

---

# Image Capture

IMPACT Briefing Paper

IMPACT project

Neil Fitzgerald, British Library

Jeanna Nikolov-Ramírez Gaviria, Austrian National Library

Released under Creative Commons - Attribution-NonCommercial-ShareAlike v3 Unported (International)

## Table of Contents

Archiving of raw master file .....	2
Questions to consider: .....	2

### Briefing paper on Image Capture

Capture is the term used in digitisation to describe the images and associated data recorded by a camera or scanner. The approach to image capture that a digitisation project takes tends to be driven in large part by the type of material to be digitised and the available budget. This includes the choice of hardware to create the images and the use to which the images will be put.

Digital images created by a scanner or camera should always be captured at a size and a colour depth large enough for all planned use. Most robotic scanners will capture images automatically as a bitmap file; that, is in a form where the file size directly relates to the dimensions and colour depth of the image. These bitmap files may be converted and compressed for different types of use and storage. See the sections on image file formats and storage later in this resource.

Equipment calibration is the term used to describe the processes by which the capturing device can be set up or adjusted for optimal operation for the capture of a set of images. For consistent operation the equipment should be used in a controlled environment (with checks and mitigations for fluctuations in temperature, humidity, etc.) applying the operating thresholds advised by the equipment manufacturer. It is therefore necessary to calibrate all the capture devices, allowing proper exchange and interpretation of colour information with each other, by reference to an agreed colour profile such as Adobe 1998<sup>1</sup> or sRGB<sup>2</sup>. This agreed colour profile should be embedded into the digital image, enhancing the reusability of the image.

Calibration of capture devices should also be performed throughout the digitisation process as a form of preventive quality assurance, helping to keep the devices operating optimally. Scanners and other visual capture devices are subject to drift, which needs readjustment over time. Colour profiling should also be carried out regularly, with intervals dependent on the anticipated performance of the capture device and the general stability of the environment. Colour profiling should always be redone after maintenance or recalibration work on the device. The use of standard colour/resolution targets and rulers should also be considered as part of the standard capture set-up and ongoing digitisation.

Before beginning capture of digital images, an institution should consider the following:

- For what use are the digital images being created?

---

<sup>1</sup> *Adobe RGB (1998) Color Image Encoding*; 1998; Adobe Systems Ltd: <http://www.adobe.com/digitalimag/adobergb.html> Retrieved 10.03.2011

<sup>2</sup> *A Standard Default Color Space for the Internet – sRGB*; 1996; Stokes, M. et al: <http://www.w3.org/Graphics/Color/sRGB.html> Retrieved 10.03.2011

- What file formats support that use (e.g. what file format best suits delivering readable OCR files quickly to the user?). What file format is best to use for long-term preservation?
- What output level of images do you need to achieve hourly, daily or weekly?
- Can the desired output be achieved within the available infrastructure?
- If not, what output can realistically be achieved within the available infrastructure and what needs to change?
- Will the images from a particular collection need specific post-processing? Can this be done in-house?
- Does it make sense to outsource some or all of the scanning work and/or post-capture processing?

Please be aware that while the source material may itself sometimes be an impediment to creating fit-for-purpose digital materials, scanners and cameras can sometimes add their own faults to the image. For guidance on choosing a scanner appropriate to a project's needs, please consult the European Commission on Digital Access<sup>3</sup> or Harvard's guide to available scanners<sup>4</sup>.

## Archiving of raw master file

Once the images have been captured they can be archived and saved in the form created by the original camera or scanner. This means that if optimised files are lost or corrupted later in the process, the original master will always be available. Historically, the master file has normally been either a lossless TIFF file or a proprietary "raw" format. A raw file is a high quality lossless image that includes metadata about the camera's make and set-up and the date and time of capture: it will usually be larger than its TIFF equivalent. Because raw files are proprietary, you may wish to consider a technology-watch process or converting to a non-proprietary format. (Note that some cameras capture directly into JPEG. While this saves storage space, it is a lossy format and should be avoided when projects are aiming at long term preservation.)

The master images should be archived within a wide-gamut colour space such as CIE Lab<sup>5</sup> or Adobe RGB 1998. The archived master files are normally only used for long-term emergency back up. If RAW files are archived, care must be taken to also archive the required software to open them and migrate to an open standard such as TIFF.

In a mass digitisation project, there is sometimes neither time nor storage space to archive a master file and its optimised equivalent. In an increasing number of projects and institutions, master images are being converted direct to JPEG2000, with this becoming the de facto archive master<sup>6</sup>.

## Questions to consider:

- Which format and colour space should the master files be in? Should the master already include some form of optimisation?
- Where will master and intermediate files be stored so they are resilient? Is it necessary to store all variants of the same image?
- What software is needed to be able to open images in the future?

---

<sup>3</sup> *Basic advice in choosing imaging devices*: <http://www.knaw.nl/ECPA/sepia/workinggroups/wp4/guidelines.html#advice> Retrieved 05.02.2010

<sup>4</sup> *Guideline: Selecting a scanner or Digital Camera*; 2007; Harvard University: <http://preserve.harvard.edu/guidelines/selectascanner.pdf> Retrieved 05.02.2010

<sup>5</sup> CIE Lab Color Space; 2009; Hoffman, G: <http://www.fho-empden.de/~hoffmann/cielab03022003.pdf> [<http://www.fho-empden.de/%7Ehoffmann/cielab03022003.pdf>] Retrieved 08.02.2010

<sup>6</sup> *JPEG2000 at the Wellcome Library*; 2010; Wellcome Trust: <http://wellcomelibrary.blogspot.com/2010/06/new-blog-jpeg-2000-at-wellcome-library.html> Retrieved 10/03/2011

- Will file naming be done automatically and by what rules?

Image file formats and storage are discussed in more detail in the next section of this document. A recent discussion of file size and long-term storage between commercial and public organisations engaged in digital archiving can be found on the Digital Curation Blog.<sup>7</sup>

### Key Terms

**Bitmap:** An image formed by a grid of pixels. A computer assigns a value to each pixel – ranging from 1 ‘bit’ of information (simply black or white) to as many as 24 bits per pixel for full colour images

**Calibration:** The technical procedures necessary to ensure that capture and display devices are operating within their performance thresholds. Calibration is necessary to ensure correct profiling of imaging devices: that is to say that there is a known and adequate relationship between the image as captured by one device and displayed by another. Within IMPACT, calibration has the additional aim of reducing unwanted visual effects (‘noise’) that can harm legibility and OCR accuracy

**Colour profile:** A technical statement of how an input, monitoring or output device will interpret and present colour in a digital image. Such devices are often tuned to slightly different colour specifications (known as colour spaces), so colour management procedures and systems must be used to ensure uniformity of colour across all devices

**Colour space:** A representation of all possible mixtures of the primary colours within a given colour model, with each variant assigned a unique place within the representation

**Colour Target:** A reference chart of colour patches with known properties, which can be used to calibrate image capture devices so that the colour displayed to the end user matches the colour of the original

---

<sup>7</sup> *Responses to Raw vs. Tiff*, 2008; Rusbridge, C: <http://digitalcuration.blogspot.com/2008/07/responses-to-raw-versus-tiff.html> Retrieved 05.02.2010