Document Name			
Document Title	Significant Properties Testing Report: Raster Images		
Work Package			
Author(s) & project role	Lynne Montague, project officer		
Date	12 February 2010	**Filename**	rasterimages-significantproperties-v11
Access	☑ Project and JISC internal		☑ General dissemination

Document History		
Version	**Date**	**Comments**
1.1	12 February 10	Copied text into revised template
1.0	02 April 09	First version of report
0.2	15 March 09	Second draft of Significant properties of raster image testing report
0.1	10 March 09	First draft of Significant properties of raster images testing report

Contributors

The following people have made direct or indirect contribution to this report: Adrian Brown, Tim Gollins, Stephen Grace and Gareth Knight.

Intended Audience

This document is written for use by the InSPECT project team, the JISC community and those interested in digital preservation.

1. Table of Contents

Project Overview

Significant properties are those aspects of a digital record that must be preserved over time in order for the Information Object to remain accessible and meaningful. The InSPECT Project is funded by JISC to investigate methods for maintaining the authenticity of digital resources across digital environments and transformation processes. It has produced a framework for the analysis of significant properties and created a set of reports that outline its application to four object types – audio recordings, raster images, structured text and e-mail – that will contribute and advance strategies for the characterisation and maintenance of significant properties over time.

Purpose of the report

This report examines the notion of significant properties as it applies to raster images, or bitmaps. It seeks to identify the significant properties of raster images that must be maintained by examining each of its constituent elements and analyzing their designated function. It goes on to examine strategies that may be utilised to maintain access to raster image assets in the long-term. Finally, it outlines a set of experiments that were performed by the project team to identify and evaluate tools that may be utilised to convert significant properties from one form to another.

1. Introduction

1.1. Overview of raster images

The three ways of representing digital images are as raster images, vector images or as metafiles, a combination of both. Raster images are the most common type of image found on the Internet[1] and they are used for the creation and storage of many types of image, including photographs, diagrams and scanned images. Vector images are composed of individual, scalable elements such as curves, lines and polygons, which have their own attributes such as colour, which can be individually edited. For the purposes of the InSPECT project, we are considering only raster images.

A raster image is composed of a rectangular array or grid of pixels, each of which represents a colour. Each pixel has one or more numbers, or colour values, associated with it and these numbers define the colour represented by the pixel. The three major components which influence how a raster image is recreated or rendered are spatial resolution, bit depth and colour space.

1.1.1 Spatial resolution

The spatial resolution of a raster image is an indication of the number of pixels that are, or should be, contained in an in it. The resolution is measured in pixels per inch (ppi), dots per inch (dpi) or samples per inch (spi) and specifies the degree of detail that the image will contain[2]. The amount of detail that may be stored in an image is proportionate to the number of pixels. For example, the scanning of an A4 document (9 x 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); the scanning of the same A4 document at 600 ppi will produce a digital image that is 5400 pixels x 7200 pixels. The latter may therefore contain details that are not found in the smaller image. The scanning of a postage stamp (1 inch x 1 inch) at 300ppi will produce a digital image that is 300 pixels x 300 pixels. Although both items are scanned at 300ppi, they produce a different sized digital image. Details of the physical dimension of the image are useful, but not essential as the image may only ever exist as a digital, rather than physical, manifestation. Generally, the higher the number of samples, the better the quality of the image because more spatial and colour information will be captured. However this will also lead to greater file sizes. Where a good master copy of an image is needed, or where it isn't known what the final size or output of the image will be, it is advisable to digitise at a high resolution in order to capture enough information for any subsequent use. Re-sampling down to lower resolutions can be done later if lower quality versions with a smaller file size are needed[3].

---

[1] Eadie, M. (2005) Preservation Handbook: Raster Images. Retrieved on March 8, 2009 from: http://www.ahds.ac.uk/preservation/Bitmap-preservation-handbook_d6.pdf
[2] Whilst the terms ppi, dpi an spi are used interchangeably, they do describe different things. See http://www.jiscdigitalmedia.ac.uk/stillimages/advice/resolving-the-units-of-resolution/
[3] http://www.jiscdigitalmedia.ac.uk/stillimages/advice/the-digital-still-image/

1.1.2 Bit Depth

The bit-depth, or colour resolution refers to the amount of colour information held in relation to each individual pixel. A higher bit depth offers a greater number of available colours. A two-colour image, often black and white, contains just 1-bit; a greyscale image typically contains 8-bits; and a full colour photograph typically contains 24-bits of information, offering 16,777,216 colours. The number of bits in an image has an effect on the file size, which increases significantly for 16- 48 bit images.

1.1.3 Colour Space

The colour space of an image refers to the method of working with colours. It is influenced by the colour model in use – a mathematical formula that allows colours to be represented as tuples of numbers. Several colour models are available, including bitonal, grayscale, indexed colour, RGB and CMYK that are used for different types of images. The bitonal colour space uses two values, black and white; greyscale offers 256 shades between black and white; Indexed colours offer a limited palette of 216 colours which may be displayed on both Macintoshes and Windows PCs in a consistent manner. Computer monitors and televisions use RGB, to create colours as a combination of Red, Green and Blue colour values. It is common for designers to work with RGB and reduce the number of colours to Indexed colour for use on the World Wide Web.

1.2. Overview of Metadata Standards

Many metadata standards relevant to the storage and preservation of raster images have been developed and it is not intended that this document provide a comprehensive look at all relevant ones.[4] Standards range from the simple generic schemas, such as Dublin Core, to highly complex, granular schemas such as the draft NISO Z39.87 standard: Data Dictionary – Technical Metadata for Digital Still Images.

Dublin Core[5] is a basic standard, which has 15 core elements, all of which are optional and repeatable. It is not image specific, being applicable to all types of digital object, but the majority of terms are suitable for use with raster images. It is a widely adopted standard but its simplicity and scope can be a disadvantage. As stated on the Dublin Core website, 'in the diverse world of the Internet, Dublin Core can be seen as a "metadata pidgin for digital tourists": easily grasped, but not necessarily up to the task of expressing complex relationships or concepts'.

PREMIS[6] is another metadata standard which applies to all types of digital object, not just images, but which augments the resource discovery metadata specified in Dublin Core with management and technical metadata under the entities Objects, Events, Agents and Rights. In March 2008 PREMIS 2.0 was released which implemented a number of changes from the previous version such as expanded rights metadata and, most notably for this project, allowing for the extensibility and increased granular structure of significant properties metadata. The aim behind this was to provide a more flexible structure within which significant properties could be defined and described.

NISO Z39.87[7] is a data dictionary specifically used to describe the elements of raster images. Originally used for object stored as TIFFs it has subsequently become applicable to all raster image formats. It is a very extensive and granular standard, with close to 200 individual elements described within it. The technical and management metadata contained within the standard can be used to supplement the more generic digital object metadata found in the PREMIS model. The data within NISO Z39.87 can be represented in the MIX XML schema[8] for ease of interchange and storage.

This standard is seen as the most wide-ranging, raster- image-specific metadata standard and therefore was used as the basis for the project team's analysis of the significant properties of raster images (see section 2.1.2 below).

---

[4] For a more comprehensive review of current standards, please see Digital Images Archiving Study :
http://www.ahds.ac.uk/about/projects/archiving-studies/digital-images-archiving-study.pdf
[5] http://dublincore.org/documents/usageguide/
[6] http://www.oclc.org/research/projects/pmwg/premis-final.pdf (PREMIS 1.0)
http://www.loc.gov/standards/premis/v2/premis-2-0.pdf (PREMIS 2.0)
[7] http://www.niso.org/kfile_download?pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-
24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVa86hjP37r_hM9t9qad1BrrORLqssvegis%3D
[8] http://www.loc.gov/standards/mix/

1.3.   Application of the Performance model

To determine the significant properties of a digital Record, a consistent, formal method of identifying the important aspects is required. The National Archives of Australia (2002) has developed a 'Performance Model'[9], which has been adopted by the InSPECT Project. The Performance model establishes the concept of the 'essence' of a digital record that contains the "characteristics that must be preserved for the record to maintain its meaning over time". The principle of the model is that the process of rendering the Information Object in a form that can be understood by a user requires some interaction between the underlying data object and interpretative software. The model is comprised of three components:

1. Source: the encoded data object that contains the text, still images, moving images, or other content for interpretation;
2. Process: the method in which the encoded data is interpreted, e.g. a software tool, an algorithm;
3. Performance: the recreation of the Information Object in a form that can be understood by the user.

A key concept in the Performance model is the recognition that the method in which the Source is processed will vary between members of a Designated Community[10] and is likely to change over time as a result of the evolving technological environment. The consistency of the visual recreation of the image is vital in order to maintain it's authenticity and therefore, establishing what the essence of a performance is that must be maintained is essential i.e. what elements of the raster image must be retained for the image itself to remain understandable and to prevent it becoming unintelligible?

The Reference Model for an Open Archival Information System (OAIS)[11], introduced the concept of representation information i.e.

'The information that maps a Data Object into more meaningful concepts. An example is the ASCII definition that describes how a sequence of bits (i.e., a Data Object) is mapped into a symbol.'

It is important at this stage to clarify the difference between the concepts of representation information and significant properties. To apply the performance model, the representation information is involved at the process stage in interpreting the source data object and rendering it as an information object or performance. The significant properties are the characteristics or essence of this information object or performance that need to be preserved over time, regardless of technological changes, to maintain its meaning. It is these significant properties that we are assessing in this project rather than the representation information used to interpret them.

---

[9] http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm2-888.pdf

[10] A Designated Community is an identified group of potential Consumers who should be able to understand a particular set of information.
http://public.ccsds.org/publications/archive/650x0b1.pdf

[11] http://public.ccsds.org/publications/archive/650x0b1.pdf

Source   Process   Performance

Digital Photo   Image Processing Software   Printed Image

Scanned
Raster Image   Display of Digital
                                                  Image
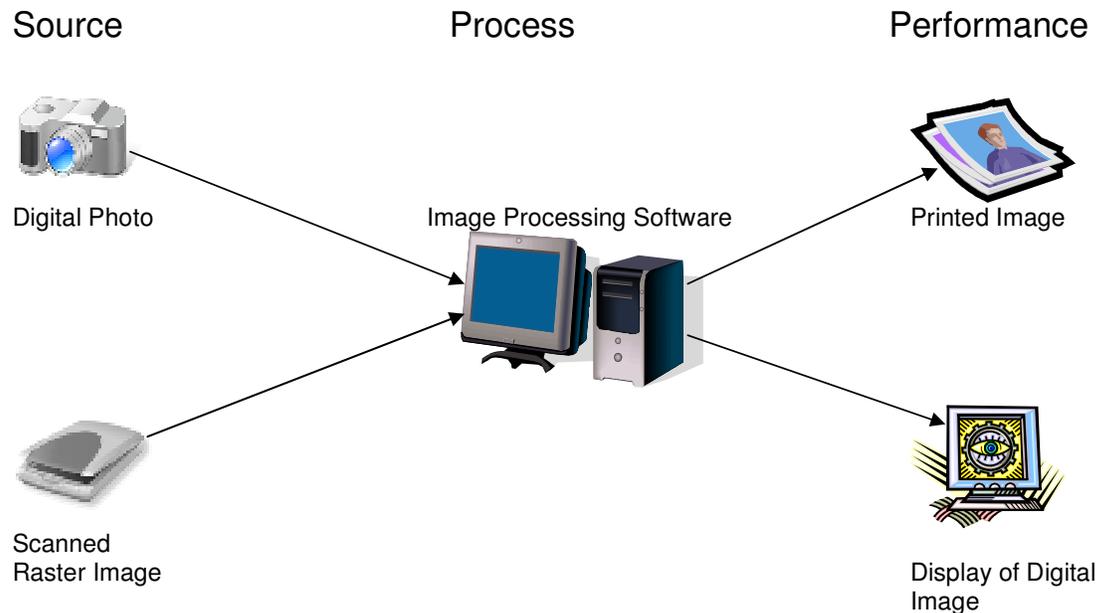
Figure 1. Application of the Performance Model to raster images

2. Testing requirements

2.1. Significant properties that must be maintained

2.1.1 Introduction

The identification of properties of a digital object that are worthy of preservation is not a simple task that can be analysed based upon a set of universal rules. A set of rules defined for one category of digital object may prove to be too restrictive when applied to unusual variations, or inappropriate for other object types. Instead, the InSPECT Project team has developed a methodology to identify factors that establish the authenticity and integrity of the Information Object through a combined technical and epistemological approach.

During the process of investigating the creation, storage and use of digital objects it was found that the classification of significant properties was influenced by four key elements:

1. The form that the creator has chosen to express an intellectual or artistic idea and the method that they have used to communicate information
2. The function for which the digital object has been created to perform or the aims and objectives that its use will achieve.
3. The method in which information is encoded and stored in a digital environment, influenced by the encoding format and data standards in use.
4. The interpretation of the audience – the intended recipient of the digital object or an unknown future user – that is accessing the information to achieve an objective.

The challenge for the curator or archivist is to identify the characteristics of a digital object that enable them to fulfil the required function of maintaining the authenticity and integrity of that object throughout the preservation process. It is possible that that person will be able to answer some, but not all, of the questions needed to be asked. For example what information did the creator of a raster image intend to communicate and who was the intended viewer of the image? A raster image may have been created as an original piece of artwork or a photograph in which elements such as the accurate rendering of colour is vitally important to the creator. Or it may be a simple diagram indicating how to construct a piece of furniture where the rendering of colour is not important but the clarity of the image is in order to understand the instructions being conveyed in the diagram.

An image that is the target of analysis is unlikely to contain all the necessary information to answer these questions, unless extensive metadata is received with it.

2.1.2 Assessment of Significant Properties

To develop a list of the properties that may be significant for establishing the authenticity and integrity of a raster image, the evaluator reviewed several specifications and standards that are widely used for the storage and description of raster images. As previously mentioned, the assessment of the significant properties of raster images in this document is based primarily on the analysis of the *ANSI/NISO Z39.87 Data Dictionary – Technical Metadata for Digital Still Images* as it is felt that this is the most comprehensive and granular standard for digital images that has been developed to date.

Within the NISO Z39.87 standard, data is organised into containers, sub-containers and data elements. The containers and sub-containers are groupings of two or more additional sub-containers or data elements that contain related information. The data elements are the lowest level of data within a container or sub-container.
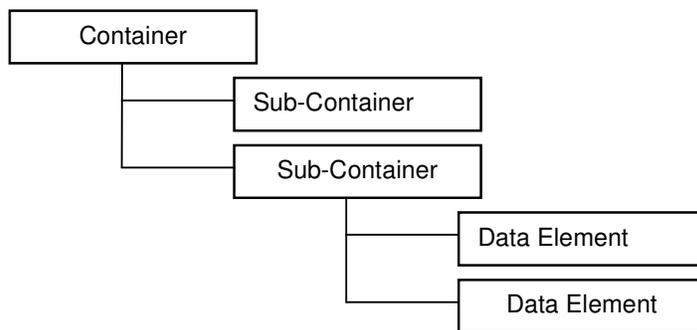
Figure 2. The organisation of data within the NISO Z39.87 standard
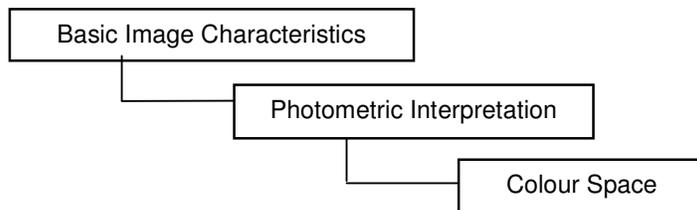
Figure 3. An example from section 7.1 of NISO Z39.87

The first part of this exercise was to assess which metadata elements from NISO Z39.87 were regarded as significant properties of raster images within the scope of the Inspect project. This was first assessed by looking at the highest, least granular, level of specification within the NISO Z39.87 standard e.g. in the example above, this would be the Basic Image Characteristics container. If, on an initial assessment, it was not felt that this high level specification related to a significant property or properties of raster images, for the purposed of the project, it was disregarded, along with any sub-containers or data elements contained within it. If however, it was deemed that this highest level of specification described a significant property or group of properties, then any sub-containers or data elements within it were also assessed.

However, as stated in the Inspect *Framework for the definition of significant properties* document[12],

'To distinguish the properties that are essential from those that are superfluous, an assessor must have a defined understanding of the function performed by each property and its contribution to the whole.'

---

[12] http://www.significantproperties.org.uk/documents/wp33-propertiesreport-v1.pdf

Author: Montague, L.   Date & Time: 16/03/2009

The NISO Z39.87 standard has nearly 200 containers, sub-containers and data elements. These range from easily understandable concepts such as image width, to very technically complex elements such as those relating to colour science. A high degree of photographic expertise is needed to understand the function performed by many of these elements and therefore defining which properties are significant and then assessing their level of significance, as non-expert, has proven very difficult for the project team.

2.1.2.1 Parameters of project

For the purposes of the InSPECT project we are considering raster images at the highest level i.e. we are considering properties which apply to all raster images rather than format- or technology-specific properties or properties that apply to specific sub-classes of raster images. However, beyond this, the parameters for the assessment of significance within the project was less defined and therefore there was an additional challenge of deciding which of the 200 categories of metadata within the Z39.87 standard should be considered as significant properties. See below under the specific metadata elements for a discussion of this where relevant e.g. under colour profiles.

2.1.2.2 Section 6 of NISO Z39.87 – Basic Digital Object information

This section considers general metadata for preservation such as format, file size and compression, which applies to all digital files, rather than technical, raster- image-specific metadata and therefore this metadata is not within the parameters of this project.

2.1.2.3 Section 7 of NISO Z39.87 – Basic Image Information

As described in the standard itself,

'The items in this section are fundamental to the reconstruction of the digital object as a viewable image on electronically interfaced displays. The standard makes no presumption about the rendered or spatial accuracy of the displayed image, only that a reasonably appearing image can be reconstructed using these elements.'[13]

Property	Description	NISO ref.	Significant?
Image Width Image Height Sampling Frequency Unit Bits per sample Samples per pixel Extra samples	The horizontal width of a digital image, in pixels.	7.1.1	Yes. Property used in the reconstruction of the image.
Image Width Image Height Sampling Frequency Unit Bits per sample Samples per pixel Extra samples	The vertical height of a digital image, in pixels.	7.1.2	Yes. Property used in the reconstruction of the image.
Colour Space	The designated colour space of the decompressed image data e.g. RGB, CMYK, YCbCr	7.1.3.1	No. The property is used by the rendering device to set the tone and colour of the reconstructed image. However, this colour space doesn't have to remain the same between migrations. Colours can be rendered

---

[13] http://www.niso.org/kfile_download?pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVa86hjP37r_hM9t9qad1BrrORLqssvegis%3D

			accurately by different colour spaces.
Colour Profiles	The colour profile associated with the digital image. This maps colour values between input, display and output devices and a device-independent colour space, with the aim that accurate and consistent rendering of the colour of the raster images should result, whatever the media.	7.1.3.2	No. It logically follows the same reasoning as above for colour spaces i.e. the colour space, and therefore the colour profile doesn't have to remain the same between migrations and therefore is not a significant property.
YCbCr	YCbCr is a specific colour space.	7.1.3.3	No. This is information related to a specific colour space. For the reasons set out above under colour space, this would not be regarded as a significant property.
Reference Black White	This is encoded headroom and footroom image data for each pixel which describes image colorimetry information for an image when the YCbCr or RGB colour spaces are used.	7.1.3.4	No. This is information related only to specific colour spaces. For the reasons set out above under Colour Space, this would not be regarded as a significant property.
Special Format Characteristics		7.2	No This format-specific and therefore outside the scope of the project.

2.1.2.4 Section 8 of NISO Z39.87 – Image Capture Metadata

This contains 'descriptive technical metadata or administrative metadata. Some of the information may be harvested from the file itself while other information will need to be provided by the institution managing the image capture process.'

'This metadata block documents selected, irreversible attributes of the analogue-to-digital conversion process that may be used for future quality assessment of the image data. By definition, image capture occurs only once. While it provides no quantitative information, per se, it can provide critical information with respect to the logistics and administrative conditions surrounding digital image data capture.'[14]

The information in this section applies only to specific types of raster images i.e. those that have undergone analogue to digital conversion, and does not include born-digital raster images. Therefore, whilst much of the metadata in this section can be seen as significant when applicable, for example information about the analogue source object, general capture information such as the date the image was created, and information about the orientation of the image when it was captured, the information does not apply to all types of raster images and is therefore outside the scope of the InSPECT project.

---

[14] http://www.niso.org/kfile_download?pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVa86hjP37r_hM9t9qad1BrrORLqssvegis%3D

Property	Description	Niso ref.	Significant?
Source Information	Information about the analog source object e.g.: Source Type Source X Dimension Source Y Dimension Source Z Unit	8.1	Yes in specific contexts but does not apply to all rasters and therefore is outside the scope of the InSPECT project.
General Capture information	This contains three data elements: Date Time Created Image Producer Capture Device	8.2	Yes in specific contexts but does not apply to all rasters and therefore is outside the scope of the InSPECT project.
Scanner Capture	This contains information about any scanner and scanner capture settings used e.g.: Scanner Model Scanner Model Number Maximum Optical Resolution Scanning System Software (S-C)	8.3	Yes in specific contexts but does not apply to all rasters and therefore is outside the scope of the InSPECT project.
Digital Camera Capture	This contains information about any camera, camera capture settings and GPS information recorded. It contains 65 sub-containers or data elements e.g.: Digital Camera Manufacturer Exposure Time Flash Focal Length Gps Latitude Gps Processing Method	8.4	Yes in specific contexts but does not apply to all rasters and therefore is outside the scope of the InSPECT project.
Orientation	This specifies the orientation of the image in relation to the placement of its rows and columns when it was saved to disk i.e. it denotes whether the image has been rotated or flipped.	8.5	Yes in specific contexts but does not apply to all rasters and therefore is outside the scope of the InSPECT project.
Methodology	Specifies the methodology used in digitising the image.	8.6	Yes in specific contexts but does not apply to all rasters and therefore is outside the scope of the InSPECT project.

2.1.2.5 Section 9 of NISO Z39.87 – Image Capture Metadata

'The operative principle in this section is to maintain the attributes of the image inherent to its quality. The title image assessment has both a present and future context: these elements serve as metrics to assess the accuracy of output (today's use) and of preservation techniques, particularly migration (future use).'[15]

---

[15] http://www.niso.org/kfile_download?pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVa86hjP37r_hM9t9qad1BrrORLqssvegis%3D

Property	Description	Niso ref.	Significant?
Sampling frequency plane	This specifies the reference plane location for which the x sampling frequency and y sampling frequency are designated.	9.1.1	Yes in specific contexts but does not apply to all raster images, as this element is used specifically for analogue-digital conversions and therefore is outside the scope of the InSPECT project.
Sampling Frequency Unit	This is the unit of measurement for the two data elements it contains:	9.1.2	Yes

This gives a value for the resolution of the image. |
| X sampling frequency | 1. X sampling frequency - The number of pixels per sampling frequency unit in the image width. | 9.1.2.1 | |
| Y sampling frequency | 2. Y sampling frequency - The number of pixels per sampling frequency unit in the image length. | 9.1.2.2 | |
| Bits per sample | This is the number of bits per component (channel) for each pixel and has two data elements:

Bits per sample value
Bits per sample unit | 9.2.1.

9.2.1.1
9.2.1.2 | Yes

This is a quantative value which can be used to evaluate image quality, tone and colour. |
| Samples per pixel | The number of colour components per pixel | 9.2.2 | Yes.

This is a quantative value which can be used to evaluate image quality, tone and colour. |
| Extra samples | The value of any extra components in each pixel. | 9.2.3 | Yes if present.

This is a quantative value which can be used to evaluate image quality, tone and colour. |
| Colour map (AKA colour table/lookup table LUT) reference | This gives the location of the colour map being used.

A colour map defines values for palletised colours (i.e. for the Palette colour space) so that pixel colour values are defined as an index into the colour map (rather than storing specific colour values for each pixel) . | 9.2.4.1 | No.

This is information related to a specific colour space. For the reasons set out above under colour space, this would not be regarded as a significant property. |
| Embedded Colour map | Where a colour map is embedded into the image itself. | 9.2.4.2 | No

Reasoning as above. |

Gray Response Curve	Specifies the optical density of each possible pixel value in terms of greyscale data It only applies to the grayscale colour space.	9.2.5	No. This is information related to a specific colour space. For the reasons set out above under colour space, this would not be regarded as a significant property.
White Point (rendering)	Specifies white point details [when the CIE 1931 XYZ colour space is being used] and is used to help display an image colour consistently on different devices http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf	9.2.7 9.2.7.1 9.2.7.2	No. This is information related to a specific colour space. For the reasons set out above under colour space, this would not be regarded as a significant property.
Primary Chromaticities (rendering)	This describes the x and y values for the chromaticities of the primary colours of the imaging process when the CIE 1931 XYZ colour space is being used.	9.2.8	No. This is information related to a specific colour space. For the reasons set out above under colour space, this would not be regarded as a significant property.

2.1.2.6 Section 9.3 of NISO Z39.87 - Target Data

These are physical points of reference about the image from the time of capture and are used as a benchmark to correct or reconstruct the source image. They can be internal to the image i.e. within the field of view at the time of capture, or external i.e. captured session to session. They are used to calibrate imaging hardware.

Property and Classification	Description	Niso ref.	Significant?
Target Data	Specifies info about targets used in digitisation.	9.3	Yes in specific contexts but does not apply to all raster images, only those which are digitised, and therefore is outside the scope of the InSPECT project.

2.1.2.7 Section 10 of Z39.87 - Change History

'Change history metadata serves the function of documenting processes applied to image data over the life cycle of an image.' Whilst section 8 is used to 'document the source, scanning system and capture settings used to create an image from an analogue source' in this section they are used to 'document the source, systems, and settings used in all subsequent digital-to-digital operations.'[16]

Property	Description	Niso ref.	Significant?
Image Processing	This contains 9 data elements and 1 sub-container e.g.: Date Time Processed Processing Agency	10.1	Yes in specific contexts but does not apply to all raster images, as these elements are used only for images that are subsequently processed and changed from the original therefore is outside the scope of the InSPECT project.
Previous Image Metadata	Any previous metadata from previous manifestation of the image.	10.2	Yes in specific contexts but does not apply to all raster images, as these elements are used only for images that are subsequently processed and changed from the original therefore is outside the scope of the InSPECT project.

2.1.3   Summary

The significant properties of raster images that need to be maintained, within the scope and definition of the InSPECT project are:

Image Width
Image Height
X Sampling Frequency
Y Sampling Frequency
Bits per sample
Samples per pixel
Extra samples

3.   Methodology

3.1.   Representation Formats

Representation format is a general term that describes the method in which information is stored. In its abstract form, a representation format may be applied to many types of information. Restrictions on the type and extent of information are imposed when handling representation formats intended for a specific purpose. To provide a simple example, a representation format for image data is unlikely to be able to contain audio. Limitations may be imposed, even if information is stored in a representation format of the correct type. Specific properties of the information content may be degraded or removed when it is stored in a representation format.

As a general rule, the representation formats used to store raster images have an introductory header section with a body which follows containing the majority of the image data. Image files can be very large and many formats use compression as a way of encoding files so that they take up less storage space. Compression can be lossless i.e. it does not loose any image quality but the file size can still be relatively large, or lossy, where image information is removed in order to reduce file size to a greater degree. With lossy compression the aim is that the information that is removed is redundant or irrelevant so that it is not discernible to the human eye, but the greater the compression, the more likely the quality will by noticeably compromised.

The ability of the representation formats to store the significant properties of raster images, as defined in this paper, is dependent on the constraints of the particular format and it is therefore possible that information about significant properties may not be transferred when converting from one format to another.

3.1.1   Common representation formats

There are hundreds of representation formats used for the storage of raster images. This section aims to give a brief overview of the most widely used, TIFF, JPEG/JFIF, GIF and PNG.

- GIF (Graphic Interchange Format) can support up to 8 bits per pixel and can utilise up to 256 different colours. It was designed by Compuserve and released in 1987. It is suited to simple images, graphics and logos and can be used for animations but is not recommended for digital photography due to it's limited colours. The current version is 89a.

- PNG (Portable Network Graphics). Due to potential patent problems with the LZW compression used with GIF, PNG was designed as a replacement. Whilst it can be utilised as an 8-bit file format like GIF, using up to 256 colours, there is also a 24-bit full-colour version and a 48-bit version. Whilst it is not regarded as such a good a format for animations as GIF[17], and is not as good at creating high-quality, small images for web use as JPEG[18], it is now widely supported and is a versatile format for creating both 8-bit web quality images and 24-bit lossless archival versions.

- JPEG. Whilst JPEG is commonly used as format name, it is actually a mode of compression used with JFIF (JPEG File Interchange Format). JPEG is a lossy format which provides the best combination of small file sizes and quality. The degree to which the file can be compressed can be chosen when saving the file, the greater the compression the lower the quality of the image. The current version is JPEG 1.02.

- TIFF (Tagged Image File Format) is widely used as the format of choice for high-quality master versions of images for archival purposes and is often seen as the archiving standard. It is a flexible format which is usually saved as 24- or 48-bit versions.

For this project, TIFF, JPEG and GIF were the formats chosen for testing as these are all supported by the Jhove tool which was chosen to do the file characterisation

---

[17] http://en.wikipedia.org/wiki/Graphics_Interchange_Format
[18] http://www.jiscdigitalmedia.ac.uk/stillimages/advice/choosing-a-file-format-for-digital-still-images/

3.2. Software tools

3.2.1 Requirements

The criteria for identification and selection of the software tools needed for this project were based upon those suitable to extract the significant properties and migrate and characterise representation formats identified in the research part of the project. .

General criteria for the selection of software tools were:

1. *Task:* Able to identify some or all properties of an Information Object that are considered to be significant;
2. *Task:* Able to extract significant properties of source format and store them in an open, well documented destination format;
3. *Environment:* Can be compiled or operated on a number of computing operating systems;
4. Distribution: Are publicly available as a full product or in demo form for testing;
5. *Legal:* Provide clear guidance on the licence for use of the software in a production environment. Particular preference given to open source licence models;
6. *Documentation:* Are well documented.

3.2.2 Software tools available

The ability to identify, extract and convert the significant properties of a raster object requires a combination of different software tools. There are many proprietary and non-proprietary tools available to process raster images. However, due to the computer security restraints inherent in working within a government department, the types of product freely or easily downloadable for use are limited and it was necessary, within the time available, to chose products already available to the project team. Therefore the Adobe tool Photoshop CS (version 8) was chosen to undertake the image conversion tasks. For one type of image conversion, GIF to JPEG, Photoshop was not suitable and the Microsoft product, Windows Picture and Fax Viewer was used as again, it was already available.

For the characterisation tasks, Jhove (version 1) was chosen as it is able to characterise and format image files in accordance with the NISO Z39.87 standard that was used to define the significant properties of raster images. Two of the formats chosen as desirable for testing, TIFF and JPEG, are supported by Jhove which was another reason for choosing it. A third format for testing, GIF, was specifically chosen because it is supported by Jhove.

- **Photoshop:** Photoshop is one of the leading image manipulation tools available. Developed by Adobe, it is able to use the colour spaces bitmap, grayscale, duotone, RGB, CMYK and Lab and can read and write raster images in many formats including TIFF, JPEG, GIF, BMP, PNG. For this project it was simply used to convert images from one file format to another. However it was unable to convert from GIF to JPEG. It is a well documented format.

- **Jhove**: Jhove (JSTOR/Harvard Object Validation Environment) is an identification, validation and characterisation tool developed by JSTOR and Harvard University Library. These actions involve being able to identify files of particular specified formats, state whether particular object examples of these formats are well formed and valid, and determine the specific properties of a particular object in a supported format. It has modules to support these actions for **arbitrary byte streams**; **ASCII** and **UTF-8** encoded text; **GIF**, **JPEG2000**, **JPEG** and **TIFF** images; **AIFF** and **WAVE** audio; **PDF**, **HTML**, and **XML**. Output from these modules is available in text and xml formats. It includes both a command line and GUI version, with the latter being used in this project.

- **Windows Picture and Fax Viewer:** Windows Picture and Fax Viewer is a Microsoft product used for image viewing but not editing. It can view the following formats; JPEG, BMP, PNG, GIF, ICO, WMF, EMF and TIFF and images can be saved in these formats. In this project it was used purely to convert GIF images to JPEG.

4. Experiment

4.1. Sample data to be analysed

To demonstrate the identification, extraction and conversion of properties in a production environment the project team obtained data samples from several sources which were used as the basis for analysis. Prior to data selection, it was established that the data should represent real-world examples, i.e. raster images created in a production environment, as opposed to raster images created in a controlled environment for analysis purposes.

It was originally intended that images for testing would be gathered from the TNA's Digital Archive. However the Digital Archive is still relatively small and it was not possible to find sufficient suitable images to build a working set of test data. This was due both to the fact that there were very few images in the required version of the sample formats, and also because the images that were suitable were from the same few sources, were very similar and therefore they were not seen as a broad enough sample. It was established that the test data should include scanned, photographed, graphic, colour and greyscale images.

Learning from this, it was felt a more random, and wider reaching, approach should be taken and the final test data was assembled from a mixture of internet searching using Google Web and Google Images, searching well established internet sites known to contain images e.g. the National Library of Scotland, searching for internally held images on the TNA intranet e.g. such as those used for publishing materials and using personal collections of images.

The process of locating these images proved to be time consuming. After considerable time spent searching for ostensibly suitable material in the right format, each image went through a format identification process in Jhove in order to formally identify the format and to clarify which version of the format the image was in. One notable problem was locating TIFF files as these are large, uncompressed files which, whilst they are of a higher quality, take up a lot of storage room and so are not generally the format of choice on the internet, where this level of quality is not required or discernible.

The final test set is made up of a mixture of scanned, photographed, graphic, colour and greyscale images as follows:

4 X GIF (Version 89a)
4 X JPEG (Version1.02)
4 X TIFF (Version 6.0)

NB. Unless stated otherwise, further mention of these three formats refers to these format versions.

4.2. Testing Environment

All software testing was performed on a CompaqEvo D510 SFF fitted with a Pentium 4 1.80 GHz CPU, 1GB RAM and installed with Microsoft Windows XP Professional (version 2002) Service Pack 2.

4.3. Experiment testing

4.3.1 Initial Characterisation

At the same time as having the format formally identified, during the finalisation of the test data, each of the test images outlined in 4.1 above, were characterised, using Jhove. This characterisation process determines a set of properties as pre-defined by the relevant Jhove module, and gives a value for each of these properties where present. Jhove states that these properties are 'the format-specific significant properties of an object of a given format'.[19] However, it should be noted that the use of significance here is not defined and differs from that defined in the InSPECT project. The Jhove concept of significant properties includes technical information such as byte order and compression scheme which would be outside of the InSPECT definition of significance because they are properties which apply to all digital objects and not just raster images.

---

[19] http://hul.harvard.edu/jhove/using.html

The property values obtained from this characterisation served as the basis for comparison with our images once they were migrated in the next stage of the experiments.

4.3.2 Migration
Each of the test objects was migrated twice, from its original format to each of the other test formats

4.3.3 Post-migration characterisation
Once each image was migrated, each of the two new format versions was characterised using Jhove and the output used as the basis for comparison with the original image to see how well properties were retained through migration.
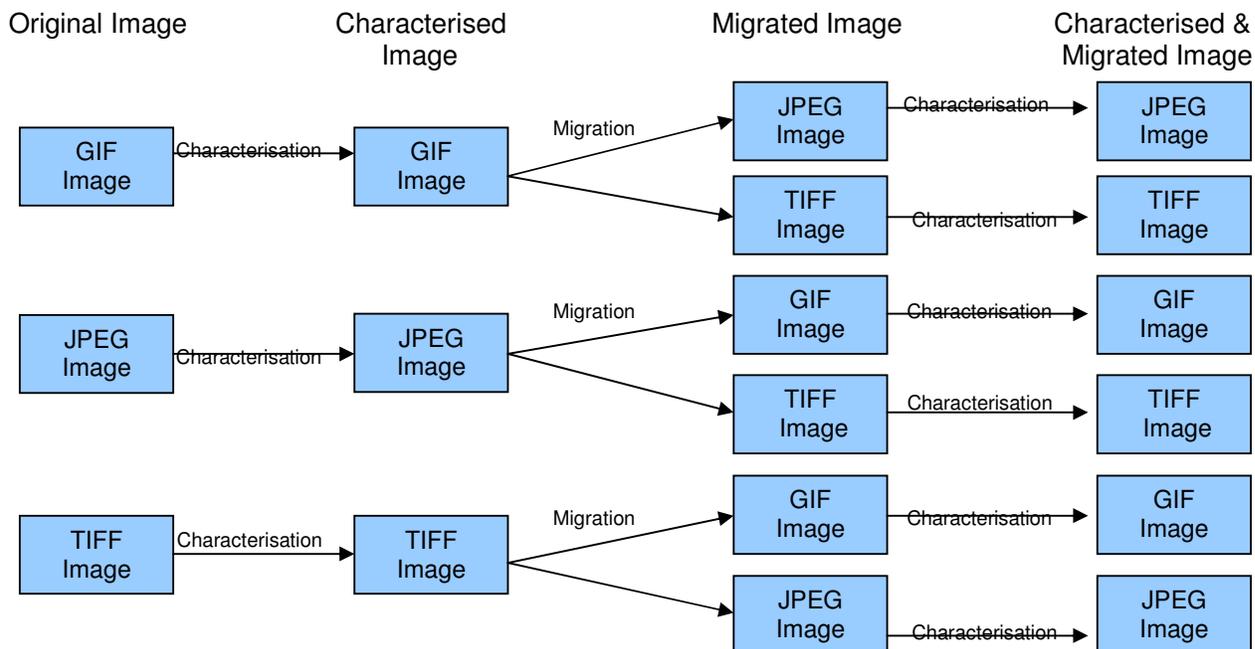
Figure 4. Illustration of Automated Experiment Procedure

4.3.4 Visual Assessment of converted images.
Once the automated parts of the process were carried out, a visual assessment of the images was carried out. Photoshop was used to open each image so that the evaluator could visually compare them.

4.4. Experiments

4.4.1 Experiment 1: Convert GIF 89a to JPEG 1.02 and TIFF 6.0 using Photoshop
The first experiment involved converting the collected GIF sample images to JPEG and TIFF using Photoshop.

4.4.1.1. Initial Characterisation

In order to compare and measure the properties of the file before and after conversion, the initial step was to characterise the original Gif images using Jhove. This simply involved selecting the GIF-hul module within the Jhove Edit menu and then opening the image from the Jhove File menu. This provides a file analysis which is formatted so as consistent with the NISO image metadata standard (NISO Z39.87) (Screengrab 1). This metadata standard was used as the basis for the project team's analysis of the significant properties of raster images (see section 2.1.2 above).
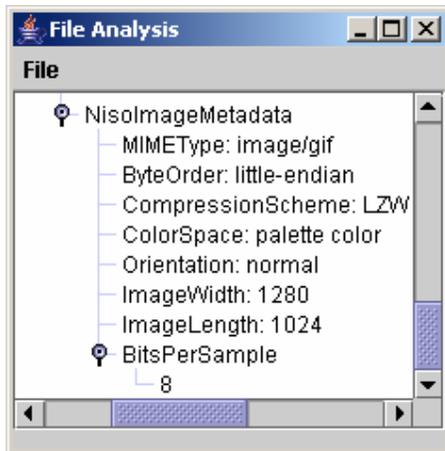
Screengrab 1. Example of the presentation of NISO metadata in the Jhove file analysis of a GIF image.

This file analysis was then saved in both text and xml format (the two available options in Jhove) and screen shots of the Jhove output were also taken as this was sometimes the easiest way to read the results, particularly as the Niso metadata for each file was relatively small and easy to view in this way.

4.4.1.2 Migration

The aim was to then use Photoshop to migrate the GIF images to both the JPEG and TIFF formats using the 'save as' option. However, whilst Photoshop gave the option to migrate a GIF to many formats, it did not give the option to save a GIF as a JPEG.
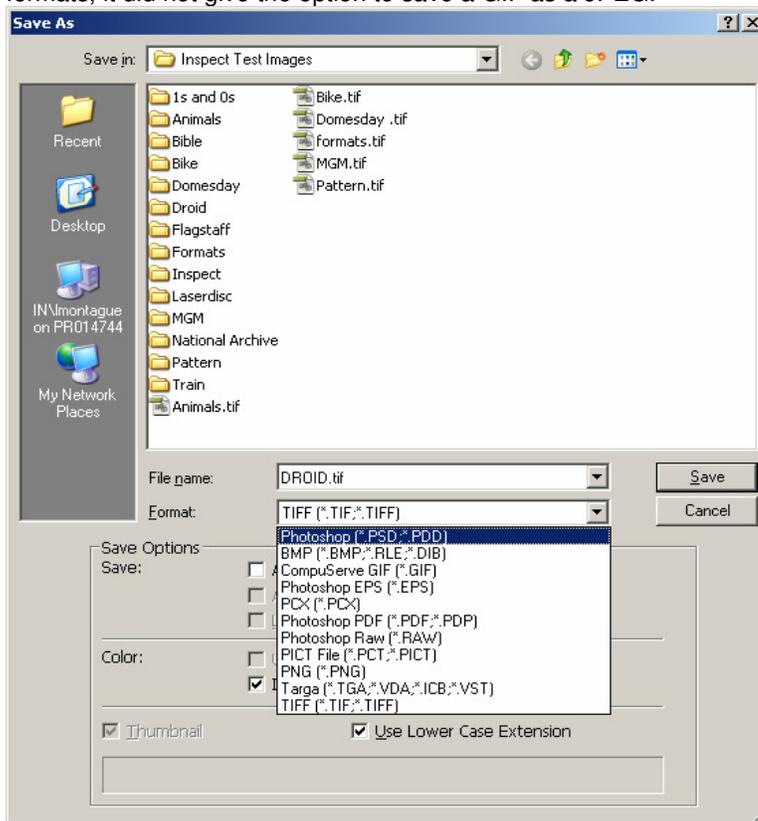
Screengrab 2. Available formats for migration of GIF image in Photoshop CS

Therefore an alternative method of conversion was sought and the Windows picture viewer, Window Fax and Picture Viewer, was used to do this migration. This simply involved choosing to save as a JPEG and pressing save. However, this was only able to save as a JPEG 1.01 not 1.02.

When saving as a TIFF the following two option screens were presented. In both cases the default options, as offered by Photoshop, were accepted except that IBM PC was chosen on the second screen instead of Macintosh as all testing was done in a PC environment.
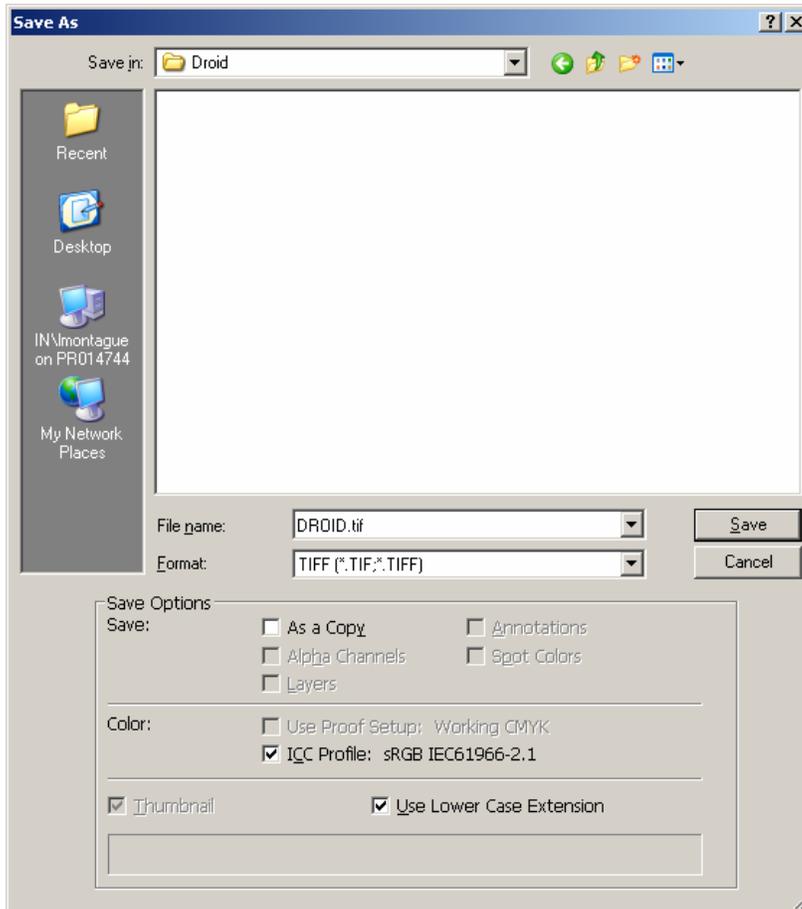
Screengrab 3. Photoshop options for saving files.
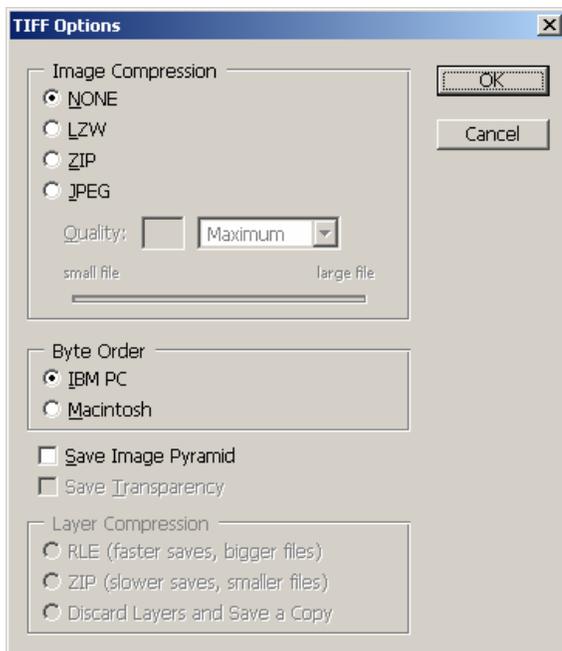
Screengrab 4. TIFF-specific options when saving a TIFF file in Photoshop

An anomaly occurred during this process in that on trying to save one of the GIF images as a TIFF file, Photoshop produced the message 'file must be saved as a copy with this selection' and therefore the 'As a copy' option was automatically checked in the in the first screen. It is unclear why in this particular case this happened.

4.4.1.3 Second Characterisations

The migrated JPEG and TIFF files were then characterised using Jhove, as in section 4.4.1.1 above, by choosing the JPEG-hul and TIFF-hul modules respectively. These characterisations were used as the basis for the comparison of properties between the original GIF and the migrated JPEG and TIFF files in order to see how the NISO metadata was converted.

4.4.1.4 Results – Significant Properties identified by Jhove for original and migrated images 1-4

	Image 1		
NISO Metadata identified by Jhove	**GIF**	**JPEG**	**TIFF**
ImageWidth	400	400	400
ImageLength	400	400	400
Bits per sample	8	8,8,8	8
Samples per pixel	-	3	1
Extra samples	-	-	
ColorSpace	Palette colour	YCbCr	Palette color
XSamplingFrequency	-	-	72
YSamplingFrequency	-	-	72

	Image 2		
NISO Metadata identified by Jhove	**GIF**	**JPEG**	**TIFF**
ImageWidth	500	500	500
ImageLength	347	347	347
Bits per sample	8	8,8,8	8
Samples per pixel	-	3	1

Extra samples	-	-	-
ColorSpace	Palette Color	YCbCr	Palette Color
XSamplingFrequency	-	-	720000/10000
YSamplingFrequency	-	-	720000/10000

	Image 3		
NISO Metadata identified by Jhove	**GIF**	**JPEG**	**TIFF**
ImageWidth	100	100	100
ImageLength	80	80	80
Bits per sample	8	8,8,8	8
Samples per pixel	-	3	1
Extra samples	-	-	-
ColorSpace	Palette Color	YCbCr	Palette Color
XSamplingFrequency	-	-	720000/10000
YSamplingFrequency	-	-	720000/10000

	Image 4		
NISO Metadata identified by Jhove	**GIF**	**JPEG**	**TIFF**
ImageWidth	419	419	419
ImageLength	259	259	259
Bits per sample	8	8,8,8	8
Samples per pixel	-	3	1
Extra samples	-	-	-
ColorSpace	Palette Color	YCbCr	Palette Color
XSamplingFrequency	-	-	72
YSamplingFrequency	-	-	72

4.4.2. Experiment 2: Convert JPEG 1.02 to GIF 89a and TIFF 6.0 using Photoshop

The second experiment involved converting the collected JPEG sample images to GIF and TIFF using Photoshop.

4.4.2.1 Initial Characterisation

As with the previous experiment, in order to compare and measure the properties of the file before and after conversion, the initial step was to characterise the original JPEG images using Jhove. This simply involved selecting the JPEG-hul module within the Jhove Edit menu and then opening the image from the Jhove File menu. This file analysis was then saved in text and xml formats and screen shots of the Jhove output were again taken.

4.4.2.2 Migration

Photoshop was then used to migrate the JPEG images to both the GIF and TIFF formats using the 'save as' option. When saving to GIF the message 'File must be saved as a copy with this selection' was given by Photoshop and the 'As a copy' box automatically checked. The 'Save' option was then chosen.
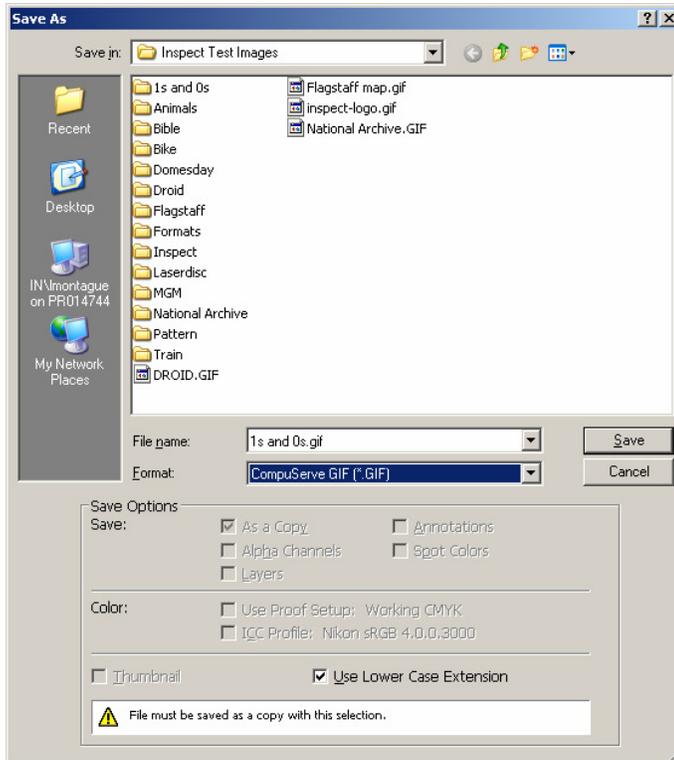
Screengrab 5. Photoshop save options when saving a JPEG to GIF

An 'Indexed Color' screen then appears in Photoshop and the default options given by the programme accepted by the tester except that the 'Preserve Exact Colors' box was checked before pressing 'OK'.
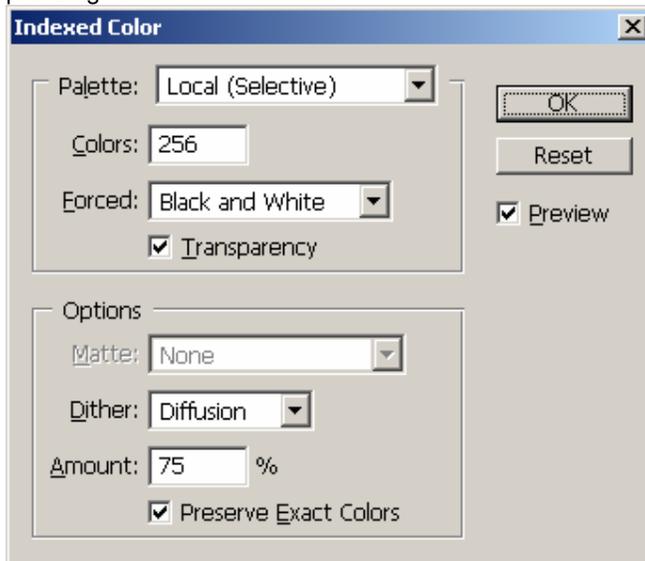
Screengrab 6. Indexed Colour options for saving GIFs in Photoshop

Finally a GIF options box appears and the default 'Normal' option accepted.

Screengrab 7. GIF-specific options for saving in Photoshop

When saving as a TIFF the same process was followed, and the same screens offered by Photoshop, as when saving a GIF as a TIFF in experiment 1. In the case of Image 8, the migration to TIFF was saved by Photoshop as a TIFF 5.0 rather than 6.0. It is not clear why this happened and seems to be a quirk of the migration process.

4.4.2.3 Second Characterisations

The migrated GIF and TIFF files were then characterised using Jhove, as in section 4.4.2.1 above, by choosing the GIF-hul and TIFF-hul modules respectively. These characterisations were used as the basis for the comparison of properties between the original JPEG and the migrated GIF and TIFF files in order to see how the NISO metadata was converted.

4.4.2.4 Results - Significant Properties identified by Jhove for original and migrated images 5-8

	Image 5		
NISO Metadata identified by Jhove	**GIF**	**JPEG**	**TIFF**
ImageWidth	2616	2616	2616
ImageLength	2554	2554	2554
Bits per sample	8	8,8,8	8,8,8
Samples per pixel	-	3	3
Extra samples	-	-	-
ColorSpace	Palette Colour	YCbCr	RGB
XSamplingFrequency	-	-	300
YSamplingFrequency	-	-	300

	Image 6		
NISO Metadata identified by Jhove	**GIF**	**JPEG**	**TIFF**
ImageWidth	600	600	600
ImageLength	800	800	800
Bits per sample	8	8,8,8	8,8,8
Samples per pixel	-	3	3
Extra samples	-	-	-
ColorSpace	Palette Colour	YCbCr	RGB
XSamplingFrequency	-	-	72
YSamplingFrequency	-	-	72

	Image 7		
NISO Metadata identified by Jhove	**GIF**	**JPEG**	**TIFF**
ImageWidth	2616	2616	2616
ImageLength	3820	3820	3820
Bits per sample	8	8,8,8	8,8,8
Samples per pixel	-	3	3
Extra samples	-	-	
ColorSpace	Palette Color	YCbCr	RGB
XSamplingFrequency		-	300
YSamplingFrequency		-	300

	Image 8		
NISO Metadata identified by Jhove	GIF	JPEG	TIFF
ImageWidth	600	600	600
ImageLength	792	792	792
Bits per sample	8	8,8,8	8,8,8
Samples per pixel	-	3	3
Extra samples	-	-	-
ColorSpace	Palette Color	YCbCr	RGB
XSamplingFrequency	-	-	72
YSamplingFrequency	-	-	72

4.4.3   Experiment 3: Convert TIFF 6.0 to GIF 89a and JPEG 1.02 using Photoshop

The final experiment involved converting the collected TIFF sample images to GIF and JPEG using Photoshop.

4.4.3.1. Initial Characterisation

As previously the original files to be migrated, in this case TIFFs, were characterized using Jhove in order to compare and measure the properties of the file before and after conversion. This involved selecting the TIFF-hul module within the Jhove Edit menu and then opening the image from the Jhove File menu. This provides a file analysis which is formatted so as consistent with the NISO image metadata standard (NISO Z39.87) This file analysis was then saved in text and xml formats and screen shots of the Jhove output were again taken.

4.4.3.2 Migration

Photoshop was again used to migrate the TIFF images to both the GIF and JPEG formats using the 'save as' option. On opening the file in Photoshop, three of the TIFF images give the following message. OK was chosen. From this point, saving as a GIF followed the same process as in experiment 2 above.

Screengrab 8. Embedded Profile Mismatch

When saving as a JPEG, the generic Photoshop save screen was offered (Screengrab 3). Then a JPEG options screen appears and the default options given by Photoshop were chosen.
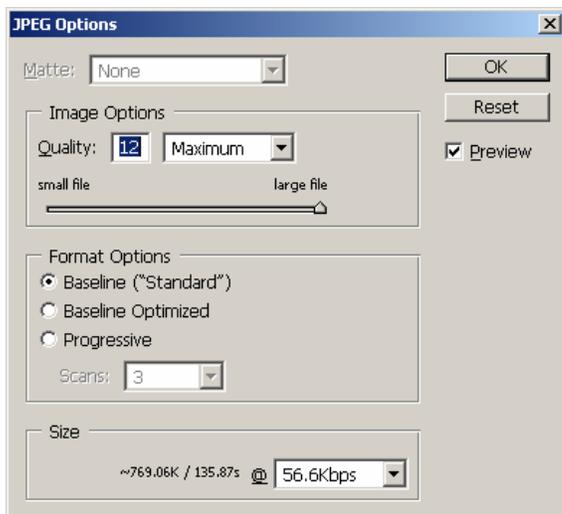
Screengrab 9. JPEG-specific options when saving a JPEG file in Photoshop

4.4.3.3. Second Characterisations

The migrated JPEG and GIF files were then characterised using Jhove, as in section 4.4.3.1 above, by choosing the JPEG-hul and GIF-hul modules respectively. These characterisations were used as the basis for the comparison of properties between the original TIFF and the migrated JPEG and GIF files in order to see how the NISO metadata was converted.

4.4.3.4 Results - Significant Properties identified by Jhove for original and migrated images 9-13

	Image 9		
NISO Metadata identified by Jhove	GIF	JPEG	TIFF
ImageWidth	1280	1280	1280
ImageLength	1024	1024	1024
Bits per sample	8	8,8,8	8,8,8
Samples per pixel	-	3	3
Extra samples	-	-	-
ColorSpace	Palette Color	YCbCr	RGB
XSamplingFrequency	-	-	300
YSamplingFrequency	-	-	300

	Image 10		
NISO Metadata identified by Jhove	GIF	JPEG	TIFF
ImageWidth	3504	3504	3504
ImageLength	2336	2336	2336
Bits per sample	8	8,8,8	16,16,16
Samples per pixel	-	3	3
Extra samples	-	-	-
ColorSpace	Palette Color	YCbCr	RGB
XSamplingFrequency	-	-	300
YSamplingFrequency	-	-	300

	Image 11		
NISO Metadata identified by Jhove	GIF	JPEG	TIFF
ImageWidth	1536	1536	1536

ImageLength	2048	2048	2048
Bits per sample	8	8,8,8	8,8,8
Samples per pixel	-	3	3
Extra samples	-	-	-
ColorSpace	Palette Color	YCbCr	RGB
XSamplingFrequency	-	-	350
YSamplingFrequency		-	350

	Image 12		
NISO Metadata identified by Jhove	**GIF**	**JPEG**	**TIFF**
ImageWidth	1417	1417	1417
ImageLength	945	945	945
Bits per sample	8	8	8
Samples per pixel	-	1	1
Extra samples	-	-	-
ColorSpace	Palette Color	YCbCr	Black is Zero
XSamplingFrequency	-	-	300
YSamplingFrequency	-	-	300

4.4.4 Visual inspection of results.

A visual inspection of the image files in Photoshop resulted in the following obvious differences being noted in the images. This was a superficial inspection where the evaluator was not an expert and it may be that further differences would be noted by a photographic professional.

Image	Colour representation of the original in bit size	Visually discernible differences in conversions
Image 1	8	No discernible difference
Image 2	8	No discernible difference
Image 3	8	No discernible difference
Image 4	8	No discernible difference
Image 5	24	No discernible difference
Image 6	24	No discernible difference
Image 7	24	The GIF conversion from the JPEG original was possibly more granulated.
Image 8	24	No discernible difference
Image 9	24	There was a noticeable visual difference between the original TIFF and the GIF conversion.
Image 10	48	There was a noticeable visual difference between the original TIFF and the GIF conversion.
Image 11	24	No discernible difference
Image 12	24	There was a noticeable difference in colour between the TIFF original and the GIF conversion.

Table 1. Visually discernible differences in conversions.

5   Conclusions

There were surprisingly few variations in the characterised migrated images when compared with the original files. In all cases, the Image Width and Height were maintained through migrations. This, or rather the ratio of Width to Height, can be seen as one of the most significant aspects of maintaining the integrity of an image.

The main differences are the constraints of the source and target formats, their ability to handle metadata or the software used the conduct the experiments. For example, colour space is not a property designated as significant by this project but the colour space metadata produced by the

experiments was included in the experiment results for interest. Whilst, in theory it doesn't matter what the colour space of a migrated image is, as long as the colours are represented correctly from the original image, the colour space will dictate the number of bits per sample i.e. the number of colours that are possible, and this is a significant property.

Nearly all migrations resulted in a change in colour space. However, it is difficult to tell whether the visual colour changes noted were down to the change in colour space, or the fact that the number of bits used to represent colour had also changed. It is also difficult to assess this effect this other than by viewing the different versions of an image and making a visual judgement. As stated above, in terms of visual assessments of the converted images, with the aim of pinpointing changes in the images due to changes in the significant properties, a photographic expert is really needed.

In terms of the results of the visual inspections of images:

1. With images 1-4 it is not surprising that there were no visually noticeable differences as the conversions are from the GIF file format which is limited to 8-bit colour representation and they are simple files. So even though the conversions are to formats with larger bit colour representations they are limited by the original information received from the GIF.

2. No visual differences were noted with JPEG images 5, 6 and 8. This would be expected with the conversion to TIFF where both formats render the image in a 24-bit format but is surprising with the conversion to 8-bit GIF where it would be expected that some colour detail my be lost, as with image 7 below, as the image is being converted from a format that potentially allows 16.7 million colours to one that allows 256.

3. With image 7 it appeared that the GIF conversion from the JPEG original was slightly more granulated than the original. This was not surprising because, as stated above, this was a conversion from a 24-bit to an 8-bit format. It may be that the difference was noticeable with this image and not images 5, 6 and 8 because it was a greyscale image and therefore the differences were more noticeable than with the other conversions. Images 6 and 8 only had a sampling frequency of 72 pixels per inch in the TIFF version which is a spatial resolution suitable for the internet and so any colour or detail changes may have been hard to notice.

4. For images 9, 10 and 12, it would be expected that there would be a visually noticeable change of colour resolution because it is a conversion between a format with large potential colour representation to one with small potential colour representation (as explained in 2 above). This was particularly pronounced with image 10 where the original TIFF which contained 48-bits of colour information. It was surprising that such a difference was not also noted with image 11. However, whilst this was a colour image, the colours represented seem to be relatively few to the human eye.

Due to the constraints of the representation formats, particularly GIF, in handling metadata, as mentioned above, the only metadata consistently recorded across all three formats were image width, image length, bits per sample and colour space (which is not deemed as significant for our purposes). Therefore the ability to measure and assess the success of a file migration in an automated way between these three formats was limited. However, it is suggested that the significant properties of image width, image length, and bits per sample which were recorded, are the most vital aspects in maintaining an authentically rendered image.

5.1 Other Issues

- Working within a government organisation produces its own additional challenges when doing this type of research work as all work has to be conducted within the standard operating procedures concerning internet and software usage. Many websites are blocked which hinders research as judging whether the site would be useful is impossible without going through procedures to get it unblocked which is time consuming and often results in it being obvious, immediately that a site is unblocked, that it isn't a useful resource. In addition, it is not possible to easily download tools to trial to see if they are suitable for a particular project. Again IT procedures need to be complied with which can make it prohibitively time-consuming when trying

to analyse and compare suitable tools. In future, these additional constraints would need to be factored into such a project.

5.2 Recommendations

- Recommend that further experimentation work is done with other migration and characterisation tools to compare results and develop tools further as necessary.

- Recommended that further experimentation with other, complex image formats be done in order to see how well the other significant properties are migrated.

- Recommend that a large sample set of test images be built up which have values for all of the significant properties (and other properties) allowable by the format for use in future tests. Some work is currently being carried out at the University of Cologne to assemble a set of test files of various digital objects as part of the Testbed workpackage in the EU-funded Planets project[20]. It is not yet known if this resource will be more widely available in the future.

---

[20] http://www.planets-project.eu/

Appendix 1: Software Tools

The project examined a number of software tools capable of analyzing representation formats used for the storage of raster images. To document the process it adopted the format adopted by the CAIRO project for its tool survey[21].

Photoshop CS

Tool Name	Photoshop CS
Source URL	http://www.adobe.com/support/photoshop, http://libpsd.graphest.com/files/Photoshop%20File%20Formats.pdf http://www.wikipedia.org http://www.informit.com/articles/article.aspx?p=169496
Formats supported	PSD, Photoshop PDF, Photoshop Raw, Large Document Format (PSB), Cineon, PNG, TIFF, OpenEXR, Portable Bitmap, Radiance, PhotoDeluxe Document (PDD), DNG, EPS, JPEG.
Technology Base	
Operating system	Cross-platform
Dependencies	
Licence	Proprietary
Category	Conversion, Viewing
Description	Photoshop is one of the leading image manipulation tools available. Developed by Adobe, it is able to use the colour spaces bitmap, grayscale, duotone, RGB, CMYK and Lab and can read and write raster images in many formats. It is a well documented format.
Output methods	
Notes	

Jhove

Tool Name	Jhove (JSTOR/Harvard Object Validation Environment)
Source URL	http://sourceforge.net/projects/jhove/
Formats supported	Arbitrary byte streams, ASCII, UTF-8, GIF, JPEG2000, JPEG, TIFF, AIFF WAVE, PDF, HTML, and XML
Technology Base	Command line and GUI. Written to conform to Java 2 Platform, Standard Edition (J2SE) 1.4
Operating system	Any Unix, Windows, or OS X platform with the appropriate J2SE installation.
Dependencies	J2SE 1.4-compliant Java Runtime Environment (JRE)
License	GNU Library or Lesser General Public License (LGPL)
Category	Identification, validation, characterisation
Description	Jhove (JSTOR/Harvard Object Validation Environment) is an identification, validation and characterisation tool developed by JSTOR and Harvard University Library. These actions involve being able to identify files of particular specified formats, state whether particular object examples of these formats are well formed and valid, and determine the specific properties of a particular object in a supported format. It has modules to support these actions for arbitrary byte streams; ASCII and UTF-8 encoded text; GIF, JPEG2000, JPEG and TIFF images; AIFF and WAVE audio; PDF, HTML, and XML. Output from these modules is available in text and xml formats. It includes both a command line and GUI version, with the latter being used in this project.
Output methods	Text, XML
Notes	

---

[21] Further details of the format can be found on p11 of the Cairo Tools Survey, located at http://cairo.paradigm.ac.uk/projectdocs/index.html

Windows Picture and Fax Viewer

Tool Name	Windows Picture and Fax Viewer
Source URL	http://en.wikipedia.org/wiki/Windows_Picture_and_Fax_Viewer, http://en.wikipedia.org/wiki/Windows_Picture_and_Fax_Viewer
Formats supported	JPEG, BMP, PNG, ICO, WMF, EMF and TIFF
Technology Base	GDI+
Operating system	Windows XP and Windows Server 2003 operating systems.
Dependencies	
License	Proprietary
Category	Conversion, Viewing
Description	Windows Picture and Fax Viewer is a Microsoft product used for image viewing but not editing. It can view the following formats; JPEG, BMP, PNG, GIF, ICO, WMF, EMF and TIFF and images can be saved in these formats.
Output methods	
Notes	

Author: Montague, L.   Date & Time: 15/063/2009