State Records
Authority NSW

Digitisation Primer

August 2023

staterecords.nsw.gov.au

NSW GOVERNMENT

Contents

1. Defining outcomes

Defining the program's intended outcomes upfront will help inform your business case, project plans and decisions (including the selection or scope of records), technical specifications and other requirements for digitisation.

Outcomes may include:

- increased access, information sharing and re-use
- improved service delivery
- preservation of information assets
- reduced physical storage.

IMPORTANT: If you intend to destroy records after digitisation, you must ensure that your public office is adequately covered by an approved retention and disposal authority. Original or source records that are not covered by relevant disposal classes or are designated as 'retain in agency' are not permitted to be destroyed after digitisation. See **Managing original paper records** and **Original or source records that have been copied (GA45)** for more information.

2. Selecting records

To make best use of money and resources, records selection should be well considered and aligned with the program's outcomes.

Selection could involve identifying records relevant to a particular business process or that otherwise relate to the program's core outcomes.

As a general rule:

- avoid selecting time expired records that are eligible for destruction or can be disposed of under Normal Administrative Practice (NAP)
- be cautious about selecting records with shorter retention periods or records that are rarely requested for access (as there may be little return on investment)
- prioritise records which are high risk/high value, frequently accessed or have long-term accountability or retention requirements.

Sentence records or evaluate retention requirements to avoid unnecessary digitisation efforts. Known retention periods can also help inform your organisation's return on investment, business case, and requirements for digitisation.

Talk to us before commencing retrospective programs involving records required as State archives.

Is digitisation the right outcome for the records?

Consider the questions below:

- Do all records require digitisation?
- Could 'on demand' digitisation deliver the intended outcome at a lesser cost?
- If records are required as State archives and no longer needed for active business, can they be transferred to the State Archives Collection instead?

As digitisation can be costly and resource intensive, public offices should weigh up business options carefully and account for the long-term management needs of the digitised images.

Records assessment

The suitability of records for digitisation also depends on the organisation's level of intellectual control over

the records and the records' physical format and condition. A records assessment may be needed to determine quantities and establish if pre-digitisation work (or a change of scope) is required. Assess:

- **Quantity** – establish the quantity of records that can be digitised within the project's budget. Note: The quantity of non-standard formats may impact what can be achieved.
- **Listings** – if indexes are missing or unreliable, then you may need to undertake preliminary data entry or checking activities to verify holdings and confirm which records are in scope.
- **Sentencing** – where retention requirements are unclear, consider sentencing the records first to avoid unnecessary digitisation efforts. Disposal dates may be useful to narrow a project's scope.
- **Physical condition** – the physical state of the records will affect the time taken to prepare items for imaging, the equipment needed and any special handling requirements. In some cases, records may require 'stabilisation' or conservation treatment in order to be safely handled by persons or equipment and/or produce a satisfactory image/copy.
- **Physical format and size** – will impact the equipment needed to digitise the records, the ease of image capture and handling requirements.

Findings from this assessment will help inform decisions, refine scope, develop a criterion for records inclusion/exclusion, and build processes for 'exception' handling (i.e. management of records that may incur additional effort or cost to prepare or digitise). Determining factors should be documented in plans, business rules or procedures.

3. Determining requirements

Requirements for digitisation programs should be determined and documented before the program begins. They should also be reviewed and amended over time if/when requirements change.

Requirements can be documented in business rules, procedures, project plans, contractual agreements, etc. as appropriate.

3.1 Physical preparation of hardcopy records

Determine requirements for:

- document/item preparation
- handling
- triaging of records in poor or unstable condition.

Time taken to prepare records will depend on their format, structure and condition. Preparation may involve:

- removal of bindings, staples, clips, post-it notes, etc.
- unfolding or flattening
- cleaning, and in some cases, conservation treatment.

Note: *When undertaking conservation activities, staff should handle records in accordance with* ***specialist advice provided by the Museums of History NSW*** *and wear appropriate personal protective equipment (PPE) as needed.*

The level of preparation required will also depend on the capture equipment used.

Additional consideration must be given to how hardcopy records should be handled, particularly if records are fragile, over-sized or non-standard in format. Records may sometimes require use of specific equipment to protect them from damage, such as use of a flatbed scanner as opposed to a scanner with a document feeder.

If your program involves records that are in poor condition and likely require additional preparation or conservation treatment, then you may need to create a process that identifies and triages records. For example, establish which records:

- can be digitised in their current state
- require use of specific or specialised capture equipment
- require standard preparation
- require basic stabilisation work
- require assessment and/or conservation treatment by a professional conservator.

3.2 Technical specifications

It is important to determine technical specifications before commencing digitisation or purchasing equipment.

Setting technical specifications ensures quality, consistency, and guarantees that the digitised records are fit for purpose. The primary goal is to capture sufficient data so that resulting images or files are of sufficient quality for their purpose and can remain legible or useable for as long as required. Where records have long term or archival value, the resulting digitised files may need to withstand time and numerous migrations.

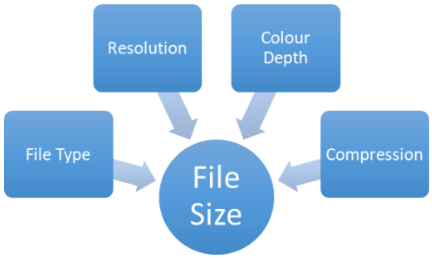
IMPORTANT: Adopting the technical specifications defined in **Digitisation specifications for paper records in public offices** will assist your organisation in meeting the copying provisions outlined in **Original or source records that have been copied (GA45)**.

Knowing your technical specifications upfront will help determine what equipment is needed and how much digital storage is required.

File size

The size and quality of a digital image is a product of its technical elements. Image resolution, colour properties, compression, output file format and embedded metadata all contribute to the quality and size of a digital image.

There is no magic bullet to reduce file size while also maintaining quality – it is a compromise. To reduce size, one or more of the technical elements will need to be decreased. Establishing technical specifications is about striking an acceptable balance.



If records have longer retention periods or are required as State archives, create a high-quality master file for preservation. Where files sizes need to be smaller for delivery purposes (e.g. via email), a lower quality version can be derived from the master file.

File formats

File formats must be suited for long-term sustainability and accessibility and be capable of meeting compression requirements. For a list of sustainable formats, consult our guidance on sustainable file formats.

To determine which file formats are most appropriate for purpose, consider:

- compression method (lossy, lossless, or uncompressed)
- ability to hold metadata
- ability for text/optical character recognition (OCR)
- page display (single vs multipage page)
- compatibility with software programs.

For digitisation programs, it is recommended to use:

- PDF/A (lossless) – for documents with long-term retention requirements or those required as State archives
- Standard PDF (high-quality lossy) – for modern/word processed records with short-term retention requirements
- TIFF or JPEG2000 (lossless) – for photographic and illustrative material.

Compression

Most common file formats use compression techniques to reduce the size of the resulting digital file for ease of storage and transmission. Depending on the file format, compression could occur by default or be (to various degrees) controlled through settings/preferences.

Lossy compression algorithms reduce file size by discarding less important information (such as slight tonal differences). Some lossy formats such as JPEG have threshold setting – the impact of the threshold can vary image to image as it is dependent on pixel content. Too much lossy compression will degrade image quality. The effects of lossy compression are irreversible and compact with every save.

Lossless compression algorithms reduce file size through simplifying the representation of the code. No data is irretrievably lost, and the decompressed object is of the same quality as the original file. Examples include LZW or ZIP lossless compression within TIFF files.

Note: *Smart compression techniques may employ additional reduction techniques such as resizing, lowering resolution, bit depth/colour reduction, etc. Any use should be after thorough testing.*

When is it okay to use lossy compression?

Lossy compression can be applied to short-term records and records digitised for reference purposes.

If the records have long-term retention requirements or are required as State archives, it is recommended to use lossless compression for master files and to create an access copy with lossy compression for everyday use.

Resolution

Resolution is a measure of the ability to capture detail (or samples) from the hardcopy record. It is most frequently quantified in pixels per inch (ppi), which describes pixel density over a given scale (ideally 1:1 – original size). The higher the ppi at the same scale/size, the better the resolution and the clearer the image.

The resolution of an image cannot be increased post-digitisation as editing software cannot take new samples from the source record, only artificially produce or re-distribute data. If a higher resolution is required, the record will need to be re-digitised. Hence, it is important to determine the correct resolution prior to undertaking any digitisation.

Consider:

- The nature of the records to be digitised. Photographs, drawings and cursive handwriting require greater resolution than typed documents.
- How the digital images will be used. Hardcopy records that are enlarged or require fine detail for viewing and printing should be digitised at a higher resolution. Hardcopy records that are reduced for viewing may be digitised at lower resolutions. If text/optical character recognition (OCR) will be applied, consider resolution requirements for the OCR software.

IMPORTANT: When selecting a capture resolution, consider the optical resolution of your capture device. Exceeding this resolution will result in interpolation (artificial data) which may degrade image quality.

Bit depth / Colour mode

Bit depth refers to the number of bits (zeros and ones) used to represent the smallest unit of information (such as a sound byte or image pixel). For digitised images, bit depth governs the range of colours or shades of grey (tonal range) represented in the resulting image.

Bit depth also determines the image's colour mode – i.e. if the resulting image will be bi-tonal, greyscale or full

colour.

Like resolution, if a higher bit depth is required, the record will need to be re-digitised.

Common colour modes and when to use them:

Bit depth	Colour mode	When to use it
1 bit	Bi-tonal (Pure black and white – no grey)	Black and white, clear high contrast word processed documents and art line only.
8 bit	Greyscale	Greyscale or black and white documents. Includes those that contain watermarks, grey shading and grey graphics.
24 bit	Colour	Documents with discrete colour used in text or diagrams and coloured documents.

Note: *8 bit colour (RGB) is not suitable for digitisation.*

Specialist equipment may also offer greater bit depths such as 16 bit Greyscale and 48 bit Colour. Larger depths provide greater tolerance for colour/tone correction during post-production editing. This is particularly necessary for preservation purposes and when reproducing images at scale.

Colour space and colour management

A colour space defines the range of colours within an image. Assigning a colour space will help software know how to render colour. Commonly used colour spaces include sRGB and AdobeRGB.

A colour space is just one component of colour management. Scanners, monitors and printers, etc. typically use different colour spectra. Colour management uses device profiling (ICC profiles) in aims of ensuring that an image looks the same across a range of different devices. Colour management is best left to experts due to its technical complexity.

Testing specifications

It is important to test your technical specifications and settings for suitability. Check that:

- the smallest detail is legibly captured (e.g. smallest type size for text, clarity of punctuation, marks including decimal points)
- details are complete (look for broken characters, missing segments of lines, etc.)
- image sharpness is sufficient compared to the source (if halos appear around the text/characters, it is likely due to over compression).

When testing specifications, view digitised images at scale (i.e. 100%).

For further information and for visual examples of the impact of technical properties on image quality, see the Getty Institute's *Introduction to Imaging*:
http://getty.edu/research/publications/electronic_publications/introimages/image.html

3.3 Equipment

Equipment for digitisation should not be selected (or procured) until you understand:

- the records you intend to digitise
- your technical specifications for image quality.

You can then assess the hardware and software to see if it can deliver the quality you require.

High level considerations for equipment selection may include:

Category	Considerations
Format	• Type and size of hardcopy material being digitised.

	• Quantity of material to be scanned (files, pages per file, quantities per batch, etc.). • If a flatbed or automatic document feeding scanner is more appropriate than format specific equipment (such as map scanners, overhead book scanners, or a camera).
Technical	• Bit depth and resolution required. • File formats the device can output to. • Image noise generated.
Page Handling	• Quantity of pages able to be loaded into the scanner at one time. • Ability to scan both sides of a document in one pass.
Output/Configuration	• Accessing the device (directly or through a software interface). • Outputting files from the device (e.g. to a network drive or via email).
Software	• Control over settings. • Ability to conduct image enhancements, if required. • Ability to crop and deskew (straighten). • Ability to automatically capture text or metadata. • Ability to apply OCR.
Operational	• Ease of use. • Time taken to scan under normal conditions. • Cost and available funds.

Numerous organisations use Multi-Functional Devices (MFDs), such as photocopiers, to operate as scanners for digitisation programs. In some cases, the functionality of an MFD will be sufficient for your digitisation needs. In others, it will not. If you are intending to use an existing MFD, you will need to check that it is capable of meeting your specifications. You will also need to determine appropriate device settings and business rules for using the MFD for digitisation purposes.

See **3.2 Technical specifications** for more information.

3.4 Digital storage

Digital storage must be considered before commencing the digitisation project. This will ensure that:

- ICT infrastructure and system architecture is able to support (and back-up) content generated from the program
- digitised images are stored using methods that guarantee their security and accessibility throughout the digitisation process.

Get your ICT team on board early to support your program's storage and security needs.

3.5 Metadata

Why collect metadata?

Metadata is data that provides information about other data. Whether descriptive or technical, metadata can help people understand digitised records.

In digitisation projects, metadata may be used to:

- find and use digital images

- link images to the business processes they document
- demonstrate that images are accurate and reliable renditions of the hardcopy records
- document the digitisation process
- document formats and dependencies to help manage images over time.

When planning your digitisation program, consider requirements for metadata as early as possible.

For more information, see **Metadata for records and information**.

File titling conventions

File titling rules should be determined to name digitised files consistently and to promote discovery.

If appropriate, file titling can follow the source records' existing file titling or numbering system. Alternatively, you can develop a standardised naming convention to identify the records, and/or employ unique identifiers consisting of alpha-numerical sequences (i.e. with a prefix and running number).

In general, file and image names should:

- be unique
- be consistently structured
- include the use of leading zeros to facilitate sorting in numerical order
- avoid special characters (e.g. tabs or symbols)
- avoid spaces (as they can cause problems across operating platforms).

File titling conventions should include provisions for each object level used (files and folders) and distinguish between master and deviate copies.

Be sure to document file titling conventions to inform future interpretation and use.

Descriptive metadata

Descriptive metadata is commonly managed alongside the digitised records, such as in business systems or EDRMS where metadata is connected to the digital files. Depending on the file format used, metadata can also be embedded into digitised files themselves for assurance.

You will need to determine which descriptive fields are relevant to your project.

Generally, this would include:

- title
- creator
- original file number
- original creation date
- document form (see advice on **Document form metadata scheme**), and
- custom fields

If the records are required as State archives, it is recommended to include the required fields for transfer. This will save effort later. See Museum of History NSW's **Transfer Fact Sheet 6 Metadata**.

Where will descriptive metadata come from?

Often descriptive metadata comes from existing systems or lists, but in other cases it may be necessary to data enter or extract descriptive metadata from the records themselves.

If hardcopy indexes and/or registers are required to make these records accessible, you will need to consider how these can be incorporated into the program.

Searchable text / Optical Character Recognition (OCR)

If your program involves word-processed documents, determine if **optical character recognition (OCR)** is a requirement of your program. This may be subject to the type and clarity of the source documents.

OCR may enable the automation of some metadata capture, particularly if your documents follow a standardised structure.

Technical metadata

Technical metadata is a by-product of the digitisation process.

Data such as:

- date and time of image creation
- imaging capture device
- software used
- operator/creator
- technical properties (file format, resolution, bit-depth, colour space, and compression)
- number of revisions or edits, etc.

can be automatically recorded during digitisation, however, equipment and software may need adjustment to ensure technical data is collected and embedded into digitised files.

Technical metadata is useful in documenting the digitisation process and can also help inform quality assurance activities.

3.6 Quality

Quality assurance and quality control measures can help ensure that:

- source records are handled correctly and are undamaged during digitisation processes
- source records are lawfully destroyed (when appropriate)
- digital images are validated against technical specifications
- digital images are accurate, complete and legible and of sufficient quality to be 'fit for purpose' with their essential characteristics preserved
- metadata to control and manage digital images is adequate
- your organisation can prove that robust and trusted systems and processes were used to produce digital images if their authenticity as evidence is ever questioned.

Quality Assurance

Quality assurance measures often include:

- setting, implementing and monitoring digitisation policy and procedures
- setting, implementing and monitoring digitisation processes
- acquiring equipment that is fit-for-purpose
- implementing a testing process and the management of quality fails, etc.
- identification and assessment of quality criteria for benchmarking.

The degree and nature of quality assurance should be defined early in a project. Quality assurance must be carried out periodically during a project in accordance with defined procedures.

Quality Control

Quality control involves checking the scanned image during or after digitisation to identify and correct any

defects or errors.

Examples of what quality control should check for, include:

Category	Criteria
Authenticity	• Are the scanned documents an accurate representation of the original/source records? • Has any content been obscured, cut off or cropped? • Do the colours accurately compare to the original? • Are pages presented neatly, consistently and in the correct orientation?
Completeness	• Is information missing? • Are all pages, annotations, attachments and enclosures captured? • Are pages represented in the correct sequence? • Have blank pages been removed?
Accessibility	• Is all information legible? • Are there any marks, lines or blotches on the image caused by the scanning process that affect legibility? • Have technical specifications been met? • Have all metadata fields been captured?

4. Managing the digitised files

Requirements for managing digitised images relate back to the outcomes originally defined for the project. They will need to be stored, backed up and managed effectively for as long as they are required.

Managing digitised files may involve:

- inclusion of digitised records into the organisation's information asset register
- ingestion or upload into a recordkeeping system
- version control, derivatives and redacted versions, etc.

For business as usual (BAU) scanning of incoming records, business rules should ensure that ingestion into your EDRMS or other recordkeeping systems result in official records being aligned to your Business Classification Scheme or retention rules.

Where source records are destroyed after digitisation, the digitised records become the official record of business and must be kept in accordance with the *State Records Act 1998* and authorised retention requirements. It is imperative to store masters in a manner that ensures their security and longevity.

If the records are required as State archives, requirements and schedules for the transfer of digitised records (as well as the physical records) should be discussed and pre-determined with the Museums of History NSW.

5. Managing source records post digitisation

How source records should be managed, and if they will be retained or disposed of, will depend on the records themselves and the intended outcomes of the digitisation program.

If source records will be retained, consider:

- re-collation requirements (i.e. if documents will be put back together, if tube clips, plastic pockets,

staples, pins, etc. will be returned)

- rehousing options, and
- transfer requirements.

If the original/source records have long-term accountability requirements or are required as State archives, post-digitisation may be the perfect time to rehouse the records into archival standard containers.

If source records will be destroyed, consider:

- when they can be destroyed (i.e. at what point quality assurance is considered complete)
- how destruction will be documented, and
- how internal authorisation to destroy the records will be granted.

For the scanning of incoming records, documentation and authorisation should be built into business rules and procedures. For retrospective programs, this may involve seeking authorisation at a batch level following quality assurance.

For source records to be legally destroyed, the conditions of **Original or source records that have been copied (GA45)** must be met.

6. Developing business rules and procedures

Requirements for all aspects of your digitisation program should be outlined in documented business rules, procedures, project plans or contractual agreements as appropriate. Requirements include:

- technical specifications
- managing digitised records (e.g. file titling conventions and where to save the records)
- what to do with the source records.

Business rules and procedures ensure that:

- the digitisation program is robust
- the digitisation program or process is standardised and will consistently produce an acceptable quality of digitised records
- records, both hardcopy and digitised, will be managed appropriately and securely
- digitised records will be actioned appropriately and in a timely manner.

Note: *An adequate level of security for records should always be maintained. This may include measures such as controlled access for authorised personnel, audit trails, and returning boxes to secure storage immediately after digitisation.*

6.1 Managing outsourced arrangements

Public offices are responsible for meeting the requirements of the *State Records Act 1998* and the standards released under the Act. If your organisation outsources the digitisation of records, it remains responsible for the management of both the source records and the digital images throughout the entire process. Therefore, you must ensure that all relevant requirements are specified in contracts with service providers.

Ensure that the following points are clearly articulated in service agreements:

- the range and type of records to be digitised
- timeframes and costs
- expectation that records are not to be altered or omitted in any way
- security and handling requirements for sensitive or personal records or for urgent requests

- roles and responsibilities of the organisation and the service provider(s)
- benchmarks for technical requirements and metadata
- quality assurance measures (including remediation required if benchmarks are not met)
- an agreed monitoring/reporting framework
- rules surrounding the return of records to the organisation
- a statement that all State records and State archives must remain in NSW unless express permission is given by State Records NSW to transfer them out of the state. See **Transferring records out of NSW for storage with and maintenance by service providers based outside of the State (GA35)**.

Service providers must be made aware of relevant standards, including:

- ***Standard on the physical storage of State records*** – which contains requirements for the protection and handling of hardcopy records, appropriate storage and environmental conditions, and the documentation of digitisation processes.
- ***Standard on records management*** – which contains requirements for metadata and the protection of digital records from unauthorised access.
- ***Original or source records that have been copied (GA45)*** – which specifies the conditions for destroying original or source records after digitisation.

The monitoring framework that forms part of the contract must ensure that all recordkeeping requirements are met throughout the term of the agreement.

If you are intending to digitise records required as State archives, you should first consult State Records NSW and Museums of History NSW to ensure that issues concerning the reproduction of archival records are discussed and that suitable parameters for the digitisation process are covered and agreed upon in the contract.

7. Initiating the digitisation program

Commence digitisation activities in accordance with policies, plans, business rules and procedures.

8. Monitoring and review

As your digitisation program progresses, monitor and review activities for effectiveness. Adjust plans, business rules and procedures if and as needed.

When a program is in its infancy, it is good practise to put safeguards in place. Implementing a day boxing procedure for BAU scanning or a more robust quality assurance process for outsourced work are common examples of applied safeguards.

State Records Authority NSW

The Mint
10 Macquarie Street
Sydney NSW 2000

Office Hours:
Monday to Friday
8.00am — 5.00pm

T: (02) 9714 3080
E: govrec@staterecords.nsw.gov.au
W: staterecords.nsw.gov.au

NSW GOVERNMENT