Digital Images

Digital Preservation Topical Note 3

DigitalPreservation**Coalition**

What are digital images?

Digital images are data files which when read by a computer render an image onscreen. Unlike an analogue image that is interpreted by the human eye, a digital image must be opened and read using a software programme. Some digital images are 'born digital', or in other words, created using digital technology like a digital camera or a mobile phone. Other digital images are '*digitized*' from a different, analogue source, such as a book or paper record.

Image Compression

In order to store digital images more efficiently and make the best use of limited storage space, images can be compressed. Image compression is either 'lossy' or 'lossless'. Lossy compression results from removing data from the image file so that the file will be smaller, therefore the image quality will be reduced. Lossless compression refers to re-writing the file more efficiently without losing any data, which reduces the file size but not as much as lossy compression. The JPEG 2000 standard, for example, allows for JP2 files to be compressed 'lossy' or 'lossless'. If an organisation uses JP2 files for preservation, or as the master copy, it will compress JP2s with lossless compression to preserve quality. If the organisation creates JP2s for access or as surrogate copies, it will likely choose lossy compression to make better use of storage space.

Types of Digital Images: Bitmap vs. Vector

Most people are familiar with 'bitmap' images (also called 'raster'), which are stored as pixels, or tiny boxes of colour arranged in a pattern to form an image. The higher number of pixels in an image, the higher the image quality will be. Bitmap images often become blurry or fuzzy if scaled up or zoomed in. Vector images, on the other hand, are stored using mathematical formulas that represent lines and curves. Vector images tend to be smaller than bitmap images and can be scaled up without losing quality because the image is simply redrawn using the mathematical formula. While vector images would seem like the best choice for preservation because they are versatile and relatively small files, they have some disadvantages. Vector images work well for creating graphs or simple icons but do not work for photographic images. In addition, vector image file types (AL, EPS) are not well supported on the web, restricting how these types of images can be shared or published.

Bitmap and Other Files Types

Bitmap images can be stored as different types of files: JPG, GIF, PNG, and TIFF are all common types. Each type of image file has advantages and disadvantages for long-term preservation. TIFF files store high quality images but are often very large and so take up a lot of storage space. JPG, GIF, and PNG images are much smaller files, however, are usually not of a high enough quality to produce derivative copies, which limits how they can be used in the future. The file standard JPEG 2000 provides an alternative that meets both requirements: a high-quality image file that can be compressed to take up less space.
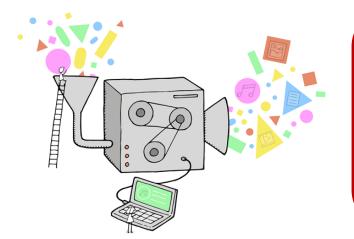
Digital Masters vs. Digital Surrogates

When an organisation digitizes an analogue record, it must decide whether to now consider the digital version to be the master copy and 'authoritative record' (see note on Preserving the Authoritative Record), and potentially disposing of the analogue version. Otherwise, an organisation might decide to digitize a record to keep as a surrogate, or access copy to 'stand in' for the original, but the analogue record will be retained. Whether a digitized image will be kept as a master or a surrogate may affect how it is digitized. For instance, a master copy to be preserved for the long-term must be as high a quality as possible to maximize how the image can be used in the future. However, higher quality images take up more storage space. By contrast, a surrogate or derivative (any copy made from the original file) does not need to be as high quality. In fact, lower quality images are smaller and therefore much easier to access and share, especially on the web.

RAW Files: A Good Preservation Choice?

In some cases, an organisation might choose to preserve RAW files. Some digitization technology, such as DSLR cameras, will capture an image in native raw formats, or RAW formats. They are considered by some to be the digital equivalent of a negative in traditional photography. Some types of RAW files provide an ideal format for preservation because they have not been processed or converted. The process of conversion can introduce flaws into TIFF or JP2s, which then become permanent if the RAW file is not retained. However, an untainted RAW file can be processed or converted at any point in the future. While some RAW file types are not well supported and therefore not sustainable over the long-term, other RAW files types like DNG images are well supported and provide a great option for preservation.

Metadata for Digital Images

Metadata associated with digitized images may refer to the original analogue record or to the digital file. Some metadata elements will need to be distinguished. For example, the digital file might have a date of creation for the digitized version as well as for the original analogue record. In addition to the metadata required about the original file, such as rights management, the digital file will also require technical metadata. Future users and records managers will need to be able to access information about how the image was created, what type of image it is, if and how it has been altered or changed, and other technical details.

> *Key Term:* **Digitization**
>
> The process of creating digital files by scanning or otherwise converting analogue materials. When digitizing paper records or other analogues materials, it is important to think about all of these factors: image type, file type, function (master or surrogate), image quality, and required metadata.

For more information on Digital Preservation visit the DPC Website: https://www.dpconline.org