Image Optimisation

IMPACT Briefing Paper

IMPACT project
Jeanna Nikolov-Ramírez Gaviria, Austrian National Library
Niall Anderson, British Library

Table of Contents

Briefing Paper on Image Optimisation

Optimisation is the term used in digitisation to refer to the processes by which original digital images are altered to fit their final purpose. Depending on the overall purpose of a project, optimisation may include the following:

- **Optimisation of image size:** cropping (to preserve storage space), orientation (to enable machine readability and OCR);

- **Optimisation of image tonality:** white balance, levels and curves;

- **Cleaning up:** removing marks, spots, correcting other image problems;

- **Addition new descriptive data**: or editing of automatically captured data, to fit with internal application standards;

- **Conversion of original image format into desired, or access, formats;**

- **Change of colour depth to required access standard;**

- **Sharpening of image**

If the creation and optimisation of digital images has been outsourced to an external supplier, the supplier must know in advance of production what the desired optimisation effects are.

Archiving of optimised master file

A growing trend in mass digitisation is to archive the optimised master file only, without saving the unprocessed master file. The minimum requirements for optimised images are that they should be cropped to the page edge (saving up to 30% in storage costs); rotated to the correct orientation (to enable OCR); with the desired colour profiles applied.

If storage permits, optimised masters should be archived in a lossless format, like TIFF or PNG or lossless JPEG2000. If the institution plans to make these images available beyond the life of the individual project, they should be stored offline.

IMPACT provides the IMPACT Storage Estimator (ISE) [http://www.impact-project.eu/taa/strat/pilot-tools/] to help with this process.

Because IMPACT deals primarily with text-based historical material, most references to optimisation in the learning resources deal with optimising for Optical Character Recognition.

Access image creation

Access files are temporary files for speedy delivery to users. They tend to be smaller in bytes and dimensions to master images, to enhance their delivery to the end user. They should be created directly from the Optimised Master Image, e.g. the master bitmap to which necessary processes such as OCR have been applied.

There are four main types of surrogate image:

• Thumbnail/Monitor images for web and multimedia presentation - normally a compressed RGB JPG or GIF, in an sRGB colour space, scaled down to the appropriate size;

• Commercial Print Images - uncompressed TIFF files within an RGB colour space;

• In-House Print Images - uncompressed TIFF files within an RGB colour space, scaled to the appropriate size. This may be identical to the Optimised Master Image;

• Master Monitor Image/Access File -usually a resized optimised master, an uncompressed (or losslessly compressed) TIFF or PNG file in Adobe RGB 1998 or sRGB colour space,

If OCR results are being presented along with the surrogate image, the image will need to be stored with an appropriate level of metadata. At its simplest, this need only be the title of the image and its provenance; but if the project intends to allow its users to search by word, article or text block, more thought needs to be given to the metadata standard and structure used. In current practice, mass digitisation projects have tended to use a combined METS/ALTO standard, discussed in detail in the IMPACT Best Practice Guide: Metadata.

Cornell University suggests benchmarks for onscreen legibility[1] including access files.

A searchable PDF of the text can also be produced with embedded OCR.

Any files created by a project will have storage and cost implications. See the IMPACT Storage Estimator (ISE) [http://www.impact-project.eu/taa/strat/pilot-tools/] for an easy way of calculating and costing necessary storage.

---

[1] *On Screen Image Quality*; 2003; Cornell Library: http://www.library.cornell.edu/preservation/tutorial/presentation/presentation-07.html Retrieved 12.02.2010