Preserving and Caring for a Community Archive: Digitisation

David Clarke, Projects Manager, Photography, Tate

Introduction
This handout is designed to support a round table discussion to be held during the Training and Practice Sharing Workshop. For the purposes of this discussion I am concentrating on the **camera** as the main method of digitisation. Other equipment can be used including flatbed scanners, film scanners and book scanners, but most subjects can be digitised or scanned using a digital camera.

1. Background to Tate Archives & Access Project – the original brief
'Plan and cost the delivery of 52,000 images delivered within a 12 month timescale. Images to be consistent, high quality, relatively high resolution ready for web delivery and of reproduction quality. Captured images to meet international metadata standards.'

By breaking down the numbers of different types of original media (eg sketchbooks, flat documents, film originals and 3D items) I was able to estimate the types of scanning/photography equipment needed, the staff required and finally a budget.

This was a project with very strictly defined parameters and the numbers of items set against the timescale of the project dictated the range of equipment and the staff required.

If however, you are able to approach your project with fewer constraints including target deadlines, there is no reason why a more affordable digitisation system can't be set up. I would say though that there are key issues to address before setting up any system and minimum standards which should be met if at all possible.

2. Planning your digitisation project
I think the first rule is to define what you want to do with your digitised images at the end of the project? Just reference images retained for identification? Or presented on a website? If so, do you want to offer viewers the ability to zoom in and read fine detail? Do you want to make prints? If so, a larger size capture image may be required in both cases. Questions like these will dictate both the level of equipment required and how to create a workflow which will meet all your needs.

Main issues to look at:

Image capture size v output size and resolution

Each image should be captured once at a size which is large enough to satisfy all requirements. For example if you wish to show the images on a website <u>and</u> want the ability to make paper prints, the image captured in the camera must be at a larger size, which in this case would be that needed to make a print. A website average image might be 800 pixels on longest side and we would generally view this on a home computer monitor which displays at 72 pixels per inch (ppi or dpi) but if printed this would be very small. A decent A4 print would need to be something like 3000 pixels on the longest side and printed at 300dpi. So to accommodate both needs we should capture at least at 3000 pixels on longest side and output the master image at 300dpi which would be resized for web purposes.

There is a case in capturing at the very largest a camera is able to achieve and keep this as the master image which is archived and preserved. This is ideal because larger images may be required in future for at present unknown applications. Always keep half an eye on long term storage costs however. Although storage is dramatically cheaper these days, the larger the image is, the more memory capacity is needed.

Memory storage is usually referred to in megabytes, gigabytes or terabytes. A simple way to work out the approximate mb size of the largest image your camera will produce and the memory needed, is as follows:
Eg. your camera is 12mega pixels (mp) the 12 mp is for each of three colour channels used to capture (red, green and blue) so 12 x 3 = 36 so… each image is 36mb

(1000 mb = 1 gb – 1000 gb = 1tb – so 1000 images at 36 mb each would need 36 gb of memory storage).

Cameras differ from make to make but setting the capture size/format is usually carried out in the camera 'Shooting Menu'. Ideally choose Raw format if available from the 'Image Quality' sub menu, this offers the best quality or Tiff or lastly Jpeg. If it's one of the latter two, set to 'Large' or 'Fine'

Camera

The most important part of a camera is its lens. This records resolution and sharpness of detail in the subject and is the one piece of equipment that should not be skimped on. A macro lens is designed to accurately copy flat artwork and is less likely to distort detail in an image. Camera mega pixel sizes (millions of pixels in each colour channel of an image) are very high now, even on mobile phones but more mega pixels do not necessarily mean more quality, just a larger size.

Tate use a 31mp camera with a very good macro lens for the Archives & Access project (images can be printed to poster size). Camera cost £15,000.
My Canon snappy camera is 12mp with a small lens and can make prints up to A3 size but the definition of detail will be low. Camera cost less than £100.

However a second hand Nikon D200 with 8mp plus a good macro lens will also make prints up to A3 at much higher quality. Camera cost about £400.
A good choice however, would be a new mid range full frame camera like the Nikon D600 including 60mm macro lens which might cost £1200 - £1600.

Image formats
Raw, Tiff, Jpeg? What are they? **Raw** is ideal for initial capture of the subject as it has most information which can be controlled and used afterwards for different purposes. The downside is that Raw files must always be processed through a 'Raw converter' and re formatted for all other purposes such as putting on a website or printing, this can be time consuming but probably worth it for the high quality. Also, Raw files are usually specific to manufacturers like Nikon or Hasselblad etc so if a manufacturer goes bust, inevitably its RAW format will quickly become obsolete making it dangerous for long term storage. **Tiff** on the other hand is a good **open format,** which means not dedicated to specific manufacturers and can be easily opened by many devices. Tiff files are largely future proof which makes them good for long term storage. **Jpeg** is a compressed image file which is good because it uses less storage space but some information is lost each time a Jpeg image is opened. Not noticeable to the human eye but over time the image will degrade.
So, Raw is good for capture, Tiff is good for long term archival storage and Jpeg is good for web use or printing sometimes.

Image tonal quality
This means making an exposure that is correct in order to achieve maximum detail in highlights and shadows present in a subject. Do tests to set an ideal exposure and use this setting consistently for a session of similar subject matter. The use of the 'automatic' exposure control does not always work as it is just an average and can be thrown by lots of highlight or shadow in a subject. Always set a 'manual' exposure.

Colour management
Colour management can be a complicated matter and open to some debate, however in its most basic sense it means to ensure that the colour of an captured image is as accurate and as consistent as possible when compared to the subject it is seeking to duplicate. One simple but effective step is to include in the first image of a session, a colour match card such as an 'X-rite colorchecker' (Formerly Gretag Macbeth) and when images are processed, a 'white balance' tool can be used in Photoshop to neutralise the mid grey tone on the card which is applied to all images in that session. Colour management also extends to computer monitors which are of course used to view, judge and control images, so they must be as accurate and consistent as possible. Monitor profiling kits can cost around £150 but are well worth the investment, and are very simple to use.

Image management
This means a system which allows several actions to take place to individual or batches of images, for example numbering, resizing, exporting in different formats, colour control (as above), adding metadata, cropping if necessary etc etc. These actions usually involve the use of a system like Adobe Photoshop but there are cheaper software systems which do similar things. I would however,

recommend the use of Photoshop if at all possible. There are further systems called Digital Asset Management (DAM) which go further in functionality like integration into collection management systems if appropriate.

Metadata
This is recorded data which in our case refers to an image and is collected either automatically by the camera (EXIF data) or manually added by Photoshop for example. It can often be a combination of the two. Metadata, although extremely boring is very important, recording essential information about how an image is created, when it's created and by whom, with what camera and at what aperture and exposure and so on. This information is a potted history of the image which remains always with the image and can be used as reference in the future, if the image is corrupted for some reason and needs to be repeated for example.

Cataloguing images
Although essentially an archive activity describing the subject being photographed, this action at its most basic is giving the image a number and/or a title which is then used to identify an image and its order/relevance within groups of images and the archive as a whole.

Computer equipment needed
Usually this falls into two categories, PC or Mac. Either will run software such as Photoshop. Some cameras can also be tethered to a computer to import images directly as they are photographed. A computer is also essential to run a flatbed scanner if being used. People who work with images tend to favour Macs as they are more intuitive to use with images and dedicated monitors tend to be easier to calibrate and profile to maintain consistent and accurate colour.

Lighting for photography
LED? Flash? Tungsten? Many archive subjects tend to be paper which can be damaged by heat, ultra violet or prolonged bright light. Of the three listed light types, LED is fast becoming a popular and affordable solution which is both cool to touch and UV free. Flash is colour consistent but can be expensive.

The standard technique for photographing a 2 dimensional object such as a paper document is with the camera pointing down located directly above the horizontal object, with a light either side angled at 45deg to the object. This way, no reflection of the light will appear in the camera image. Never use flash or an LED light attached to the hot shoe above the lens of the camera as the light will bounce back of the object into the camera and cause flare. Inexpensive copy stands can be purchased, some with incorporated lights.

Documentation
One easily overlooked detail which is important, is to document exactly what your actions are during the project. This would be in fine detail and concentrate on the workflow eg. equipment, formats and lenses. Explain exactly what you are doing and the physical procedures involved. Many projects change and evolve during the course of time so it's best to write the documentation as you go, or even towards the final part of the project. The documentation will help inform future

project planning. Also, many projects are ongoing so staff will change over a period of time so a project journal ensures ongoing consistency.

3. Preservation

Why?
To keep your assets safe for future use and reference and to avoid the need to re photograph or re scan your archive which would be expensive in time, equipment and effort.

How?
Store original digitised master images to maintain safety and accessibility for many years to come. Digital data is like food in that it will go off (Corrupt) if stored in the wrong environmental conditions. Digital data likes to be cool and dry and consistently controlled within small tolerances of change. It's also extremely important to make duplicate copies of the master images and store them in different geographic locations in case of fire or flood.

Storage media
This has changed dramatically over the last 20 years, Floppy disc, Jazz drive, Zip drive, PhotoCD, CD, Tape, External hard drive, DVD and Server. Now we have the Cloud and there will be more I am sure as technical progress moves forward. The Tate currently uses a combination of server with a double duplicate tape back up and previously CD and DVD. Whichever you choose, it's important to research new technology regularly and if necessary, migrate to different storage media to avoid obsolescence. The common view is that this process should take place every 5 to 7 years.

Formats
As mentioned before, it is essential to store master images in an uncompressed stable open format. An open format means it is widely used throughout the world, not tied to one manufacturer, can be opened and used by many devices and software systems so effectively is as future proof as is possible. There are many different formats used for data storage and there are some like Jpeg 2000 which are gaining popularity as a stable long term storage format, however at the moment I would choose Tiff as the ideal format for our purposes.

Conclusion
Hopefully this discussion has helped inform and show that setting up a digitisation system and the workflow need not be as daunting as sometimes imagined. Planning is the key area to concentrate on.