DIGITISATION GUIDELINES

GIMP 2.8 Version
2nd edition

Contents Page

Introduction

What is digitisation?

Digitisation is the conversion of analogue material into a digital format to be made available through a variety of media. Digitisation in the heritage sector enables the preservation of materials and increases access to collections through publication on the Web.

What is the purpose of these guidelines?

This guide has been created as an overview to the various components that make up a digitisation project. They include selection and preparation of collection materials, copyright clearance, creation of metadata, scanning, web delivery, digital archiving and preservation.

The People's Collection

The People's Collection is the Welsh Government's commitment to providing an interactive digital resource that celebrates the life, heritage and culture of Wales. The website provides an easily accessible platform for material digitised by both repositories and the wider public and allows access to collections from across Wales.

www.peoplescollectionwales.co.uk

Preparing to Digitise

Before you begin to digitise it is worth giving careful consideration to the material digitised as this may affect the duration and cost of the project.

Selection of Materials

This is the first step in the creation of a digital archive. It identifies items or collections for digitisation based on criteria such as:

- Historical merit – Their relevance and importance as historical items may influence whether the use of resources is justified
- Demand – Material that is particularly popular may benefit from being digitised so that it can be viewed online or by several people at once
- Condition of originals – If the originals are fragile, digitisation may prevent further damage by reducing handling. Conversely, if the material is too fragile, it may not be able to undergo digitisation.
- Feasibility of image capture – Is it within your technical capabilities to digitise and is the necessary equipment available?
- Rights – Do you own the rights to the material that is to be digitised? No material should be digitised without first clearing any associated rights.
- Potential financial benefit – If you own the copyright, digitisation could provide income via print or online.
- Human resources – Do you have the necessary resources or will volunteers or new staff need to be recruited?

Keeping Track

Rather than attempting to digitise your entire collection and storing the images as such, it is advisable to break it down into manageable batches of no more than 100 items each. Each batch can be named or numbered depending on preference.

Equipment

Any digitisation project will require some essential pieces of equipment:

HARDWARE: The most common piece of equipment for digitisation in heritage repositories is the Flatbed Scanner. They allow the digitisation of photographs, correspondence and slides and also transparencies if using a transparency hood/attachment. Scanners come in sizes to manage A4 and A3 originals. You get what you pay for; low cost scanners are unlikely to provide suitable results so it is better to invest in the highest quality within budget.

SOFTWARE: Image management software is essential within a digitisation project. Most commonly used is Adobe Photoshop which comes in a number of versions. There is also a stripped-down version for less demanding uses called Photoshop Elements. Open-source alternatives such as the GNU Image Manipulation Program (GIMP) are also available.

Module 1: Copyright

Aims and objectives

This module aims to explain the importance of copyright in the digitisation workflow and to make you aware of the main issues and problems that can be faced when embarking on the digitisation of any collection.

Please note that this module does not give any legal advice; it is merely advisory and legal advice should be sought for more specific copyright or intellectual property rights issues.

The information given here is extracted from various sources and reproduced with permission.

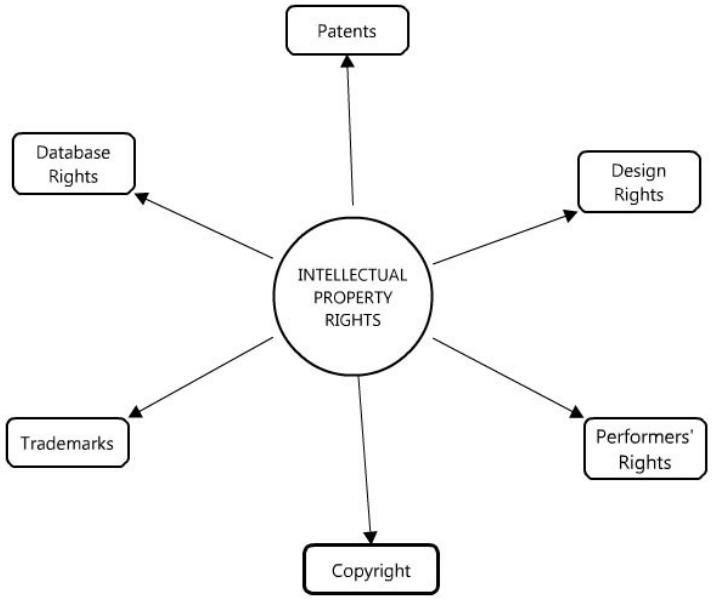
The module will provide:

1. an introduction to intellectual property rights (IPR) and copyright
2. an idea of where copyright issues should stand within the digitisation workflow
3. some tips and advice on how to go about clearing copyright
4. some practical tasks to identify the rights owners in various items
5. an example workflow for clearing rights

What is intellectual property?

'The term "intellectual property" covers a range of legal protections for things created by the human mind.' (JISC IPR Toolkit, 2009)

These include:



The type of intellectual property that you are most likely to encounter when digitising historical material is **copyright**.

What is copyright?

- Copyright is a *property right* in certain types of *work*
- Copyright is automatically owned when the work is created – it does not need to be registered
- The work must be the original product of skill, labour and judgement by the author
- The copyright remains in the work for a *fixed duration*
- Copyright determines that the owner controls the use and reproduction of a substantial part of the work
- Copyright cannot be applied to facts or ideas
- Copyright is subject to certain, defined, exceptions

What is a 'work'?

- A literary work is composed of words or symbols – there is no requirement of literary quality – it could be a business letter or a novel
- An artistic work may be 2D or 3D and can consist of text. Again, there is no requirement of artistic quality
- Maps, charts & plans are protected as artistic not literary works
- There are several other kinds of work including music, dramatic works, films and sounds recordings
- A single item, such as a film or piece of music, may be composed of more than one copyright

How do I know if a work is in copyright?

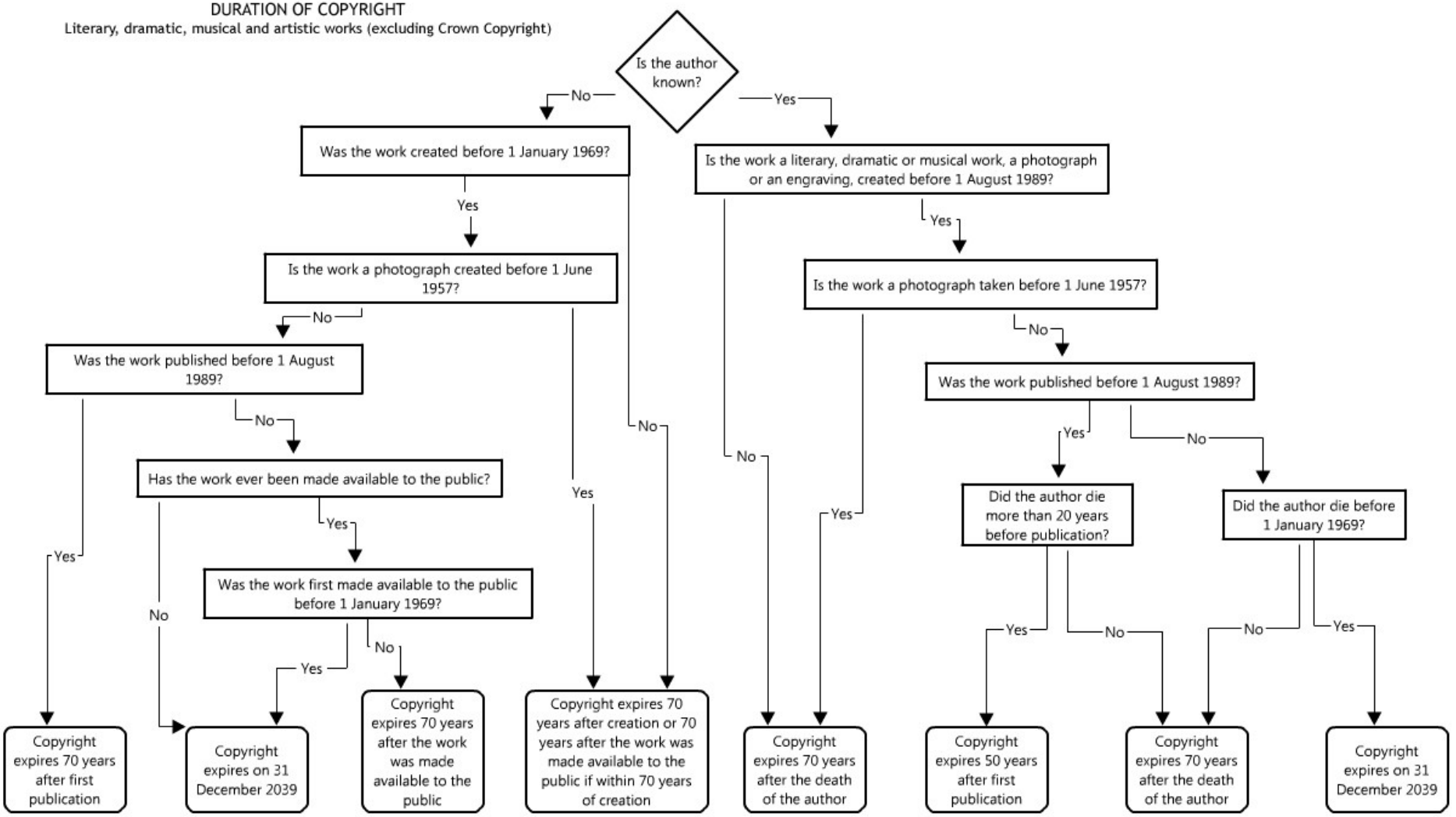
There are many complicated rules regarding the duration of copyright. It is generally assumed that the lifetime of copyright is 70 years after the death of the creator but this varies greatly and each item must be considered individually.

Unpublished material

- Works created before 1 August 1989 remain in copyright until 31 December 2039
- Works created after 1 August 1989: copyright lasts for life of the author plus 70 years

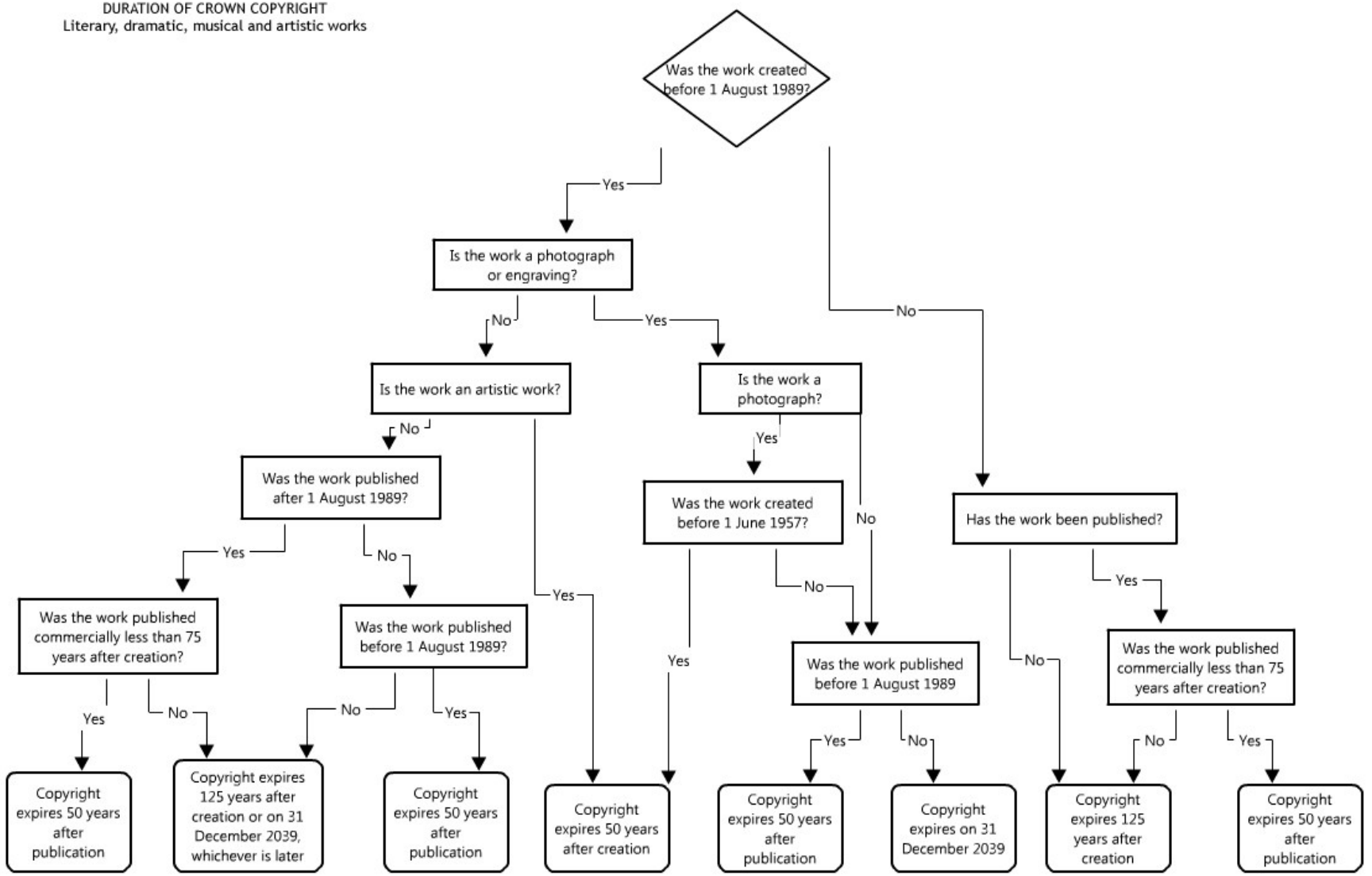
The **Copyright Duration Charts** on the next pages are useful references in determining the duration of copyright of a work.

These charts are the work of Tim Padfield, Information Policy Consultant and Copyright Officer for The National Archives, and have been reproduced with his kind permission.



Reproduced courtesy of Tim Padfield, The National Archives

DURATION OF CROWN COPYRIGHT
Literary, dramatic, musical and artistic works



Reproduced courtesy of Tim Padfield, The National Archives

The importance of copyright in digitisation

The digitisation and publication of material on the Web has complicated the issue of copyright. The availability of the digitised item and the potential for reproduction and commercial exploitation means that the usual exceptions for archives and heritage institutions no longer apply in an online environment.

Copyright should be central to any digitisation project because when an institution embarks up on such a project, it essentially becomes a publisher.

Restricted use of copyright material includes:

- copying
- publication
- communication to the public by wire or wireless means (incl. the Internet)
- exhibition (but not of artistic works)

Copyright within the digitisation workflow

Copyright infringement occurs at the point on which the work is copied. Therefore, any copyright issues must be resolved and permission granted from rights holders for the digitisation and publication of the material *before* it is digitised.

How to go about clearing copyright

1. Allow plenty of time for the rights clearance process, it can be time-consuming
2. Take a look at the 'Duration of Copyright' chart to determine whether copyright applies to the work
3. If the work is still in copyright and the rights holder is known, take the following steps:
    a. Identify the rights that will require clearance, and for how long the rights will be needed.
    b. Contact the rights owner and request permission for the reproduction of the material, clearly stating the ways in which you intend to make it available. This permission should always be received in writing before the material is reproduced.
    c. Examples of rights clearance letters and permission forms can be found on the Strategic Content Alliance (SCA) website: http://sca.jiscinvolve.org/ipr-publications/

4. If the rights holder is not known, the work is an 'orphan work' and a different process must be carried out:
    a. Attempt to locate the rights holder. You should carry out reasonable searches and document any correspondence or search terms used. This is known as 'due diligence'.
    b. If the rights owner is found, request permissions as above
    c. If no rights owner can be traced, a risk assessment should be made as to the risk of digitising and publishing the material.
    d. 'Mitigating' strategies such as 'Notice and Take-down' policies should also be implemented.

Displaying copyright online

A copyright statement should be included alongside images displayed on a website. This should include the copyright symbol, the name of the rights holder and the start date of the copyright. For example:

© National Library of Wales (1993)

Copyright policies

Copyright policy or statement – It is a good idea to include a statement of how copyright was dealt with in the digitisation process. It should stress that 'best endeavours' were made to discover the rights holders of the items on display. In instances where the rights holder could not be found or the work was unattributed, the institution requests that anyone with information regarding the item contact them. Anyone believing that they are the rights holder should follow the 'notice and take-down' policy (below).

'Notice and take-down' – This is the procedure whereby notice is given to an organisation or institution that unlawful content, such as copyright infringing material, may be displayed on their website. The material can then be removed, thus reducing liability for making the content available. A clear 'Notice and Take-Down' policy should accompany a copyright statement in order to ensure that aggrieved rights holders can quickly seek redress.

* These should not be viewed as providing a legal defence. Publishing material where the rights holders have not been consulted is infringement of copyright and illegal.

Licensing

One of the primary concerns when publishing digital resources on the Web is that other users can copy and use them without the copyright holders' permission.

As it is virtually impossible to prevent users from using the resources if they are determined to do so, one response has been to use licences as a way of letting others know exactly under what terms the digital resources can be reproduced (e.g. non-commercially, with credits to the licensor, etc).

The benefit for the licensor is that each resource is clearly labelled and licensed and there can be no excuse for misuse. Licensing also cuts down on administration time and costs.

There are many different types of licence schemes, the most recognisable being Creative Commons and the Creative Archive Licence.

Moral rights

Co-existing with copyright in most images are the moral rights of the creator or author. They include:

- the right to be identified as the author or director
- the right to object to derogatory treatment of a work
- no false attribution of a work
- the right to privacy of certain photographs and films

Moral rights last as long as copyright, except the right to object to false attribution, which only lasts for 20 years after the death of the creator or author.

Module 2: Metadata

Aims and objectives

This module provides an introduction to both technical and descriptive metadata. At the end of this module you should be better able to record technical metadata for preservation and archival purposes and descriptive metadata.

What is metadata?

- Metadata facilitates discovery both internally within an institution or database and externally, allowing users access to collections.
- Metadata is often described as 'data about data'
- Metadata can be 'intrinsic', existing within the resource in the form of technical information, or 'extrinsic' or 'explicit', being created to accompany the resource in the form of more descriptive information.

> "Metadata is key to ensuring that resources will survive and continue to be accessible into the future." (NISO, 2004)

Why do we need metadata?

- Metadata facilitates discovery both internally within an institution or database and externally, allowing users access to collections.
- Without metadata, a resource can lose its meaning or significance and important information is lost.
- Metadata should be a primary consideration within any digitisation workflow and is as essential as technical standards to the production of a valuable digital resource.

2.1: Technical metadata

What is technical metadata?

Technical metadata is information about how a digital object was captured, its format and storage etc. Accurate technical metadata helps a repository manage digital objects over time and keep them usable as well as keeping a record of the copyright in the digital image itself. Digital materials, including digital images are classified as literary works because they are constructed of binary code. It is suggested that you at the very least record the fields below for batches and individual items.

BATCH DETAILS	
Repository name	
Digitisation batch ID	
Digitisation Staff	
Date scanned	
Current location	
Back up location and format	

ITEM DETAILS	
Capture device (make and model)	Epson 10000XL flatbed scanner
Capture details (software used)	Epson Smart Panel
Manipulation History	Master files created in RGB TIFF format, opened in PhotoShop 7.0, converted to Apple RGB, checked for sharpness, rotation, skew, histogram, levels, curves and colour balance.
Resolution of Master	300 ppi at 100%
Compression	Uncompressed
Bit depth of master	24 bit
Colour profiles	Master file has RGB profile embedded and web file has sRGB profile embedded
Greyscale Patch	Macbeth Greyscale

2.2: Descriptive metadata

What is descriptive metadata?

- Descriptive metadata allows the identification, location, retrieval and management of resources.
- It should answer the questions: "What? Where? Who? When?"
- It should also provide information that allows you to place it with other similar resources.

Metadata for the digital age

You may be familiar with metadata for cataloguing and archival purposes but for digitisation DESCRIPTIVE metadata is key. Descriptive metadata accompanies the digital image and provides information and context for the user making it a digital resource, not just a digital file. The more information provided in descriptive metadata, the more useful a resource it becomes for the user.

Standards

Metadata should use a common structure, a set of fields that can be stored on paper, in a spreadsheet, or in an electronic file.

DUBLIN CORE is the most commonly used standard for web and descriptive metadata. It is a fairly open structure which provides a basic subset of fields which leave room for interpretation by individuals. There are fifteen main fields:

1. Title
2. Creator
3. Subject
4. Description
5. Date
6. Rights
7. Identifier
8. Coverage - Location
9. Publisher
10. Contributor
11. Type
12. Source
13. Format
14. Relation
15. Language

But we consider the following as essential:

1. Title – a meaningful title which tells the user what the item is
2. Creator – the original author or creator of the item
3. Subject – subjects related to the contents of the item
4. Description – the more detailed, the better!
5. Date – the date associated with the item, usually the creation date of the original
6. Rights – all information regarding rights issues
7. Identifier – the accession / catalogue number
8. Coverage / Location – the place-name most commonly associated with the item

Please note that the unique **filename** of the digitised item must also be included. This is essential to tie the metadata to the image.

There is more information in the examples that follow.

2.3: Some examples of descriptive metadata

Example 1: A Photograph

Dublin Core Element	Example: Photograph	Explanation
Filename	gtj11507	The unique alpha-numeric filename given to the digital image
Title	The Welsh Language Society's first protest on Trefechan Bridge, Aberystwyth, 7 January 1963	Should answer **who**, **what**, **where**, **when as effectively as possible.** This is the information that will allow people to find the item so it's important to be as comprehensive as possible.
Creator	Geoff Charles	The original creator of the item – the 'author'. If not known, enter 'Unknown'.
Subject [Tags]	Demonstrations (LOC), Cymdeithas yr Iaith Gymraeg, Protest (GTJ), Welsh Language (NLW), Societies (NLW), Politics (NLW),	Tags are words associated with the item and link items together online; they aid the search process. They can be taken from subject vocabularies e.g. Library of Congress [LOC] or can be made to suit your own needs as seen here. They should be separated by commas.
Description	Cymdeithas yr Iaith Gymraeg was established in August 1962 at a meeting held during the Plaid Cymru Summer School in Pontarddulais. The society held its first official protest on 2 February 1963 in Aberystwyth. Some 40 members and supporters sat in the middle of the road on Trefechan bridge, stopping the traffic for half an hour.	This should be any contextual or background information to add to the item. It could also be a transcript of text that is difficult to read, or of an audio or video file etc.
Date	1963-01-07	If the exact date is not known the following formats should be used: Nearest century: 19??-??-?? Nearest decade: 191?-??-?? Date range: 1939/1945 If partial date is known e.g. month and year but no day: 1922-06-??
Rights	Copyright: The National Library of Wales	In this instance, copyright has been assigned to the NLW, any information on rights holders and the terms of use should be included here.
Identifier	GCH19031	This is the repository/accession number
Location / Coverage	Trefechan, Cardiganshire	This should include the place name and the pre-1974 Welsh county (and the country if not Wales). The place name should be taken from Elwyn Davies's *Gazetteer of Welsh Place-Names*.

Example 2: Multi-part Items - Items with several pages

Dublin Core Element	Example: Book	Explanation
Filename	gtj10510	The unique alpha-numeric filename given to the digital image In this instance, this would be the Parent ID [see explanation below]
Title	Letter from Daniel Horton Davies to his wife, on his release from a German prisoner of war camp, 2 December 1918	Should answer **who**, **what**, **where**, **when as effectively as possible.** This is the information that will allow people to find the item so it's important to be as comprehensive as possible.
Creator	Daniel Horton Davies	The original creator of the item – the 'author'. If not known, enter 'Unknown'.
Subject	First World War [1914-18], World War One, WW1, Soldiers, Soldier, Prisoners of War, Letter, Letters.	Tags are words associated with the item and link items together online; they aid the search process. They can be taken from subject vocabularies e.g. Library of Congress [LOC] or can be made to suit your own needs as seen here. They should be separated by commas.
Description	The letter reads as follows: 'My Dearest Edyth I daresay you have received the telegram I sent yesterday telling you that I am free once more. I might be in England before this letter but I dont know yet so I am going to be on the safe side by sending it. We (1200 of us) left our camp...[continues]	This should be any contextual or background information to add to the item. It could also be a transcript of text that is difficult to read, or of an audio or video file etc.
Date	1918-12-02	If the exact date is not known the following formats should be used: Nearest century: 19??-??-?? Nearest decade: 191?-??-?? Date range: 1939/1945 If partial date is known e.g. month and year but no day: 1922-06-??
Rights	Owner: Ceredigion Archives Creator: David Horton Davies	Any information on rights holders and the terms of use should be included here. It is possible to recognise the rights of more than one e.g. the owner and creator as in this instance.
Identifier	CER00023	This is the repository/accession number
Location / Coverage	Aberystwyth, Ceredigion.	The location should be the one associated with the item. If it is not known, the location of the repository may be used. This should include the place name and the pre-1974 Welsh county (and the country if not Wales). The place name should be taken from Elwyn Davies's *Gazetteer of Welsh Place-Names*.

If the image in question (see example 2) forms part of a sequence of digitised images, the subsequent pages should be associated with the cover or title in the spreadsheet using the image filenames. The cover image filename is then referred to as the 'Parent ID', and the subsequent filenames, associated to the Parent ID, provide the Page Order.

Module 3: Preparation

Introduction to scanning

The easiest way to digitise 2D material is by using a flatbed scanner. Using a good standard of scanner, high quality images can be captured for the purposes of archival preservation as well as display.

There are a number of factors that affect the quality of a scan and there are a number of ways in which to ensure that the scan is of the highest possible quality.

The process of inputting these settings and controlling these factors is known as 'calibration'.

Why calibrate a scanner?

- To ensure that you achieve the best, most accurate, colour reproduction from your scan
- Calibration sets the imaging device to a known state, ensuring that it provides consistent results each time it is used
- Making the original scan as accurate as possible avoids having to manipulate the image later e.g. with GIMP or Adobe Photoshop. Doing so saves time and prevents the distortion of the original scan.

What are the factors that determine the quality of a scan?

- **Resolution**: This is the number of pixels that form a digital image, and determines its quality. It is expressed as ppi (pixels per inch) or dpi (dots per inch). As a general rule, the more detail in the original object, the higher the capture resolution should be. It is recommended that a resolution of 300dpi be used as a minimum standard. This should provide adequate detail for most photographs, postcards, posters and prints.
  Note: the resolution of the image will need to be lowered for publication on the Web.

Here is a table of suggested image capture resolutions:

Originals	Resolution	Notes
Photograph	300 dpi	This is the minimum standard resolution.
Letters and line art	600 dpi	The higher resolution will aid legibility.
Illustrations and Maps	300 dpi	The minimum standard of resolution should be adequate but it may need to be increased if the detail is finer.
35mm slides and negatives	1200 dpi	You will need a transparency adapter for scanning slides or negatives on a flatbed scanner.

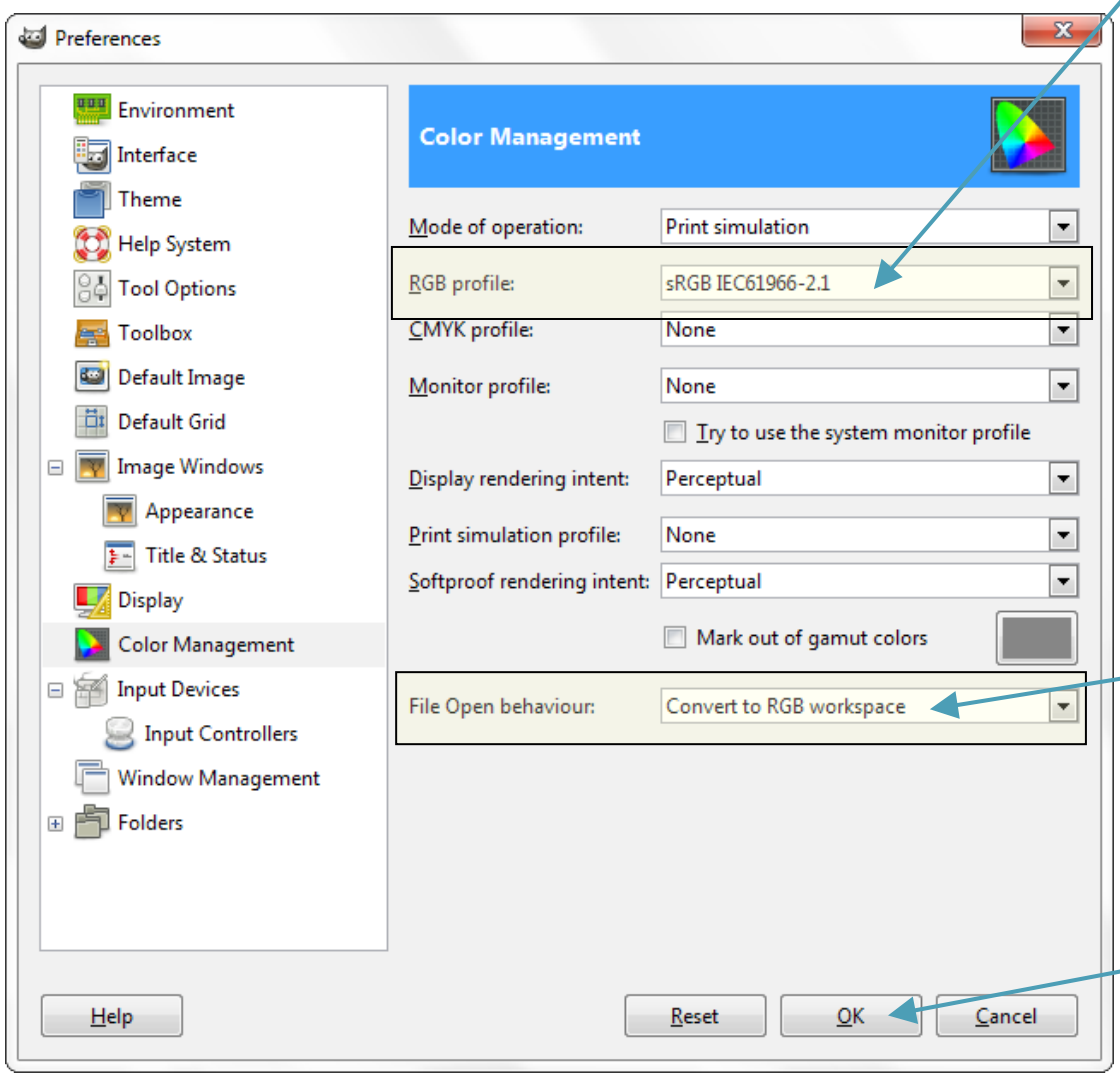
- **Colour depth or bit depth**: This is the term used to describe the number of colours that can be represented digitally. Bit depth ranges from 1-bit colour which displays 2 colours – black and white, to 32-bit and above which display billions of colours. We recommend that you capture at 24-bit colour or 'truecolour' which best mimics the real world, producing over 16.7 million colours.
- Save the images as uncompressed files. Tiff is the best format for the master file as it:
  -retains all the information that was created by the capture device
  -retains any capture device colour management information
  -uses no compression
  The TIFF format gives high quality files that are ideal for storage and archival purposes
- Set up an efficient file-naming and storage system (see Module 6: Archiving)

3.1 How to prepare the scanner

The first time you open GNU Image Manipulation Programme (GIMP) you need to set the Colour Settings:

3.1.1 Launch GIMP

3.1.2 Go to Edit > Preferences. Choose Colour management from the options list on the left.



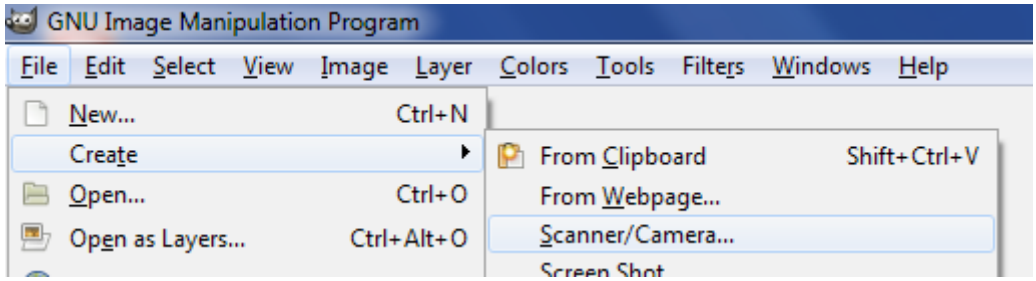
3.1.3 **RGB Profile:** Choose sRGB IEC61966-2.1 from the drop-down list.

NOTE: If this option does not show, click **Select colour profile from disk...** A new window will appear. Click the **colour** folder on the left hand side. Choose the option starting with **sRGB** from the list.

3.1.4 **File Open behaviour:** Choose Convert to RGB workspace from the drop-down list

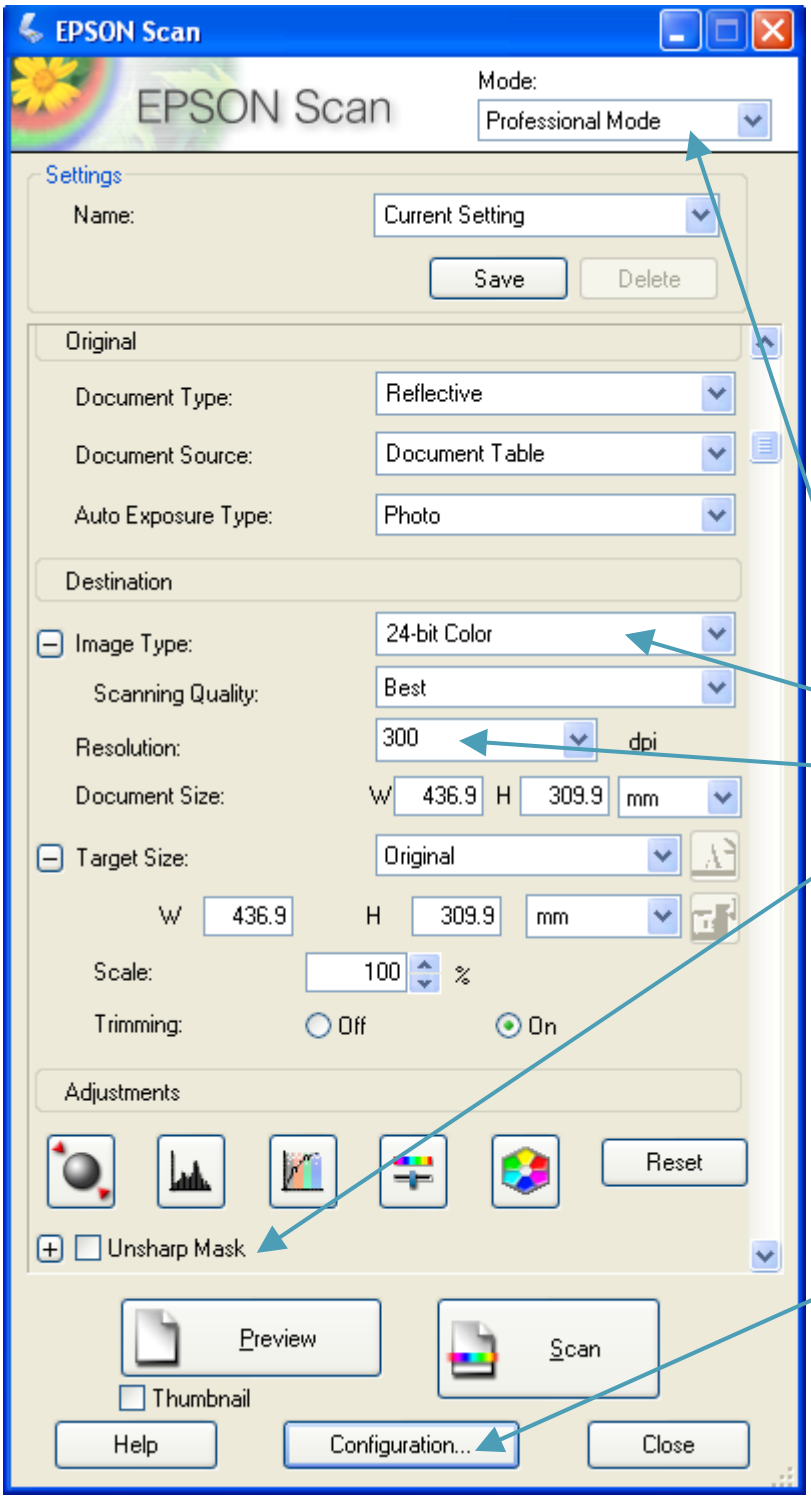
3.1.5 Select OK to commit the changes

3.1.6 Next, launch the scanner software: select File > Create > Scanner/Camera...



3.1.7 In the window that appears select your source e.g. Epson Expression 10000XL

The Epson Scan window will appear:



3.1.8
The basic settings should correspond to those in the diagram.

Note in particular:

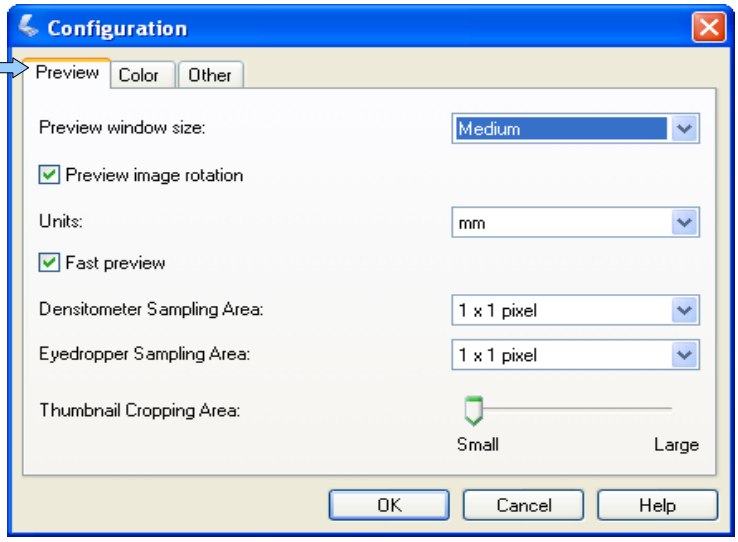
- Mode
- Image Type
- Resolution
- Unsharp mask is not ticked

Note:
Unsharp mask option might not be visible, scroll down to view it

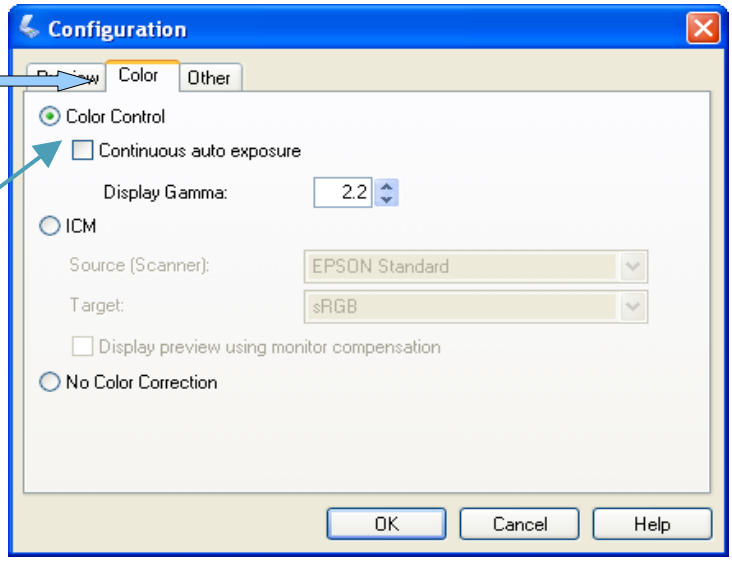
3.1.9
Check that the Configuration settings are correct. Click on **Configuration** to open the dialogue box

3.1.10 Ensure that the settings **under all three tabs** of the Configuration window appear the same as these screen shots. The settings should be saved by the software and so it is unlikely that they will need altering.

Tab 1 : Preview

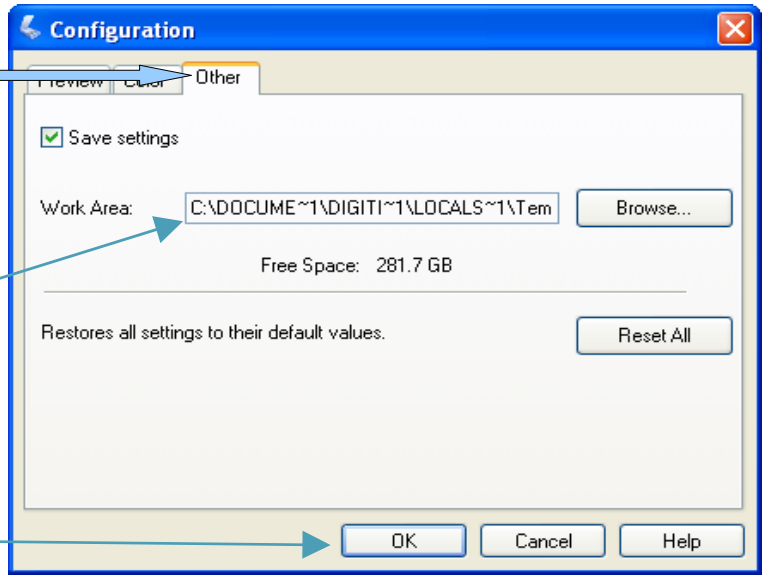


Tab 2 : Colour



Note:
Make sure that '**Continuous auto exposure**' is NOT ticked

Tab 3 : Other



Note:
This should begin with C:\ but the rest of the file drive may appear differently in your window.

3.1.11
Press OK to return to the Epson Scan window.

3.2: Preparing to scan a collection of items

The above steps have ensured that your scanner is capable of capturing the image to the required quality. Next, you need to enable your scanner to capture the original item as accurately as possible.

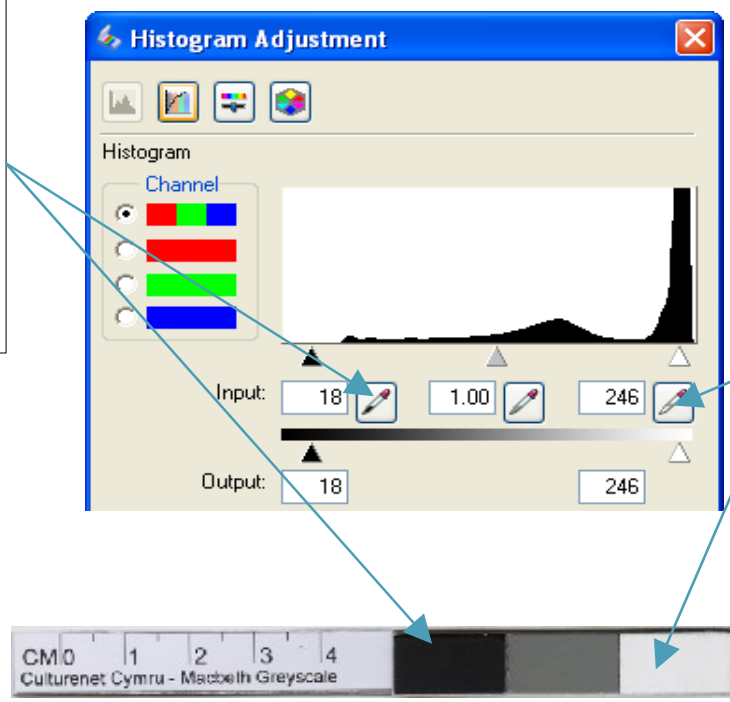
This is done by determining the colour value of black and white and telling the scanner to scan within those points, then measuring the values after scanning. This ensures that the digital image is as accurate as possible. A greyscale patch is used to do this as the black and white on the patch are of known value. By fixing the black and white points before scanning, the greyscale patch can be measured in GIMP to ensure accuracy.

This can be done by taking the following steps:

3.2.1 Take a preview scan of the greyscale patch: place the greyscale face down on the scanner bed and press 'preview' in the Epson Scan window.

3.2.2 Once the preview is done, click the histogram adjustment button in the Epson Scan window and a Histogram Adjustment window will appear.

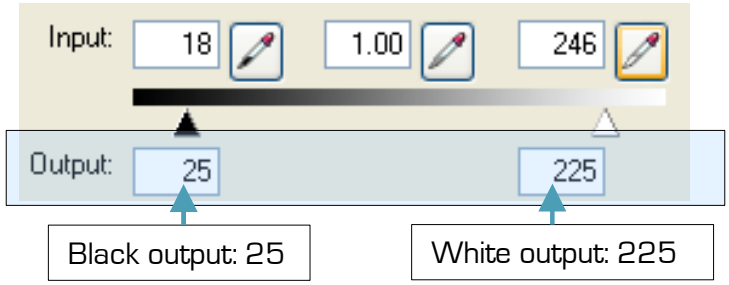
3.2.3
Select the left pipette on the black end of the scale. Then in your preview window, click on the black section of the greyscale patch.



3.2.4
Do the same with the pipette on the right at the white end of the scale and click on the white section of the greyscale patch in the preview window.

3.2.5 In the Histogram Adjustment window, enter the following values in the **output** windows:

Note: Do **not** change the values in the **Input** boxes



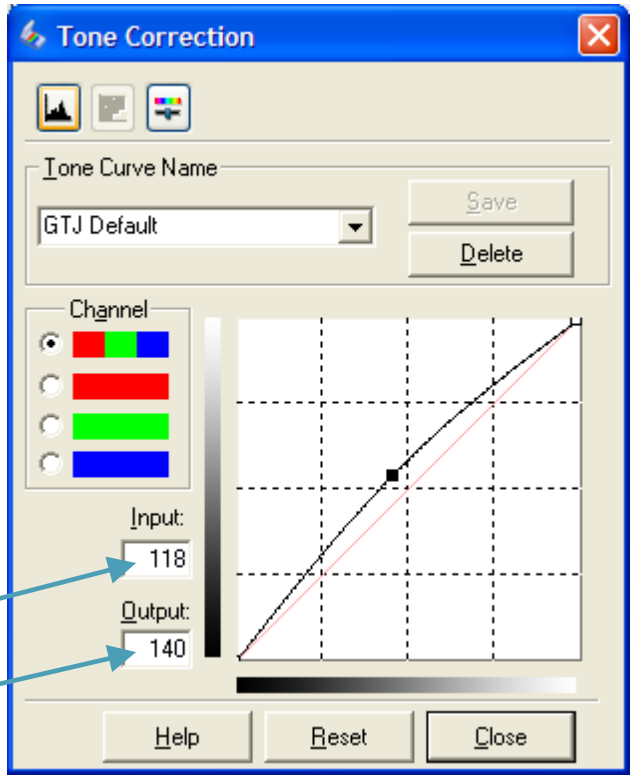
3.2.6 Click **close** to save the histogram adjustment settings.

Explanation:
Doing this predetermines the white and black values allowing us to know whether or not the scanner is producing an accurate representation of the original.
Another way of ensuring that the accuracy of the digital image is by making a slight adjustment to the tone curve, this brings out the optimum colour and brightness of the original.

Tone curve correction:
3.2.7 Click the 'tone correction' icon next to the histogram adjustment icon on the Epson Scan window:

A window like this will appear:

Note: You may need to click the cursor in the gridline to activate the input boxes.



3.2.8
Enter the following values:

Input – 118

Output – 140

3.2.9 Type your project name in the Tone Curve Name box and press Save.
3.2.10 Close the window to return to the Epson Scan window.
3.2.11 Save all these settings by clicking 'save' at the top of the settings panel on the Epson Scan window – the software will automatically choose a name e.g. 'Setting 1' so remember which is the correct setting.

3.3: A test scan

Before scanning the batch of items, you need to check that the settings are correct:

3.3.1 Place an item on the scanner bed, allowing around an inch between the greyscale patch and the item.
3.3.2 Click preview to ensure that the image is straight etc.
3.3.3 If happy with the preview, press SCAN.
3.3.4 Once the scan is complete, close the Epson Scan Window and the image should appear in your GIMP window.
3.3.5 Make sure you close the Epson window. GIMP will remain frozen until you do this.
3.3.6 To check that the settings are correct you need to perform a basic Quality Control (QC) process.

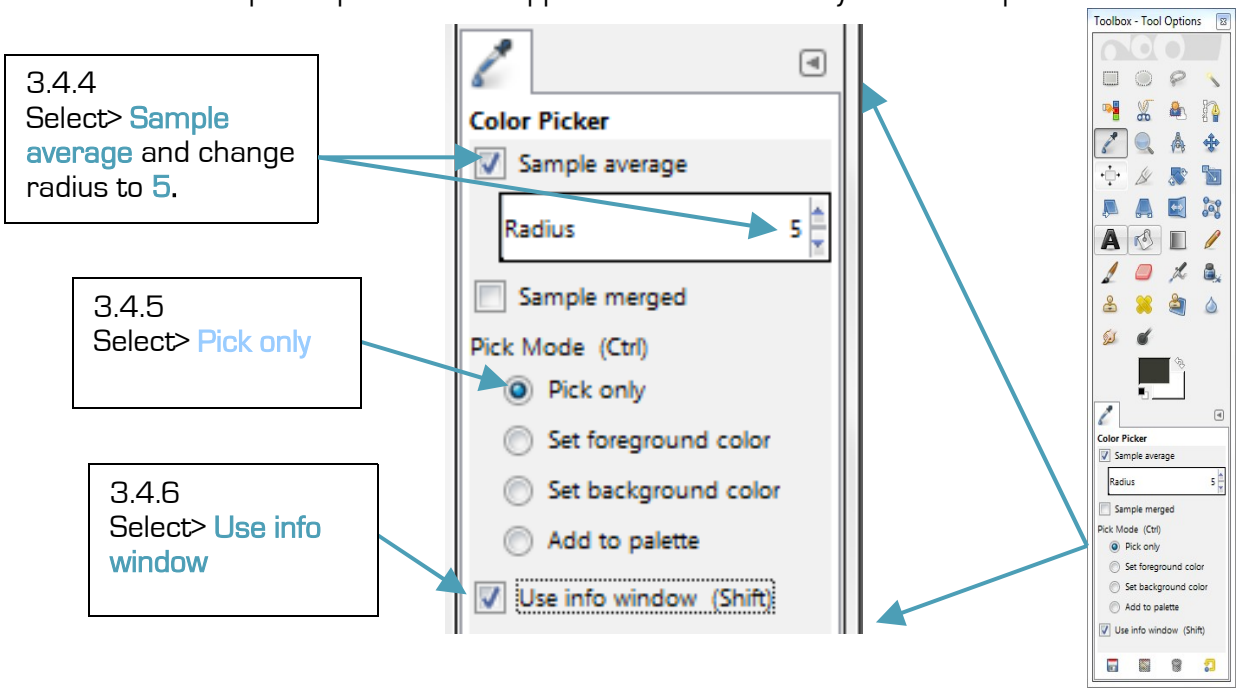
3.4: Basic QC process

3.4.1 Viewing your image in GIMP, zoom out and maximise the window to check that the entire greyscale patch has been digitised and that nothing has been lost in the scan.

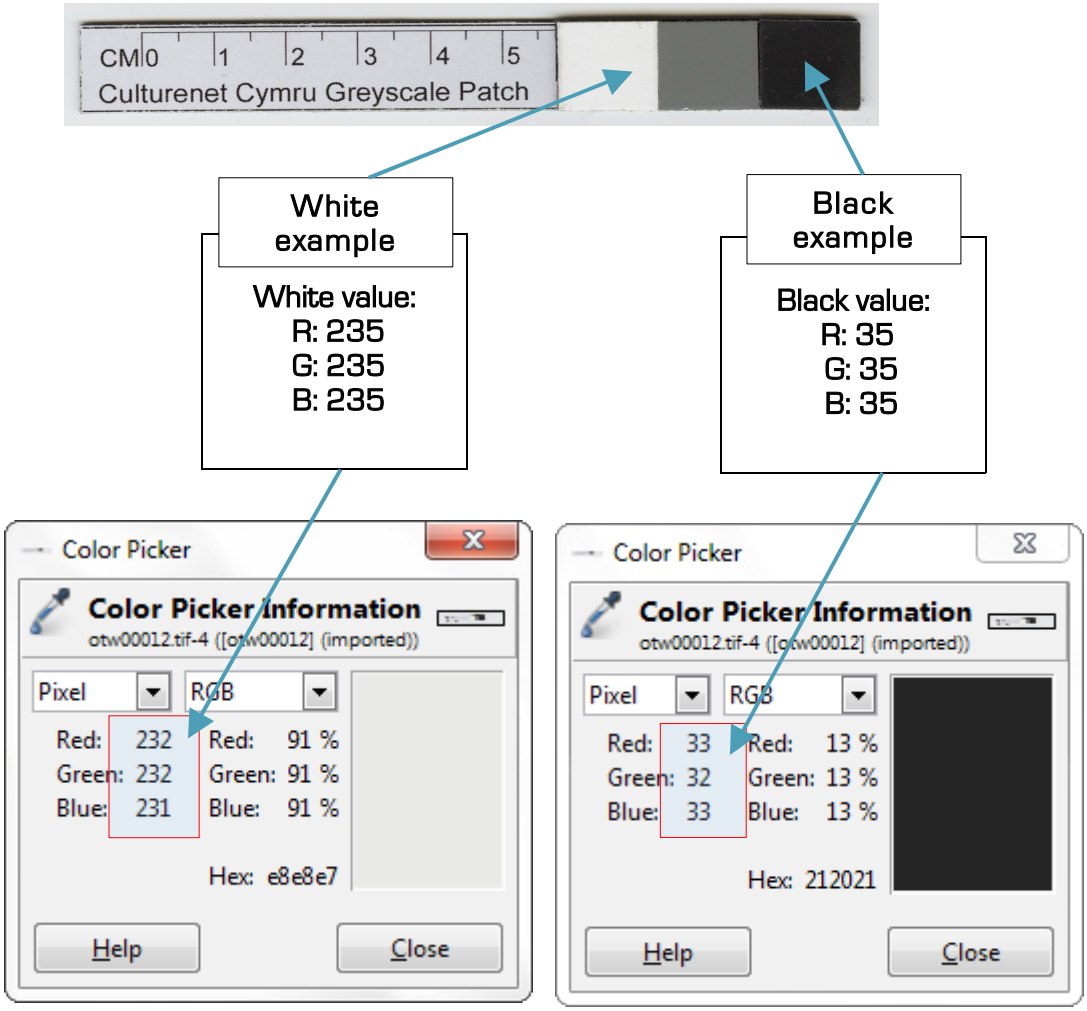
Note : The zoom tool can be found in the toolbox 

3.4.2 Then, zoom in to check that all the important detail is clear and not blurred. Any small print should be clearly legible.

Now you need to check the black and white levels on the greyscale patch:
3.4.3 Open the colour picker [tools>colour picker].
The colour picker panel should appear at the bottom of your toolbox panel:



3.4.7 Click the eye-dropper on the white section of the greyscale patch. A Colour Picker Information window will appear. The RGB readings should be **within 15 each side** of the values shown below. Do the same with the black section of the greyscale patch.



3.4.8 If you do not get a reading within 15 each side of the above values, the scanner settings are incorrect and you need to go through the process again (See Module 3: Preparation)

Module 4: Scanning

4.1: Scanning a batch or collection

First, create a new folder on your desktop to store the scans, named e.g. 'master_images' or 'postcard_collection_masters'.

4.1.1 Open GIMP, select File> Create >Scanner/Camera...
4.1.2 When the Epson Scan Software launches, ensure that the chosen setting is the one with your stored settings.
4.1.3 Place the item to be scanned face down on the scanner bed alongside the greyscale patch and press preview.
4.1.4 If you are happy with the preview image, press scan.
4.1.5 Once the scan is complete, close [not minimise] the Epson Scan Software window and the image should appear in the GIMP window.

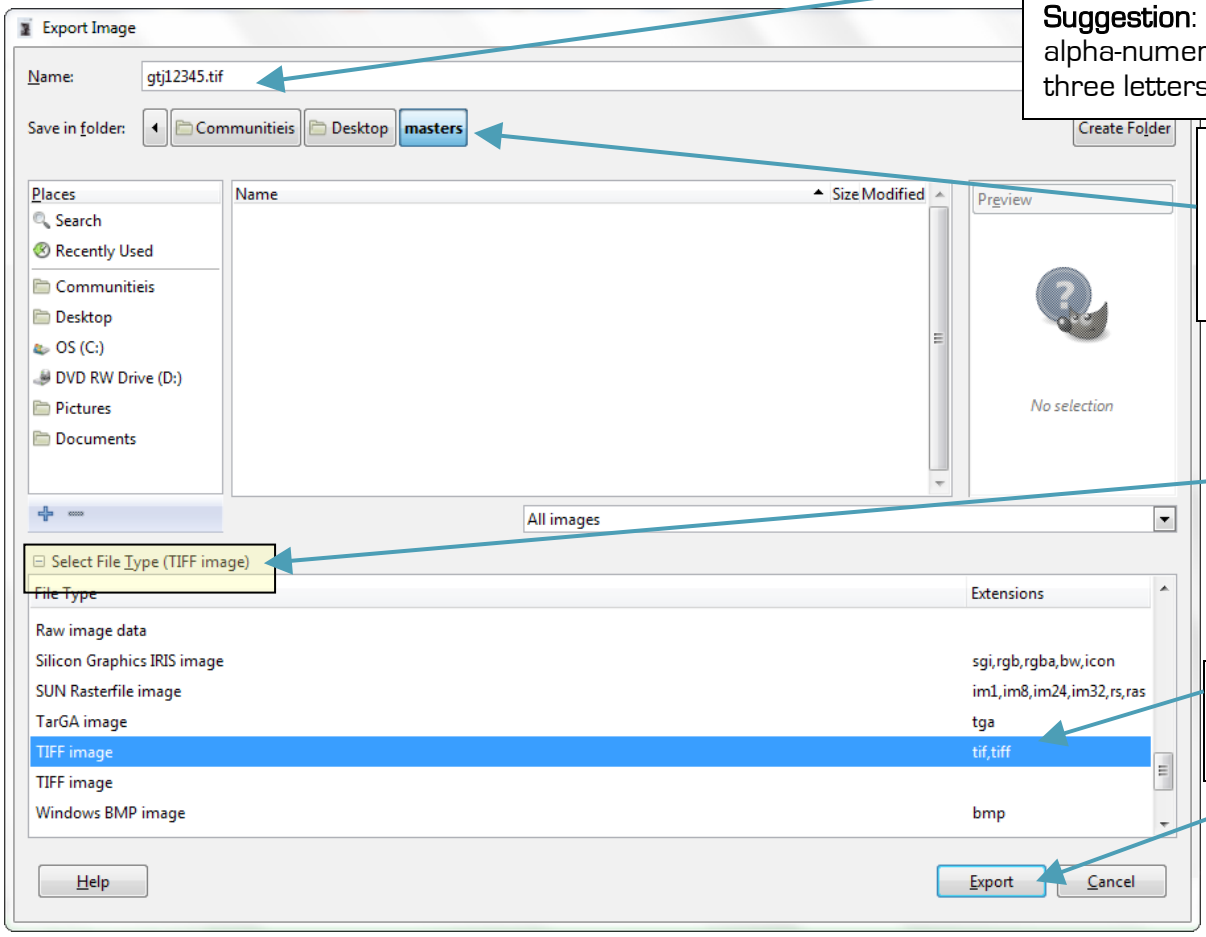
4.1.6 Carry out the basic QC process outlined above (section 3.4).

If you are happy with the quality of the scan, save the image as shown below (4.2).

4.2: Storing the scanned image – Master files

4.2.1 In GIMP select File > Export . . .

4.4.1 When the 'Export Image' window appears you need to:



4.2.2
Ensure that the image has a unique filename.

Suggestion: Use an unique alpha-numeric filename of three letters and 5 numbers

4.2.3
Save in appropriate folder e.g. 'masters'.

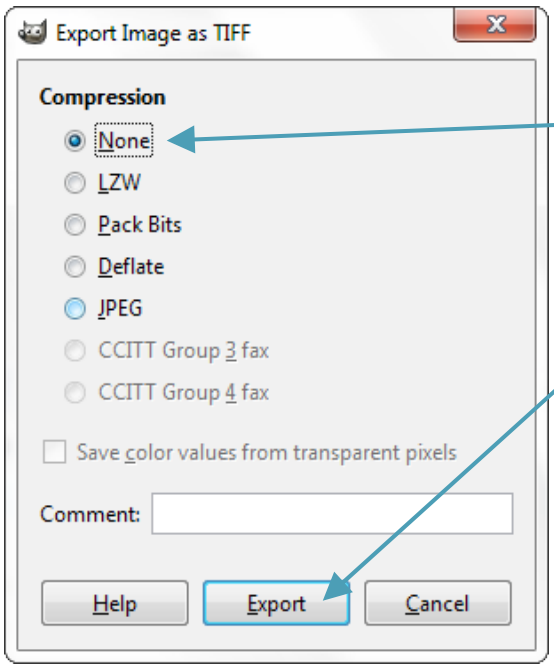
4.2.4
Expand the file type selection by clicking on it

4.2.5
Select TIFF file format.

4.2.6
Press Export.

Note: This is known as a 'master image' – it has not been altered in any way and should be kept for archival purposes.

This window might appear:

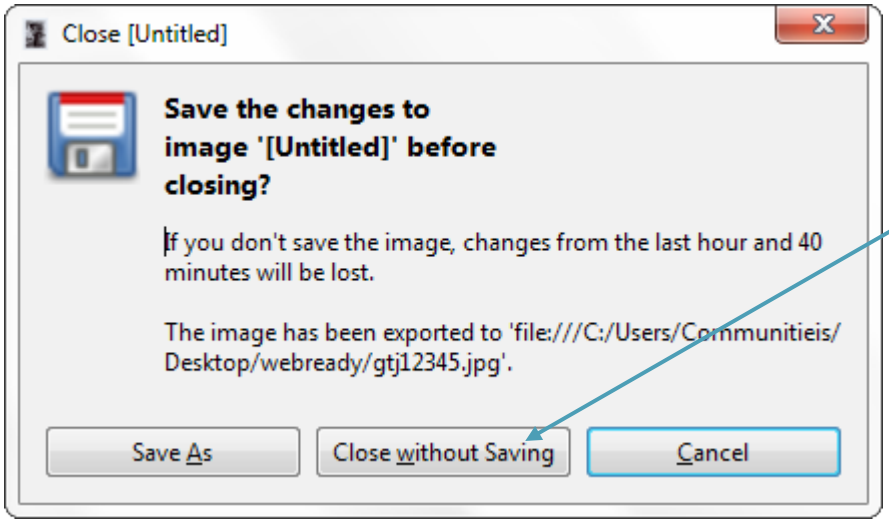


4.2.7
Ensure that None is selected under Compression.

4.2.8 Press Export

4.2.9 To close the image after saving, go to File > Close

This window will appear:



4.2.10
Click Close without Saving

(you have already saved a copy by exporting it)

4.2.11 To scan your next item, return to section 4.1.1

Module 5: Web Prep

Aims and objectives

This module will show how to prepare digital images for display on the web. It will:

- Explain the principles of web display
- Provide a guide to preparing images for the web – using Adobe Photoshop, Adobe Photoshop Elements or GIMP software
- Introduce uploading to the People's Collection website

Why do images need to be prepared for the Web?

The original or 'master' images you have were scanned at a resolution of 300dpi and stored as tiffs. This means that they are very large, high quality files that are unsuitable for the web.

For display on the web, images need to be:

- Of lower resolution, 72dpi is usually recommended
- Smaller in size
- A compressed file, usually JPEG

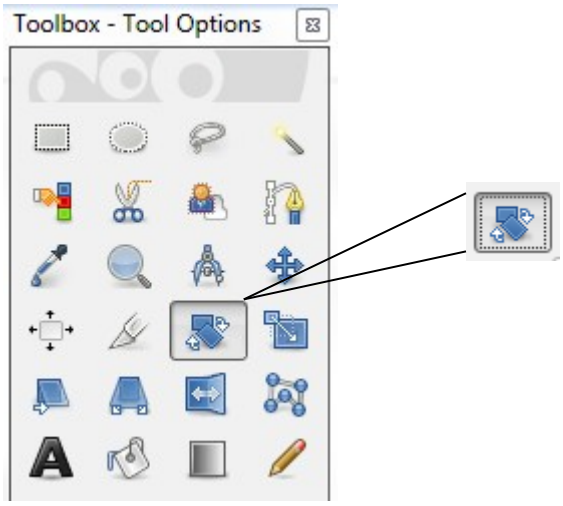
These alterations are made using software such as Adobe Photoshop or GIMP and the web-ready images are stored in a separate file using the same filename as the master image.

Storing the images

Your web-ready images will need to be stored in a separate folder but will retain the same filename. Create a folder in which to store your images for Web display.

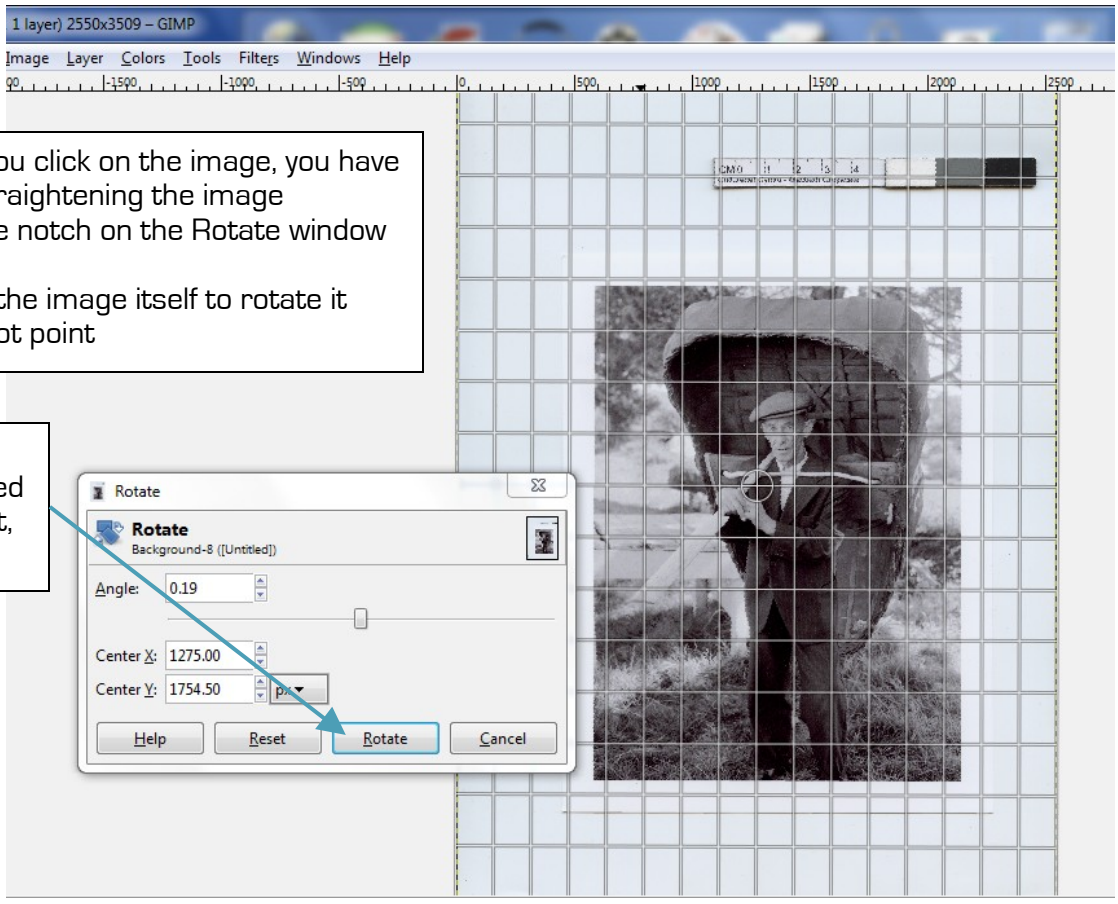
5.1: Preparing images for display on the Web

5.1.1 Open GIMP.
5.1.2 Select File > Open and open the folder where your master images are stored and select your image.
5.1.3 Should the image require rotating, select Image > Transform and the required rotation e.g. Rotate 90° clockwise.
5.1.4 Should the image require straightening, select the straighten tool from the toolbar:

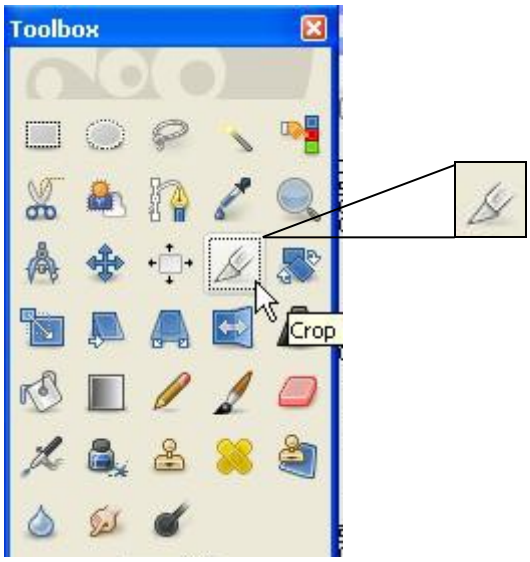


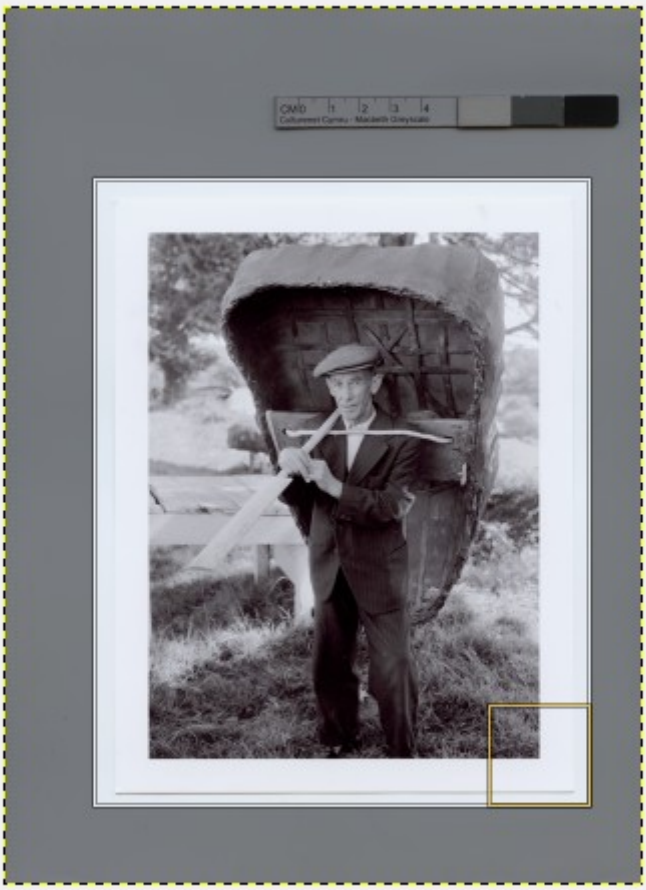
5.1.5 Once you click on the image, you have a choice of straightening the image
a) by using the notch on the Rotate window
OR
b) clicking on the image itself to rotate it around its pivot point

5.1.6
When you are satisfied your image is straight, press **Rotate**



5.1.7 Crop the image to remove the greyscale patch by selecting the crop tool from the toolbox:





5.1.8
After creating the crop box you have an option to either:

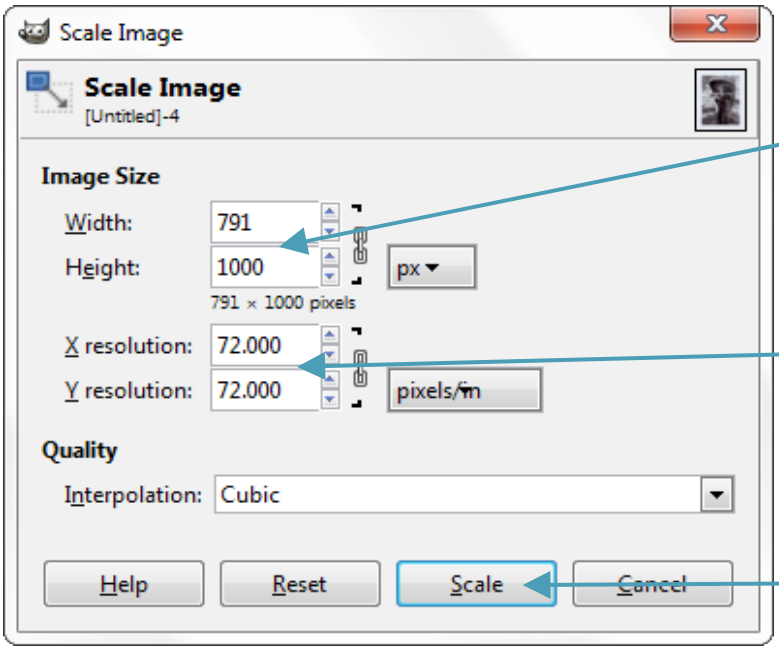
a) Commit to cropping by clicking on the image

OR

b) Click off the image to undo the crop box

5.1.9 Once you are happy with your cropped image, select Image > Scale Image . . .

This window will appear:



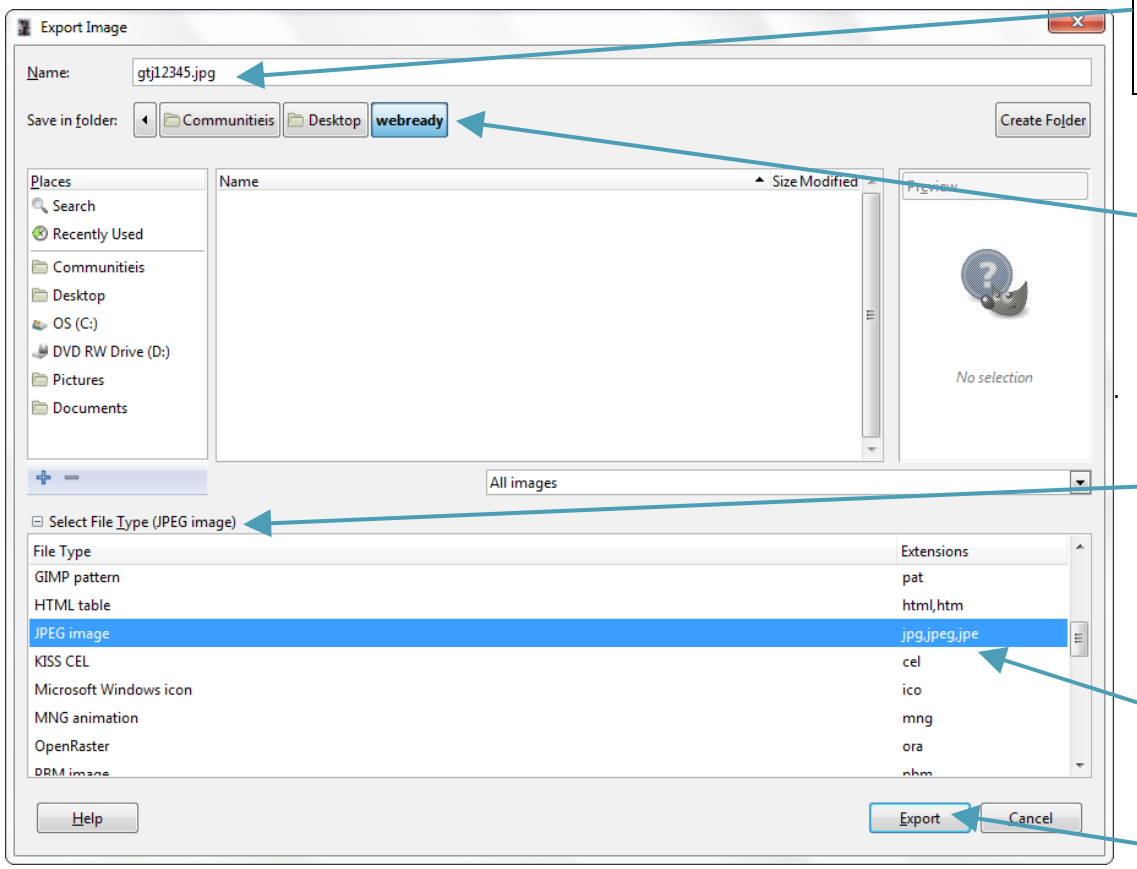
5.1.10
Image Size:
Resize the longest edge of the image to 1000 pixels

5.1.11
Change the resolution to 72

5.1.12
Press Scale

5.1.13 To save the image, select File > Export . . .

This window will appear:



5.1.14
Keep the same flie-name as your master image

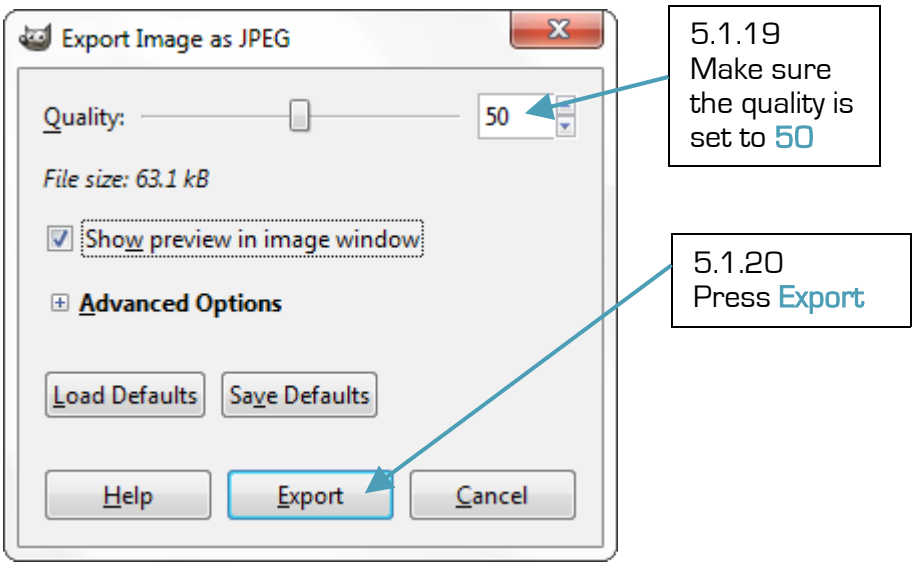
5.1.15
Choose the appropriate folder in which to save your webready image

5.1.16
Expand the file type selection by clicking on it

5.1.17
Select JPEG file format.

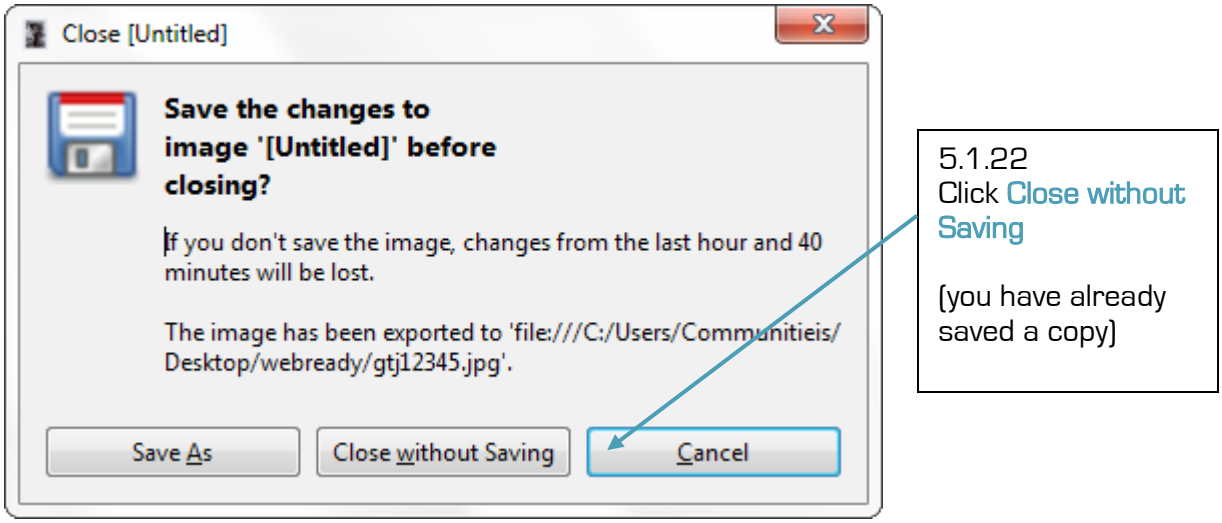
5.1.18
Press Export.

This window might appear:



5.1.21 To close the image after saving, go to File > Close

This window will appear:



5.1.23 Repeat the process from section **5.1.2** to make other master files web-ready.

Module 6: Archiving

Here are some of the simple measures you could put in place to ensure a high standard of digital preservation.

6.1 Name and file appropriately

6.1.1 Keep file names short but meaningful and unique (e.g. **digi_guide.doc**)
6.1.2 Put common words at the end of a file name rather than at the beginning (e.g. **digi_plan_v03_final.doc**).
6.1.3 Do not use capital letters as this can cause problems when moving files between different computing environments, use underscore rather than leaving a space (e.g. **scan_guide.doc**).
6.1.4 When including a number in a file name always give it a two digit number to maintain the numeric order. This helps retrieve the latest record number (e.g. **cnc _01 .pdf**)
6.1.5 If using a date in the file name always state the date 'back to front' YYYYMMDD (e.g. **20090701_paper-a.doc**).
6.1.6 If you include a name in a file use family name first followed by the initial e.g. (**jones-s.doc**).
6.1.7 Order the elements in a file name in the most appropriate way for retrieval during the course of every day business.
6.1.8 Establish a version control system (e.g. **corp-plan2009-2010v3draft.doc, corp-plan2009-2010v4final.doc**).
6.1.9 Ensure that access to files is password controlled.

6.2 Back up files

6.2.1 Hard disks fail therefore design a simple backup routine and decide what should be backed up, also check backups annually.
6.2.2 Make copies on servers, trustworthy on-line vaults or on removable media such as external drives, CDs, DVDs.
6.2.3 Store at least 1 copy off site, 2 on site and in different formats.

6.3 Look after your hardware and software

6.3.1 Replace your hardware before it fails by having an ongoing 5 year hardware cycle.
6.3.2 Select simple formats that can be supported by multiple applications, consider using open source software with well established specifications with a sustainable community.
6.3.3 Save high quality master images (min. 300dpi) in TIFF format.

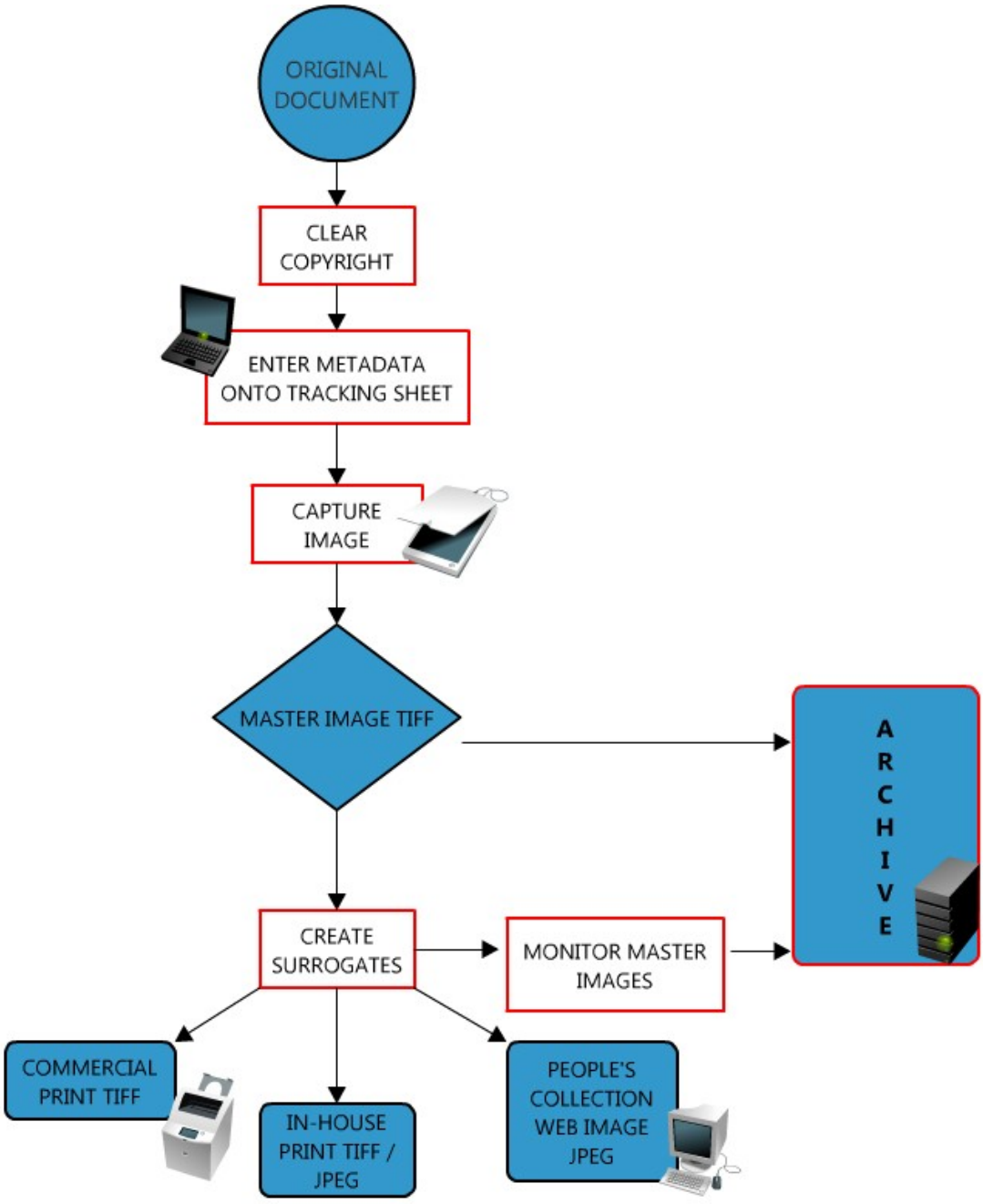
6.4 Manage your systems

6.4.1 Ensure files are backed up before updating hardware and software.
6.4.2 Ensure anti virus software e.g. **AVG** is installed and updated regularly.

6.5 Keep up to date

6.5.1 Keep in contact with digital archivists who can update you on developments in digital preservation (e.g. The National Library of Wales).

Appendix 1: Workflow



Appendix 2: Glossary

Pixel
Each digital image is comprised of pixels. A pixel (pix = picture, el = element) is the smallest item of information in a digital image, it's usually a tiny dot. Each pixel is a digital representation of a typically provide part of the original image – more pixels original. Therefore more-accurate representations of the (PPI) the greater the resolution. the greater the number of pixels per inch clarity of the image, usually referred to as

Software
A general term to describe various programmes that are used to operate computer and related devices.

Greyscale patch

A greyscale patch is a printed strip of graduated tones. The patch is used to check the quality of the scan or digital photograph. It is a quick and easy way to see if a scanner has been calibrated correctly and that it records the true back and true white of an image.

Scanner calibration
Scanner calibration ensures that your scanned image will display correctly on screen and in print. Scanner calibration verifies and corrects the colours that the scanner sees versus what is actually on the document. Scanner calibration should on average be performed on average over 1000 scans. To fix, check or correct the graduations of a scanner refer to **how to calibrate the scanner**.

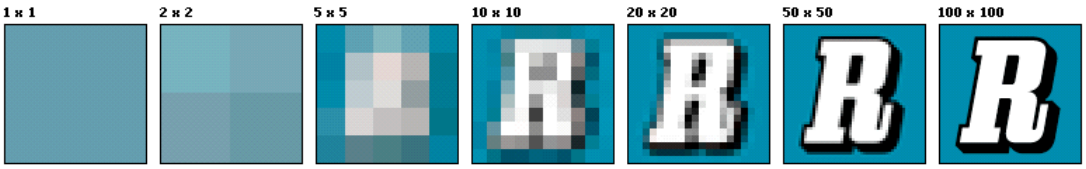
Colour scanning
Uses multiple bits per pixel to represent colour, 24 bits per pixel is called true colour level, and it makes possible a selection from 16.7 million colours.

Colour depth or bit depth
This is a measurement of the number of bits used to define each pixel. It is the term used to describe the number of colours that can be represented digitally. Bit depth ranges from 1-bit colour which displays 2 colours – black and white, to 32-bit and above which display billions of colours. We recommend that you capture at 24-bit colour or 'truecolour' which best mimics the real world, producing over 16.7 million colours.

24-bit colour
24-bit colour images are composed of (3) 8-bit colour channels. Each colour channel, similar to an 8-bit greyscale image, contains up to 256 colours. When combined, the red, green and blue channels provide up to 16.7 million colours. 24-bit colour is also known as True Colour and Photo-realistic Colour.

32-bit colour
32-bit colour images have 4 colour channels of 8 bits each - one channel each for red, green and blue, plus 8 bits of greyscale data to provide higher detail.

Resolution
Image resolution describes the detail the image holds. The number of pixels in a digital photo is commonly referred to as its image resolution.



Compression
Compression is used to reduce file size. Compressed file formats include jpeg and gif.

Quality control
Is an important component in every stage of digitisation to guarantee the integrity and consistency of the digital image.

Alpha channel
An 8-bit channel reserved by some image-processing applications for masking or additional colour information.

Alphanumeric
A set of characters composed of letters and numbers.

Appendix 3: Plan

DIGITISATION PLAN INFORMATION			
Digitisation Plan ID			
Title			
Start Date		End Date	
Institution			
Digitisations Plan Manager & contact details			
Partners	Other Institutions		
	Volunteers		
	Education providers		

DOCUMENT NAME			
Document Title			
Author & role			
Date		Filename	

DOCUMENT HISTORY		
Version	Date	Comments

OVERVIEW OF PLAN

1. Background

2. Aims and objectives

3. Overall approach

4. Plan outputs

5. Plan outcomes

6. Stakeholder analysis

Stakeholders	Interest / stake	Importance

7. Risk analysis

Risk	Probability (1- 5)	Severity (1- 5)	Score (P x S)	Action to prevent / manage risk

8. Standards

Name of standard	Version	Notes

9. Intellectual property rights

RESOURCES

10. Partners

11. Resources for digitisation
Staffing (to include volunteering) and equipment

DETAILED DIGITISIATION PLAN

12. Work packages

	month	1	2	3	4	5	6	7	8	9	10	11	12
Work packages	1												
	2												
	3												
	4												
	5												
	6												
	7												
	8												
	9												
	10												

Month 1 =

13. Evaluation

Why not digitise all items at 1200 ppi?

The human eye can not distinguish the difference between an image that has been scanned at 300 ppi compared to an image scanned at 1200 ppi but the 1200 ppi image will allow you to blow it up to a larger print without loosing quality. The down side however is that 1200 ppi image is a far larger file. This is an issue when storing files.

What is the difference between dpi and ppi?

	
ppi (square pixels per inch) - when zooming into an image the pixels are visible as you can see from the image	dpi (dots per inch) refers to the number of dots in a print

Why is black represented with a K in CMYK?

The 'K' in CMYK stands for *key* since in four-color printing cyan, magenta, and yellow printing plates are carefully *keyed* or aligned with the *key* of the black *key plate*.

What is the difference between Creative Commons Licence and Creative Archive Licence?

The Creative Archive Licence is heavily inspired by the Creative Commons Licence but includes additional features that meet the needs of public service organisations in the UK. The two most obvious are the UK-only and the No Endorsement requirements. The Creative Archive Licence is also designed to protect the Licensor's right not to have a work treated in a derogatory or objectionable way.

www.bbc.co.uk/creativearchive/faqs.shtml

Appendix 5: Useful Links

InterPARES 2 Project Creator Guidelines
www.interpares.org/display_file.cfm?doc=ip2(pub)creator_guidelines_booklet.pdf

Collections Link – the national collections management advisory service
www.collectionslink.org.uk

Strategic Content Alliance – IPR advice and publications
http://sca.jiscinvolve.org/ipr-publications/

JISC Digital Media (formerly TASI) – digitisation advice and training services
www.jiscdigitalmedia.ac.uk

Renaissance East Midlands Simple Digitisation Guide – pdf digitisation guide
www.renaissanceeastmidlands.org.uk/news_and_events/simple_guide_to.html

Dublin Core Metadata Initiative Usage Guide
www.dublincore.org/documents/usageguide

UKOLN – digital information management advice
www.ukoln.ac.uk

Creative Archive Licence
www.bbc.co.uk/creativearchive

Creative Commons Licence UK
www.creativecommons.org/international/uk

Further reading:
Tim Padfield, *Copyright for Archivists and Record Managers (3rd Ed.)*, London. 2007.