Destruction of source information after digitisation

August 2018

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

New Zealand Government

Document details

Document Identifier: 17/G13

Version	Date	Description	Revision due
0.1	Feb 2017	Development Draft	
1.0	Mar 2017	Publication	Mar 2020
2,0	Oct 2017	Legislation reference changes due to repeal of the ETA 2002, replaced by the CCLA 2017, Part 4	Mar 2020
2.1	Aug 2018	Editorial correction	Mar 2020

Contact for enquiries

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

Licence

Contents

1 Introduction 4

2 Legal overview 4

3 Decision workflows 5

3.1 "Public records" 5

3.2 "Local authority records" 5

4 Meeting the conditions of section 229(1) of the CCLA 6

4.1 Integrity 6

4.1.1 Complete 6

4.1.2 Unaltered 6

4.1.3 Intrinsic value 7

4.2 Useable 9

5 Recommended Minimum Technical Specifications 10

5.1 Glossary of technical terms 10

5.2 How to use these technical specifications 11

6 Quality Assurance 11

Appendix 1 Flowcharts 13

Decision flowchart for "public records" 13

Decision flowchart for "local authority records" 14

Appendix 2 Quality assurance checklist 15

Example of a quality assurance checklist from Hutt City Council Archives 15

1 Introduction

Under the Contract and Commercial Law Act 2017 (CCLA) and the Public Records Act 2005 (PRA), public offices and local authorities have conditions to meet before destroying the source information after converting to electronic form (referred to in this Guide as "digitisation").

This guide explains how to create electronic information of sufficient quality to replace the source information and enable long term retention and digital preservation (if and when required), allowing the source information to be destroyed. When the conditions of the CCLA are met, the electronic form of the information becomes the authoritative record.

This guide provides relevant guidance for all public offices and local authorities on meeting the conditions of the CCLA and includes the minimum recommended technical specifications and quality assurance considerations.

In conjunction with this guide, public offices must refer to the *Authority to retain "public records" in electronic form only* (the Authority*)*, before destroying source information. For local authorities, the Authority is not relevant because it is issued under section 229(2) of the CCLA which applies to "public records" only. Therefore, local authorities should refer to this guide only. Flowcharts giving an overview of the decision making processes are contained in Appendix 1.

The capture and management of born-digital information, and digital preservation, are not within scope of this guide nor is detail about planning, organising and managing digitisation projects. For guidance on specialised formats, such as audio-visual, contact Archives New Zealand.

The standard *AS/NZS ISO 13028:2012, Information and documentation – implementation guidelines for digitization of records* is recommended guidance for digitisation processes and policies.

2 Legal overview

The PRA framework for information management in the New Zealand public sector ensures the preservation and accessibility of information required for long-term retention. Under the PRA, public offices and local authorities must create and maintain full and accurate records of their activities and, in the case of "public records" and "local authority protected records", ensure their ongoing accessibility.

Under the PRA, public offices and local authorities must not dispose of "public records" and "local authority protected records" without authorisation from the Chief Archivist. However, where the conditions of section 229 of the CCLA are met, a public office or local authority is able to keep an electronic form of a record, and destroy the source record, without seeking the specific authorisation of the Chief Archivist.

Section 229(1) of the CCLA provides that a legal requirement to retain information that is in paper or other non-electronic form is met by retaining the electronic form of the information if:

(a) the electronic form provides a reliable means of assuring that the integrity of the information is maintained; and

(b) the information is readily accessible so as to be usable for subsequent reference.

Section 229(2) of the CCLA provides that section 229(1) only applies to a "public record" if the Chief Archivist has approved the retention of that record in electronic form. The Chief Archivist has given

general approval to retain "public records" in electronic form only in the Authority, subject to certain categories of exclusions.

3 Decision workflows

3.1 "Public records"

Step 1. Does section 229 of the CCLA apply?

Section 229 of the CCLA applies to all "public records" and "local authority records" except those that are listed in the Schedule to the CCLA ("*Enactments and provisions excluded from subpart 3 of Part 4*"). These enactments and provisions (i.e. legislation, regulations) are not covered by section 229 of the CCLA. The legal requirement for managing the records covered by the Schedule cannot be met by electronic means. Public offices are responsible for keeping up-to-date with any amendments to the Schedule.

Step 2. Is the "public record" covered by the Authority?

In the Authority, the Chief Archivist gives general approval to retain, after digitisation, "public records" in electronic form only. The Authority outlines the following categories of information that are excluded from general approval. Examples are given in section 4.1.3 of this guide.

- Unique or rare information, information of importance to national or cultural identity or information of historical significance
- Unique or rare information of cultural value to Maori and their identity
- All information created prior to 1946.

Note: If a public office wishes to digitise information that is within one of the above categories, and destroy the source information, it must first contact Archives New Zealand to seek authorisation from the Chief Archivist.

Step 3. Have the conditions of CCLA section 229(1) for integrity and usability been met?

For details about meeting these conditions see pages 6-9.

If all three steps above are satisfied, the digitised information can be retained and the source information destroyed in accordance with section 229 of the CCLA. No authorisation from the Chief Archivist is required.

3.2 "Local authority records"

Step 1. Does section 229 of the CCLA apply?

Section 229 of the CCLA applies to all "public records" and "local authority records" except those that are listed in the Schedule to the CCLA ("Enactments and provisions excluded from subpart 3 of Part 4"). These enactments and provisions (i.e. legislation, regulations) are not covered by section 229 of the CCLA. The legal requirement for managing the records covered by the Schedule cannot be met by electronic means. Local authorities are responsible for keeping up-to-date with any amendments to the Schedule.

Step 2. Have the conditions of CCLA section 229(1) for integrity and usability been met?

For details about meeting these conditions see pages 6-9.

If the two steps above are satisfied, the digitised information can be retained and the source information destroyed in accordance with section 229 of the CCLA. No authorisation from the Chief Archivist is required.

If either step is not satisfied, the authorisation of the Chief Archivist under section 40(3) of the PRA is required before a source "protected record" can be destroyed.

4 Meeting the conditions of section 229(1) of the CCLA

4.1 Integrity

There are two requirements of section 229(1) that must be met before information can be destroyed after digitisation. The first is that –

- The electronic form provides a reliable means of assuring that the integrity of the information is maintained

The integrity of information will be maintained only if the information has remained **complete** and **unaltered** other than the addition of any endorsements or immaterial changes that arise in the normal course of communication, storage or display (section 221 of the CCLA) and digitisation does not destroy any cultural, historical or other value intrinsic to the source form of the record. Further detail is provided below.

4.1.1 Complete

The Chief Archivist's view is that information will be **complete** if the digitisation process:

- Successfully digitises all of the information within the range or aggregation for digitisation
- Fully and accurately reproduces all of the information from the source, for instance:
  - Digitised information is legible and accurately represents what was contained in the original. The faintest element needs to be legible
  - All annotations, attachments and enclosures are captured
  - Pages should be un-cropped and not skewed
  - Colour is reproduced to the required extent. For instance, handwritten coloured annotations and maps marked with colour are essential information; although a coloured logo, letterhead or invoice may be less important
  - Optimisation only includes essential enhancements to improve legibility and quality of otherwise indistinct or faded elements in the source.

4.1.2 Unaltered

For information to be **unaltered** means that:

- The system in which the digitised information is captured, stored, and managed has adequate security, access and file integrity controls to ensure the information remains inviolate (tamper-proof)

- Files created by digitising the information can only be changed by users with appropriate privileges, and the changes must be logged. Disposal (including destruction) can only be actioned by authorised users with the appropriate system privileges. Disposal actions must be authorised and documented.

- Files created by digitising the information are checked for integrity by using checksums at least annually. This is to identify any unintentional changes in the digitised record bit stream.

- An audit trail must be generated and available to capture the details of any changes. Audit trails must log the usage of the information (create, read, update or edit, delete, dispose, etc.), to the extent required. This metadata helps attest to authenticity and reliability and must be maintained for at least the same period as the information itself.

If an alteration causes a loss of information, either at the time of digitisation, or subsequently, then the conditions of section 229(1) will not be met.

Acceptable **endorsements or immaterial changes** arising in the normal course of communication, storage, or display include:

- Optimisation or enhancements to improve legibility and quality during the digitisation process
- The creation or editing of metadata, including:
  - Descriptive and contextual metadata (recordkeeping metadata)
  - Administrative or process metadata accrued during use of the digitised information (or the logging of auditable events)
  - Technical metadata contained in the file which documents the details of the digitisation, including capture device, software, and operator
- Content indexing, such as optical character recognition (OCR) that enables full-text searches of the digitised content
- Electronic stamps, annotations, mark-up, or redaction. These are alterations (endorsements) to the information, providing that it is readily apparent that these do not permanently or irreversibly obscure the original information, they are immaterial

- Manipulation, measuring, plotting, overlays, and similar functions in image viewers. This is a function of the viewing software that will not alter the underlying information.

4.1.3 Intrinsic value

Intrinsic value refers to the qualities and characteristics of the original item inherent in that item and which contributes to the archival value. **Information with intrinsic value should be preserved in its source paper or non-electronic form**.

With digitisation, it may not be possible to preserve all of the qualities and characteristics of the source information. The source information may have intangible, historic, cultural, iconic, symbolic, talismanic, or aesthetic qualities and attributes, as well as being unique or rare or having significant monetary value.

Below are some examples of information that are most likely to have intrinsic value.

Unique or rare information, information of importance to national or cultural identity or information of historical significance

- Items made from rare materials, such as parchment, volumes with unique form or binding, glass negatives
- Uniqueness or rareness, such as mint issue stamps, rare publications, seals, wax seals
- Items of artistic or cultural significance, or with aesthetic qualities, such as watercolour sketches, photographs, hand-coloured maps, architectural drawings, illuminated manuscripts
- Original documents of general and substantial public interest due to a direct association with historically significant people, places, things, issues, or events, such as the Treaty of Waitangi, or documents relating to a Prime Minister, or Governor-General
- Primary establishment documents significant for the establishment or continuing legal basis of an agency or institution, the functions or powers of government, or the formulation of the highest levels of legislation, constituent documents, such as charters or articles of incorporation, signed minutes, seals
- Documents with an individual as the subject which have significance or value to that individual or others as evidence of their ancestry or heritage and which may contain original photographs, handwriting, such as immigration information, service dossiers
- Significant awards, warrants of appointments, honorifics

Unique or rare information of cultural value to Māori (land and people) and their identity

- Items containing whakapapa (genealogy)
- Items that tīpuna (ancestors) have signed or written
- Items where the content relates to connections to the land, this includes hapū and whānau knowledge and evidence provided for this purpose and includes urupā (burial grounds)
- Maps that show areas of significance including iwi place names, waterways, topographical and geographical features relating to connection to land
- Te reo Māori content (handwritten form)

All information created prior to 1946

This aligns with the existing Archives New Zealand general disposal authorities. Prior to 1946 sources of information are limited and thus have inherent archival value.

Examples

19th Century field books (1872, 1880) created by the Railway Department. These field books were used by J. H. Henderson, the Chief Engineer. The field books contain annotated notes throughout made by J.H. Henderson. One of the field books contains plans and diagrams of various railway machinery, fencing, tunnels, bridges, cattle stops, and level crossings. The book still has the original metallic pencil and a bookmark attached to it. Archives New Zealand assessed that these field books have intrinsic

value by reason of age and historical associations. The physical characteristics also contribute to its archival value.

Leather bound cash registers (1871-1933) created by the University of New Zealand. These registers were originally offered for sale at auction and later gifted to Archives New Zealand. Financial transaction records (i.e. general ledger, cash books) are usually not archival. However, these registers record the establishment of the university and the contemporary relationship between the university and its affiliated colleges. These registers also complement Archives New Zealand's existing holdings. For these reasons as well as their age and rarity, Archives New Zealand assessed that the registers have intrinsic value that contributes to their archival value.

4.2 Useable

The second requirement of section 229(1) is that –

- The information is readily accessible so as to be **usable** for subsequent reference

The Chief Archivist considers that useable information is information that can be located, retrieved, presented and interpreted within a reasonable time period. A usable record should be connected to the business process or transaction that produced it. Linkages between records that document related business transactions should be maintained (*ISO 15489-1:2016 Information and documentation – Records management – Concepts and principles*).

To be **usable** the digitised information should be:

- Locatable and retrievable. Adequate, persistent and searchable descriptive metadata should be present in order to retrieve and understand the information. Metadata must support understanding of the physical form of the source information
- Stored and managed in systems that are designated as approved, corporate repositories or information systems (including line-of-business applications; EDRMS, ECMS, or other digital asset management or preservation systems). These systems should:
  - Be operated and administered by skilled staff and allow only authorised staff to dispose of information (including deletion)
  - Have been tested and proven to be reliable over several years or versions of software
  - Log activity to the detail required
  - Provide adequate storage and the reliable operation of data protection (including backup and disaster recovery), consistent with normal, prudent IT practice.

5 Recommended Minimum Technical Specifications

Bit depth		Resolution	File Format	Compression
8 bit	Greyscale or bi-tonal	300 ppi	PDF/A TIFF JPEG 2000	Lossless compression
24 bit	Colour	300 ppi	PDF/A TIFF JPEG 2000	Lossless compression

These technical specifications are applicable to the digitisation of text and photographic prints.

5.1 Glossary of technical terms

Colour resolution or bit depth

Bit depth: number of bits (zeros or ones) used to describe the colour of each pixel. Bit depth can range from 1 bit up to 48 bits. Greater bit depth allows a greater range of colours or shades of grey to be represented by a pixel.

Bi-tonal: Black and white, two tone scan.

Greyscale: Black and white in addition to a range of intermediate greys, where 8 bits are required to describe each pixel.

24 bit colour: enables the storage of 8 bits of information for each red, green and blue component of each pixel. 24 bit colour, also known as True Colour, enables millions of colour variations.

Resolution

A measure of the ability to capture detail in the original work, and only one aspect of quality in an image. It is often quantified in pixels per inch (ppi). The optimum resolution depends on the nature of the documents being scanned and the smallest meaningful element in them being able to be rendered legible in the scanned image.

File Format

PDF/A standards apply to long-term archiving of electronic documents. PDF/A is highly suited to long term information and records management because it is entirely self-contained. PDF/A differs from PDF by prohibiting features ill-suited to long-term archiving, such as font linking (as opposed to font embedding) and encryption. The most widely used standards for PDF archiving are PDF/A-1a and PDF/A-1b (for less stringent requirements). A PDF/A file may be marginally larger than the original PDF file it was created from. Fonts are embedded in a PDF/A file and more information is stored in the metadata. Colour profiles also add to overall file size, but the added file size is usually minor and is justified by the extra information embedded.

TIFF or Tagged Image File Format is a common file format for graphics and photographs. TIFF, with lossless compression, has been extensively used and remains a default preservation imaging format. The considerable size of lossless TIFF images can present significant storage issues.

JPEG 2000 is an image coding system that uses state-of-the-art compression techniques; its architecture is useful for many diverse applications, including Internet image distribution, security systems, digital photography, and medical imaging.

Compression

Compression is a set of algorithms designed to reduce the size of an image for storage or transmission.

Lossless is when no information is irretrievably lost. Objects where lossless compression was applied are identical to the original source after they are un-compressed, i.e. no bits (or information) is lost.

5.2 How to use these technical specifications

Technical specifications help ensure the legibility and usability of the digitised information for as long as is required. These specifications are primarily for information that has long term/high value to the organisation or is required to be transferred to Archives New Zealand. In these cases, they are highly recommended as the minimum needed if it is planned to destroy the source information.

As a general rule, the highest technical specifications that can be realistically supported should be used. For information that is being digitised and is covered by a disposal authority and identified as ultimately for transfer to Archives New Zealand, the expectation is that these specifications, or better, will be applied. If information is not covered by a current disposal authority then these specifications should be treated as the default to cover the possibility of information of archival value being included.

If an organisation decides that these recommended minimum technical specifications are not suitable for its particular digitisation work, then the choices are:

- Use a more rigorous set of technical specifications or
- Use a less rigorous set of technical specifications and keep the source.

For information that has low value and is transitory or facilitative, organisations should decide the required level of quality for digitised information sufficient to support their business needs. To determine the right technical specifications for the organisation's digitisation work, all relevant factors should be analysed to ensure functional equivalence requirements of the CCLA are met.

6 Quality Assurance

Quality assurance involves checking the operation and output of digitisation processes against agreed benchmarks to ensure that the benchmarks are met. This ensures that the digitised information has integrity, is unaltered, complete and authentic, and is able to 'do the job' of the source information.

Quality assurance must be completed before the digitised information is used in business processes and before the source information is destroyed.

The level of quality assurance depends on whether or not the digitised information will replace the source information. If the source is to be destroyed, then higher levels of quality assurance are needed.

Organisational policies and procedures for digitisation and destruction of source information should include roles, responsibilities and delegations dependent on the value of the information and associated risk.

Quality assurance procedures are to be documented and built into the ongoing operation of the digitising process, and not just a check on output. The procedures are, at minimum, to address the following issues:

- any acceptable variations from normal procedures
- scanner operation quality control
- verification that digital output matches the quantity of the source input
- extent and frequency of sampling of digitised images
- criteria for checking image quality
- frequency and criteria for checks on metadata
- checks that digitised information is (re)producible in its original format
- short-term retention of the source information in case re- re-digitising is required due to faults in the original scanning
- processes for re-digitising and
- operator training

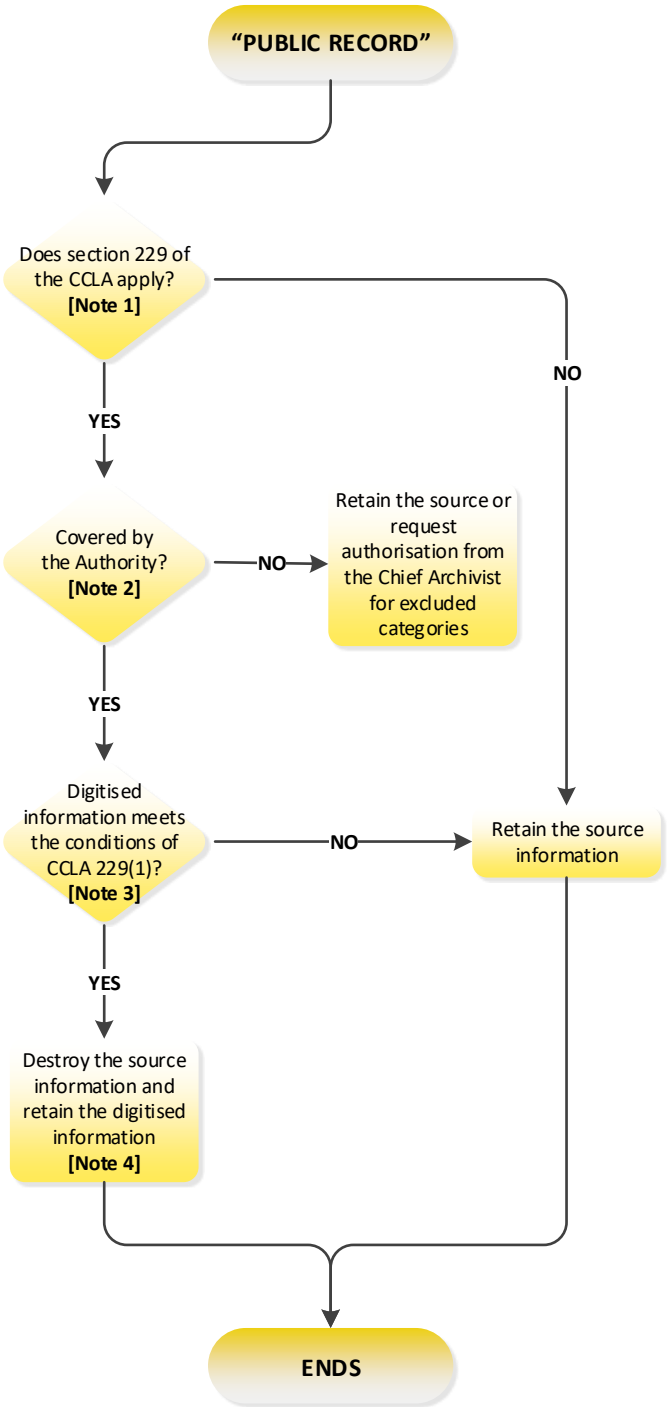
If outsourcing digitisation, relevant documentation about benchmarks should be agreed with service providers. Checklists can be useful and an example from Hutt City Council Archives is contained in Appendix 2. If outsourcing includes disposal, organisations are still responsible for ensuring that procedures for disposal are in line with their responsibilities under the PRA.

Periodic review of quality assurance procedures is important to ensure that benchmarks and quality assurance measures continue to meet business needs and changes to legislation.

Appendix 1 Flowcharts

These flowcharts are applicable for organisations that have an approved quality assurance process for digitisation.

Decision flowchart for "public records"



NOTES

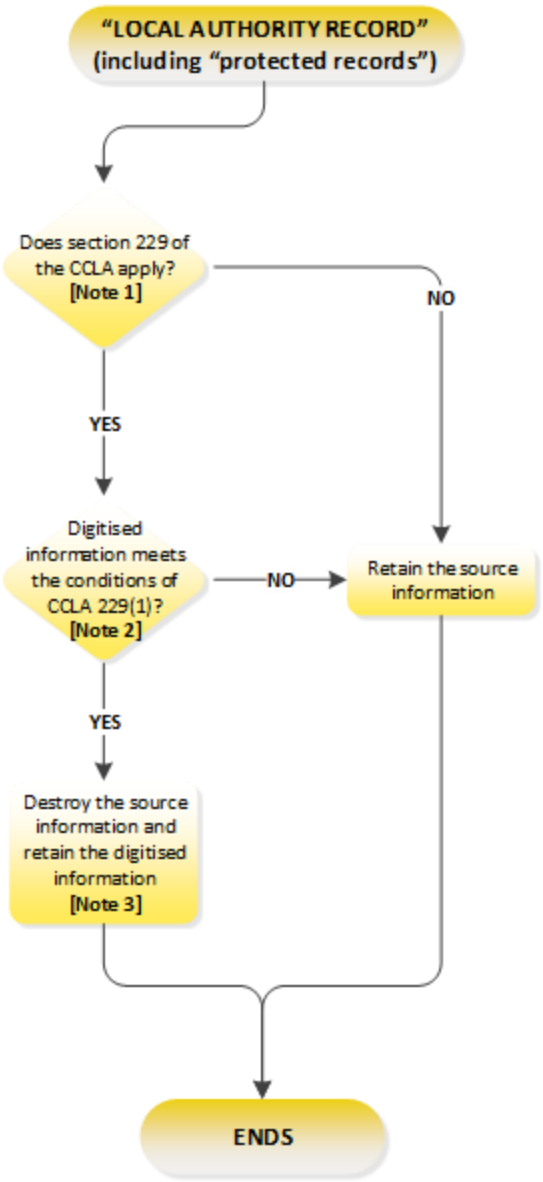
Note 1:
To be excluded it must be listed in Schedule 5 to the Contract and Commerical Law Act 2017 (CCLA)

Note 2:
The *Authority to retain public records in digital form only*

Note 3:
Section 229(1)(a) – *the electronic form provides a reliable means of assuring that the integrity of the information is maintained*
Section 229(1)(b) – *the information is readily accessible so as to be usable for subsequent reference*

Note 4:
The *Authority to retain public records in digital form only* applies

Decision flowchart for "local authority records"



NOTES

Note 1:
To be excluded it must be listed in Schedule 5 to the Contract and Commercial Law Act 2017 (CCLA)

Note 2:
Section 229(1)(a) – *the electronic form provides a reliable means of assuring that the integrity of the information is maintained*
Section 223(1)(b) – *the information is readily accessible so as to be usable for subsequent reference*

Note 3:
There is no legal requirement to retain source "protected records" if the conditions of the CCLA are met.

Appendix 2 Quality assurance checklist

Example of a quality assurance checklist from Hutt City Council Archives

Hutt City Council Archives outsourced digitisation of information that was to be used as part of LIM reports. The pilot project returned a large percentage of information which failed their quality assurance check. However, once they established a quality assurance checklist for the service provider to follow and sign off, there was a 100% pass rate on the following batches (see checklist below).

SCENARIO	**ACTION**	**QA** Y/N
File Preparation	Remove staples and attach paper clips or clippies for larger documents, flatten dog ears so no text obscured, Unfold **everything.** Where there is a stapled note to a page or plan and no text obscured – leave the staple in place and use a piece of mylar or a mylar sleeve to cover the staple. Cellotaped notes to a page and no text obscured-leave in place. Cellotaped notes to a page and no loose-position on page and use a mylar sleeve. Attachments – covering original text and showing a change in wording – remove staple and clip in exact same place. Pages or plans in pieces-leave in place and scan as one.	
age Compression	No compression of images.	
Target Pages	Target Pages for Council seals.	
Folders and Pockets		
File in a Cardboard Folder	Remove contents from the file then Scan Front of File Folder then contents then Back of File Folder. This will prevent loss of brightness due to distance from Light/sensor..	
File in a Plastic Pocket	Remove contents from plastic pocket then Scan Front of Pocket then contents then Back of File Pocket.	
Content		
Text	Scan all pages with text. This includes **both** sides of a plan where there is a file number only written on the back in pencil. Note please open plan up to scan the reverse side.	

	This includes **both** sides of a note which has been made on paper with old text on the back. This includes post-it notes with handwritten notes e.g. “Send to John”	
Envelopes in the File	Remove contents from the pocket then scan the envelope front, then contents then scan the envelope back.	
Stapled or Paper Clipped Document	Remove staple/paper clip, then scan in order found. This includes scanning cube notes as one page. Then reattach with a paper clip.	
Page with Attachment Showing a Change in Text	Scan page with attachment in place first, then lift attachment and scan the page.	
Page with a Stapled Note Not Obscuring Text	A4: Leave staple in place and use a mylar sleeve to cover staple. This includes Minute pages with stapled notes. A3+: Leave staple in place and use a piece of mylar to cover staple.	
Page with a Separated Cellotaped Note	Position note on page using the cellotape marks to line up and use a mylar sleeve to hold in place.	
Page Order	Scan in the order you find it. Only change if there is a clear numbering system and a page is out of order.	
Dog Ears	Major dog ears and/or where covering text rectify. Minor dog ears and/or not covering text will rectify, but there is some reasonable limitations	
Damaged Page or Plan in Pieces	Scan as one not as the separate pieces. Where an object requires repair to scan, or cannot be scanned due to disintegration, scanning company will flag to the HCCA’s attention and advice.	
racer Cards	Scan if text written on it.	
Rotation/Orientation	Orientate the scanned TIFF image as Right Reading. Where text appears in multiple directions favour orientation of majority of text, or the Title when in doubt.	
Cropping and Skewing	Use automated cropping (mostly for maps). When skewed this can be tricky. Look out for in the QA process.	
Metadata	Metadata in Keywords only.	

Rebinding	Ensure all documents are on pins and connectors are attached and hooked.	
Folders scanned that were already digitised	HCCA will remove folders already digitised prior to box delivery to the scanning company and provide a tracking sheet per box.	
TIFF and PDF/A	Capture as TIFFS and create PDF/A 1b.	