

# JULAC Digitisation Guideline

*Prepared by Digital Initiatives Committee*

## Purpose

- To ensure a minimum standard conducive to the flexible use of digitised objects and long-term preservation.
- To ensure our practices follow international best practices as much as possible while adapting to the local environment.

## Scope

- This guideline mainly covers the digitisation of textual and pictorial material, i.e. books, artworks, posters, postcards, photographs, and negatives and a variety of audio-visual materials, including but not limited to Audio recordings (analogy and digital); Video recordings (analog and digital); Film reels; Oral history interviews; Lectures and presentations as well as Multimedia artworks.

## Principles

- **Fit for purpose** – each institution may apply discretion to apply different digitisation standards for (1) preservation of items with intrinsic value, such as rare books; or (2) reproduction of mass-produced items for purposes other than preservation.
- **Be faithful to the original** when capturing. Do not use sharpening techniques to improve on the original (i.e., master copy), unless without sharpening significant information will be lost for the user. Ensure that the digitized AV materials are preserved at the highest quality possible while also considering access needs and formats suitable for online delivery.
- **Capture at the highest possible standard at the time if affordable** – especially for items with intrinsic value, allowing subsequent multiple re-use of the images or content.
- **Use generic, non-proprietary file format** as much as possible for the master copy.
- **Include a minimum set of metadata** with the digitised items for item discovery purposes. Consider how and where the digitised content will be used when planning the digitisation project (website, database creation, etc.); Technical metadata is essential for AV materials. Include information (such as: Original format and carrier, Duration, Audio channels and sample rate, Video resolution and frame rate, Codec used for access copies, Date and time of digitisation. Equipment used. Any restoration or processing performed, etc)

## Digitisation Output

- **Master copy** – also called Preservation copy, which captures the most information from the original item. Uncompressed or losslessly compressed TIFF is the recommended format for master copies. Do not use multi-page TIFF. When there is a need to use a colour target, i.e. for important visual materials (e.g. paintings), the master copy should include the colour target. Capture audio in lossless formats like FLAC or WAV and video in formats like ProRes or FFV1. Aim for high resolutions suitable for the specific format and content.
- **Co-master** – If a colour target is used, a co-master can be created with the target cropped out. The co-master should be in the same file format as the master copy.

- **Access copy** – also called Presentation copy, Derivative copy or Service copy
  - created from the master copy, but small enough in file size for online delivery purposes
  - can be images or textual. Thumbnails can also be considered access copies
  - In case the master copy contains a colour target, the access copy should be created from the co-master (i.e. with the colour target cropped out)
  -
- **Text file** (applicable for print materials) – used to hold OCR'ed text or checksum, in UTF-8 encoding plain text format.
- **Captioning and Subtitling** (applicable for video materials) Consider adding captions or subtitles to video materials to enhance accessibility and searchability.
- **Checksum** (in MD5 format) – checksums for each master copy file is highly recommended to ensure file integrity. Checksums should be stored in a text file, named in a way that is identifiable with the associated master copy.
- **Note on file names**
  - Avoid the use of uppercase letters
  - Use underscore ( \_ ) as separator between words
  - Include numerical values as part of the file name if digitised items have innated sequential orders (e.g. page number)
  - Apply an alphabetical prefix to minimise filename duplication with other digitisation projects
  - Use unique file name for a digitisation project, even if files are stored in different folders

## Capture Standards

### Standard for Textual Materials

(books, serials, pamphlets, newspaper clippings, etc.)

- Resolution – at least 300 dpi; 400 dpi is recommended if OCR is to be performed
- Colour – Adobe RGB 24-bit colour
- Border – Up to 2 mm border recommended especially for rare books
- De-skew – the scanned image should be de-skewed as much as possible, especially if OCR is to be performed
- In a book/journal, a page in a different orientation to the rest should be captured as is for the master copy. For access copy, consider rotating the page for better readability.
- When necessary, apply physical intervention for enhancing legibility (e.g. placing a piece of paper underneath a very thin page to improve text legibility)
- When necessary, apply physical intervention to aid in capturing the object's original form (e.g. placing a paper underneath an irregular-shape page)
- For books/journals, digitise one page per image, except for situations when scanning a two-page spread or centrefold separately on two images create visual discontinuity or readability issues.
- For books/journals, maintain a consistent width-height aspect ratio for all images from the same volume, except for attachment or loose-leaf.

- Attachments or loose-leaf material in books/journals should be captured as is to maintain contextual information. If no specific contextual information needs to be maintained, append it to the end of scans.

#### Standard for OCR of Print Materials

- OCR can be undertaken at the item, page or article level, depending on resources and significance of the material
- When OCR'ing, ensure the book chapter/journal article title and author information is 100% correct

#### Standard for Pictorial Materials

(posters, postcards, photographs, negatives, etc.)

- Dpi – 400 dpi for posters and postcards; 600 dpi for photographs and prints; 1200 dpi for negatives and slides
- Colour – Adobe RGB 24-bit colour
- Border – Up to 2 mm border recommended
- Colour targets should be used for items with unique and subtle qualities such as original paintings

#### Standard for Audio Materials:

- Audio Quality – Capture audio at a high bit depth and sample rate to preserve nuances in sound quality.
- Bit Rate – Aim for encoding in the MP3 format with bitrate 128 kbps for high-quality audio recordings.
- Mono/Stereo – Capture audio in mono or stereo based on the original recording and intended use.

#### Standard for Video Materials:

- Video Quality – Capture video at a high resolution and bitrate to maintain visual fidelity. Video file encoded in the MP4 format with a video bitrate of 800 kilobits per second (kbps) .
- Resolution – Depending on the source material, consider capturing video at HD (720p), Full HD (1080p), or higher resolutions.
- Frame Rate – Maintain the original frame rate of the video during digitization to preserve the intended viewing experience.

## Reference

- FADGI (Federal Agencies Digital Guidelines Initiative): <https://www digitizationguidelines.gov/?loclr=blogsig>
- Europeana Digitization Handbook: <https://europeana.github.io/fste-digitization-handbook/>
- IFLA Guidelines for Planning the Digitization of Rare Book and Manuscript Collections <https://repository.ifla.org/handle/20.500.14598/454>
- Digital Preservation Handbook: Fixity and checksums <https://www.dpconline.org/handbook/technical-solutions-and-tools/fixity-and-checksums>
- Metamorfoze: <https://www.metamorfoze.nl/english/digitization>
- NARA Guidelines for Digitizing Archival Materials for Electronic Access <https://www.archives.gov/files/preservation/technical/guidelines-1998.pdf>

- Library of Congress <https://www.loc.gov/preservation/care/scan.html>
- ISO <https://www.iso.org/standard/52391.html>
- IASA TC-04: The International Association of Sound and Audiovisual Archives (IASA)  
<https://www.iasa-web.org/tc04/audio-preservation>