



Government of **Western Australia**
Department of **Culture and the Arts**
State Records Office of Western Australia



Digitization Specification

**A Recordkeeping Guideline for
State Organizations**

**State Records Office of WA
Perth, Western Australia**

June 2016

(updated May 2018)

TABLE OF CONTENTS

DEFINITIONS	3
PURPOSE.....	5
BACKGROUND	5
SCOPE.....	5
DIGITIZATION PROCESSES	6
TEXT DOCUMENTS.....	9
PHOTOGRAPHS	10
MAPS AND PLANS	11
AUDIO RECORDINGS	12
VIDEO AND FILM RECORDINGS.....	13

DEFINITIONS

Refer also to the Glossary of Terms produced by the State Records Office of Western Australia.

Bit – basic unit of computing used to measure information storage capacity. A binary digit having a value of 1 or 0 is used.

BMP – Bitmap file format, usually uncompressed.

Compression – methods of encoding data so that fewer bits are required for storage.

DPI – Dots Per Inch. A measurement often used interchangeably with PPI but refers specifically to measurement of the resolution for computer printers.

GIF – Graphics Interchange Format. Uses a patented lossless compression format and is being replaced by PNG.

JPEG – commonly used lossy compression format for photographs and images.

JPEG2000 – International standard for lossless compression of images.

Long Term File format means a digital file format that is likely to be accessible for greater than two generations of technology.

Lossless compression – computing algorithms that allow image data to be compressed for storage, while at the same time ensuring that the exact original data can be reconstructed for use.

Lossy compression – provides for approximate reconstruction of compressed image data.

MPEG - an international standard for encoding and compressing video images.

Normalization means a form of migration to open standards based formats as part of a digital preservation system (NAA. 2009 <http://www.naa.gov.au/about-us/organisation/accountability/operations-and-preservation/digital-preservation-policy.aspx>)

Open standards “are standards made available to the general public and are developed (or approved) and maintained via a collaborative and consensus driven process. "Open Standards" facilitate interoperability and data exchange among different products or services and are intended for widespread adoption.” (ITU, 2005)

PDF/A – Portable Document Format developed by Adobe Systems Inc, and is suited for archival storage of electronic information. It is defined by ISO 19005 - 1:2005. PDF/A – 1b conforms to the minimum requirements of the standard, and is

the format created by digital scanning. Records that are converted, normalised or migrated to PDF, should meet the requirements of either PDF/A -1b, or the higher PDF/A – 1A.

PNG – Portable Network Graphic. A lossless compression format defined by ISO/IEC 15948:2004.

PPI - Pixels Per Inch. A measurement of resolution for computer display.

RGB – Red, Green, Blue colour model used by most visual display technology, scanners and cameras.

Scan format – describes the most commonly available scanning formats available on digital copying devices.

Scan resolution – describes the number of dots per inch and bit resolution for visual scans and sampling rate and bit resolution for audio scans.

Scan type – describes whether the digitized version should be in colour or black and white, and the management system to be used in assessing the digitized version on a digital device.

SVG – Scalable Vector Graphics. A W3C endorsed language using XML to define two dimensional images and graphics.

TIFF – Tagged Image File Format. It includes header data and image data, which may or may not include lossy compression formats. TIFF single images are generally widely supported, but multi-page TIFF documents are not.

W3C – World Wide Web Consortium.

XML – Extensible Markup Language.

PURPOSE

The purpose of this Specification is to assist State organizations in creating high quality digital reproductions of records, and in choosing appropriate file formats for them.

This Specification should be read in conjunction with the General Disposal Authority for Source Records, *SRC Standard 7 - State Archives Retained by Government Organizations* and *SRC Standard 8 - Managing Digital Information*.

BACKGROUND

In addition to keeping born digital records, government organizations are digitizing records created or received in analogue format in order to provide access to them over time. Most digital file formats and media are fragile and will need to be migrated into alternative formats to ensure the ongoing preservation and usability of the content.

Where the digitized version of a record is intended to stand in place of a source record, the authenticity, reliability, integrity and usability of the digitized record must be ensured, regardless of the required retention period. Scanning at higher resolutions will allow organizations to create derivative access copies for online use, while also ensuring that the digital file is useable for good quality printing and optical character recognition.

This Specification details technical requirements for digitization and long term file formats for digital records.

SCOPE

This specification applies to all State organizations and is intended to be of use to records management, digitization and information technology professionals.

DIGITIZATION PROCESSES

State organizations undertaking digitization must determine:

- what records are to be digitized;
- how the digitized versions are to be created, used, stored, maintained and manipulated;
- whether the source records are to be retained; and
- the costs and risks associated with all stages of the process.

State organizations must have approved policies and procedures for digitization including:

- **Preparation of records**
 - Records are checked to ensure they are physically suitable for digitization. For example, paper records are checked to see if they are free of tears or creases, that staples and pins are removed without damage to the document, and attachments are identified.
 - Paper based documents such as maps and plans are scanned at original size. Where this is not possible, a resized scan may be made for access purposes only, and the source record must be retained.
 - Audio visual records are checked for damage to the carrier and damage to the media.
 - Audio visual records such as cassettes or video tapes are digitized fully, including both sides of a tape where appropriate.
 - Records that are too large or too fragile to be scanned are identified and removed from the scanning process.
- **Creation of metadata**
 - Appropriate metadata is created for each record and captured (with the document) in a recordkeeping or business information system.
- **Calibration and choice of equipment**
 - The choice of equipment for scanning (digitizing) will depend on the sort of material that is being digitized, and the length of time it is being retained and the purpose for which it is required.
 - Equipment is calibrated in accordance with relevant international standards and calibration checks are undertaken at regular intervals, in accordance with the manufacturer's specifications.

- A master calibration target is maintained for the life of the equipment, and calibration checks are measured against this target.
- Calibration targets include colour and density checks, as well as fine lines for detail. Such targets may be included with equipment purchased for scanning, purchased separately, or may simply consist of a printer test page.
- Settings on equipment are correct and will produce files that meet the minimum technical specifications in this document.
- **Use of scanning targets**
 - A colour or greyscale test target, and measurement scale (such as a measuring tape or rule) along the horizontal or vertical axis or both, is made for each type of record scanned (eg: A1 maps vs A4 letters).
 - These targets are not calibration targets, and are intended to ensure that material may be colour and tonally corrected on a wide range of viewing environments, and accurately reproduced as to size. This is particularly important for engineering drawings, maps and plans.
- **Verification of the digitized record**
 - The digitized record is checked, preferably by a second staff member, to ensure that the source record has been completely digitized, that is, all pages of a document have been scanned (including both front and back if there are markings on both sides), or a cassette tape is recorded on both a and b sides.
 - Source records are re-digitized when necessary.
- **Retention periods**
 - The minimum required retention period of the records is identified.
 - Where the minimum retention period of the records is longer than ten (10) years the digitized version is converted to or saved in an approved long-term file format.
 - Where the minimum retention period of the records is less than ten (10) years the digitized version may be saved in any file format which the agency is able to access as this digitized version should not achieve technological obsolescence in that time. Organizations must monitor that their digital files remain accessible for the full period they require them.

For further information on appropriate procedures for scanning records see AS/NZS ISO 15801:2014 *Electronic imaging - information stored electronically – recommendations for trustworthiness and reliability.*

TEXT DOCUMENTS

Digitization of text documents must be done to a level that ensures not only that the text remains legible, but that any additional markings on the document are also reproduced and retain their meaning.

Technical specifications:

Size	Up to A0 - scan to original size
PPI	Minimum 300 ppi
Colours	Minimum 8 bit colour
Black & White	Minimum 1 bit
Greyscale	Minimum 8 bit greyscale
Long term file formats	PNG, JPEG, JPEG2000, PDF/A with embedded text or PNG, JPEG or JPEG2000 (for multipage documents)

Specifications for reproductions of receipts/proof of purchase records

Organizations may create reproductions of receipts or other proof of purchase records using devices such as a mobile phone camera, where the reproduction is transferred into a financial management system.

If the device used to create the reproduction of the receipt is unable to meet the calibration and scanning processes and/or specifications outlined above, or if the reproduction specifications cannot be determined, the reproduction **must** be of sufficient quality to allow the accuracy and integrity of the reproduction to be verified.

The officer making the reproduction must ensure that all essential information in the receipt has been captured.

NB: Where a scanning target is not included with the original document, a target check must be performed on the scanner whenever it is used or at least monthly, and saved to an EDRMS or business information system with the date of target check.

PHOTOGRAPHS

Digitization of photographs must be done to a level that ensures that all elements in the photograph are able to be clearly discerned when shown on screen at 1:1 scale.

Technical specifications:

Size	Up to A0 - scan to original size
PPI	Minimum 600 ppi
Colours	Minimum 16 bit colour
Black & White	Minimum 16 bit greyscale
Long term file formats	PNG, JPEG, JPEG2000, PDF/A with embedded text or PNG, JPEG or JPEG2000 (for multipage documents)

NB: Where a scanning target is not included with the original document, a target check must be performed on the scanner whenever it is used or at least monthly, and saved to an EDRMS or business information system with the date of target check.

MAPS AND PLANS

Digitization of maps and plans must be done to a level that ensures all elements in the map or plan are clearly discernable when shown on screen at 1:1 scale, and that all types of crosshatching or other markings are distinguishable from one another.

Technical specifications:

Size	Up to A0 - scan to original size
PPI	Minimum 300 ppi Maps and plans with fine lines and detail will require high resolution
Colours	Minimum 16 bit colour
Black & White	Minimum 8 bit greyscale
Long term file formats	PNG, JPEG, JPEG2000, SVG, PDF/A with embedded text or PNG, JPEG or JPEG2000 (for multipage documents)

NB: Where a scanning target is not included with the original document, a target check must be performed on the scanner whenever it is used and saved to an EDRMS or business information system with the date of target check.

AUDIO RECORDINGS

Digitization of audio recordings must be done to a level that ensures that the entire recording is clear, and all parts of the medium which was recorded onto are digitized so that there is no question of data being lost (eg: entire length of tape from both sides must be digitized).

Technical specifications:

Speed	Audio must be digitized at original speed
Compression	Lossless only
Channels	Minimum Stereo, or as appropriate if recording contains more than 2 channels
Long term file formats	BWAV, WAV, AIF

VIDEO AND FILM RECORDINGS

Digitization of motion picture recordings must be done to a level that ensures that the entire recording (both audio and video) is clear, and all parts of the medium which was recorded onto are digitized so that there is no question of data being lost (eg: entire length of reel or video must be digitized).

Technical specifications:

Speed	Video and film must be digitized at original speed
Compression	Lossless only
Channels	As per original (eg: 1280x720px)
Long term file formats	MPEG-2, MPEG-4
Audio	AAC

For further information please contact:

State Records Office of WA

ph: 9427 3600

email: sro@sro.wa.gov.au