



RASTER IMAGES PROCEDURES (VERSION 1.120)

DIGITAL ARHIVISTS
ARCHAEOLOGY DATA SERVICE
<https://archaeologydataservice.ac.uk/>

| | |
|-------------------------|--|
| Created date: | 26 January 2012 |
| Last updated: | 18 December 2019 |
| Review Due: | 31 March 2021 |
| Authors: | Stewart Waller, Jen Mitcham, Jon Bateman, Michael Charno, Tim Evans, Kieron Niven, Ray Moore, Jenny O'Brien, Paul Young, Teagan Zoldoske, Digital Archivists |
| Maintained by: | Digital Archivists |
| Required Action: | |
| Status: | Live |
| Location: | https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml |

1. Purpose of this document

1.0.1 This documents current ADS procedures for production of dissemination and preservation copies of raster images. It contains a list of current dissemination and preservation formats and how to migrate files to required formats. More information on this data type, can be found in the G2GP for Raster Images http://guides.archaeologydataservice.ac.uk/g2gp/RasterImg_Toc .

2. Formats¹

2.0.1 Please note for geo-referenced rasters please consult the Data Procedures for GIS²

| Offered format | Accepted | Preservation | Presentation | Notes |
|---|----------|--|--|--|
| Uncompress ed Baseline TIFF v.6 .tif | YES | Uncompressed Baseline TIFF v.6 .tif | Portable Network Graphics .png (lossless compression @ c.5) or Joint Photographic Expert Group .jpg .jpeg ¹ | All tiffs should be assessed for type at ingest for non-standard features e.g. geo, multipage or pyramidal content. If EXIF & IPTC metadata is significant then this should be preserved during conversion, or if regarded as significant by the archivist and following negotiation with the depositor. |
| Portable Network Graphics .png | YES | Uncompressed Baseline TIFF v.6 .tif | Portable Network Graphics .png | As above. |
| Joint Photographic | YES | Uncompressed Baseline TIFF v.6 .tif | Joint Photographic | As above. |

¹ Generally, we can take most raster formats. Files are preserved as uncompressed tif and disseminated as png or jpg. Choice of jpg or png for dissemination has to be decided by the Digital Archivist - according to the G2GP "PNG has become the format of choice for lossless presentation... While the format offers certain advantages over JPG (e.g. less visible artefacts) it is not recommended for use with digital photographs." If a file arrives as a jpg (or jpeg2000) disseminate in that original format. See also notes on preserving embedded metadata below.

² Internal access only.

Raster Images Procedures (Version 1.120)

| Expert Group .jpg .jpeg | | | Expert Group .jpg .jpeg | |
|--|---|---|---|--|
| Graphics Interchange Format (Compuserve) .gif | YES | Uncompressed Baseline TIFF v.6 .tif | Portable Network Graphics .png | |
| Bit-Mapped Graphics Format (Microsoft) .bmp | YES | Uncompressed Baseline TIFF v.6 .tif | Portable Network Graphics .png | |
| PhotoCD .pcd | NO | | | |
| Photoshop (Adobe) .psd | NO | | | |
| CorelPaint .cpt | NO | | | |
| Adobe Digital Negative .dng | YES | Adobe Digital Negative .dng and Uncompressed Baseline TIFF v.6 .tif | Adobe Digital Negative .dng and Joint Photographic Expert Group .jpg .jpeg | Check input and output file sizes match. XnView seems to shrink the output in batch mode, use ImageMagick. |
| JPEG2000 .jp2 / .jpx | YES | Uncompressed Baseline TIFF v.6 .tif | JPEG2000 .jp2 / .jpx | |
| Portable Document Format .pdf | YES, but only if not available in original format | Uncompressed Baseline TIFF v.6 .tif | Joint Photographic Expert Group .jpg .jpeg or Portable Document Format .pdf | For where drawings, photos, scans etc. have been deposited as PDF files but would have a data type of image. Ideally these should be deposited in their original format (due to the loss of information resulting from the |

| | | | | |
|-----|----|--|--|---|
| | | | | creation of the PDF. Preservation format is at the discretion of the archivist. |
| RAW | NO | | | |

3. Documentation / Metadata

3.0.1 Alongside the standard metadata for files, the following additional documentation is required for any raster images. The current metadata template is available from the Guidelines for Depositors.³

3.0.2 Currently the no additional metadata is requested for raster images.

| Element | Description |
|---------------------------------|--------------------------------------|
| | |
| Supporting Documentation | |
| Copyright/GDPR Clearances | These are very important for movies. |

3.0.2 This table is derived from the G2GP

https://guides.archaeologydataservice.ac.uk/q2gp/RasterImg_Toc.

3.1 Associated metadata

3.1.1 It is important that any copyright information/permissions are stored alongside the requisite file in a suitable preservation format.

3.2 Embedded metadata⁴

3.2.1 In an ideal world we should ask for such metadata to be supplied separately as XML or TXT. If we do receive this then this data should be preserved and disseminated with the relevant files (see below for notes on storage). However, it is most likely that we will receive digital images - either taken with a digital camera or scanned - with embedded metadata but not in an additional format.³ Work on the G2GP has shown that embedded metadata is often created automatically (usually technical information and things like dates). So, if it is noted that a file has embedded metadata, that is not also supplied as XML/TXT, a decision should

³ <https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml>.

⁴ From the Guides to Good Practice: "Embedded metadata can also be seen in certain cases as a significant property of an image and, where relevant should be preserved with the file or exported to a separate plain or delimited text or XML file to be stored alongside the image. Although it is possible to preserve JPEG EXIF within the TIFF tag structure it is better held in a separate file, avoiding the risk of loss or corruption during later migration and making the metadata more easily accessible. Extraction of EXIF fields is relatively straight forward, with a number of free tools available."

Raster Images Procedures (Version 1.120)

be made about the value of this information. Generally, if embedded metadata is seen to be valuable, it should be preserved (see notes on storage below). As noted in the Guides, there are numerous free online tools to extract this information, such as <http://www.artstor.org/global/g-html/download-emet-public.html>.

3.2.2 When disseminating files, it is best practice to try and include the metadata in the dissemination format, thus allowing for more advanced reuse of the file. JPEG keeps this information, PNG generally does not. Therefore choose your dissemination strategy appropriately, disseminated alongside the relevant files (see below for notes on storage).

3.2.3 Information on extracting this metadata can be seen below.

4. Accessioning checks

4.1 Checks

- Do we have the necessary documentation (see below)
- Images are suitable for deposition i.e. they do not contain inappropriate content (e.g. children, people, etc). Copyright permissions should be assumed, but if questionable then this should be checked with the depositor.
- TIFF files are uncompressed.
- Embedded metadata (e.g. EXIF, and other types) in file. See section below, if regarded as significant by the archivist and following negotiation with the depositor.
- It is quite common for raster content to be deposited as a single PDF file (or multiple images in a PDF file). In this case we should ask for the original (i.e. non-PDF file. If this is not possible, then we will have to proceed on a "best-efforts" basis
- Geo-rectified raster images - guidance on these is provided in the GIS procedures

4.2 Significant properties

From the G2GP: "The significant properties of raster images are discussed in detail in the InSPECT Significant Properties Testing Report on Raster Images:

- Image Size and Resolution - conversions should ensure that the original resolution and image size remains the same in the preservation file format. In addition it is important that, when converting files to a new format, lossy compression is not applied to the image.
- Bit depth and Colour space - converted files should ensure that the bit depth and colour space of the original image are supported in preservation formats and that images are not degraded when converted.
- Embedded metadata such as EXIF and IPTC data (see below)

4.3 File-naming

4.3.1 Where possible files should retain the same name as the original. On occasion (and normally for dissemination), it may be necessary to create different versions of the same file. In these cases a logical naming strategy should be used, and should be accompanied by explanation in the Processes section of the CMS.

4.3.2 Extracted metadata should also be named consistently, for example

`original_name_exif_meta.csv`

4.3.3 All files and metadata should be placed in the appropriate location as outlined below.

5 How to convert files

| Starting Format | Procedure | End Format | Checks |
|--|---|--|---|
| Joint Photographic Expert Group .jpg .jpeg | XnView - https://www.xnview.com/en/apps/ | Uncompressed Baseline TIFF v.6 .tif | <ul style="list-style-type: none"> • Bit depth and colour space of the original image are retained • Original resolution and image size remains the same - images are not degraded • Embedded metadata retained (if appropriate) |
| Portable Network Graphics .png or Joint Photographic Expert Group .jpg .jpeg | XnView - https://www.xnview.com/en/apps/ | Portable Network Graphics .png (thumb) or Joint Photographic Expert Group .jpg .jpeg (thumb) | <ul style="list-style-type: none"> • should be a maximum width OR maximum height of 125px - so that landscape and portrait images have the same surface area⁵ |
| Portable Network Graphics .png | XnView - https://www.xnview.com/en/apps/ | Portable Network Graphics .png (preview) | <ul style="list-style-type: none"> • should be a maximum width OR maximum height of 750px - so that landscape and |

⁵ All images for download should have a thumbnail (for raster images, these should be jpgs). Thumbnails should be renamed with the prefix "thumb" and store in the web folder under /images/thumbs. Please note, these images are not part of the AIP and should never be stored

Raster Images Procedures (Version 1.120)

| | | | |
|--|--|--|---|
| <p>or</p> <p>Joint Photographic Expert Group .jpg .jpeg</p> | | <p>or</p> <p>Joint Photographic Expert Group .jpg .jpeg (preview)</p> | <p>portrait images have the same surface area⁶</p> |
| <p>Tagged Image File Format .tif (multi-page)</p> | <p>XnConvert - https://www.xnview.com/en/xnconvert/</p> <ol style="list-style-type: none"> 1) In the input tab, add the files you need to convert 2) Ignore the actions tab and go to output 3) Select 'rename' from the drop-down list next to 'where output files already exist' 4) Tick 'convert all pages from multipage file' 5) Tick 'preserve metadata and colour profile etc.' 6) Click 'Convert' | <p>Uncompressed Baseline TIFF v.6 .tif (individual images)</p> | |
| <p>Joint Photographic Expert Group .jpg .jpeg</p> | <p>Extracting metadata using Adobe Bridge:</p> <ol style="list-style-type: none"> 1) Download the file BarredRock CSV Extract.jsx attached to this page and save it to the Scripts folder (mine is at Applications>Adobe>Bridge CS3>Startup Scripts) 2) Start Adobe Bridge 3) Navigate to the folder of images you want to extract metadata from, and select all images (so they're highlighted) 4) From the toolbar at the top of the screen select "Scripts>Export Metadata .." 5) You should be presented with a screen that allows you to export metadata by category (e.g. EXIF) and select the fields you want to include: | <p>Comma Seperated Values .csv</p> | |

⁶ In ALL cases, for example for large (bytes and pixels wise) a preview image will need to be created. The interface should present this as a "preview" via highslide, the user can then decide whether to download the full file. These should be stored in the web folder under /images/preview/. Please note, these images are not part of the AIP and should never be stored with the actual data.

Raster Images Procedures (Version 1.120)

| | | | |
|--|---|--|--|
| | <ul style="list-style-type: none"> • Make sure 'Include filename' and 'Include Column names' are selected • If IPTC metadata is present, Select the Schema "IPTC Core" and include all fields that need to be preserved • Click OK, a csv will be created in the same directory as the images - (follow procedures for file naming and storage) • If EXIF metadata is present, Select the Schema "EXIF" and include all fields that need to be preserved • Click OK, a csv will be created in the same directory as the images - (follow procedures for file naming and storage) | | |
|--|---|--|--|

6 Storage

6.1 Storing data

6.1.1 Data should be stored in appropriately named folders, as described in the ADS Repository Operations manual.⁷ Any directory structure from the SIP should be retained in the AIP. In some cases editing/restructuring may be necessary, but such restructuring should be recorded in the Processes section of the CMS.

```

/preservation
  /{original_structure}/
    original_name1.tif
    original_name2.tif

dissemination/
  /{original_structure}/
    original_name1.jpg
    original_name2.jpg

```

6.2 Storing metadata

6.2.1 File and embedded metadata (copyrights, documentation, etc) should be stored in an appropriate archival format with the preservation/dissemination files in a "documentation" folder within the requisite folder. Any metadata extraction should be recorded in the Processes section of the CMS.

```

/preservation/
  /{original_structure}/
    myimage.jpg
  /documentation
    myimage_exif_meta.csv
    ADS_image_metadata.csv

```

6.2.2 For dissemination, embedded metadata can be kept in the files, unless supplied separately by the depositor

```

/dissemination/
  /{original_structure}/
    myimage.jpg
  /documentation
    ADS_image_metadata.csv
    myimage_metadata_permissions.pdf

```

⁷ <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>.

7. Creating and linking objects in the OMS tables

7.0.1 See Match Objects Overview for general overview {internal access only}
 see also CMS-OMS TableStructure for MOS data requirements {internal access only}

8. Tech watch / things to note

| Item | Person | Date |
|---|--------|------|
| Deposit trends | | |
| Future formats | | |
| Future tools to things (links) | | |
| what's not working but can't currently be fixed | | |

9. Archival notes

| Item | Person | Date |
|---|--------|------|
| Not sure we should be accepting multi-page tifs as they are only costed as one file, and generally do not contain sufficient metadata for each individual image | | |

10. References

- Adobe digital negative - <http://www.adobe.com/products/dng/main.html>
- DPC Technonology watch report on JPEG 2000 - <http://www.dpconline.org/knowledge-base/tech-watch-reports>