.e., the quality and consistency of sentencing decisions) and current digital health (i.e., the identification of unique file formats and potential digital preservation issues).

The extract analysis report concludes with a recommendation for your organisation to either:

- undertake more transfer preparation and repeat the eligibility selection, readiness assessment and test extract steps, or
- proceed with planning and preparation for a full extract, or
- postpone the proposed transfer.

To document this analysis, we will use the *Extract analysis report template (23/Fm9)* available on our website.