

SECTION C: DIGITIZATION BEST PRACTICES AND GUIDES

Digitization is informed by best practice acquired through experience on projects. Below is a list of best practices and guides from institutions that have undertaken digitization projects. There are comprehensive guides that encompass multiple formats as well as format specific guides for audio and images.

C1: DIGITIZATION RESOURCES

Comprehensive

Audio, Video, and Image Digitization: Technical Specifications and Best Practices, May 2016 (Sustainable Heritage Network)

URL: <http://sustainableheritagenetwork.org/digital-heritage/audio-video-and-image-digitization-technical-specifications-and-best-practices> Last access: January 12, 2018.

This document provides basic specifications for audio, video, and image digitization. While exact specifications should be decided on a project by project basis, the following recommendations should be sufficient for most audio, standard definition video tape, and photographic and textual documents.

Complete Guide to Writing Data Management Plans, November 2017 (University of Maryland)

URL: <https://www.lib.umd.edu/data/dmp-guide> Last access: January 12, 2018

This guide outlines a writing strategy for creating a data management plan based on requirements common to many funding agencies. Some of the advice in this guide also applies to data sharing plans or data availability statements required by journals and certain funding organizations.

Digitization and Archives, October 2002 (Preservation Committee, Canadian Council of Archives)

URL: <http://www.cdncouncilarchives.ca/digitarc.html> Last access: [January 12, 2018](#).

This guide contains a listing of principles specific to digitization in archives and a decision tree for digitization projects.

Digitization Activities: Project Planning and Management Outline, November 2009 (Federal Agencies Digitization Guidelines Initiative)

URL: <http://www.digitizationguidelines.gov/guidelines/DigActivities-FADGI-v1-20091104.pdf> Last access: [January 12, 2018](#).

This document attempts to define work activities relating to the digitization of original cultural materials by outlining a generic sequence of high-level activities for planning and management purposes.

Digitization Best Practices, Version 1.0, August 2009 (University of Colorado Digital Library)

URL: <https://www.cu.edu/digitallibrary/cudldigitizationbp.pdf> Last access: [January 12, 2018](#).

This document offers an introduction to digitization, provides links to resources containing more information, and describes the recommended digitization parameters for collections in the CU Digital Library.

Digitization Best Practices (York University)

URL: <https://www.library.yorku.ca/web/collections/digitalscholarship/digitization-best-practices/> Last access: January 12, 2018.

This website provides a high-level overview of considerations for digitization projects, as well as links to further resources.

Digitization Guidelines (University of Virginia Library)

URL: <http://guides.lib.virginia.edu/content.php?pid=40437> Last access: [January 12, 2018](#).

These guidelines were developed to increase interoperability and accessibility, ensure a consistent, high level of quality, and decrease the likelihood of rescanning and promote long-term preservation of digital resources.

Digitization Guidelines (U.S. Federal Agencies)

URL: <http://www.digitizationguidelines.gov/> Last access: January 12, 2018.

The materials and information appearing on this site have been formulated collaboratively by a group of federal agencies to articulate a common set of guidelines, methods, and practices for the digitization of historical content in a sustainable manner.

Framework of Guidance for Building Good Digital Collections, 2007 (National Information Standards Organization)

URL: <http://framework.niso.org/> Last access: [January 12, 2018](#).

This framework provides an overview of some of the major components and activities involved in the creation of good digital collections as well as a framework for identifying, organizing, and applying existing knowledge and resources to support the development of sound local practices for managing good digital collections.

Guide to Digital Projects, Winter 2004 (MINITEX/LSD Joint Standards Review Task Force)

URL: <http://www.minitex.umn.edu/Communications/Standards/Digitization/Guide.pdf> Last access: [January 12, 2018](#).

This guide provides guidance to Minnesota Digital Library participants and to all Minnesota cultural heritage institutions that undertake digitization projects.

JISC Digital Media

URL: <https://www.webarchive.org.uk/wayback/archive/20160101151305/http://www.jiscdigitalmedia.ac.uk/> Last access: [January 12, 2018](#).

A comprehensive website offering practical advice on digitization for a wide range of digital media. The project ceased operating in 2016, however their website is still available for reference through the UK Web Archive.

Putting Content Online: A Practical Guide for Libraries, 2006 (Mark Jordan)

URL: <https://www.elsevier.com/books/putting-content-online/jordan/978-1-84334-176-5> Last access: [January 12, 2018](#).

This book focuses on practical, standards-based approaches to planning, executing and managing projects in which libraries and other cultural institutions digitize material and make it available on the web.

Audio

Audio Archives

URL: <http://www.theaudioarchive.com/index.htm> Last access: January 12, 2018.

The resources section of this website provides helpful advice on audio metadata, file size calculators, standards and best practices.

Digital Audio Best Practices, Version 2.1, October 2006 (CDP Digital Audio Working Group)

URL: <http://www.mndigital.org/digitizing/standards/audio.pdf> Last access: [January 12, 2018](#).

This document provides guidelines and a set of practices for cultural heritage institutions interested in converting analog audio recordings to digital formats. Recording audio directly into digital formats is also addressed.

Sound Directions: Best Practices for Audio Preservation (Mike Casey, Indiana University and Bruce Gordon, Harvard University)

URL: <http://www.dlib.indiana.edu/projects/sounddirections/papersPresent/index.shtml> Last access: [January 12, 2018](#).

This publication presents the results of research and development carried out by the Sound Directions project with funding from the National Endowment for the Humanities in the U.S.

Audio Tape Digitisation Workflow (Training for Audiovisual Preservation in Europe)

URL: <http://www.jazzpoparkisto.net/audio/> Last access: [January 12, 2018](#).

This guide provides step-by-step instructions on how to digitize open reel tapes. Much of the workflow can also be applied to cassette tapes, particularly when looking at the A/D conversion section.

Capturing Analog Sound for Digital Preservation, March 2006 (Council on Library and Information Resources and Library of Congress)

URL: <http://www.clir.org/pubs/abstract/reports/pub137> Last access: [January 12, 2018](#).

This report documents a roundtable discussion regarding best practices for the digital preservation of analog audio. It also briefly discusses how to inspect for and what to do in the case of deterioration of analog carriers.

Digitizing Speech Recordings for Archival Purposes (Plichta, B. & Kornbluh, M.)

URL: http://www.historicalvoices.org/papers/audio_digitization.pdf Last access: [January 12, 2018](#).

This paper details best practices for digitizing speech recordings including rationale for why these practices work best and how to choose what works best for your collection.

EBU Technical Standards

URL: <http://tech.ebu.ch/publications/tech3285> Last access: January 12, 2018.

Provides the technical specifications for metadata for the WAVE and Broadcast Wave (BWF) formats.

Open Language Archive Community

URL: <http://www.language-archives.org/> Last access: January 12, 2018.

This website is a central location for accessing current practice for the digital archiving of language resources, and for information on the developing network of interoperating repositories and services for housing and accessing such resources. The metadata standards used by the Open Language Archives community is found at <http://www.language-archives.org/OLAC/metadata.html>

Guidelines on the Production and Preservation of Digital Audio Objects. Second Ed., 2009 (IASA Technical Committee)

URL: <http://www.iasa-web.org/tc04/audio-preservation> Last access: [January 12, 2018](#).

This manual details best practices of audio digitization from various analog formats into digital formats. The manual also highlights storage procedures and addresses smaller scale approaches to audio digitization for smaller archives.

The State of Recorded Sound Preservation in the United States: A National Legacy at Risk in the Digital Age, August 2010 (Council on Library and Information Resources and The Library of Congress)

URL: <http://www.clir.org/pubs/reports/pub148/pub148.pdf> Last access: [January 12, 2018](#).

This report outlines the history and current state of sound preservation in the United States. It includes a very informative chapter with a basic overview of audio digitization and some of the technical issues that need to be taken into consideration

Images

CDP Digital Imaging Best Practices, Version 2.0, June 2008 (Bibliographical Center for Research)

URL: http://mwdl.org/docs/digital-imaging-bp_2.0.pdf Last access: [January 12, 2018](#).

This document is intended to serve as a guide for practitioners in cultural heritage institutions that seek to create images from physical objects.

Image Digitisation Manual 2007 (Picture Queensland)

URL: http://www.slq.qld.gov.au/_data/assets/pdf_file/0006/63879/SLQ_PQ_image_digitisation_manual_-_complete.pdf Last access: [January 12, 2018](#).

The manual provides information on the key aspects of planning and conducting a digitization project.

Guides to Quality in Visual Resource Imaging, (Digital Library Federation)

URL: <http://old.diglib.org/pubs/dlf091/dlf091.htm> Last access: [January 12, 2018](#).

These guides were written for those who have already decided what they will digitize and what purposes the digital images will serve and cover project planning, scanner selection, imaging system set-up, and the resulting digital masters. The website has not been maintained since 2010.

Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images, June 2004 (National Archives and Records Administration)

URL: <http://www.archives.gov/preservation/technical/guidelines.html> Last access: [January 12, 2018](#).

These guidelines define approaches for creating digital surrogates for facilitating access and reproduction.

Technical Guidelines for Digitizing Culture Heritage Materials: Creation of Raster Image Master Files, September 2016 (Federal Agencies Digitization Initiative FADGI – Still Image Working Group)

URL:

http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf Last accessed: [January 12, 2018](#).

This document represents shared best practices followed by agencies participating in the Federal Agencies Digitization Guidelines Initiative (FADGI) Still Image Working Group for digitizing cultural heritage materials.

C2: STANDARDS

The following tables contain standards summarized from various industry standards that should be followed for all digitization projects if possible. By adhering to these accepted standards, we are able to

- Ensure that the digital files created through digitization are of high quality and meet national and international standards
- Maintain the integrity and longevity of the digital files for long term digital preservation

These standards are subject to change as technology and practice evolve. Furthermore, each digitization project is unique in its setting and goals. The ultimate objective is to have a preservation master copy that is a faithful reproduction of the original from which additional copies can be made.

Manuscripts and printed text

| | Preservation and Access Master | Print Access | Screen Access | Thumbnail |
|--------------------|---|--------------------------------------|--------------------------------------|--------------------------------------|
| File format | TIFF and TXT or PDF/A with OCR | JPEG, PNG or PDF with OCR | JPEG, PNG or PDF with OCR | JPEG or PNG |
| Resolution | 300 – 600 dpi | 150 – 300 dpi | 150 dpi | 150 dpi |
| Bit depth | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale | 24 bit colour RBG or 8 bit grayscale |
| Dimensions | 3000 – 6000 pixels across the long edge | 3000 pixels across the long edge | 800 pixels across the long edge | 200 pixels across the long edge |
| Compression | Uncompressed | Lossless compression | Lossless compression | Lossless compression |

Photographs

| | Preservation and Access Master | Print Access | Screen Access | Thumbnail |
|--------------------|---|--------------------------------------|--------------------------------------|--------------------------------------|
| File format | TIFF | JPEG or PNG | JPEG or PNG | JPEG or PNG |
| Resolution | 300 – 600 dpi | 150 – 300 dpi | 150 dpi | 150 dpi |
| Bit depth | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale |
| Dimensions | 3000 – 6000 pixels across the long edge | 3000 pixels across the long edge | 800 pixels across the long edge | 200 pixels across the long edge |
| Compression | Uncompressed | Lossless compression | Lossless compression | Lossless compression |

Film, negatives, and slides

| | Preservation and Access Master | Print Access | Screen Access | Thumbnail |
|--------------------|---|--------------------------------------|--------------------------------------|--------------------------------------|
| File format | TIFF | JPEG or PNG | JPEG or PNG | JPEG or PNG |
| Resolution | 800 – 1200 dpi | 150 – 300 dpi | 150 dpi | 150 dpi |
| Bit depth | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale |
| Dimensions | 4000 – 6000 pixels across the long edge | 3000 pixels across the long edge | 800 pixels across the long edge | 200 pixels across the long edge |
| Compression | Uncompressed | Lossless compression | Lossless compression | Lossless compression |

Graphic art

| | Preservation and Access Master | Print Access | Screen Access | Thumbnail |
|--------------------|---------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| File format | TIFF | JPEG or PNG | JPEG or PNG | JPEG or PNG |
| Resolution | 600 - 800 dpi | 150 – 600 dpi | 150 dpi | 150 dpi |
| Bit depth | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale |

| | | | | |
|--------------------|---|----------------------------------|---------------------------------|---------------------------------|
| Dimensions | 6000 – 8000 pixels across the long edge | 6000 pixels across the long edge | 800 pixels across the long edge | 200 pixels across the long edge |
| Compression | Uncompressed | Lossless compression | Lossless compression | Lossless compression |

Maps

| | Preservation and Access Master | Print Access | Screen Access | Thumbnail |
|--------------------|---|---|--------------------------------------|--------------------------------------|
| File format | TIFF | JPEG or PNG | JPEG or PNG | JPEG or PNG |
| Resolution | Less than 36 inches on the long edge: 600 dpi Greater than 36 inches on the long edge: 300 – 400 dpi | Less than 36 inches on the long edge: 300 dpi Greater than 36 inches on the long edge: 150 dpi | 150 dpi | 150 dpi |
| Bit depth | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale | 24 bit RBG colour or 8 bit grayscale |
| Dimensions | 6000 – 8000 pixels across the long edge | 6000 pixels across the long edge | 1078 pixels across the long edge | 200 pixels across the long edge |
| Compression | Uncompressed | Lossless compression | Lossless compression | Lossless compression |

Audio recordings

| | Preservation and Access Master | Screen Access |
|--------------------|--|---|
| File Format | WAV, BWF or AIF (Apple) | MP3 |
| Sample Rate | Spoken language: 44.1 kHz Music and ambient sounds: 96 kHz | 44.1 kHz |
| Bit Depth | 24 bit | 16 bit |
| Comments | Highest recommended current quality, Standard for DVD/HD audio, Requires conversion to 16 bit and 44.1 kHz for most consumer audio devices | Lowest frequency range acceptable, Maximizes storage space, May not provide sufficient quality for future formats |

Video recordings

| | Preservation and Access Master¹ | | Screen Access² |
|----------------------------|---|------------------------------|----------------------------------|
| File format | QuickTime .mov | File format | .mov |
| Codec | UYVY | Codec | QuickTime H.264 |
| Bit depth | 10 bit | Frame size width | 640 |
| Frame size width | 720 pixels | Frame size height | 360 |
| Frame size height | 576 pixels | Pixel aspect ratio | Square |
| Frame rate | 25 frames per second | Frame rate | 23.976 |
| Frame type | Progressive | Field Output | Progressive |
| Frame aspect ratio | 4:3 | Pixel depth | 24 |
| Pixel aspect ratio | 1:1 | Spatial quality | 75 |
| Colour space | YCrCb | Min. Spatial quality | 25 |
| Chroma sub sampling | 4:2:2 | Key frame interval | 30 |
| Audio component | Uncompressed stereo audio | Temporal quality | 50 |
| Compressor | uncompressed PCM | Min. temporal quality | 25 |
| Bit depth | 16bit / 24bit | Average data rate | 1.331 Mbps |
| Sample rate | 48KHz | Maximum data rate | 1.331 Mbps |

¹ To be truly a preservation master, video should be in an uncompressed state. However, in a raw, uncompressed state, 1 minute of video uses up to 1GB of storage. Another file format (codec with wrapper) suitable for preservation is JPEG 2000 with the MXF wrapper. JPEG 2000 offers lossless compression and reduces the file size by 3:1.

² These standards are applicable to born digital video as well. Born digital video captured at these standards can then serve as the Preservation and Access Masters.

| | | | |
|---------------------------|----------------------|----------------------|-------------------------------|
| Number of channels | 2 | Audio Encoder | AAC, Stereo (L R), 48.000 kHz |
| Audio interleave | 1 sec | File size | 599.04 MB/hour of source |
| File size | 93 GB/hour (approx.) | | |

GLOSSARY OF TERMS

dpi stands for dots per inch, a measurement of resolution for a digitized document (the higher the dpi, the better the tonality of the image.) The dpi setting of the scanner relates to the final pixel size of the scanned image.

8-bit refers to method of storing image information in a computer's memory or in an image file, such that each pixel is represented by one 8-bit byte.

grayscale refers to an image in which the value of each pixel is a single sample composed exclusively of shades of grey.

JPEG stands for Joint Photographic Experts Group and refers to a type of graphics file format commonly used for images, photographs, etc.

PDF stands for Portable Document Format and is Adobe's proprietary file format.

OCR stands for Optical Character Recognition. It is the electronic translation of scanned text into machine-encoded text. OCR makes it possible to edit the text, search for a word or phrase, etc.

24-bit RGB refers to 24 bits per pixel in which three 8-bit integers between 0 and 255 represent red, green and blue intensities.

TIFF stands for Tagged Image File Format and refers to a type of file format for storing images.

96 kHz 24-bit refers to sample rate for audio. It means that a sample at 24 bits is taken 96,000 times per second.

BWF stands for Broadcast Wave Format. It is a standard used by the broadcast industry whereby metadata can be added to Wave files.

WAV stands for Waveform Audio File Format. It is an audio file format standard for storing an audio bitstream.

MP3 is a digital audio encoding format using a form of lossy data compress