

Standard

Information Asset Digitisation and Preservation

Overview

This guide outlines the preservation standards applied by the University when digitising hardcopy original information assets in line with the University's Records and Information Management Policy. These standards must be adhered to for the capture of digitised information assets into an approved system such as Content Manager Electronic Document and Records Management System (EDRMS) where hardcopy original information assets are intended for disposal under General Disposal Schedule 21 v5 (GDS 21) (as amended) - for the disposal of hardcopy information assets after digitisation. The standards are based upon those employed by the National Archives of Australia but specifically tailored to suit University requirements.

1. Hardcopy source documents can be destroyed following digitisation and capture, if the Director of Library Services (or delegate) has certified that:
 - a. the digitised process uses compliant practices that ensures the digitised counterpart is a true and accurate representation of the hardcopy source, and
 - b. the digitisation is captured into an approved system that manages the digitised counterpart in a way that ensures the information asset's ongoing authenticity and reliability.and ensures the ability of the:
 - c. system to manage the digitised counterparts in a way that ensures the information asset's ongoing authenticity and reliability, and
 - d. business area to manage the digitisation process in a way that ensures a true and accurate record is retained as evidence of a business activity or transaction.
2. This practice excludes permanent value hardcopy source documents created prior to 1 January 2005. These cannot be destroyed and must continue to be maintained on a hardcopy file but can be digitised for ease of access and reference.

Contents

Overview	1
Information Asset Digitisation Preparation	3
Minimum Mandatory Digitisation Standards (Per Content Type)	3
Documents (Paper Documents under A3)	3
Large Format Documents (over A3, including maps, plans and other large-format information assets)	4
Photographs	4
Photographic negatives and transparencies	4
Bound Volumes (includes books, diaries, ledgers, registers)	4
Endorsed Digitisation Device List	5
Mandatory Digitisation Verification Checklist	5
Capture of Digitised Information Assets into Content Manager	5
Glossary - Valid Retention Formats for Digital Information Assets	6
GDS 21 Source Information Asset Disposal Validation	8
Hardcopy Source Information Asset Retention	8
Disposal Validation Approach	9
Hardcopy Information Asset Digitisation and Disposal Verification Process	10

Information Asset Digitisation Preparation

To prepare your hardcopy original information asset for digitisation:

1. Find a desk space clear of any clutter and other documents as you don't want to impair the integrity of the hardcopy original information asset by accidentally mixing other documents into it.
2. Remove the hardcopy original information asset/s from the file (if applicable).
3. Using a staple remover (easy glide is preferred over claw-style), remove all staples from the document. As you do this, lift each page separately and turn it over to make sure there are no other staples hidden further in the document. Continue removing staples, paperclips, bulldog clips, and other fasteners and turning the pages over neatly stacking them along the clean side of the long edge.
4. If any documents have post-it notes or stickers obscuring text, move these to a blank part of the document or the reverse side of the document if possible. If there are no blank spaces, take a clean A4 page and add the post-it notes to the page and add to the stack of documents.
5. If there are documents that are smaller than A4 (for example a business card or eftpos receipt) stack these in the scanner appropriately so they can be picked up by the rollers, or alternatively, tape smaller items to a piece of A4 paper for ease of scanning.

Minimum Mandatory Digitisation Standards (Per Content Type)

Documents (Paper Documents under A3)

The below minimum mandatory standards apply to the digitisation of a physical original information asset with text or graphics, either for which colour is not present or is non-essential:

- 300ppi (pixels per inch)
- Bi-tonal (black and white / greyscale)
- Format: PDF or multi-page TIFF
- No compression
- Optical Character Recognition (OCR) for PDF searchability where practical

Documents where colour is present and vital to the context of the physical original information asset, or with low contrast (including faded text or coloured backgrounds):

- 300ppi (pixels per inch)
- 8 bit (24-bit) colour/sRGB
- Format: PDF
- No compression
- Optical Character Recognition (OCR) for PDF searchability where practical

Large Format Documents (over A3, including maps, plans and other large-format information assets)

- 400ppi (pixels per inch) minimum
- 600ppi recommended if fine detail is present on images
- 2540ppi recommended for aerial photography
- 24-bit colour/sRGB - Adobe RGB recommended for rare or significant documents, artworks or posters
- Format: PDF
- No compression
- Optical Character Recognition (OCR) for PDF searchability where practical

Photographs

- PPI (pixels per inch): Variable upon size:
 - Smaller than A6: 2700ppi
 - A6 to A4: 900ppi
 - Over A4: 600ppi
- 48-bit colour/sRGB - Adobe RGB recommended for rare or significant documents, artworks or posters
- Format: JPEG or PDF
- No compression
- Optical Character Recognition (OCR) for PDF searchability where practical

Photographic negatives and transparencies

- PPI (pixels per inch): Variable upon size:
- Smaller than 5"x7": 5000ppi
- Above 5"x7": 1600ppi
- 48-bit colour/ProPhoto RGB
- Format: PDF or single-page TIFF
- No compression

Bound Volumes (includes books, diaries, ledgers, registers)

- 300ppi (pixels per inch) minimum; recommend 600ppi where fine detail is present
- 24-bit colour/sRGB
- Format: PDF
- No compression
- Optical Character Recognition (OCR) for PDF searchability where practical
- Capture covers, spine and all pages
- Capture the entire page wherever practical

Endorsed Digitisation Device List

The below listing of endorsed University digitisation devices denotes those capable of meeting minimum mandatory digitisation standards per content type:

Device Manufacturer	Device Model	Device Type
Fujitsu	fi-6400	Desktop scanner
Kodak	S3100	Desktop scanner
Epson	V700 photo	Flatbed and photo scanner
Ricoh	IM C3000	Multi-function device
Ricoh	IM C4500	Multi-function device
Ricoh	IM C6000	Multi-function device

If the device you are using is not on the above list, check the specifications on the manufacturer's website to make sure it complies with the minimum mandatory digitisation standards.

Mandatory Digitisation Verification Checklist

Once digitised, the digital rendition of the file will become the source information asset, replacing the hardcopy original. It is imperative that you take the time to open your digital rendition of the information asset and verify each page has been captured, is legible and can be easily interpreted.

To ensure the integrity of the digitised information asset and confirm it as a one-to-one rendition of the hardcopy original, scroll down through each page in the document and compare the image on the screen with the document in front of you. Take special care to note:

- Have you captured all the information on the page?
- Is it all legible? Zoom into sections of pages to verify legibility.
- Are there any post-it notes or other stickers blocking text?
- Does the page need to be in colour for full context? For example, photographs, or a graph with a legend comparing the blue column with the red column.
- Are there page numbers on each page? If so, check that the page numbers are sequential.

If you find any text that is not legible, rescan that page and insert it into the digitised document.

Capture of Digitised Information Assets into Content Manager

Following successful digitisation and validation, the digitised information asset can be captured into the appropriate file to which it relates in Content Manager (or other approved system) using all mandatory titling and metadata conventions.

Glossary - Valid Retention Formats for Digital Information Assets

The below formats have been assessed as capable of maintaining the integrity of contextual information asset value over time, and are approved for the creation and capture of digital information assets in systems:

Information Asset Type	Valid Format
Audio	Moving Picture Experts Group MPEG 1/2 Audio Layer 3 (.mp3) Moving Picture Experts Group MPEG-4 (.mp4) Free Lossless Audio Codec (.flac) WAV using an LPCM codec (.wav, .bway, .bwf)
Computer Aided Design (CAD)	Drawing eXchange Format (.dxf) Drawing (.dwg) Standard for the Exchange of Product Data (.stp, .step, or .p21) Portable Document Format (.pdf/e)
Data	Comma Separated Values (.csv) Tab-separated Values (.tsv) Extensible Markup Language (.xml) JavaScript Object Notation (.json or .jsn) Software Independent Archiving of Relational Databases (.siard)
Documents (text)	Plain text (.txt) Portable Document Format (.pdf) Microsoft Word (.doc, .docx)
Electronic Publications	Portable Document Format (.pdf) Electronic Publication (.epub)
Email	Microsoft Outlook (.msg or .pst)
Encapsulation	7-Zip (.7z) ZIP (.zip) GZIP (.gzip) Tape Archive (.tar)
Geospatial	ESRI Shapefiles Geopackage GEOjson (.json) Digital Elevation Model, Geography Markup Language (.dem,.gml) Keyhole Markup Language (.kml or .kmz) GEOtiff (.tiff) BigTiff (.tiff) Enhanced Compression Wavelet (.ecw) JPEG2000 (.jp2) Scalable Vector Graphics (.svg), Log ASCII Standard (.las) Image Format (.img)

Information Asset Type	Valid Format
Images	Joint Photographic Experts Group (.jpg, .jpeg) Tagged image file format (.tif, .tiff) Digital Negative (.dng) Portable Document Format (.pdf)
Presentations	Microsoft PowerPoint (.ppt, .pptx) Portable Document Format (.pdf)
Spreadsheets	Microsoft Excel (.xls, .xlsx) Comma separated values (.csv) Tab-separate values (.tsv) Portable Document Format (.pdf)
Video	Moving Picture Experts Group MPEG2 (.mpg, .mpeg) Moving Picture Experts Group MPEG4 (.mp4, .m4v, .m4a, .f4v, .f4a)
Websites	HyperText Markup Language (.htm, .html) or Extensible Markup Language (.xml) together with supporting files (.css, .xsd, .dtd) Web ARChive format (.warc) Internet archive (.arc)

GDS 21 Source Information Asset Disposal Validation

GDS 21 does not authorise the destruction of hardcopy information assets, once digitised, in all circumstances. In some instances, there may be information assets in hardcopy format that cannot be destroyed.

Hardcopy Source Information Asset Retention

Hardcopy source information asset class	General Disposal Schedule (GDS) retention period	Retention of physical original information asset after digitisation and capture
Permanent retention information assets created prior to 1 January 2005.	Permanent	Permanent on hardcopy file (digitised rendition can be used for ease of reference).
Permanent retention information assets created on or after 1 January 2005.	Temporary	Destroy after digitised and verified rendition is saved into Content Manager or other approved business system.
Temporary retention information assets.	Temporary	Destroy after digitised and verified rendition is saved into Content Manager or other approved business system.
Temporary retention information assets that are over 50 years at; a) the time of scanning; or b) the point at which the source information asset retention period expires.	Refer to State Records for appropriate disposal authority	To be determined by appropriate disposal schedule.
Information assets not included in any current, approved disposal authorities.	Unknown	Retain until an appropriate disposal schedule is established and adopted by State Records Council.

Disposal Validation Approach

To ensure you do not destroy an information asset inappropriately, ask yourself these questions:

1. Is this hardcopy information asset not covered under a current general disposal schedule (GDS) or another disposal determination? NOTE: If the information asset is being saved into Content Manager it will be covered under an existing GDS.
2. Does the hardcopy information asset have a permanent retention value under an existing GDS and was created before 1 January 2005?
3. Does the hardcopy information asset possess evidence that cannot be reproduced in the digital version?
4. Does the hardcopy information asset have any intrinsic value (e.g., a document with a seal)?
5. Is the information asset restricted or excluded through the Electronic Communications Act 2000 and Regulations?

If you answered YES to any of the above questions, you will be required to retain the hardcopy information asset in addition to the digitised version.

Additionally, you should also ask:

1. Are these information assets part of, or likely to be part of, a legal discovery process?
2. Is there any possibility that destruction of the hardcopy information asset could be prejudicial to the interest of the University, its clients, or the State?

If you answered YES to either of the above questions, seek further advice from the Information Management team and/or Legal before destruction of the hardcopy information asset can occur.

Hardcopy Information Asset Digitisation and Disposal Verification Process

