



The Photographic Collections Network Toolkit

Digitisation of Photography Archives – Scanning, Labelling, Information & Data Management

Contents

1. Introduction	2
2. Before you start	2
3. Planning	3
4. Equipment.....	5
4.1 Scanning versus photography	5
4.2 Buying a scanner	5
4.3 Scanner quality.....	6
4.4 Digitising using a camera.....	7
5. Scanning.....	8
5.1 Be Prepared.....	8
5.2 Scanning Workflow	8
5.3 Scanning tips	9
5.3.1 Scanning Photographs/Contact Sheets/Polaroids.....	10
5.3.2 Scanning Glass Plates	10
5.3.3 Scanning Slides and Film.....	10
5.4 DSLR film scanning	10
5.5 Additional Processes	11
5.5.1 Labelling / naming.....	11
5.5.2 Colour and tonal correction	12
5.5.4 Image Size / Resolution.....	13
5.5.5 Image retouching for dust, scratches and other faults	14
5.5.6 Metadata.....	15
6. Information Management & Data Storage.....	18
7. Further Information and Resources.....	20

1. Introduction

This toolkit is designed to give you the framework and skills to digitise your collection of photographic works and related ephemera. Do not consider this as a way to replace a physical or analogue collection. It is one of a series of tools through which to arrange physical and digital photographs and other images into an organised collection. It makes equal use of ideas of conservation, preservation* and digitisation.

*For clarity, "[preservation](#)" is stabilising a deteriorating artefact so it slows down the deterioration. "Conservation" is a more active and usually specialist intervention.

There are many reasons to digitise photographic collections:

- For the protection of light sensitive photographic materials so they are handled less frequently;
- For ease of searching through large quantities of material;
- Digitised archives and collections can become part of more widely accessible resources such as public online databases and catalogues.

It takes a certain kind of person to digitise a collection successfully. One can train oneself to act like such a person; or hire one if need be. But levels of care, of methodological steadiness, of patience and precision are essential and to be cultivated. Successful digitisation is a wonderful activity in its own right, as well as a useful underpinning to other activities. But a slapdash programme or a poorly thought-out one is worse than no help.

2. Before you start

It is important to consider these questions:

- Evaluate the material in your collection and the resources you have available for the project. Planning is really important.
- Decide what you need to digitise and develop a workflow specific to that. If you have a large collection it will be difficult to digitise everything.
- Are there any particularly fragile items? These might need to be treated differently.

- What equipment do you need; will you be scanning or photographing your collection?
- Will you be digitising the collection yourself or will you outsource services?
- Where and how will you be storing the digital files?
- Do you have an existing cataloguing system or will you be cataloguing the collection from the ground up?
- Are you going to undertake any conservation / preventative conservation while you are digitising and organising the collection?

This toolkit brings together advice from a number of sources. Every archive and collection is different so if you have any particular concerns that are not covered here then please seek advice. The PCN has ready access to further information on many of the topics outlined here.

3. Planning

It's important to evaluate the entirety of your collection before you begin. There are many solutions to digitisation depending on your time and budget:

- **Why are you digitising the collection?**

This is an important question to begin. Is the aim of digitisation to create a database of records or to create a collection of high-resolution print-ready files? In most cases a good solution is to have the capacity to do both and scan important items (i.e. finals or select files) at a very high resolution, and others at a lower resolution that will make the digitised collection easier to navigate and use.

- **How much time do you have?**

It takes time to scan and correct images. Be realistic about the amount of time you have and try and establish a streamlined workflow to ensure you get the most from your time.

- **What is your budget?**

Do you have the budget to buy new equipment such as a scanner or to outsource the bulk scanning of items to a company? Or can you make use of equipment you may already have such as a camera? Will you need to pay for software or any other equipment?

- **What skills do you have or do you need training?**

You may already have an understanding of the skills required for digitisation and data management but additional training might be useful. There are some links for training at the end of this document. Be realistic about the match of your skills to the skills required.

- **What additional data will you be adding to your files?**

You will probably be adding metadata (data about data) to your files through the scanning process or through the use of software such as Adobe Bridge, Photoshop or Lightroom. It is important as far as possible to build in from the outset all the types of data which may be required later, even if some are not an immediate priority. What data will you be adding? Is there information about intellectual property rights that you should be including?

- **Will you be building a digital database or rehousing the collection?**

How will you organise the digital files and physical items once you embark on this project? This includes developing a system of labelling. Depending on the number of digital files you have, you might want to turn this into a database, or develop a system of organisation on a hard drive or server. Methods of preventative conservation will be covered in another toolkit.

- **How will you back up the digital files?**

Once you have your digital files it's important to consider that they need looking after as your original images do, although in very different ways. Do you need to back them up? There are different options for this.

4. Equipment

4.1 Scanning versus photography

In this toolkit we are going to focus on scanning as the primary method of digitisation. With some scanners you will have the flexibility to scan prints, contact sheets, negatives, transparencies, slides and glass plates. So, if you have a variety of materials it will be worth checking that the machinery you use allows the flexibility you need.

If you have items that are particularly delicate or sensitive it will be better to capture them using photography; pages of photo albums or daguerreotypes, for example, are better suited to being captured as an object rather than as an image.

4.2 Buying a scanner

When buying a scanner ensure that it meets your needs. You will need to consider:

- What size scanner do you need?
- At what resolution does it scan? You will need a high-resolution scanner to be able to digitise slides and film. Look for a “native” resolution of around 4000dpi or more.
- Is it compatible with your computer?
- Does it have the capability/adaptors to scan other materials like slides, film, plates etc.
- Does it include software, or do you need additional software? And if so, does it allow you to undertake corrections such as colour, dust and scratches, or add metadata?

Here is a useful buying guide for a scanner:

<https://www.bhphotovideo.com/explora/photography/buying-guide/film-scanners>

4.3 Scanner quality

Flatbed scanners are a well-developed technology and even relatively inexpensive flatbeds can provide good results. The differences are speed, so-called “native” resolution, and “noise” – that can be visible if the contrast is increased. The poorest flatbed scanners can have varying brightness across the scanning area, or noticeably reduced sharpness, and these are to be avoided.

For scanning film there are significantly different levels of quality dependent on price. To explain why, consider a very dark or “dense” piece of film, say a slide that has been underexposed. If you hold it up to the ambient light, you can see some of what is in it. But if you shine a bright light through it, you can suddenly see a lot more. Photographic film hides a lot of information in these dense areas. Worse scanners will struggle to “reach into” the densest area of a slide or negative. Film scanners are designed to achieve a particular density range, often labelled D-max; and the higher this number, the better (though it doesn’t tell you anything about the scanner’s other qualities such as sharpness and evenness).

How flat does a scanner hold the film? This is another critical area. Even quite expensive scanners often have disappointing results in this regard.

The “bit depth” of the scan is also important for film scanners. This defines the number of brightness levels the scan is converted into. This is important if you have a very low contrast negative. A high-bit scan allows you to “stretch” the tones without introducing posterisation – visible jumps from one brightness level to the next.

Resolution and sharpness are important if you are scanning small format film for a large output. The “native” resolution is a key figure – this is the actual resolution of the scanner rather than an interpolated figure. Like D-max, this figure is only a partial measure of quality.

Infra-red scan. Some scanners carry out a separate scan of the infra-red channel and this can then be used to remove dust and scratches very effectively.

There are many good websites giving you more information, but it's worth drawing attention to the different types of scanner:

- **Flatbeds.** Some flatbed scanners can scan film materials. The Epson V series (V550 to V850) and Canon 9000F Mark ii are probably the best known low(-ish) cost flatbed film scanners at the time of writing. They will give you a useable scan, if slightly lower in quality than most dedicated film scanners. On the plus side, you can scan almost any shape and size of film.
- **Film scanners.** Good quality dedicated film scanners are made by Plustek, Reflecta and Nikon. One brand is worth singling out: Hasselblad Flextight scanners (previously called Imacon). Rather than trying to flatten the film, they use its natural curve. They are expensive (well into five figures) but among the best.
- **Drum scanners.** For the very best scans and important material you might seek out a drum scanner – or more likely send film away to be drum scanned, as it's a tricky professional process.

4.4 Digitising using a camera

If you need to digitise images within books or photographic objects you may need to photograph the items rather than scan them. To do this you will need:

- Camera
- Two lights
- Backdrop
- Tripod
- Table
- Remote shutter release switch

You will need a slightly different workflow for this method.

In recent years it has become viable to digitise negatives and slides using a digital camera. For more details on that see below under DSLR film scanning.

5. Scanning

5.1 Be Prepared

If you will be digitising your collection yourself, to begin ensure you have the correct materials and equipment:

- Gloves for the protection of your photographic materials (these can be cotton or nitrile)
- Pencil for notes (do not use ink)
- Chinagraph pencil if you need to mark glossy surfaces. Can be removed with a paper towel
- Acetate to cover the scanner bed if you are scanning transparent, fragile or sharp items
- Polyester pockets (or any other transparent conservation material) might also be useful for scanning delicate materials without handling them
- Masking tape in case you need to secure items
- Brushes, microfibre cloths, staple remover and air dust blower might be useful but not essential.

5.2 Scanning Workflow

Below is a very basic workflow that you can follow:

1. Prepare the materials for scanning: remove dust from the scanner bed and the item(s) you are scanning
2. Scan
3. Naming/Labelling

4. Undertake any post-production required: colour correction, cropping, dust and scratch removal
5. Add additional metadata if required
6. Save files as a high resolution uncompressed TIFF and a lower resolution compressed JPEG. This gives you options for future use.

Note: you may be able to undertake points 2 to 5 in one action using scanning software.

5.3 Scanning tips

- Scan in the order that you retrieve items.
- Be prepared to scan multiple items in one pass of the scanner as this will save time. This might add some post production if you'd like to separate files afterwards.
- Think about the maximum size you are likely to need your scans. Although it's tempting to go for the best possible settings on the scanner, this will slow you down and might not be necessary. It may make more sense to scan more material at a smaller size unless it is an important item. If you're not sure, a good compromise is to scan at a size that allows the image to be printed to A4 size. Printers often want a resolution of up to 300 pixels per inch – at that resolution, the pixels will be effectively invisible. So go for A4 at 300 ppi, which is equivalent to a digital file size of 25MB. There's more on ppi and resolution below.
- Always scan as an uncompressed TIFF file; you can compress to JPEG afterwards. The JPEG format achieves compression by throwing away information (so-called "lossy" compression), slightly degrading the image – so it's not a good primary format for storing scans. You do not in general want your computer choosing which information you can afford to lose.
- If you are short on space, it is also possible to use LZW compression with TIFF files. This is lossless compression (i.e. the image is not degraded) that can still reduce files sizes by 30-50%. Don't use any other type of compression on TIFFs.

- The shiny side of transparent materials should not be in contact with the glass scanner bed during a scan, as this introduces some unwanted patterns called Newton's rings. This can be avoided using adaptors that hold the film a little away from the glass; or acetate or "anti-Newton" glass between the film and the scanner glass.
- If there is anything alongside your photographic material that holds additional information, like the back of a photograph, or a sleeve, scan this as well. This is what metadata used to look like before we coined the word. It will be so useful for archivists and curators in the future!

5.3.1 Scanning Photographs/Contact Sheets/Polaroids

- Scan multiple items at once if you can
- Place items within a polyester sleeve if fragile

5.3.2 Scanning Glass Plates

- Scan emulsion side down
- Use acetate so that the glass doesn't come into direct contact with the scanner bed (to avoid Newton's Rings)
- Be careful when covering the plate in case of breakage

5.3.3 Scanning Slides and Film

- You will need to use an adaptor for your scanner so that the film doesn't come in direct contact with the scanner bed.

5.4 DSLR film scanning

With recent improvements to digital cameras, many people have started to digitise negatives and slides using a digital camera. If you have the right equipment and knowledge, this is usually quicker and sometimes better quality than using a scanner.

You will need:

- DSLR Camera and remote shutter release
- Macro lens
- Lightbox
- Tripod that allows you to point the camera downwards

- Something to hold the film – an old film holder from an enlarger or scanner, or some Anti-Newton glass.

If you are photographing negatives, after the “scan” the next part of the process is to convert them to positive using software. It takes a fair bit of skill and experimentation but once you have established a good workflow it can be much faster than scanning. If you are interested in this technique Google “DSLR film scanning.”

5.5 Additional Processes

Through the scanning process you may also be able to undertake corrective processes, add metadata and label files at the same time.

There is more detail about the processes below, though not all of them will be necessary in every case.

5.5.1 Labelling / naming

The labelling of your digital files is really important. Develop a convention that you will use throughout your archive/scanning process. Scanning software can set the file name automatically so that you do not have to change each time. Make this as detailed as you need (within reason) depending on the amount of information you have. It is much easier to do this while you scan than to change later.

Items to include are:

- DATE (either made, scanned or printed – I would advise writing this as YYYYMMDD or YYYYMM or YYYY)
- PROJECT / IMAGE NAME (you could also give projects a code)
- PHOTOGRAPHER
- FINAL (if this is a ‘final image’)
- FILE NUMBER

An example of this could be:

1985_BlackpoolTower_JohnSmith_Final_1

Try to have no breaks in the file name, and avoid punctuation apart from hyphens or underscores. There is always a dot (full stop) then the three or four-character file format at the end – this allows any computer to access the file. So the above example would be stored as:

1985_BlackpoolTower_JohnSmith_Final_1.tif

Technically the maximum number of characters in file name is 255, but they get a bit unwieldy over 70-80 characters.

5.5.2 Colour and tonal correction

If you are scanning from prints, and using a calibrated monitor, it is fairly easy to tell visually if the scan looks similar to the original print. It might then be possible to make corrections. However, when scanning large numbers of photos for a collection, it's better for speed and quality to get your scanner set up to give fairly accurate colour and tones, and limit any tweaking to small corrections. The exception is if you know you are going to print or send out a file.

5.5.3 Bit depth

Although many scanners operate at a higher bit depth for quality (see above), best practice would be to ensure your scans end up as a minimum:

- Colour: 24 bit RGB colour
- Black and white: 8 bit greyscale

RGB means “red green blue” – in other words three colour “channels”. Each pixel has a red, green and blue component. Combined, these can make up most visible and almost all printable colours. Printers convert the RGB file to CMYK (cyan magenta yellow black), to match the inks; however this means a file size is a third bigger; and the conversion process also affects the colour information in the file. Always scan and store as RGB.

5.5.4 Image Size / Resolution

Terms used:

- PPI: Pixels per Inch: the size of your image
- DPI: Dots per Inch: this refers to how your image is printed

As stated above, when scanning, focus on the final image size. In other words, select your dpi resolution to end up with a file at the size you need. There is little point in scanning at too high a resolution, as it's slower, takes up more space, and might not improve quality. If you scan at too low a resolution, the quality will be reduced.

Printing out images normally needs files that are around 300ppi at the printed size. In other words, an image that is reproduced 1 x 1 inch would need to be 300 x 300 pixels to be printed at optimum quality; 2 x 2 inches needs 600 x 600 pixels – etc.

I would recommend scanning your material as an **uncompressed (or LZW compressed) TIFF at A4 size (or as close as possible), with another compressed JPEG copy.**

So a 35mm slide or negative, which is roughly 1 x 1.5 inches, needs to be scanned at around 2400dpi to be printed at A4 size, or at about 3600dpi to reproduce at A3 size.

An enprint (6x4 inches), or half-plate negative, should be scanned at 600 dpi to reproduce around A4 size.

For simplicity, a helpful indicator for scanning is megabyte size (MB) which for a good quality TIFF would be:

- A4: 25MB

- A3: 50MB

Once you have scanned all of your material as a TIFF, you can if you like bulk convert your files to JPEG so that you have a lower resolution copy. To do this you can use the Image Processor in Adobe Photoshop. More information can be found [here](#).

5.5.5 Image retouching for dust, scratches and other faults

This is optional: if you would like to clean up dust and scratches you can undertake this either through your scanning software or in post-production in image editing software such as Adobe Photoshop. You can follow a tutorial [here](#).

If you find yourself having to retouch a lot of dust spots, there is a clever Photoshop shortcut [outlined here](#) that can be used very quickly.

There is interesting debate on the extent to which dust, scratches and faults should be corrected for scanning of photography archives. In the era of film, most professional photographers would have expected to remove dust and scratches before showing a photograph in public or selling it, and would be disconcerted if they thought that later scans of their work would include these “faults” as an indication of authenticity, as sometimes happens now. However studios then frequently went on to carry out far more significant retouching of backgrounds and even image features.

Perhaps a reasonable compromise now is to retouch dust that might have been introduced during storage or at the time of scanning and minor spots or scratches if they are typically retouched on vintage prints. However major faults such as folds, water damage and other deterioration might be left in as an indication of the state of a work. It is not appropriate to cover every eventuality here; but just to be aware that there is a complex debate in this area and that many possible treatments have their adherents and opponents.

5.5.6 Metadata

Metadata is data that describes other data. With digital image or other media files, metadata is stored in text format within the file, but not part of the visible image. It can be added automatically by certain software and equipment (cameras and scanners) or added by users for purposes of captioning, cataloguing and searching.

In the case of digitised images it can typically be used to store descriptive information about the photograph, its creator and current status e.g. copyright, as well as aiding the search process.

It is possible to add metadata at any time to a media file, for example while scanning, and/or afterwards using software such as Adobe Photoshop, Bridge or Lightroom or through cataloguing software. Decide on your scheme and framework depending on your use. It is unusual to fill in all possible fields (i.e. boxes). As a minimum **we recommend you fill in information such as the date, creator or author, and copyright notice**. Any other descriptive information will help you or others find the image in a search, and know what can be done with it. The most commonly used fields are **title, description, keywords, credit and copyright status**.

Good metadata is very important for digital archives. Without good descriptive information it's very difficult for an image to be found later. Search tools generally default to look through the complete metadata, to find text matches, but can be set to search only particular fields. So it's good practice to put the right sort of data in the right fields. Try and imagine someone searching for this image, and what terms they might type in.

Description and keywords are commonly-used fields. The description should be a full caption describing what an image shows. Then you can add keywords to refer to important visible aspects of the photo, or as synonyms of words in the

description. An example description might be: "Film still from the 1952 movie Singin' in the Rain. The actor Gene Kelly, as the character Don Lockwood, swings round a lamppost holding a folded umbrella in the pouring rain." Then your keywords for the same image might be: singing, street, light, streetlight, lamp, post, doorway, night, nighttime, downpour, raining, film, star, musical, dance, dancing, broly, smile, smiling, raincoat, hat.

With that metadata added, the following text searches will bring up this image (among others):

Film star rain

1952 Musical

Movie streetlight night

But if you wanted to find an image of Gene Kelly not smiling, you could search for:

Gene Kelly -smiling

The minus in front of the word smiling would mean a search would exclude this image, since the metadata contains that word.

Again, keywording is an area in which different positions have their supporters. Different libraries and archives have different conventions in this respect – some will add in the shape of the image, the dominant colours, and the moods depicted. But the aim is always the same, to find as many images which fit the search as possible, and as few as possible which don't.

Metadata Standards

There are different standards of Metadata. The most commonly used are Dublin Core for archive/cataloguing and IPTC for photographs. Each of these uses a particular subset of fields. Sometimes, confusingly, different software has different names for the same data field. For example what Photoshop calls

the Document Title, Headline and Description, the cataloguing application Media Pro calls the Product, Title and Caption.

For more information on metadata see www.photometadata.org and in particular [this page](#) lists all the fields and how they should be used.

More information about the types of Metadata

What follows is an outline of technical vocabulary that will not be useful to all. What is useful is to know that specialists regard these distinctions as important:

- **Descriptive Metadata** enables identification, location and retrieval of information resources by users, often including the use of controlled vocabularies for classification and indexing and links to related resources.
- **Technical Metadata** describes the technical processes used to produce, or required to use a digital object.
- **Administrative Metadata** is used to manage administrative aspects of the digital object such as intellectual property rights and acquisition. Administrative Metadata also documents information concerning the creation, alteration and version control of the metadata itself. This is sometimes known as meta-metadata!
- **Use Metadata** manages user access, user tracking and information regarding multiple versions.
- **Preservation Metadata**, amongst other things, documents actions which have been undertaken to preserve a digital resource such as migrations and checksum calculations.

6. Information Management & Data Storage

Through the digitisation process you will be amassing a large number of digital files that will need to be stored or organised. In this regard, develop a system that makes sense for your time, budget and skills. Here is a useful resource for naming and organising files from the [University of Cambridge](#).

Organising Files

You will have set up a naming convention while scanning or documenting your files. It is also important to have a wider system for the entirety of your archive. It is important that you:

- Are Consistent, Clear and Concise
- Try to avoid too many hierarchical levels

Any arrangement of your files will be dictated by the collection itself and how the material is organised in its physical state. Try to keep the digital files together as they are together in their physical state.

Depending on the nature of the collection, as the top level of organisation one might use either:

DATE or CREATOR

You will then need a folder structure within this top level of organisation, usually organised by project, photographer or box number. This to suit whether the archive is of the work of a single photographer or a collection of the work of several.

You may wish to also create a sub category of different media.

```
1986 > 1986_PROJECT-NAME > BOX 1 > 1986_PROJECT-  
NAME_PHOTOGRAPHER_1.TIFF
```

Databases

There is a range of database or collections management software that you could use to build a collection:

- Databases: These are applications for organising files. If you have a good organisational system on your storage system then this may not be required but databases are a useful tool particularly if you'd like to make information public in the future through a website catalogue. Off-the-shelf solutions are applications like Filemaker Pro, or you can make use of an open source and free to use system such as [Omeka](#), [Collective Access](#) and [Open Archive](#).
- Collections Management Systems: If you have a large collection of objects that need to be tracked then it may be useful to use a collections management system as used in museums, galleries, libraries and archives such as TMS and QI. The Collections Trust has a useful resource for selecting the right [software](#).

Be aware that databases and their file formats come and go; and once they are obsolete, it will be hard to access your data. Choose a file format with a track record.

Digital Preservation

If you are not scanning a large collection I would recommend scanning to a good quality hard drive, and ensure that you have two copies of the drive which are stored in two locations. Other methods you could follow could be to:

- Set up a server in your location
- Cloud based storage
- Tape back-up: this will create an analogue back-up of digital files.

For more information about digital preservation see: [Digital Preservation Coalition](#)

7. Further Information and Resources

You could also make use of the resources that are available through the following organisations and groups:

[The Collections Trust: Resources and Guides](#)

[Jisc: Resources and Training](#)

[DCMS: Culture is Digital Strategy](#)

[Museums Computer Group](#)

[Digitisation at the National Archives, 2016](#)

[University of Leeds Digitisation Toolkit](#)

[Copyright in Photography Factsheet](#)

[Tate Archives & Access Toolkit](#)

Written by Kay Watson, 2017
Editing by Paul Herrmann and Francis Hodgson
© The Photographic Collections Network

Kay Watson

Kay Watson is a researcher, curator and archivist based in London. She is currently Assistant Digital Curator at Serpentine Galleries and a PhD student at Birkbeck, University of London examining distributed modes of programming in photography organisations and institutions in the 1980s and 1990s.