Digital transfer process

A step-by-step guide

March 2023

Te Rua Mahara o te Kāwanatanga
ARCHIVES NEW ZEALAND
Part of the Department of Internal Affairs

New Zealand Government

Document details

Document Identifier: 23/G24

Version	Date	Description	Revision due
0.1	Nov 2022	Development Draft	
0.2	Jan 2023	Final draft	
1.0	Mar 2023	Final publication version	Mar 2026

Contact for enquiries

Government Recordkeeping Directorate
Archives New Zealand
Phone: +64 4 499 5595
Email: rkadvice@dia.govt.nz

Licence

Introduction

We manage transfers of digital information and records (digital records) on a case-by-case basis in stages, some of which are broken down into steps. These stages are similar to the stages of a transfer of physical information and records, except more work must be done beforehand to prepare and assess the transfer readiness of the digital records and consequently, the feasibility of the transfer.

The process for transferring digital records to us consists of five stages:

Stage 1. Transfer initiation

Stage 2. Transfer planning

Stage 3. Transfer preparation

Stage 4. Transfer

Stage 5. Post-transfer.

Many of the steps in the stages are also similar to transfers of physical information and records, for example, the use of a Transfer Agreement and an Access Authority. However, a distinctive characteristic of digital transfers is the repetitive nature of some of the steps. There are likely to be many digital health checks done and issues needing to be resolved throughout the process. Test extracts for pre-transfer analysis are also necessary, similar to the sample checks that are done for transfers of physical information and records. This means that flexibility is essential when planning timeframes for a digital transfer.

We are continually developing guidance and identifying tools and methods that public sector organisations can use to assist in the digital transfer process, and we encourage organisations to provide feedback to assist us with shaping this guidance and sharing their experiences with others.

Stage 1: Transfer initiation

This stage can help organisations identify digital records that are eligible for transfer to us and assess their transfer readiness.

Step 1: Identify eligible digital records

We have developed methods for processing digital records into Rosetta (which stores and preserves the Government Digital Archive) and making them accessible through our Collections search tool. However, we need digital records to have certain eligibility characteristics in order to be transferred successfully.

If a public sector organisation considers that they have a set of digital records that meet the characteristics listed below, please contact us via rkadvice@dia.govt.nz. We will work with organisations on a case-by-case basis for digital transfers as our tools and methods continue to evolve.

1.1 Disposal

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.

If a set of digital records meets the following characteristics, then it may be a possible candidate for digital transfer:

- all records are of archival value and have a disposal action of 'Transfer to Archives New Zealand'
- all records are considered to have been accurately sentenced
- all records have met their minimum retention period and are no longer needed to meet administrative, legal or accountability requirements, and
- all records are the 'authoritative' version.

Ideally only digital records that have been sentenced as 'Transfer to Archives New Zealand' are included in the transfer. This means that your organisation needs to have a current disposal authority. Depending on your systems and knowledge about the digital records, you will need sufficient resources to sentence the records accurately against this disposal authority. Similarly, records should have met the minimum retention period prescribed in the disposal authority, however due to the volatile nature of digital records, we are open to discussing earlier transfer.

If digital records have reached the end of their recommended retention period but are still required for business use by your organisation, this indicates the records are not yet ready for transfer. You should seek advice on deferral of transfer of the digital records by contacting us on rkadvice@dia.govt.nz.

Although there is no one product or provider that can offer a fully automated sentencing process with a guarantee of 100% accuracy without human input, eDiscovery software offers ways to use computer or technology assisted review to:

- understand holdings of digital records at a high level
- reduce the amount of information to review
- extract meaning by categorising and clustering information, and
- identify personal and sensitive information.

PLEASE NOTE: It is very important that all digital records transferred are the 'authoritative' version and that your organisation is are aware of their post-transfer responsibilities (see Step 10). All remaining in-house copies of the records must be destroyed once the transfer has been fully processed and confirmed by us. Digital records that have been transferred to the custody of the Chief Archivist are public archives and must not be downloaded from Collections and resaved by your organisation. The dangers of this are the potential loss of authenticity, reliability, and integrity in these public archives as the 'authoritative' version.

1.2 Access

Under section 43 of the Public Records Act 2005, you are required to undertake a thorough sensitivity review of your digital records prior to transfer to determine whether their public access status is 'open' or 'restricted'.[1] This is not the same as having an internal 'open-by-default' policy for staff access to digital records. Once fully processed by us, digital records classified as 'open' will be viewable immediately online by members of the public. Any access conditions and their duration on digital records classified as 'restricted' must be made in consultation with the Chief Archivist and documented in an Access Authority (see Step 6).

Digital records that are known or believed to have password protections on them must also be identified ahead of transfer.

1.3 Size

From experience we have found that the complexity of technical assessment and analysis required by your organisation (see Step 2) and by us (see Step 5) is not related to total size of the transfer. There can be a high number of technical issues found in a small transfer, or only a few issues found in a transfer of thousands of consistent digital records. Please note however, that the quantity of issues may influence decisions by your organisation and us on whether to proceed with a full transfer or not.

1.4 Transfer metadata and file formats

We will not accept digital records for transfer without information that describes what those records are. This information is metadata. A metadata file or list that includes metadata for all the records must accompany the transfer. At a minimum, we expect you to provide the mandatory metadata elements required by *the* Information and records management standard (16/S1).

Although we prefer XML, we currently have no fixed requirements on the format of the transfer metadata file (TMF) and we will accept Excel, CSV or TXT. However, it is important that the TMF is structured consistently, and that your organisation has the expertise to understand the metadata and can assist us to understand it.

It is important for us to be made aware of any specialist file formats, or any deliberate format modification that may have taken place: for example, if Microsoft Word documents have been migrated to Adobe PDFs. However, there are no restrictions on the file formats that we will accept in digital transfers, in fact we strongly prefer to have the original format - you are not required to transform your digital records into other file formats for transfer to us.

Step 2: Assess your transfer readiness

The following tools and methods, which your organisation can use to assess your digital records in readiness for transfer to us, can also be useful to understand how well you are managing your records for digital continuity.

You are welcome to contact us for specific guidance and advice if you are considering the use of these or other similar tools.

2.1 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.

[1] This is different from the national security classification of RESTRICTED.

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)[2] 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 SQLint[3] or Demystify[4] can be used to discover more details about the record set intended to be transferred. They can be used (among other things) to:

- quality check the accuracy and consistency of content sentencing (for example, showing timelines based on last modification dates; duplicate records; summary of file formats), and
- locate obvious sensitive, non-business related and/or draft material by listing potentially problematic words or characters in file folder names. This will assist organisations in identifying and managing any access risks.

[2] 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).

[3] 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).

[4] Demystify is a way to analyse DROID CSV and Siegfried 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).

2.2 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 you 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 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 of the digital records included in the transfer, the TMF must also include the following information so the transfer can be processed by us:

- a checksum value (generated using one of the methods outlined below)
- a 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).

For more information on TMF, see Step 4.3.

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). For more information on checksum values, see Step 4.4.

We will identify any other technical metadata we need to enable the digital preservation of the records. You can also provide any extra metadata that you need to add value and enable discovery of the records.

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.

Stage 2: Transfer planning

Transfers of digital records to us can be unique, complex and take place over extended periods of time. We will collaborate with your organisation to agree on roles and responsibilities when planning a transfer of digital records and their metadata to the Government Digital Archive, and when transferring legal custody for the records to the Chief Archivist (see also Stage 4).

Step 3: Agree roles and responsibilities

A *Digital Transfer Management Plan Agreement*, similar to that used for transfers of physical information and records, will be discussed and agreed to. Its purpose is to ensure that both your organisation and us have a common understanding of key aspects of the process and are informed of factors which might affect the transfer at any stage. It is intended to guide both the preparation and transfer of the test extract (see Step 4) as well as the formal transfer of the full extract of all eligible digital records and their accompanying metadata (see Step 7), and to facilitate communication.

See the *Digital Transfer Management Plan Agreement template* (23/Fm8) available on our website.

Stage 3: Transfer preparation

In order for us to determine whether a digital transfer is feasible, you must first create an extract or copy of a test or sample set of eligible digital records (see Step 1) 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 (see Step 7).

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

Step 4: Create an extract

4.1 Identify extract requirements

We will meet with you to create a list of requirements for the test extract of eligible digital records and their metadata. These 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/Fm4) may be used which is available on our website.

4.2 Export extract

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

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

4.3 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 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[5] 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

[5] 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).

- Date last modified
- File/Folder and
- Checksum values (see Step 4.4)

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 records.

4.4 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 organisations 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 TMF (ideally both). Providing the checksum for each transferred record allows us to validate that record and make sure that all the 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.

4.5 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 as the download option from a secure organisation system/environment or Government Cloud services.

Before transporting the extract to us, you must check that all the records are synchronised or have been copied correctly. We recommend using the tool 'rsync'[6] (Remote Sync) for copying and synchronising of digital records. This tool preserves the integrity of the records and their accompanying TMF, and we are happy to assist you in its use.

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

[6] 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).

Step 5: Pre-ingest analysis

Once the extract is received by us, several analytical processes (both automated and human) are run over the digital records and their TMF to identify any issues that may affect ingest or transfer into the Government Digital Archive.

5.1 Integrity check

We use the checksums supplied by you in Step 4 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)* available on our website.

5.2 Content analysis

We use a number of automated tools such as DROID, SQLint and Demystify (see Step 2.1) to assess the capability and accuracy of the sentencing or selection of eligible digital records against your organisation's disposal authority(ies) and/or our General Disposal Authorities, GDA6 and GDA7.

These tools are also used to check for any duplicate records and obvious sensitive, non-business related or draft material, as well as missing records, 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 Step 2).

5.3 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 records, and system files which might need to be excluded from the transfer after discussing with you.

5.3 Metadata analysis

We 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 Step 4.3.

5.4 Analysis results

We consolidate the 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 their 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 Steps 1, 2 and 4, or
- proceed with planning and preparation for a full extract, or
- postpone the proposed transfer.

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

Stage 4: Transfer

Once we have reached agreement with you to proceed with a full transfer of all eligible digital records and their accompanying metadata to the custody of the Chief Archivist, we will collaborate on formal transfer planning and preparation. We will provide any further requirements necessary for a full extract which will undergo the same creation and analysis processes as the initial test extract (see Stage 3).

Step 6: Plan a formal transfer

This process is basically the same as planning for a transfer of a test extract (see Step 3) but will include completion and signing of a Transfer Agreement, and preparation and signing of an Access Authority if required.

6.1 Sign a Transfer Agreement

We will prepare a Transfer Agreement which you will need to sign before the digital records and their metadata are formally accepted into the custody of the Chief Archivist.

See our website for more information about transfer agreements as part of our Transfer Process.

6.2 Complete and sign an Access Authority if required

All digital records that are 25 years or older must be classified as either ‘open’ or ‘restricted’ access, regardless of where the records are held. If digital records are transferred to us before the 25-year limit (which we strongly recommend), their access status must be determined as part of the formal transfer requirements.

If any of the digital records in the proposed transfer is classified as ‘restricted’ access, you will need to complete and sign an Access Authority form before the records are formally accepted into the custody of the Chief Archivist.

For more information about access classifications, see our guidance on Access decisions (16/G6).

Step 7: Prepare a full extract

This process is basically the same as Steps 1,2 and 4, but you will need to identify, assess and extract or copy a full extract of all eligible digital records and their accompanying TMF for transfer to us via a hard disk drive or other secure transport method available.

Step 8: Pre-ingest analysis

This step is the same as Step 5, and depending on our analysis findings, we may recommend that you repeat Step 7 if issues are identified that may affect ingest or transfer of the digital records and their TMF into the Government Digital Archive.

Step 9: Ingest into the Government Digital Archive

When you have addressed any content, technical, metadata and accessibility issues identified in our pre-ingest analysis, we will ingest the full extract of digital records and their accompanying TMF into the live environments of the Government Digital Archive, and formally accept the transfer.

9.1 Metadata mapping and ingest

We will map your TMF to the metadata in Collections and Rosetta (which collectively form the Government Digital Archive) before ingesting your digital records and their TMF into the production environments of both systems.

At this stage, we may seek additional information from you, for example, to assist in this mapping or if you want extra metadata to add value and enable discovery of the records.

9.2 Formal acceptance of transfer

We will register and return copies of the signed Transfer Agreement and Access Authority (if required) to you, and formally accept transfer of the digital records and their accompanying metadata as public archives into the custody of the Chief Archivist.

See our website for more information on Transfer.

Stage 5: Post-transfer

Once your digital records have been formally received as public archives into the custody of the Chief Archivist, it is your responsibility to destroy any in-house copies.

Step 10: Destruction of all in-house copies

We will notify you when the transfer is complete, and the digital public archives and their metadata are discoverable via Collections. Once this notification is received, you must destroy all remaining copies and versions of the digital records held in-house.

PLEASE NOTE: This step is extremely important and necessary to ensure the authenticity, reliability, and integrity of the digital public archives as the 'authoritative' version.