

Preserving Raster Images

Data Types Series

Artefactual Systems and the Digital
Preservation Coalition



DPC Technology Watch Guidance Note

July 2021



Digital Preservation Coalition

The Data Type Guidance Note Series

Each Guidance Note in the Data Types series is designed to provide a primer on the current state of community knowledge about data types commonly encountered by those seeking to preserve digital holdings. Digital preservation is about keeping information findable, usable, and trustworthy over the long-term. The best approach for any repository will vary according to the scope and content of its holdings, available resources, and the expectations of its funders and users. There are however, broadly applicable good practices that have been established as a result of many years of research, practical implementation, and consensus building. These are presented here as a starting point, along with additional resources for further exploration.

This series of Data Type Guidance Notes has been authored by staff at Artefactual Systems in collaboration with the Digital Preservation Coalition. These notes have been developed in conjunction with the UK Nuclear Decommissioning Authority.

Digital preservation is an evolving field and continues to change and develop in response to external drivers and fresh challenges. New formats, standards, and examples of good practice will emerge over time and the information contained within this report will need to be updated. We welcome comments and feedback to: info@dpconline.org.

1 Overview

Raster image is a term used for an image made up of a grid of pixels ([SAA, 2020](#)). Pixels (also known as dots) are:

- The smallest unit in a raster image ([Wikipedia, 2020c](#)).
- A representation of colour, which is made up of bits. The number of bits used to represent the colour of a pixel is known as bit depth.

Raster images can be created through activities such as digital photography and digitization. Both processes use photo sensor elements to create a grid of pixels. Raster images are different from vector images, which use points connected by lines and curves to form shapes instead of pixels. For more information on vector images, see the Guidance Note on *Preserving CAD*.

Each file format has specific properties that influence its appearance, preservation, or accessibility:

- **Bit depth** or **colour depth** defines the number of colours a pixel may represent. This is expressed in how many bits are used to define the colour, from 1-bit to 48-bit. For example, 24-bit colour depth represents true colour, or what the human eye can see ([Wikipedia, 2020b](#)).
- **Colour space** defines how colours are expressed as numbers. The most common colour space is RGB, which models colours based on three channels: Red, Green, and Blue ([Brown, 2008](#)). Each channel stores a colour numerical value that ranges from 0 to 255. Colour spaces are modelled numerically and standardised through what is known as **colour profiles**. sRGB, AdobeRGB(1998), and CMYK are examples of common colour profiles ([ASMP, 2015b](#)). Colour profiles can be converted from one to another, but such conversions can alter the quality and display of colours ([ASMP, 2015b](#)).
- **Compression** algorithms can be used to reduce the storage size of a raster image. They compress the information within the file through a lossless or lossy method. Lossy compression reduces quality to achieve smaller file sizes, which leads to an irreversible degradation of image quality. Lossless compression reduces size while retaining all original information, thereby preserving image quality.
- **Resolution** is expressed in dots per inch (DPI) or pixels per inch (PPI) ([ADS, 2009](#)). Resolution is the fixed number of pixels an image contains. Images viewed on display screens can have a lower image resolution (~72dpi) but image resolution for printing needs to be much higher (~300dpi) or the print will appear pixelated ([Wikipedia, 2021](#)).

2 Preservation challenges

Relative to many other data types, most raster file formats are simple and well documented. Software support is generally very good for raster image file formats. Preservation problems tend to arise mainly from the use of proprietary file formats or from image properties.

2.1 Image properties

There are many image properties (described above), each of which can have an impact on preservation. These include colour space, bit depth, resolution, compression, and embedded metadata. When preserving raster image files, consideration must be given to how those properties might affect image quality.

2.2 Variety of formats

There are a large number of raster image file formats, some of which are not widely supported by rendering applications. Formats that may cause preservation or access challenges include proprietary formats such as RAW formats from camera manufacturers and complex formats not widely supported by standard image viewing software (such as JPEG2000).

2.3 Volume

Raster images are typically used to store the outputs from digitization activities. Digitization programmes often produce large quantities of high quality, sometimes uncompressed images and this can lead to challenges for those charged with managing and preserving the resulting outputs.

3 File formats

There is no single perfect format for the preservation and future use of raster images. Decisions made on file formats should be dependent on the features and functionality to be preserved and the future use cases to be supported. Note that the table below does not provide an exhaustive list of formats 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.

File format	Extensions	Brief summary
Tagged Image File Format (TIFF)	.tif .tiff	TIFF is the most commonly used preservation format for digitization (FADGI, 2016). TIFF supports different compression schemes such as the lossless LZW algorithm (LC, 2009b).
Joint Photographic Experts Group 2000, JPEG 2000 (JP2)	.jp2	<p>Supports both lossy and lossless compression, but some compression standards it uses are proprietary and not widely available. JP2 was considered a challenger to TIFF as an image preservation format when it was introduced over a decade ago. It has impressive lossless compression and small file sizes, thus provides a cost saving option over TIFF (LC, 2011).</p> <p>The International Image Interoperability Framework (IIIF, n.d.) and mass digitization have led to JP2 being selected as an access format by several heritage organizations. JP2 also supports spatial random access, which is useful for serving data over the Internet.</p> <p>However, the format is more complex than TIFF, uses a number of proprietary compressions standards, and is supported by few affordable software options. These factors, as well as issues that can arise when converting JP2 files to other formats have slowed uptake of JP2 as a preservation format (Buckley, 2008; Kyle, 2015).</p>

Graphics Interchange Format (GIF)	.gif	A widely used access format that supports animation. Uses LZW lossless compression, which was under patent until 2004 (Brown, 2008). GIF files have limited colour depth (LC, 2006). Sometimes used for preservation if other versions of image files are unavailable (LoC 2020-2021).
Joint Photographic Experts Group (JPEG)/ JPEG File Interchange Format (JFIF)	.jpg .jfif	A very widely used access format. JPEG is the algorithm and JFIF is the format specification but JPEG is most commonly used as the format name (Brown, 2008). The patent on its compression algorithm expired in 2004. Altering a JPEG can lead to image degradation when the changes are saved. Sometimes used for preservation if other versions of image files are unavailable (LoC 2020-2021).
Portable Network Graphics (PNG)	.png	Uses a lossless compression for storage. While very common on the Web, software doesn't always support reading its metadata (LC, 2009a). PNG supports image transparency. Sometimes used for preservation if other versions of image files are unavailable (LoC 2020-2021).
Photoshop Document Format (PSD)	.psd	Developed by Adobe (2020) as a working format for images in Photoshop (Wikipedia, 2020a). PSD supports the use of image layers. PSD is a proprietary format and free rendering applications are not available.
Windows Bitmap (BMP)	.bmp	A proprietary format created to be device-independent. Applications for rendering and documentation are widely available.
Digital Negative (DNG)	.dng	A format aimed primarily at storing RAW camera images (Wikipedia 2021), published by Adobe (LC, 2012). DNG files can often be quite large compared to files in other formats.
Other RAW formats	various	RAW files contain minimally processed data from the image sensor of a camera or scanner. Each major camera manufacturer has its own format, and proprietary software is usually needed to render, process, and export the RAW files into another file format.

4 Metadata standards

Image formats embed metadata which includes technical information and may also support description, rights, and location information:

- Exchangeable Information File Format (EXIF) is a standard for capturing creation metadata from cameras and scanners ([ASMP, 2015c](#)). A file using EXIF should not be altered after

capture as the metadata includes information about provenance, which helps establish authenticity.

- International Press Telecommunications Council (IPTC) is a standard that contains fields for descriptive metadata such as photographer, description, keywords, copyright, and location information ([ASMP](#), 2015c).
- eXtensible Metadata Platform (XMP) is a wrapper standard from Adobe Systems Inc. used to store different kinds of metadata together with the image, including IPTC, Dublin Core, and more ([ASMP](#), 2015d). XMP metadata are encoded in [Extensible Markup Language](#) (XML) for interoperability ([ASMP](#), 2015d). XMP is supported by TIFF, JPEG, PSD, and DNG formats.

5 Tips for creators

5.1 Born-digital

For creators of born-digital images, there are good practices that can assist long-term preservation and access.

- The [American Society for Media Photographers](#) (ASMP, 2015a) makes format recommendations for still photography, and provides advice about file management, metadata creation, and data validation.
- Consider using ubiquitous file formats with either no compression or lossless compression for long-term retention.
- Use file naming conventions to maintain the relationships between raw working formats, final works, and other related files.
- Avoid repeatedly editing images saved in lossy formats (such as JPEG), as this will cause image degradation each time the files are opened, edited, saved, and closed.
- Creating metadata such as copyright details, creator name, and other descriptive metadata improves searchability and accessibility.

The [Library of Congress](#) (2020-2021) outlines recommended descriptive metadata elements to either embed in digital photographs or host in an XML-based file alongside the image. The guidelines include requirements for including technical metadata.

- FADGI provides guidelines on embedding metadata in TIFF images ([FADGI](#), 2009).
- *Z39.87-2006 (R2017): Technical Metadata for Digital Still Images* ([NISO](#), 2017) is a standard that provides information on how to represent technical metadata in an XML-based file.

5.2 Digitized

For the digitization of heritage material and artworks, there are a number of good practice standards to follow. These provide guidance on file types, resolution, image size, compression, colour space, bit depth, and more:

- *Technical Guidelines for Digitizing Cultural Heritage Materials* ([FADGI](#), 2016).
- *Metamorfoze Preservation Imaging Guidelines* ([van Dormolen](#), 2012).
- *ISO/TR 19263-1: Best practices for digital image capture of cultural heritage material* ([ISO](#), 2017).

There are some useful practices for creating digitized images for long-term preservation:

- Create and maintain documentation about the digitization project and its workflow. Ingest this documentation with the digital images into a digital repository to allow future users to understand what decisions were made about the creation process.

- Validate file formats if possible (see the Characterization section, below). Validation ensures images are constructed according to a format standard, making them easier to preserve and provide access to.
- Balance quality, storage, and sustainability considerations when selecting a preservation format as well as the needs of current and future users of the resource.

6 Tips for archivists

6.1 General guidance

The following resources provide guidance on preserving and providing access to raster images:

- [Library of Congress](#) (2020-2021) *Recommended Formats Statement: Still Image Works*.
- [Archaeology Data Service](#) (2019) *Raster Image Procedures*.
- The [International Council on Archives](#) (2016) provides a range of guidance notes on preservation practices, management of photographic archives, and digitization projects.
- [State Library New South Wales \(SLNSW\)](#) (2018) and Woodruff Library, Atlanta University Center ([Educopia](#), 2018) provide guidance of digitized and born-digital image workflows.
- [England and Hanson](#) (2017) from John Hopkins University describe automated processing procedures. This guidance is useful for born-digital or legacy digitized image appraisal.
- [Princeton University Library](#) (2020) has guidance on born-digital materials for special collections and archives.
- [SLNSW](#) (2019) outlines acquisitions and appraisal guidance for born-digital photographs.

6.2 Acquisition and appraisal

Perform selection and appraisal while considering the context of the digitized images:

- Images digitized for publications and exhibitions are often different from images digitized for preservation. They may be in a colour space ideal for printing, and they may have been retouched or otherwise manipulated. For these reasons, they are sometimes more ephemeral than images digitized explicitly for long-term preservation.
- Images digitized on demand for users are often not retained for long-term preservation, since they can be lower quality than other digitized images.
- Raster images from recent digitization programmes have usually followed well-documented procedures and result in uniform, standardized files; they often do not require further appraisal.
- Where legacy digitized images do not meet current standards, consider re-digitizing if time and budget permit.

6.3 Creation and editing tools

Cameras and scanners have built-in features that create, edit, and convert digital images. There are also non-manufacturer specific tools such as the [Silverfast](#) scanning tool. These tools manage scanner settings to create images from analogue photographs or negatives, converting them to various file formats.

- Photoshop and [GNU Image Manipulation Program/GIMP](#) (2020) are popular editing and creation programs.

- [ImageMagick](#) (1999-2021) is an open-source command line tool that can also be used for editing images. It is often bundled with other software programs for image modification and migration.

6.4 Preservation action

- File format migration is often unnecessary for images from recent digitization programmes which are typically of high quality and uncompressed or losslessly compressed.
- Where file format migration is necessary, pay close attention to potential issues caused by loss or damage to colour space or profile during migration. [Hibberd](#) (2017) and [Ledoux](#) (2019) discuss some of the potential issues colour space can cause as formats are converted.

6.5 Characterization

Characterization can be useful to identify file formats, extract metadata, identify broken or encrypted content, or check conformance to profiles or standards. Tool support and effectiveness can vary considerably for different file formats.

- Identify file formats with a tool such as [DROID](#), [FIDO](#), or [Siegfried](#) that uses the [PRONOM file format registry](#).
- Open source validation tools such as [JHOVE](#), [Jpylyzer](#), and [DPF Manager](#) can be used to validate file formats such as TIFF, JPEG and PNG. Visit the Community Owned digital Preservation Tool Registry ([COPTR](#), 2014) for information on more validation tools for raster images. Note that validation, even of simple formats, remains challenging. Tool output is often difficult to interpret and results between different tools inconsistent ([Tunnat](#) 2017).
- Perform additional quality assurance and validation checks if digital images were produced by an outsourced vendor to ensure they conform to the specifications given to the vendor.
- Tools such as ExifTool, Photoshop and Glimpse can be used to modify or add metadata. This should typically be done only in compliance with documented organizational policies.
- Extracted technical metadata can be stored as XML per *Z39.87-2006 (R2017): Technical Metadata for Digital Still Images* ([NISO](#), 2017).
- The [Sustainable Heritage Network](#) (2016) provides advice for managing image metadata

7 References

Adobe (2020) *File Formats*. Available at:

<https://web.archive.org/web/20201111185401/https://helpx.adobe.com/photoshop/using/file-formats.html>

American Society of Media Photographers [ASMP] (2015a) *Best practices*. Available at:

<https://web.archive.org/web/20201001055645/https://dpbestflow.org/links/32>

ASMP (2015b) *Best practices: Color Space and Colour Profiles*. Available at:

<https://web.archive.org/web/20201001053019/https://dpbestflow.org/color/color-space-and-color-profiles>

ASMP (2015c) *Best practices: Metadata*. Available at:

<https://web.archive.org/web/20201001052102/https://dpbestflow.org/metadata>

ASMP (2015d) *Best practices: XMP*. Available at:

<https://web.archive.org/web/20201001051309/https://dpbestflow.org/metadata/xmp>

Archaeology Data Service [ADS] (2020) *Raster Images Procedures*. Available at:

https://web.archive.org/web/20210112140010/https://archaeologydataservice.ac.uk/resources/attach/ADS_Raster_Procedures_v1-115.pdf

ADS (2009) *Guides to Good Practice -- Raster Images*. Available at:

https://web.archive.org/web/20201024201852/https://guides.archaeologydataservice.ac.uk/g2gp/RasterImg_Toc

Bodleian Libraries (2020) *Validation*. Available at:

<https://web.archive.org/web/20201221181203/https://libguides.bodleian.ox.ac.uk/digitalpreservation/validation>

Brown, A. (2008) *Digital Preservation Guidance Note 4: Graphics file formats*. Available at:

<https://web.archive.org/web/20180328103029/http://www.nationalarchives.gov.uk/documents/information-management/graphic-file-formats.pdf> [accessed 21 December 2020]

Buckley, R. (2008) *JPEG 2000: a Practical Digital Preservation Standard?* Available at:

<http://doi.org/10.7207/twr08-01> [accessed 21 December 2020]

COPTR (2014) *Category: Validation*. Available at:

<https://web.archive.org/web/20200101024945/https://coptr.digipres.org/Category:Validation>

Educopia (2018) *OSSArcFlow As-Is Workflow: Robert W. Woodruff Library, Atlanta University Center*.

Available at: <https://educopia.org/as-is-workflow-atlanta-university-center-robert-w-woodruff-library/>

England, E. and Hanson, E. (2017) *Automating Digital Archival Processing at Johns Hopkins University*. The Signal. Available at:

<https://web.archive.org/web/20201108022836/https://blogs.loc.gov/thesignal/2017/05/>

Federal Agencies Digitization Initiative [FADGI] (2011?) *Guidelines: Embedded Metadata in TIFF Images*. Available at:

https://web.archive.org/web/*/http://www.digitizationguidelines.gov/guidelines/digitize-tiff.html

Federal Agencies Digitization Initiative [FADGI] (2016) *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Files*. Available at https://web.archive.org/web/20201124214052/http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf

GitHub (2020) *Glimpse Image Editor*. Available at: <https://web.archive.org/web/20201107230350/https://github.com/glimpse-editor/Glimpse>

Harvey, P. (2020) *ExifTool by Phil Harvey*. Available at: <https://web.archive.org/web/20201112030540/https://exiftool.org/>

Hibberd, L. (2017) *JP2 and colour profile limitations: a positive conclusion and findings*. Available at: <https://web.archive.org/web/20201028204850/https://www.dpconline.org/blog/jp2-colour-profile>

ImageMagick Studio LLC (1999-2020) *ImageMagick*. Available at: <https://web.archive.org/web/20201130220427/https://imagemagick.org/index.php>

IIIF (n.d.) *International Image Interoperability Framework*. Available at: <https://web.archive.org/web/20210127041441/https://iiif.io/>

ISO (2017) *ISO/TR 19263-1:2017 Photography -- Archiving systems -- Part 1: Best practices for digital image capture of cultural heritage material*. Available at: <https://web.archive.org/web/20190113053603/https://www.iso.org/standard/64220.html>

Kyle, R. (2015) *JPEG2000 and Digitisation: Expert round table*. TownsWeb Archiving Blog. Available at: <https://web.archive.org/web/20201020051328/https://blog.townswearchiving.com/2015/10/jpeg-2000-and-digitisation-expert-round-table>

Ledoux, T. (2019) *JPEG Got the Blues*. Open Preservation Foundation Blog. Available at: <https://web.archive.org/web/20201207200417/https://openpreservation.org/blogs/jpeg-got-the-blues/>

Library of Congress. (2020-2021) *Recommended Formats Statements: Still image works*. Available at: <https://web.archive.org/web/20201111202454/https://www.loc.gov/preservation/resources/rfs/stillimg.html>

Library of Congress (2012) *Adobe Digital Negative (DNG), Version 1.1*. Available at: <https://web.archive.org/web/20201105231910/https://www.loc.gov/preservation/digital/formats/fdd/fdd000188.shtml>

Library of Congress (2011) *JPEG 2000 Part 1 (Core) jp2 File Format*. Available at: <https://web.archive.org/web/20201018171939/https://www.loc.gov/preservation/digital/formats/fdd/fdd000143.shtml>

Library of Congress (2009a) *PNG, Portable Network Graphics*. Available at: <https://web.archive.org/web/20201111202543/https://www.loc.gov/preservation/digital/formats/fdd/fdd000153.shtml>

Library of Congress (2009b) *TIFF, Revision 6.0*. Available at: <https://web.archive.org/web/20201127012640/https://www.loc.gov/preservation/digital/formats/fdd/fdd000022.shtml>

Library of Congress (2006) *GIF Graphics Interchange Format, Version 89a*. Available at:
<https://web.archive.org/web/20201103114334/https://www.loc.gov/preservation/digital/formats/fdd/fdd000133.shtml>

NISO (2017) *ANSI/NISO Z39.87-2006 (R2017) Data Dictionary - Technical Metadata for Digital Still Images*. Available at:
<https://web.archive.org/web/20200929143149/https://niso.org/publications/ansiniso-z3987-2006-r2017-data-dictionary-technical-metadata-digital-still-images>

Princeton University Library (2020) *ADAPT Born-Digital Procedures*. Available at:
<https://web.archive.org/web/20201021195010/https://library.princeton.edu/special-collections/workflows/born-digital/adapt-born-digital-policies>

Research Libraries Group (2004) *Automatic Exposure: Capturing Technical Metadata for Digital Still Images*. Available at:
https://web.archive.org/web/20170830033423/https://www.oclc.org/content/dam/research/activities/automaticexposure/ae_whitepaper_2003.pdf

SAA (2020) *Raster graphic*. Available at:
<https://web.archive.org/web/20201202143745/https://dictionary.archivists.org/entry/raster-graphic.html>

State Library New South Wales [SLNSW]. (2019). *Digital Preservation at the Point of Acquisitions: Collection Born-digital Photographs*. Available at:
<https://web.archive.org/web/20200116220235/https://www.sl.nsw.gov.au/blogs/collecting-born-digital-photographs>

SLNSW (2018) *Workflows: a language everyone can understand*. Available at:
<https://web.archive.org/web/20201028034308/https://www.sl.nsw.gov.au/blogs/workflows-language-everyone-can-understand>

Sustainable Heritage Network (2016) *Metadata for Images*. Available at:
https://web.archive.org/web/20210112141230/https://www.sustainableheritagenetwork.org/system/files/atoms/file/07_MetadataforImages_0.pdf

The National Archives (2020) *The Technical Registry: PRONOM*. Available at:
<https://web.archive.org/web/20201111032324/http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

The National Archives (2008) *Digital Preservation Guidance Note 5: Image Compression*. Available at:
<https://web.archive.org/web/20200420162501/https://www.nationalarchives.gov.uk/documents/information-management/image-compression.pdf>

Tunnat, Y. (2017) *TIFF format validation: easy-peasy?* Available at:
<https://web.archive.org/web/20210415181451/https://openpreservation.org/blogs/tiff-format-validation-easy-peasy/>

van Dormolen, H. (2012) *Metamorfoze Preservation Imaging Guidelines*. Available at:
https://web.archive.org/web/20201202141453/https://www.metamorfoze.nl/sites/default/files/publicatie_documenten/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf

Wikipedia (2021) *Pixelation*. Available at:
<https://web.archive.org/web/20210125004651/http://en.wikipedia.org/wiki/Pixelation>

Wikipedia (2020a) *Adobe Photoshop*. Available at:

https://web.archive.org/web/20210102105524/https://en.wikipedia.org/wiki/Adobe_Photoshop

Wikipedia (2020b) *Color Depth*. Available at:

https://web.archive.org/web/20201203060637/https://en.wikipedia.org/wiki/Color_depth

Wikipedia (2020c) *Pixel*. Available at:

<https://web.archive.org/web/20201231091616/https://en.wikipedia.org/wiki/Pixel>