



# Preservation Handbook

## Bitmap (raster) images

---

Author	Michael Eadie
Version	1.3
Date	11 <sup>th</sup> July 2005
Change History	

---



## Definition

The most common type of image found on the web, the raster image, is made up of a mosaic of picture elements known as pixels. These pixels are the squares seen on the image when it is zoomed. Each pixel stores colour information about the image, and all the pixels taken in total make up the image in its entirety.

This handbook provides an overview of issues concerning the creation and long-term preservation of digital still images. It will provide an overview of the nature of digital still images; guidelines on best practice in the creation of digital still images; information on avoiding common pit falls; and an understanding of how the AHDS will preserve the resulting digital objects. In this handbook still images will be referred to as raster images. Raster images are also known as 'bit-mapped' or 'bitmap' images. The terms will be described below.

## Description

The four main defining features of a raster image are:

### 1. File Format

The file format of the image relates to the type of computer code that is used to structure the raster image. There are many choices for the type of format to choose for raster images (see Common Formats). Different file formats offer differing levels of compatibility and compression. For archival purposes we prefer non-proprietary formats without compression.

If a high resolution object is created, it is considered best practice to store at least two versions of the object: the master high resolution uncompressed file that is archived at point of capture, and a surrogate compressed version derived from the master for dissemination purposes. The uncompressed master or archival version should, wherever possible, be stored in an open source file format.

The two types of compression available to the digitiser are called 'lossless' and 'lossy'. Lossless compression enables a reduction in file size without losing any information within the file, hence *loss/ess*. It does this using algorithms (or software routines) that reduce the number of bits used to represent data in a file, thereby reducing its size while retaining all the original information. Lossless compression typically achieves space savings of up to thirty percent. Graphical Interchange File Format (GIF) is a common type of lossless compression used for still images. GIF uses a form of compression called LZW compression after its creators: Lempel, Ziv, and Welch), but it is limited to 256 colours.

Lossy compression, on the other hand, reduces file size with a corresponding loss of data. It works by eliminating information from the file that the program deems superfluous. The lost information is either unnoticeable to the user, or can be recovered during decompression by extrapolation of the existing data. Lossy is probably the appropriate type of compression for data that is originally analogue rather than digital, such as video and sound clips, and continuous-tone greyscale or colour images. Joint Photographic Experts Group (JPEG) is a common image file format that uses lossy compression.

### 2. Resolution

The resolution of an image concerns the number of pixels held within the digital file, and is measured in pixels per inch (ppi). The more pixels stored per inch, the greater the density of the colour information, and therefore the greater the detail evident in the image. The appropriateness of the resolution chosen depends on the intended *purpose* of the digital image. However, note that resolution or ppi is only an indicator of image size, and therefore 'quality', when we know the dimensions of the original analogue object. For example, scanning an A4 document (9 × 12 inches) at 300 ppi will produce a digital image that is 2700 pixels x 3600 pixels (the dimensions of the original multiplied by the ppi). Scanning a postage stamp that is 1 inch x 1 inch in size, will produce a digital



image that is 300 pixels x 300 pixels. Both these images are scanned at the same 300ppi resolution, but produce vastly different sized digital images. A more accurate way to refer to the size of a raster image therefore is to use its pixel dimensions.

### 3. Bit Depth

Bit-depth refers to the amount of colour information held with each individual pixel. In a greyscale image there are usually 8 'bits' of information in each pixel, in a colour image 24 bits are standard, although some hardware will now scan at 48 bit. The number of bits, or information, also increases the file size, and for storage purposes it is worth considering what the most important information is e.g. do you really need a black-and-white photograph scanned in RGB colour or will greyscale do. See below for more information.

### 4. Colour Space

The colour space of an image refers to the method of working with colour based on colour models. The most common colour models are bitonal, grayscale, Indexed colour, RGB and CMYK. The bitonal mode uses only two colour values, black and white, to display images. The grayscale mode offers 256 shades of grey that range from white to black. Indexed colour is the limited palette of 216 colours that web designers are able to use which display on both Macintoshes and PCs. It is more usual to work in RGB and then only convert to Index colour mode if you think it will really effect your users e.g. if a logo has to display as a certain shade of red. Computer monitors and TV screens use the RGB model to display a mixture of red, green and blue colour values. The CMYK model refers to the printing colours of cyan, magenta, yellow and black.

The colour model effects the file size of the image, since the more bits and bytes, the larger the file size. Images with only black or white pixels (bitonal – 1 bit per colour channel) will therefore have the smallest file sizes, grayscale images quite small file sizes (8 bits per colour channel) whereas RGB images with 24-bits (8 bits per colour channel) or more will have much bigger file sizes.

Another issue to consider is that it can be difficult to convert from RGB to CMYK. This is normally done when you need a commercial company to print an image for you. When you convert to CMYK you need to be aware of the 'gamut'. The term gamut refers to the range of colours that the combination of CMYK inks print. Some colours may be 'out of gamut' and therefore can't be printed accurately. This can be resolved with certain programmes, such as PhotoShop, that provide an 'out of gamut' warning, and give you options to replace a colour with one that is in the gamut. It is best to work in RGB and then keep a copy as RGB and convert another copy to CMYK if needed for printing purposes.

## Additional Information

The main sources of information referred to in compiling the handbook were:

- Creating Digital Resources for the Visual Arts: Standards and Good Practice  
< <http://ahds.ac.uk/visualarts/creating/guides/index.htm> > Last checked 04/04/2005
- Technical Advisory Service for Images (TASI) website  
< <http://www.tasi.ac.uk/> > Last checked 04/04/2005



## Technical Environment

The tables below list the suitable AHDS deposit formats. These are defined according to the criteria below.

### Preferred Deposit Formats

Preferred deposit formats include formats that the AHDS recommend as best practice, our preferred preservation formats and popular formats for which there is good software support (especially export options) and we can successfully preserve the identified significant properties. Cost and likely future support for necessary software should also be taken into account.

### Acceptable Deposit Formats

Formats that the AHDS can *probably* successfully preserve given our current software and skills.

### Problematic Deposit Formats

Any formats that will be *very difficult* to ingest and preserve either, a) due to expense of, or difficulty of obtaining, appropriate software, b) requirements for highly specialised knowledge that the AHDS does not have in-house and cannot contract, or c) over reliance on software or hardware specific functionality that cannot be replicated in different technical environments.

## Common Formats

Format	File Extension	Notes
Uncompressed Baseline Tagged Image File Format v.6	.tif	Preferred Deposit Format This is the De facto standard and is widely used. Creators should be aware that there are various types of TIFF available, including compressed versions that are not suitable as preservation formats. The uncompressed baseline v.6 is the only suitable TIFF for preservation purposes. Also note that the compression used by the TIFF (LZW) is based on proprietary software and is unsuitable for long term preservation.
Portable Network Graphics	.png	Preferred Deposit Format The Portable Network Graphics file format (or PNG, pronounced 'ping') is another alternative. It is supported by the World Wide Web Consortium (W3C) and is expected to be released as ISO/IEC International Standard 15948. PNG is still not widely used and it has taken some time for Web browsers and image



JPEG 2000	.jp2	<p>application software to support it. At time of writing (Summer 2005) PNG is now supported by Adobe Photoshop v.7 and above and by most later Web browsers.</p>
		<p>Problematic Deposit Format</p> <p>JPEG 2000 has been an international standard since December 2000. However, there has been a relatively slow take up of the format in the industry and support remains patchy. A plug-in is usually required to view JPEG 2000 images. At time of writing we would be wary of accepting JPEG 2000 as an archival format; however, the likelihood is that this will change in the near future.</p>
Joint Photographic Experts Group	.jpg	<p>Problematic Deposit Format</p> <p>This is a popular file format for Web publishing and ideally suited to images where the picture is a photograph. However because JPEG uses “lossy” compression, which means that some of the data that makes up the image is lost in the compression process, it is not suitable for archival purposes.</p>
Graphic Interchange File Format	.gif	<p>Problematic Deposit Format</p> <p>GIF is not recommended as a preservation file format for digitised photographs, slides or other forms of complex images. However, GIF is a popular file format for Web graphics and publishing, and is especially useful for images where the picture is line art, has large areas of solid colour or uses a limited colour palette. GIF uses “lossless compression” (LZW), so there is no image degradation or blotchy colour.</p>
PhotoShop Document	.psd	<p>Acceptable Deposit Format</p> <p>When you are working in Adobe Photoshop it is possible to save the image as a PhotoShop Document. This file format can also be</p>



MrSid	.sid	<p>converted into other file formats within Adobe PhotoShop.</p> <p>Problematic Deposit Format</p> <p>MrSid is a proprietary format intended for the storage and delivery of very large images, particularly geo-spatial images. At time of writing it is not suitable as an archival format but since it uses the same wavelet compression as JPEG 2000 it can potentially deliver the same functionality and may be supported in the future.</p>
Bitmap	.bmp	<p>Problematic Deposit Format</p> <p>As this is a proprietary standard graphics file format for Windows, Bitmap is not a preferred file format for long term preservation.</p>
DjVu	.djvu	<p>Problematic Deposit Format</p> <p>DjVu is useful for compressing documents with a combination of text and images, rather than just individual images. It is unsuitable for deposit because it is a proprietary format and overall the compression is very “lossy”, which therefore involves degradation of the image.</p>
PixellLive/VFZoom	.pfz or .vfzm	<p>Problematic Deposit Format</p> <p>PixellLive is used to encode types of images typically stored as rasters, although the encoding used is vector-based. VFZoom is an earlier version of Celartem’s PixellLive format. Both formats use Genuine Fractals, which break the image into small shapes (fractals), which are described mathematically and can be redrawn at a larger scale. Whilst they are effective as scaling and display formats they are not suitable for archival storage.</p>
CorelDraw	.cdr or .cvt or .cdx	<p>Problematic Deposit Format</p> <p>This vector graphics package can hold both raster and vector data. It is used by the Ordnance Survey and is particularly good for handling</p>



Digital Negative .dng

text. However because it is a proprietary format it is not suitable for deposit.

#### Problematic Deposit Format

At time of writing this is a format which is receiving increasing interest. Adobe has been looking at the possibility of standardising the 'RAW' format used by most one-shot digital cameras, creating, in effect, a 'Digital Negative'. Adobe is marketing this as "a new unified public format for raw digital camera files". The DNG format is fully able to read all image tagging (EXIF & IPTC) from the original RAW file and transfer it to the new DNG image. Adobe have also released a free software utility the Adobe DNG Converter, which will convert many proprietary RAW format images into the new .DNG file format, compliant with the Digital Negative Specification.

The main problem with using any RAW format for preservation has been the proprietary nature of each manufacturer's RAW images. Therefore a cross industry supported 'RAW' standard would immediately make this an acceptable preservation file format. However at the moment the use of the DNG format is dependant on camera manufacturers' support and until this format is established Adobe recommend that you still archive the original camera-specific raw file alongside the DNG file.

---

## Additional Information

- Technical Advisory Service for Images (TASI) website provides a fuller list of proprietary file formats  
< <http://www.tasi.ac.uk> > Last checked 11/07/2005
- Adobe website with more information on the Digital Negative format  
< <http://www.adobe.com/products/dng/main.html> > Last checked 11/07/2005



## Ingest Checklist

The following is a list of actions that should be taken during ingest.

### Level 1 (Essential)

A basic set of technical metadata is required for every image deposited. This should include at a minimum a unique identifier in the form of an image file name.

- FileName (unique identifier)

### Level 2 (Preferred)

A preferred basic set of metadata in addition to the image file name (unique id) includes:

- Creator
- DateOfCreation
- RightsOwner
- FileType (archival master)
- FileFormat (e.g. TIFF)
- FileFormatVersion
- FileSize (in bytes)
- SpatialResolution (measured in pixels per inch (ppi))
- Dimensions (measured in pixels)
- ColourSpace
- BitDepth
- CaptureDevice
- CaptureSoftware

### Level 3 (Best Practice)

Use the following guidelines no matter what storage format is used:

- Never use re-writeable media
- Don't burn at top speeds
- Do make two copies
- Do buy known brands for storage media
- Do protect the discs using covers
- Avoid buying cheaper spindle (cover-less) packs of CDs
- Store CDs at an even temperature
- Do use the verify function on your writer every time.
- Don't 'over-burn' or attempt to put too much information on the media
- Don't do something else on the machine while you are burning data to disc

### Inform Depositor

That we cannot guarantee the preservation of image files which are in a problematic file formats. That they should contact us as early as possible to discuss deposit.

# Preservation

## Significant Characteristics

For preservation purposes, the significant characteristics of digital still images are: the true resolution, which is the actual resolution without the use of interpolation or other methods; and the file format. To ensure preservation the file format should be a “preferred deposit format”. The resolution used depends on what is ‘fit for purpose’, but the project should be clear about its requirements and this will give an idea of how high the resolution should be e.g. whether in-depth analysis of the images is needed, or if they are just to provide a basic record of an object.

## Technique

The long term preservation of raster images is best planned at the outset of digitisation, at the point of creation. This means that the photographer or scanning technician should create an archival master version of each image produced at time of capture. This ‘master’ copy should be the version from which the various surrogates (e.g. optimised for delivery over the web, saleable copies for printing etc) are derived.

There are two main methods for working with archival, or ‘Master’, images:

### Master Method 1

The master archival image should be as close to the scanned or camera ‘RAW’ file as possible. It should have no post production optimisation carried out upon it. Examples of optimisation would be things like: un-sharp masking, cropping or colour correction. The master archival version should be preserved in its unaltered state and any future work should be carried out on copies of this master.

### Master Method 2

The downside of Method 1 is that projects will often also have the need to preserve an optimised (or usable) version of the master as well. In effect preserve two copies of every image produced. This can significantly add to the cost of production, time taken and storage space required. In practical terms the project can, therefore, forego the concept of two archival versions and keep one, the uncompressed optimised image that is intended for re-use, as the sole archival copy (Master Method 1). This is still an acceptable approach to preservation, and much will depend on individual project requirements.

## Validation of Exported Data

To ensure that the digital object has been preserved correctly:

Set in place a Quality Assurance procedure, whereby a selection of images and their metadata are checked regularly, this will hopefully highlight any errors sooner rather than later.

Also Quality Assure a sample of your storage media regularly. Ideally you should be looking to open and check every CD that is produced to ensure that no degradation of the medium has occurred and image files have not corrupted.

## Problems and Issues

Essentially there are two ways of storing image metadata. Traditionally it has been stored in a database separate from the image itself. The Visual Resources Association Core Categories v.3.0 provides a good metadata schema for cataloguing not only the digital image, but its surrogate analogue (e.g. slide) and original analogue (e.g. chair) forms as well. This is fine when the image



remains with its data in, say, an online catalogue. But it does present problems when the image is downloaded and removed from its original setting.

The second method, which overcomes this issue, is to store information embedded in the image file header itself. The two common standard formats used in the industry for embedding file information within an image are: International Press Telecommunications Council (IPTC) and EXchangable Image File (EXIF). Another common format, Photoshop's File Info, is based on a subset of IPTC.

We would not recommend relying on embedded information alone. There remain various issues regarding the use of the common formats and their ability to be read in common image application software. For example older versions of Photoshop (pre v. 7) can read IPTC information via its File Info function, but only v. 7 and later can read EXIF formatted tags.

However the editing of embedded tags is possible in Image Management Systems such as Canto Cumulus (IPTC), Extensis Portfolio (EXIF/IPTC), Thumbsplus (IPTC), FotoStation (IPTC) and iView (EXIF/IPTC). Imatch (EXIF/IPTC) also allows for bulk editing of information and has the ability to add new user defined fields.

Also it is worth keeping in mind the relatively short lifespan of whatever medium you choose to store your images. The diagram (see Additional Information) shows in detail the preferred medium of the day, over the past 20 years or so. Sometimes the most basic things can make a difference, for example we recommend using CD marker pens for writing on CDs, but these things can be easily

## Additional Information

- Diagram 1: Timeline illustrating the changes in common storage media from the TASI website.  
< <http://www.tasi.ac.uk/advice/delivering/digpres.html> > Last checked 04/04/2005
- Image Tagging: Embedding Metadata in Image Files  
< <http://www.tasi.ac.uk/advice/delivering/facttagging.html> > Last checked 04/04/2005