
Wits Digitisation Centre

GUIDELINES FOR DIGITISATION

TABLE OF CONTENTS

Introduction.....	2
Overview of Scan Types.....	3
File naming conventions.....	4
Some things to keep in mind when designing a file naming scheme:	4
File names should at least: Be unique	4
Quality control.....	5
Assessing Image Quality	5
Quality Control Checklist	6
Metadata	7
Minimum Scanning guidelines.....	8
Manuscripts, Printed Text, Photographs, 35 mm. Slides, & Sheet Music.....	8
Rare Books.....	8
Graphic Art	9
Maps.....	9

Introduction

The Wits Digitisation Centre Guidelines for Digitisation were developed for the use of all library digitisation projects. These guidelines are designed to provide a set of *minimum* specifications for preservation-quality digitisation of printed text, manuscripts, photographs, slides, rare books, sheet music, graphic arts, and maps. Guidelines for audio and video are not included. They provide a baseline for creating digital images that are of sufficient quality for long-term preservation. These specifications should not be seen as a replacement for project-specific analysis of factors such as purpose and use that might dictate higher level capture standards.

This document provides an introduction to the following aspects of your digitisation project:

- Deciding whether your project will produce content for long-term preservation
- Identifying various uses of the materials and types of images needed
- Developing file naming conventions
- Choosing hardware and software for scanning
- Establishing quality control procedures
- Determining storage needs
- Understanding the project's metadata needs

The guidelines also include specifications for the file format, resolution, bit depth, dimensions, and compression for a variety of types of two dimensional works. These specifications are provided for preservation/service masters; full size high resolution; full size low resolution; medium size; and thumbnail images.

These guidelines are designed to provide an overview of the issues that will need to consider when starting a digitisation project and to present the context necessary to understand the digitisation specifications given. They are not intended to be an exhaustive treatment of all aspects of digitisation or to provide a tutorial on digitisation.

Are you digitising materials for short or long term

Short term	Long term
Small scale: A few scans, limited in scope	Large scale: Thousands of scans for a curated collection
Ephemeral: Content that changes, is updated frequently, or only used for a limited time	Enduring: Content that is stable and will have enduring value
Limited audience: Material for use solely within the Library or a Library department	Wide audience: Content is significant and has research interest beyond the University.
Short term examples: <ul style="list-style-type: none">- Library web pages- LibGuides- Content scanned for ILL or Reserves- Training materials- Blogs- Library publicity	Long term examples: <ul style="list-style-type: none">- Electronic dissertations and theses- institutional Repository items- Historical Papers collections

Are you scanning for the web or for print?

If for PRINT Follow suggestions for Full-High Resolution

If for WEB Follow suggestions for Full-Low Resolution

Overview of Scan Types

When planning a scanning project, it should be remembered that images are used in a variety of ways, even within one specific project. A well-constructed database of images, for example, may have as many as four different copies of each image that were created in order to provide an ideal level of display and access. Each copy is distinguished by its size, something dictated by the specific use of the image.

- *Preservation/Service image* – this is considered the master image. It should be at the highest size and quality; probably beyond what your anticipated need requires. The preservation image will be stored once finished and will be kept safe to serve as a backup to the images in active use. The service image is a working copy of the preservation image. The service image serves as the source image for all of your other copies of this image.
- *Full size image* – the full size image is the best quality image that you intend for your users to have. Depending on policies and rights you may choose to allow the user to have a high-resolution or low-resolution image. The difference generally is based on the ability to print the image at a large size. High resolution full-size images can produce a photo quality print at dimensions up to 8" (20cm) x 10" (25cm). Low resolution full-size images will be suitable for projecting, screen viewing, and small prints. At the larger sizes they will not print well.
- *Medium size images* are also known as screen-size images. They are intended to serve as an image that may be viewed on a computer screen but will not project well or print well at a large size. The advantage to using medium size images is that they display well on a computer screen and take far less storage than full size images. They can provide a very modest degree of protection against misuse of the image if your intention is for the image to be viewed solely on screen.
- *Thumbnail images* are used as very small surrogates for a larger version of the same image. Thumbnails are often used as links to the larger size images. They are a convenient way to display many images on one page, allowing a user to see the contents of a database, a web page, or a folder at a glance. Thumbnails are not good as a print source, nor are they suitable for detailed onscreen viewing. They are large enough to be recognizable and to lead a viewer to a larger image.

When planning a scanning project, it is wise to anticipate your intended use for the images as much as possible and be certain that your digital image will be appropriate for your anticipated need. You should create your preservation/service images at the largest size that you can store and maintain. Technologies and intended usages may change with time. You may find additional uses for the images that require larger sizes, and it would be far preferable to re-use your original scan rather than scanning the same materials multiple times.

File naming conventions

Effective file names are essential for the stability and sustainability of digital storage and access systems and for ensuring interoperability between systems. A good file naming scheme can help you connect the various parts of a digital object, tie together metadata and images, and track individual files throughout the digitisation process.

Some things to keep in mind when designing a file naming scheme:

- Each file name must be unique
- Think long term: how will this name scale as you add digital material to your collections?
- File names should provide context: names could include codes for department or collection.
- Keep file names simple for readability.
- Self-explanatory file names make it easier to understand the context of files as they make their way through digitisation workflows.
- The more complicated the file name, the higher likelihood of human error when entering the name.
- Consider including the system's unique digital object ID in the name of the individual files that make up that object.
- File names are not metadata: let your metadata describe the digital object. Use file names to connect metadata to digital images.
- File names will outlast the current project staff.

File names should at least: Be unique

- Use lowercase letters of the Latin alphabet and the numerals 0-9
- Have no spaces between characters
- Avoid punctuation marks other than hyphens and underscores.
- Have no more than 31 characters (the fewer the better)
- Have a single period between the file name and the three letter extension

Quality control

Quality control (QC) is an important part of any digitisation project. QC encompasses procedures and techniques to verify the quality, accuracy, and consistency of digital images. The goal of any scanning project should be to “capture once, use many times.” Digitisation is expensive, time-consuming, and requires extensive handling of original materials. A digitisation project should therefore focus on creating high-quality master images from which many derivative images can be created for specific uses.

Consider the goals of the project

The first step is to define the goals of the project as the quality control criteria will depend on them. For example, if the goal is to create a faithful reproduction of the original, the digital images should look as close as possible to the original material. However, if the goal of the project is to create the best quality digital images regardless of the condition of the originals, then it will not be part of your quality control criteria to make sure that the digital images accurately represent the originals.

Identify Your Products

Identify the products to be evaluated. These might include master and derivative images, printouts, accompanying metadata, and converted text or OCR’ed files.

Develop a Consistent Approach

To measure quality and judge whether the products are satisfactory, clearly define baseline characteristics for "acceptable" and "unacceptable" digital products.

Control the QC Environment

The impact of image-display conditions on perceived quality is often underestimated. Given an improper environment, even a high-quality image may come across as unsatisfactory. Factors that may affect on-screen image quality include viewing conditions, human characteristics, monitor calibration, and colour management.

Assessing Image Quality

The key factors in image quality assessment are resolution, colour and tone, and overall appearance.

Resolution

Resolution is the key factor in determining image quality for textual materials and other distinct, edge-based representations. Resolution attributes to inspect are legibility, completeness, darkness, contrast, sharpness, and uniformity.

Colour and Tone

For colour, grayscale, and some monochrome images, colour and tone reproduction are significant indicators of quality, complementing the "detail" provided by resolution. The goal behind assessing colour and tone appearance is to determine the extent to which a digital image conveys the same appearance as the colour and tone ranges of the original document (or intermediate used). Tone and colour assessment may be highly subjective and changeable according to the viewing environment and the characteristics of monitors and printers.

Overall Appearance

Image quality is cumulative, affected by a range of individual factors--capture system performance, resolution, dynamic range, and colour accuracy. The final evaluation should be made on the overall image, appreciating all the individual factors that contribute to quality.

Quality Control Checklist

	Master digital image is a faithful representation of the original (if that is the goal)
	File name is correct
	File format is correct
	Bit depth is correct See: file, properties, details
	Image is correct size/resolution in long dimension
	Image is not rotated or backwards
	Image is not skewed or off-centered
	Image has clean edges, clear contrast, and legible text
	No broken figures (illustrations, maps, etc.)
	No moiré patterns (wavy lines or swirls, usually found in areas where there are repeated patterns)

Metadata

Metadata is “structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.” (NISO, Understanding Metadata) Good metadata is essential for the management and preservation of digital objects. Without metadata, information about a digital object can only be obtained from its filename, file extension, and perhaps its directory structure. This data will not provide enough information for the creator, user, or manager of the digital object to discover and identify the object, much less understand what the object is, how it was created, who can use it, and what computing environment is necessary to access it. There are many different types of metadata that together support the discovery, evaluation, selection, access, navigation, management, and preservation of digital objects:

Descriptive metadata provides information about the intellectual content and physical format of the object. It supports identification, discovery, and selection. Examples of descriptive metadata elements include title, author, abstract, extent, and subject terms.

Structural metadata is data about the components (e.g., individual files) that make up a complex digital object, and the relationships among those components. For example, a four page pamphlet may be represented by four thumbnail images, four full-size access images, four archival master images, and an XML file of the encoded text. Structural metadata provides a means of identifying each of these files and their roles, as well as tracking their sequence and/or hierarchical relationships. It supports navigation and reconstruction of the object in an online environment.

Administrative metadata supports the short- and long-term management of a digital object. It includes the following subtypes:

- **Technical metadata** includes information about the format and creation of a digital object, such as data on hardware and software, compression ratios, and encryption keys. Technical metadata is necessary for current and future access to the object. As such, it is an important part of preservation metadata.
- **Preservation metadata** is “the information necessary to carry out, document, and evaluate the processes that support the long-term retention and accessibility of digital materials.” (PREMIS) It includes information about the object’s creation, any changes to it, its chain of custody, and technical requirements for access.
- **Use metadata** is information about the use of the digital object, for example usage statistics and search logs.
- **Rights metadata** is data about intellectual property rights, such as copyright details, terms of use statements, and license agreement information.

Minimum Scanning guidelines

Manuscripts, Printed Text, Photographs, 35 mm. Slides, & Sheet Music

	Preservation/Service	Full Size -High resolution	Full Size -Low resolution	Medium	Thumbnail
File format	TIFF	JPEG	JPEG	JPEG	JPEG
Resolution	300-600 ppi	300-600 ppi	150 ppi	150 ppi	150 ppi
Bit depth	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale
Dimensions	3000-6000 pixels across the long dimension	3000-6000 pixels across the long dimension	3000-6000 pixels across the long dimension	600 pixels across the long dimension	150-200 pixels across the long dimension
Compression	Preservation copy uncompressed/LZW lossless compression okay for service	LZW lossless compression	LZW lossless compression	LZW lossless compression	LZW lossless compression

Rare Books

	Preservation/Service	Full Size -High resolution	Full Size -Low resolution	Medium	Thumbnail
File format	TIFF	TIFF	JPEG	JPEG	JPEG
Resolution	400-600 ppi	400-600 ppi	150 ppi	150 ppi	150 ppi
Bit depth	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale
Dimensions	3000-6000 pixels across the long dimension	3000-6000 pixels across the long dimension	3000-6000 pixels across the long dimension	600 pixels across the long dimension	150-200 pixels across the long dimension
Compression	Preservation copy uncompressed/LZW lossless compression okay for service	LZW lossless compression	LZW lossless compression	LZW lossless compression	LZW lossless compression

Graphic Art

	Preservation/Service	Full Size -High resolution	Full Size -Low resolution	Medium	Thumbnail
File format	TIFF	JPEG	JPEG	JPEG	JPEG
Resolution	600 - 800 ppi	600-800 ppi	150 ppi	150 ppi	150 ppi
Bit depth	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale
Dimensions	6000-8000 pixels across the long dimension excluding mounts and borders	6000-8000 pixels across the long dimension excluding mounts and borders	6000-8000 pixels across the long dimension excluding mounts and borders	300 pixels across the long dimension excluding mounts and borders	150-200 pixels across the long dimension
Compression	Preservation copy uncompressed/LZW lossless compression okay for service	LZW lossless compression	LZW lossless compression	LZW lossless compression	LZW lossless compression

Maps

	Preservation/Service	Full Size -High resolution	Full Size -Low resolution	Medium	Thumbnail
File format	TIFF	TIFF	JPEG	JPEG	JPEG
Resolution	600ppi	600ppi	150 ppi	150 ppi	150 ppi
Bit depth	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale	24 bit colour or 8-bit grayscale
Dimensions	6000-8000 pixels across the long dimension	6000-8000 pixels across the long dimension	6000-8000 pixels across the long dimension	1078 pixels across the long dimension	150-200 pixels across the long dimension
Compression	Preservation copy uncompressed/LZW lossless compression okay for service	LZW lossless compression	LZW lossless compression	LZW lossless compression	LZW lossless compression