

Nick Hiley

An introduction to digitisation.

Digitisation is wonderful for increasing access to images, but what can't those images show?

What cannot be digitised?

Texture	The detailed information about the surface characteristics of the object. How does it feel?
Weight -	The physical presence of the object. The characteristics of the material from which it is made.
Scale -	The size of the object in relation to the viewer, and to other objects. Every digitised object looks to be the size of your screen.
Fragility	The age and decay of the object - its history expressed in its fragility.
Smell	If smell was a part of the original experience of the object, it will be lost. You can digitise film, but not the smell of nitrate film in a carbon-arc projector, or the smell of the cinema auditorium.
Aura	People still want to see the original item, and will not go to an exhibition of digital copies.

The answer is that a vast proportion of the information we derive from objects cannot be transmitted digitally. So, what can be digitised? Digitisation is good at increasing access, and at providing indications of texture, weight, scale, fragility and aura. But it is best at reproducing content.

Content comes in two forms:

1. The content specific to a particular medium, such as an oil painting.
2. The content held in a non-specific medium, such as the text on a typewritten document.

The digitisation of these two types of material will be quite different. For example, the digitisation of the content of an oil painting will need to capture as much information as possible about the colour of the paint, and about the detailed brushstrokes used to create the image. It may also need to record information relevant to conservation, such as the fine cracking of the varnish applied over the paint. The process of digitisation must not damage the painting in any way, and the treatment of the painting will not change after digitisation.

By contrast, the digitisation of a typewritten document might be regarded simply as moving the text from one medium to another - in fact from an outdated storage medium to a more modern and flexible one. There is no need to capture any information about the paper

Nick Hiley

medium - colour and detail are irrelevant. With the digitisation of business archives, the paper record might even be discarded after digitisation, so care and conservation may not matter. Even with historical archives, the typewritten document may be put into deep storage, and never again be produced for readers.

The British Cartoon Archive has used both of these approaches. With the digitisation of artwork we might produce a 100Mb TIFF image scanned at 600 dpi, with a colour chart included in the image. This allows us to record a large amount of information in a form that will last. The artwork is then carefully stored and monitored, and if necessary conservation measures are taken.

In contrast, British Cartoon Archive's digitisation of newspaper cuttings from the 1960s involved preserving only the information about the black lines of the cartoon printed on the paper. This can be done at a much lower resolution with compressed files of perhaps 500Kb - 0.5% of the size of the other images. We had no interest in capturing any information about the decaying woodpulp paper, and had to accept that, once digitised, the cuttings would have to be left to decay, with no further conservation. This process of digitisation involved simply moving the black lines of the cartoon from one storage medium to another.

The creation of digital images is part of an amazing transformation in the way that information is handled and processed, and offers unimagined levels of access. But it is nevertheless true that digital images involve a significant loss of information.

For example, the text of a late nineteenth-century newspaper can be digitised and OCR'd, and the content moved away from the unstable acidic woodpulp-paper storage medium. But you will lose all information about the physical experience of reading a late nineteenth-century newspaper - the size of the sheet, the nature of the paper and its weight, the imprint of the type, the quality of printed illustration. All these changed over time, and are vital to the history of the press in this period, but all this change will be lost with digitisation.

We have to accept that the physical encounter with objects is a vital part of learning, and is an essential part of generating interest in archival material. Digitisation is a valuable tool, which most successfully moves content from an outdated and relatively inflexible storage medium, to a flexible storage medium. It is largely carried out to record the informational content of objects, and less often to create an archival surrogate.

How do we start thinking about digitising images?

The first thing to remember is that the digital image has to be created with flexibility in mind. The basic rule of digitisation is often said to be "Capture once, use many times". What this means is that you should try to digitise an object without a specific use in mind. The aim is to create a high-resolution master file, from which digital derivatives can be created to meet many different needs. This should avoid unnecessary handling of the original records.

As an example of this, the British Cartoon Archive's 100Mb TIFF files of artwork are not designed for anything other than long-term storage. We would not expect to distribute them in any form, or to supply them to anyone for reproduction - they simply contain too much

Nick Hiley

information. However, we do use them to generate a range of smaller digital image files in other formats, which can be distributed on the web, and supplied for reproduction.

For book reproduction we would usually create an image from the archival master that would easily print up to A4 in size. This might be a 1Mb JPEG, created essentially by taking the 100Mb TIFF and throwing the vast majority of the information away. But if someone wants to print only a detail of the artwork at A4 size, we still have the ability to return to the archival master and derive that detail from it, as a 1Mb JPEG. If we had limited ourselves to producing only 1Mb JPEGs of the artwork at the outset, we would have had to rescan it to produce this detail.

“Capture once, use many times” is thus the ideal, but in most circumstances you’ll find yourselves working out a compromise. It is, after all, costly and difficult to make and store 100Mb image files, just in case they might be needed. If a collection of valuable cartoon artwork is on loan to us, it might be a high priority for us to make 100Mb files for long-term use, after the collection has been returned to the owner. But what of a collection of printed cartoons owned by the British Cartoon Archive, from which it is unlikely that anyone will ever request detailed images, and where we could always go back to the originals if they did? In those circumstances we might compromise, take a risk on occasional re-scanning from the stable originals, and digitise for use.

The National Archives at Kew has also compromised in this way, for reasons of cost. Their archival masters are relatively small, compressed digital files, which could not satisfy every future demand. But their digitisation is principally aimed at moving written text from one storage medium to another, and they took this decision to reduce storage costs.

How to digitise.

There are many different types of hardware, and different file types, in use for the digitisation of images. But one basic question is whether to scan or photograph.

Scanning is easy and convenient, with no need to worry about tripods and image distortion, lighting control or exposure. However, it is worth remembering flat-bed scanning could be the most destructive thing that has happened to an object in the last hundred years.

All handling causes damage, and that may be a vital consideration in digitising fragile materials. You should treat all archives as if they are fragile - just remember that a 100 year old document may be 300 year old document that is only one-third through its effective life. I was discussing digitisation with an academic who said that the material he was hoping to digitise was “only 100 years old”, but that attitude will help to ensure that it doesn’t get any older.

Flat-bed scanning is convenient, but damaging. The alternatives are overhead scanning, or photography. As a rule you should think of photographing anything that is larger than the glass on your scanner, or which will not sit flat easily, or is too fragile for the handling required by flat-bed scanning.

Nick Hiley

James Baker has posted a blog under this link

<http://britishlibrary.typepad.co.uk/digital-scholarship/2013/05/on-metadata-and-cartoons.html>