Digital Preservation of Multimedia Materials

LANT Technical Guidelines

NORTHERN TERRITORY GOVERNMENT

Document title	Digital Preservation of Multimedia Materials
Contact details	Library & Archives NT – Assistant Director Collections and Content
Approved by	Senior Director
Date approved	8 June 2023
Document review	Two to three years
TRM number	N/A

Version	Date	Author	Changes made
1.0	4 May 2023	Lilly Ho	Creation
1.1	7 July 2023	Lilly Ho	Proofreading, checking on textual format

Acronyms	Full form
AV	Audio-visual
AWS	Amazon Web Services
FADGI	Federal Agencies Digital Guidelines Initiative
LANT	Library & Archives NT
LC	Library of Congress
OAI-PMH	Open Archives Initiative's Protocol for Metadata Harvesting
OH	Oral History
PREMIS	Preservation Metadata Implementation Strategies
RFS	Recommended Format Statement
TS	Territory Stories

Contents

1. Background

The purpose of this document is to ensure that multimedia archival materials (images and audio-visual materials) within Library & Archive NT (LANT) holdings are created and preserved to a consistent standard. These guidelines are also to be used as the minimum technical requirements of digital materials within the LANT repository Territory Stories (https://territorystories.nt.gov.au/).

This document is written under the assumption that only popular material types and file formats are collected. The standards for Web Archiving and 3-D digital objects are not included in this document but will be developed in the next version. Copies for preservation and access purposes are also discussed. It is recommended that this document be reviewed every two to three years in order to ensure the currency of ISO/IEC standards[5]. In light of rapidly changing digital preservation technologies, it is also suggested that the minimum technical requirements of archival materials are reviewed on a regular basis. If necessary these requirements should also be increased in order to maintain compatibility with modern application software. This process is particularly important regarding access copies.

2. About Territory Stories (DSpace)

Territory Stories (TS) is built on a DSpace services which support the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH) as a data provider. OAI support was implemented using OCLC's OAICat open-source software to make DSpace item records available for harvesting.

DSpace, an OAIS-compliant digital repository[15], can apply filters or transformations to files (bitstreams), creating new content. Filters are included that extract text for full-text searching, and create thumbnails for items that contain images. The media filters are controlled by the dspace filter-media script which traverses the asset store, invoking all configured MediaFilter[9] or FormatFilter classes on files (bitstreams).

DSpace supports all kinds of digital file formats[2, 10]. Here are some examples:

Material Type	File Formats	Extensions
Audio (Sound)	• Basic audio • Audio Interchange File Format • MPEG Audio • RealAudio file • Broadcast Wave Format	• au, snd • aif, aifc, aiff • abs, mpa, mpega • ra, ram • wav
Still Image	• Graphics Interchange Format • Joint Photographic Experts Group/JPEG File Interchange Format (JFIF) • Portable Network Graphics • Tag Image File Format • Microsoft Windows bitmap • Kodak Photo CD image	• gif • jpeg, jpg • png • tif, tiff • bmp • pcd
Textual	• Cascading Style Sheets • Hypertext Markup Language • Plain Text • Rich Text Format	• css • htm, html • asc, txt • rtf

	• Extensible Markup Language • Portable Document Format	• xml • pdf
Video (Moving Image)	• Audio Video Interleaved • Moving Picture Experts Group • Material Exchange Format • Video Quicktime	• avi • mpe, mpeg, mpg • mxf • mov, qt

Territory Stories applies the MediaFilter to transform the file formats of 'still' images from TIFF to JPEG. The current version of Territory Stories is DSpace 6.4. S3BitStore approach (Amazon Web Services, AWS S3) is chosen for storing the digital contents.

3. Digital File Formats

There is no single perfect format for the preservation and future use of multimedia material. Decisions made on file formats should be dependent on the features and functionality to be preserved and the future use cases to be supported.

3.1. Sustainable File Formats

The Library of Congress (LC) recommends the important features of the digital content format descriptions and defines the sustainability of digital formats[7]. The tables below provide a selective list of preferred and acceptable formats that are suitable for preservation and access. The most suitable format for preserving the important features and functionality of a file may be the original format that it was created in. It is recommended that careful research and analysis is carried out before migrating files to a new format.

Categories	Extensions	Format Properties	LC Preference and Experience
Audio (Sound)	MP3	Format Category: encoding Other facets: unitary, binary, symbolic	MP3_ENC in MP3_FF, used extensively as a service format for American Memory. Used as the **accepted format** for electronic registration of sound recordings by the U.S. Copyright Office in the CORDS online registration support system (late 1990s, early 2000s), and likely to be used in successor copyright-related systems. LC's general preference for preservation-oriented recorded sound is WAVE_LCPM. For compressed sound, **MP3 is acceptable, especially at data rates of 128 Kb/s (mono) or 256 Kb/s (stereo) or higher.**

Audio	WAVE	Format Category: file-format Other facets: container-wrapper, binary, structured	As of March 2021, LC has over 150 TBs of files with the .wav extension in digital storage, although the version of WAVE is not specified. For the LC's archival master format for reformatted mono and stereo analog sound recordings, see WAVE_BWF_1. LC Recommended Formats Statement (RFS)[6] lists **WAVE as a Preferred format for Audio-Media-independent (digital)**. The RFS does not specify a version of WAVE. See also WAVE_BWF_1.
Still Image	JPEG	Format Category: encoding Other facets: unitary, binary, sampled	LC activities have created and archived extensive numbers of JPEG images. These are generally baseline sequential (some may be progressive) and are generally reduced-data derivatives of uncompressed master images. LC's general preference for still image "masters" has been for **uncompressed bitstreams**, and rich metadata is always welcome. Thus **TIFF_UNC_EXIF and TIFF_UNC have been the preference**. In 2011-12, however, LC began actively exploring the use of JPEG 2000 compression for master images; see JP2_FF. Meanwhile, for images only available in lossy compressed form, **JPEG is acceptable, especially if wrapped as JPEG_EXIF (rich metadata).** LC RFS includes **JPEG as both a preferred and accepted format for accompanying image/text files for audio.** The RFS does not specify a version of JPEG.
Still Image	TIFF	Format Category: file-format Other facets: container-wrapper, binary, unstructured, sampled	LC leads the Federal Agencies Digital Guidelines Initiative (FADGI)[3] Still Image Working Group and follows its recommendations for Technical Guidelines for Digitizing Cultural Heritage

			Materials which include the use of **TIFF as a format for primary or master files in many content categories for the LC's digitization workflows**.
Textual	PDF	Format Category: file-format, encoding, family Other facets: unitary, binary, symbolic	LC uses it as service format, including for some scanned historical materials, primarily to support convenient **downloading and printing**.
Textual	PDF/A	Format Category: family, file-format, encoding Other facets: unitary, binary, structured, symbolic	**PDF/A (Archivable PDF) for long term preservation**. LC RFS includes PDF/A as a preferred format for textual works in digital form, electronic serials, and digital musical scores as well as an acceptable format for 2D and 3D Computer Aided Design vector images. **Documents compliant with PDF/UA are preferred when available**.
Textual	PDF/X	Format Category: file-format, encoding Other facets: unitary, binary, structured, sampled	LC RFS includes **PDF/X (Press-ready PDF) as a preferred format for textual works in digital form, electronic serials, and digital musical scores**. The RFS also includes PDF/X as an **acceptable format for other graphic images - digital and 2D and 3D Computer Aided Design vector images**.
Video (Moving Image)	MPEG4	Format Category: file-format Other facets: container-wrapper, binary	LC RFS lists **MPEG-4 as an acceptable viewing proxy format for Video-File-Based and Physical Media**.
Video (Moving Image)	MXF	Format Category: file-format Other facets: container-wrapper, binary	LC RFS lists **MXF as a preferred format for Video-File-Based and Physical Media.** The RFS does not specify a version or profile or MXF. For **preservation reformatting**, the LC's Packard Campus for Audio-Visual Conservation has chosen **losslessy-compressed JPEG 2000 encoded video wrapped in MXF**. For web access: MPEG-4_AVC (profile unknown).

Video (Moving Image)	MOV	Format Category: file-format Other facets: container-wrapper, binary, sampled	LC RFS lists **ProRes 4444 (XQ), 4444 or 422 HQ in a QuickTime wrapper as a preferred format for Video-File-Based and Physical Media.** Regarding compressed video, bitstreams in MPEG-2 or -4 formats. Regarding uncompressed or losslessly compressed video, MXF_OP1a_JP2_LL is preferred by specialists at the LC's Packard Campus for Audio-Visual Conservation.

The Library of Congress' preferences of digital file formats are:

Categories	Content Details	Preferred*	Acceptable*
Audio (Sound)	Media-independent - Digital	1. Final production /release version of content rather than pre-production version 2. Highest native resolution PCM WAVE file of final version produced (44.1 kHz / 16 bit or higher) in addition to Compact Disc (CD) when both are produced 3. WAVE file with embedded metadata (**Broadcast WAVE**) rather than without embedded metadata (LC will specify fields)	1. Uncompressed file of final release version 2. Highest resolution compressed version in a major standard compression scheme 3. Lossless compression scheme rather than lossy compression scheme
Still Image	• Photographs - Digital • Other Graphic Images - Digital	**1. TIFF (*.tif)** 2. JPEG2000 (*.jp2) 3. PNG (*.png) **4. JPEG/JFIF (*.jpg)**	1. Photoshop (*.psd) 2. JPEG2000 Part 2 (*.jpf, *.jpx) 3. Digital Negative DNG (*.dng) 4. Proprietary Camera Raw formats (*.nef, *.crw) 5. GIF (*.gif)
Textual	• Digital • Electronic Serials	1. XML-based markup formats (EPUB3, BITS, TEI, DocBook) 2. Page-layout formats: **PDF/A** (ISO 19005-compliant), PDF (highest quality available)	1. Other structured or markup formats (XHTML, HTML, DOCX) 2. Page-layout format (PDF web-optimized) 3. Other formats (RTF, plain text)

Video (Moving Image)	Video - File-based	Final production version with the original production resolution and frame rate (i.e. 1080p24; 720p60, etc.) and file-based format that was delivered to the content distributor. 1. Interoperable Master Format (IMF) consisting of • Essence files as **MXF** tracks including video, audio, data and dynamic metadata essences • Composition playlist • Packaging data XML files (asset map, packing list, volume index) 2. ProRes • **QuickTime (.mov)** container • 4444 (XQ), 4444 or 422 HQ codecs 3. MPEG-2 • Compliant with ISO/IEC 13818 4. XDCAM • MXF container • HD422, SHD422, HD codecs	FFV1 (version 3) in Matroska (.mkv) container only for content without closed captions and/or timecode information. Viewing proxy such as 1. Recordable DVD 2. Recordable Blu-ray disc **3. MPEG-4 (.mp4)**

** In the order of preference*

3.2. Long-term Management of Files and Metadata

The three main digital material types that LANT collects are still images, moving images and audio recordings. The original formats of these materials are most commonly photographic negatives, photographic transparencies, movie film tapes, audio cassettes, Video Home System (VHS) tape cassettes, U-matic tapes, compact discs (CD) and Digital Video Discs (DVD).

For long-term storage, Federal Agencies Digital Guidelines Initiative (FADGI) recommends storage and management of digital image files and associated metadata in an ISO16363-compliant digital repository and trustworthy digital repositories such as DSpace.

3.3. Migrating files to a new format

If file migration is considered, it is generally easier to migrate or normalize digital objects before depositing them into the repository. Ideally, copies of the digital object in both its original and migrated formats should be preserved if resources allow.

Territory Stories DSpace automatically processes Still Images from TIFF into JPEG format on harvesting. Common examples of file migration that are applied at LANT are:

Categories	Original Format	Migrated Format
Audio (Sound)	MPEG1 and MPEG2 Layer 2 Audio[11]	WAVE (preservation), MP3 (access)
Still Image	HEIC[4] (likely acquired in archiving websites)	TIFF (preservation), JPEG (access)
Moving Image (video)	AVI[1], MKV[8]	MXF (preservation), MPEG4 (access)

4. Preservation and Access Copies

Preservation copies are the digital copies of analog or digital resources that are used for digital preservation purposes. They are normally of the highest quality and resolution that was possible at the time that the original materials are preserved. This is in contrast to access copies which are available for use by researchers.

Access copies are the digital copies of analog or digital resources that provide access while the original materials are being kept secure. This is in contrast to preservation copies which are used for digital preservation purposes.

4.1. Minimum Requirements

It is difficult to be specific about access copies as it depends on the quality of the original files and the place you are going to make the copy available. The following provide a guideline, but there are decisions that the person producing the access copy may need to make. The converter LANT uses has built-in presets that help produce the most appropriate files. LANT also considers the file size when producing access copies and this may reduce quality because the resulting file size is impractical for an access copy.

The preservation and access copies specifications that LANT recommends for compatibility with Territory Stories are:

Content	Preservation	Access
Category	**Specification / Resolution / Quality**	**Specification / Resolution / Quality**
Audio (Sound)	• WAVE • 48kHz (sample rate), 24 bit (bit size), 2304kbps (bitrate) or higher • 2 (Stereo)	• MP3 • At least 128 kbps (bitrate), 16 bit (bit size), 48kHz (sample rate) • Notes: 98kps is acceptable for non-music sound track
Still Image	• TIFF • 6000 pixels on the longest side or higher • 600 x 600 pixel per inch (ppi) resolution or higher	• JPEG • At least 1640 pixels on the longest side • At least 300 x 300 ppi resolution
Textual	• PDF/A • B/W resolution 600 dots per inch (dpi), or higher • Colour resolution 1200 dots per inch, or higher	• PDF • At least B/W resolution 300 dpi • At least Colour resolution 600 dpi
Video (Moving Image)	• Broadcast files (production version) • MFX OP-1a with a 10 bit – VBR Bit depth/rate • Video Codec shall be MJPEG2000 and the audio stream shall be BWF 48 kHz, 16 bit PCM encoded • RF 22-28, bitrate 4500 for 2160p 4K Ultra High Definition4 or higher	• MPEG4 container • Web optimised – indexes at the beginning (critical) • H.264 Coded, Framerate same as source • At least RF 20-24, bitrate 3000kbps for 1080p Full High Definition3

For all of the above, files must contain no measures such as digital rights management technologies or encryption that control access to or prevent use of the digital work.

5. Other Copies

LANT uses a network drive as a short-term procedure ensuring bit-stream preservation (preserving the ones and zeros that make up a digital file), but this does not address things like the availability of software to access the file, obsolete file formats, questions of rights, and issues of authenticity and provenance.

A policy is recommended to include the digital preservation metadata that are required to describe, manage and preserve digital resources over time. This will ensure long-term management of and access to electronic files. The PREMIS[12, 13] (PREservation Metadata: Implementation Strategies) data model takes into account the content that is recorded in standard metadata and provides for the creation of additional metadata specific to the needs of digital preservation.

6. References

[1] Audio Video Interleaved File Format, https://www.loc.gov/preservation/digital/formats/fdd/fdd000059.shtml#local

[2] DSpace Supported Formats, https://dspacecris.eurocris.org/help/formats.jsp

[3] Federal Agencies Digital Guidelines Initiative, https://www.digitizationguidelines.gov/

[4] High Efficiency Image File Format, https://www.loc.gov/preservation/digital/formats/fdd/fdd000526.shtml#local

[5] ISO Standards, https://www.iso.org/standards.html

[6] Library of Congress Recommended Formats Statement, https://www.loc.gov/preservation/resources/rfs/format-pref-summary.html

[7] Library of Congress Sustainability of Digital Formats, https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml

[8] Matroska Multimedia Container, https://www.loc.gov/preservation/digital/formats/fdd/fdd000342.shtml

[9] Mediafilters for Transforming DSpace Content, https://wiki.lyrasis.org/display/DSDOC7x/Mediafilters+for+Transforming+DSpace+Content

[10] Metadata and Bitstream Format Registries, https://wiki.lyrasis.org/display/DSDOC7x/Metadata+and+Bitstream+Format+Registries

[11] MPEG-1 and MPEG-2 Layer II Audio Encoding, https://www.loc.gov/preservation/digital/formats/fdd/fdd000338.shtml#local

[12] PREMIS - Digital preservation, https://ufs.libguides.com/c.php?g=1113411&p=8118665#:~:text=Preservation%20Metadata%20Implementation%20Strategies%20(PREMIS,by%20the%20Library%20of%20Congress

[13] PREMIS: Preservation Metadata Maintenance Activity, https://www.loc.gov/standards/premis/

[14] Technical Guidelines for Digitizing Cultural Heritage Materials, https://www.digitizationguidelines.gov/guidelines/digitize-technical.html

[15] Trustworthy Repositories Audit, https://en.wikipedia.org/wiki/Trustworthy_Repositories_Audit_%26_Certification

7. Appendix – LANT Technical Guidelines

The preservation and access copies specifications that LANT recommends for compatibility with Territory Stories are:

Content	Preservation	Access	Storage
Category	**Specification / Resolution / Quality**	**Specification / Resolution / Quality**	**P / A / T***
Audio (Sound)	• WAVE • 48kHz (sample rate), 24 bit (bit size), 2304kbps (bitrate) or higher • 2 (Stereo)	• MP3 • At least 128 kbps (bitrate), 16 bit (bit size), 48kHz (sample rate) • Notes: 98kps is acceptable for non-music sound track	AWS (P), TS (A), Network drive (T)
Still Image	• TIFF • 6000 pixels on the longest side or higher • 600 x 600 pixel per inch (ppi) resolution or higher	• JPEG • At least 1640 pixels on the longest side • At least 300 x 300 ppi resolution	
Textual	• PDF/A • B/W resolution 600 dots per inch (dpi), or higher • Colour resolution 1200 dots per inch, or higher	• PDF • At least B/W resolution 300 dpi • At least Colour resolution 600 dpi	
Video (Moving Image)	• Broadcast files (production version) • MFX OP-1a with a 10 bit – VBR Bit depth/rate • Video Codec shall be MJPEG2000 and the audio stream shall be BWF 48 kHz, 16 bit PCM encoded • RF 22-28, bitrate 4500 for 2160p 4K Ultra High Definition4 or higher	• MPEG4 container • Web optimised – indexes at the beginning (critical) • H.264 Coded, Framerate same as source • At least RF 20-24, bitrate 3000kbps for 1080p Full High Definition3	

** P – Preservation copies; A – Access copies; T - Temporarily*