Basically all the equipment comes in different levels: Consumer (low end), Prosumer (mid range/mix), Professional (high end)

Flatbed Scanners

Place document face down on glass plate called a platen; Lid keeps out light

Pros:

- Can be good for photographs in good shape
- Good for slides [but need a transparency adapter for the lid]
- Often come with software
- Adjustable settings [such as resolution and color]

Cons

- Slow [8.5 x 11 color image @ 300 dpi can take 25 seconds or more]
- Glass plate can get hot
- With books, not great and can damage the binding; pushing on platen to flatten [ex. Some models, ***like the one on the slide***, have a platen which goes all the way to the edge and may even have an angled side for better book scanning]
- Not a good idea to scan documents larger than the platen, as it may cause damage
- Brittle materials be easily damaged
- Platen easily dirtied, dust & debris

Low-end vs. High-end

- Different types and numbers of imaging sensors.
Low end models sometimes use CIS [contact image sensor] for scanning. This sensor records the information directly from the document to the sensor. Not good for archival purposes, better for an office environment. Most all scanners now use CCD [charge coupled device] technology, which records a mirror image of the photograph or document. High end scanners often have multiple CCD sensors for better quality.

- High end models have higher resolution [ex. Epson Expression 10000XL - 2400dpi ***on left***]
Often listed as "2400 x 9600" – lower # is true dpi, higher # is what the scanner can create through interpolation [when the scanner combines two adjacent pixels and inserts a blended pixel between them]

- High end models have better dynamic range [ability to capture highlights and shadows from pure white to total black] and tonal range [the number of tones of color].
Dynamic range is represented by manufacturers with a "D" followed by a number from 0 to 4. You want at least 3.2 for high quality scans. Tonal range is listed in "bits" and is usually expressed as 24-bit or 48-bit color. You want a scanner that is at least 24-bit color.

- High end models often last longer and have lights that do not heat up.

Overhead/Planetary Scanners

The imaging device is mounted above the scanning bed.

Overhead scanners use similar imaging technology as flatbed scanners, while Planetary scanners use a digital camera.

The light source is sometimes integrated into the scanning head so that the light passes across the document as the image is captured, but it may also be mounted separately above the scanning bed [usually 2 on either side].

They come with or with out a glass plate that lays on top of the document. The plate may be hinged so scanning can be done without it.

Some models may have a monitor attached and be self contained [where everything is stored on a hard drive inside the unit], while others need to be hook up to a computer.

Usually have a USB port to transfer files and some, if connected to a network, can email the files.

High end models generally have an adjustable scanning bed to better accommodate books.

High end models can capture image in numerous formats, especially tiff. Low end may only do jpeg or pdf.

<u>Pros</u>

- Good for fragile documents
- Good for books, photographs, maps
- Can be relatively fast [if using a camera]
- Adjustable settings [much like flatbed]
- Built in editing/compiling software [for self contained models]

<u>Cons</u>

- Can be slow [if using flatbed type sensors]
- Without a glass plate, the curvature of books can show producing distortion
- Even with the glass plate, without an adjustable base you can have the same problem [some scanners come with software that can digitally fix this]
- If using the glass plate, must be periodically cleaned

V-shaped Book Scanners

Evolved from the planetary scanner

They use a V-shaped book cradle to keep the book open at an angle, putting less strain on the binding.

Use a pair of digital cameras to capture each page and has integrated lighting. Each company uses different cameras and lighting.

Has a V-shaped glass plate that rests on top of the book to flatten pages [uses very little pressure]

Usually have a monitor to view what has been scanned.

Like other scanners, has a variety of settings and can save to various file formats.

Automatic ones can greatly speed digitization process. [Google] They have numerous sensors to detect the edges of the page and use blown air or vacuum, usually in conjunction with a robotic arm, to turn page.

While some claim they are safe for damaged books, it’s not the ideal choice.

<u>Pros</u>

- Less strain on binding
- Fast
- No curvature [like with the overhead/planetary scanners]

<u>Cons</u>

- Expensive
- Can not scan fold out pages
- Mainly just for books

Large Format Scanning System

Generally consist of a digital camera, lighting system, and a platform or easel.

Come in a variety of shapes and sizes and are generally custom built.

Example:
Rutgers University set-up (***picture on the right***)
Phase One P45+ digital back, mounted on a Hasselblad H2 body. It's a 40 Megapixel system, coupled with an iMac Workstation and Phase One software.
The vacuum unit, table and arm are all one piece, made by Tarsia Technical industries in New York.

Pros

- Can capture virtually anything, especially good for larger materials such as maps
- High quality
- Fast [images are captured instantly]
- Minimal contact with materials

Cons

- Needs well trained staff to operate
- Can be very expensive
- Lighting can get hot (tungsten), depending on the kind of lighting used
- Large (takes up a lot of space)

Camera
Often the most expensive part of the system (Phase One - $25,000-$50,000)

Various types of cameras, but you’ll want a medium or large format camera. Large format camera can cover about 16 times the area, and thus 16× the total resolution, of a 35 mm frame.

Some popular models are Phase One, Hasselblad, and Canon.

In order to capture the highest quality images, a digital back, or scanback, is needed.

Digital Back is a device that attaches to the back of a camera in place of a film holder and contains an electronic image sensor. This lets cameras that were designed to use film take digital photographs. These backs can be expensive [$5,000+] and are primarily built to be used on medium and large format cameras.

Pro:

- Cameras with features not available on digital cameras can be used to make digital images.
- Reduces image noise
- Helps produce the highest quality image

The resolution of digital camera backs (in 2011, up to 80.1 megapixels) is higher than any digital SLR (in 2011, up to 60.5 megapixels) and captures more detail per pixel.

Lighting

- LED - low power, long life, no heat, no UV, no IR
- HID (High-intensity discharge lamp) – large, have built in cooling fans, use reflective surface to help evenly distribute light, have filters to reduce UV and IR, heavy
- Fluorescent – low power, low heat (built in fan), need filter for reflective materials, can cause odd coloration of image [green/blue] <-need to adjust white balance on camera or afterwards on computer

Platform/Easel

Platform lays flat [more common]; Easel is upright [sometimes used for artwork]

Where you put the document you want to digitally capture.

Should generally be black

Can be simple [a table] or more advanced [like a vacuum platform].

Vacuum platform is beneficial, especially with maps, because it can help flatten wrinkled or warped materials. A vacuum easel can help keep materials in place.

Monitor

CRT

Pros

- Better overall display quality
- Better display features
- Flexible resolutions
- Wide viewing angle
- Superior brightness & contrast
- Smoother slanted lines
- Infinite color depth

Cons

- Bulky/heavier
- Higher power consumption

LCD

Pros

- smaller/lighter
- energy efficient
- last longer
- less glare from overhead lights
- larger screen without huge price tag

Cons

- Poor color depth [on low end models]
- Can suffer from pixel distortion in other than native resolution
- Narrow viewing angle

Dot pitch – the space between the pixels; measured in millimeters

Whatever kind of monitor, this should be small, .28mm or less

Larger dot pitch can make things look grainy

Calibration

Most monitors have some calibration built in, but it is generally not very accurate.

Calibration done by just looking at the monitor is subjective; consistent repeatable results are not possible with visual calibration methods.

Colorimeter [aka Spectrophotometer] is a device that is connected to the computer and placed on the monitor's screen. It measures the light [colors] on the screen and in conjunction with software calibrates the monitor to display colors accurately. Mostly automated, but there may be some user interaction. Should be done about once a month. Best done in dark room.

Computer/Operating System

Regardless of operating system, you want as much RAM as possible. 4gb is usually adequate, but 8gb or even 16gb is better. Should note that Windows XP can only use 3gb. Only 64-bit systems can effectively use more than 4GB of RAM.

Ideally have a 2.0 GHz or faster processor

Both Windows and Mac can run the same image editing software, but Macs have better bundled software and often can perform tasks more quickly

Windows

- More affordable
- You can customize, but too many choices of different parts can lead to hardware conflicts that slow things down
- Imaging software tends to run a bit slower on Windows

Mac

- Tend to work better even compared to Windows machines with similar specs due to the fact that Apple directly sub-contracts hardware production to Asian original equipment manufacturers, thus maintaining a high degree of control over the end product.
- Last longer

Mac tend to be more friendly, virus free, stable & highly optimized for performance. With Mac, you won't need to worry about hardware compatibility issues.

Windows 7 is starting to bridge the performance gap

The hard drive size is important, but if not permanently storing files on the computer (which you shouldn’t do), get what best suits your needs. A drive that runs at 7200RPM will have less lag time.

Make sure computer specs are adequate to deal with the imaging software being used. Don’t get just the minimum needed.

Software

Scanning software/driver is almost always included with scanner. With cameras you may have to purchase separately.

Image editing software: Adobe Photoshop Creative Suite is the most widely used

- allows for simple tasks like cropping or de-skewing
- more complex tasks such as color correction using curves, levels, or histograms
- automate batch processes using actions, like setting the color profile or rotating

Photoshop Elements is sort of like a Photoshop CS light. It can do the basic editing (cropping/de-skewing), but does not have the advanced color management capabilities.

Capture One Software has most of the same features as Photoshop and is especially good for working with images in RAW format. This software is included free with Phase One cameras and digital backs.

Aperture is another popular program that is similar to Photoshop CS, but it is only available for Macs.

GIMP (GNU Image Manipulation Program) is a free program that can perform many of the same tasks as Photoshop, though much of its Photoshop-like functionality involves installing separate plug-ins and scripts.

Image Validation Software - JHOVE [JSTOR/Harvard Object Validation Environment]

- Files are analyzed and checked for being well-formed (consistent with the basic requirements of the format) and valid (generally signifying internal consistency).
- Supports: AIFF, ASCII, Bytestream, GIF, HTML, JPEG, JPEG 2000, PDF, TIFF, UTF-8, WAV, and XML.

Image Metadata Editing – ExifTool

- A free command line program for reading, writing, and editing image, audio, and video metadata.
- Works with just about every file type

Adobe Bridge allows for batch renaming and image rotation, view thumbnails, view image properties, and edit metadata. Works in conjunction with Photoshop.

Pre-Capture Set Up

When selecting materials to be digitized, consider:

- Research value
- Access [is it something that is frequently used or that people don’t generally have access to]
- Legal issues [copyright]
- Cost
- Proper equipment
- Trained staff
- Ability to group together materials that aren’t together
- Will OCR be used
- Condition of original item
- Rare or unique

Issues/Problems

- Text obscured by folds or creases
- Paper or photographs that are cockled and need to be flattened
- Tears in paper or text area that are longer than ¼”
- Books with loose joints, detached spines, or boards
- Restricted bindings that open less than 90 degrees
- Acidic or otherwise damaging housing that needs to be replaced
- Difficult formats, such as scrolls, accordion books, palm leaves, panoramas, or oversized items

Risks

- Use extra care with materials in poor condition.
- Limit light and heat exposure
- Keep all materials in their original order
- Ensure that no materials are misplaced

Good to have an area set aside for materials to be digitized, those in process, and finished work.

Adequate work space

Keep scanning area clean and the floor uncluttered [to avoid tripping], no bags or coats [tripping hazard and for security]

No food or drinks; Use only pencils

Targets, color bars, and grayscales

Used when using a digital camera

Targets and color bars are used to measure resolution, tonal range and color fidelity.

Color targets are a scientifically created set of color patches with established numerical values and neutral gray. They help to ensure that the imaging system you are using is effective and accurate and performing in a consistent level of quality over time.

Including color bars allows for later color adjustment, though it does not necessarily have to be used with every scan.

Targets are a way of predicting image quality and help ensure that the imaging system you are using is producing the best quality image it can and is operating at a consistent level of quality over time.

Resolution targets allow projects to measure the level of detail a particular piece of equipment can capture. Helps to ensure that your camera is meeting your needs.

Some come with software to assist in calibration/focus.

Color bar should be close to, but not touching materials being digitally captured.

Bit Depth

Bit depth measures the number of colors (or levels of gray in grayscale images) available to represent the color/gray value in the original work.

Bitonal – 1 bit per pixel [should only be used for typed documents] 2 colors

Grayscale – 8 bits per pixel [alright for some documents and black and white photos] 256 colors

Color – 24 bits per pixel [8 bits for each color: Red, Green, and Blue] True color -> 16,777,216 colors / 48 bit -> 281 Trillion colors [24 or 48 are the best choice because they most accurately represent what the item looks like]

Informational value – only concerned with the information on the page

Artifactual value – usefulness or significance of an object based on its physical or aesthetic characteristics

Color Modes

RGB – used on the monitor
CMYK – used on printers

As a result of this items which are displayed on a computer monitor may not completely match the look of items which are printed.

RGB can be converted to CMYK, but it is not exact

Color Space and Profile

International Commission on Illumination (CIE) developed one of the first mathematically defined color spaces, known as the 1931 XYZ color space. [***on slide - image with sRGB and Adobe RGB***]

Two of the most commonly used working spaces in digital photography are Adobe RGB 1998 and sRGB IEC61966-2.1.

sRGB is an RGB color space that approximates the color gamut of most computer monitors and is the standard color space for displaying images on the internet. sRGB's color gamut encompasses just 35% of the visible colors specified by CIE. Although sRGB results in one of the narrowest gamuts of any working space, sRGB's gamut is still considered broad enough for most color applications.

Adobe RGB 1998 was designed (by Adobe Systems, Inc.) to encompass most of the colors achievable on CMYK printers, but by using only RGB primary colors on a device such as your computer display. The Adobe RGB 1998 space encompasses roughly 50% of the visible colors specified by CIE — improving upon sRGB's gamut primarily in blues and greens.

For Grayscale – choose either Gray Gamma 1.8 or 2.2 [2.2 recommended for WDL]

- 1.8 originally designed for Apple’s OS and 2.2 for Microsoft
- Due to modern color management on computers, both display accurately on either OS
- The important thing is to embed the color space profile if you are distributing your images so that recipients can view it accurately.

Resolution

Resolution determines the quality of an image.

Higher resolution [PPI] images contain a more accurate representation of the original. But as resolution increases, image quality will level off.

There is no one “perfect” resolution to scan all collection materials.

- Resolution should be adjusted based on the size, quality, condition and uses of the digital object.

The combination of PPI and size of the original object determine the resolution needed to accurately capture as much information about the original object as is available.

The deep zoom feature on WDL will more easily show flaws of low resolution images.

Since higher resolutions are capturing more information, file sizes also increase. [see example]

If storage space is an issue, this formula will help determine the total size of the files before scanning.

Just because a program says an image file is a certain PPI (say 400 PPI), doesn’t necessarily mean it is 400 PPI at its original size.

We recently received image files for a book that was supposedly scanned at 470 PPI (an strange PPI number). Photoshop listed its dimensions as 97mm x 64mm (about the size of an iPhone screen), although the metadata we received had the book dimensions as 433mm x 285mm. So in effect, the book was scanned at ¼ of its original size and the PPI at 100% was about 105 PPI. I’m not exactly sure how this happens, but it is something to be mindful of while digitizing.

File Formats

Master

- The archival file from which derivative (or access) files are created.

- RAW - An image file that contains unprocessed data.
  - Digital Single Lens Reflex (DSLR) cameras and some high-end scanners allow users to capture images in a RAW or native file format that is unique to each manufacturer.
  - The proprietary nature of these files is of concern for the long-term preservation and access of these digital files.
  - Must be converted to an open standard format such as JPEG or TIF [33% smaller than TIF].
  - Processing RAW files creates an additional step in the imaging workflow and may require sophisticated photographic skills or expertise.

- TIF - Is the format of choice for archival and master images.
  - It is versatile, widely accepted, open standard image format and considered the professional image standard.
  - It covers most color spaces, metadata is supported, and it can compressed without loss of quality or data [LZW]
  - Due to the large file size of TIF images, they are not suitable for web delivery.

- JPEG2000 - Is a wavelet-based standard for the compression of still digital images. [about 30% the size of a TIF, but can be reduced further]
  - It was developed by the ISO JPEG committee to improve on the performance of JPEG while adding significant new features and capabilities to enable new imaging applications.
  - JPEG 2000 is being used increasingly as an archival image format.
  - Not necessarily a good idea
    - Limited color spaces available [could impact future migration to a new format, but JPEG committee working on it]
    - Older JPEG2000 files may have lossy compression
    - Not widely supported by imaging programs and can make editing more time consuming

Access

- JPG – Uses lossy compression [data is lost]
  - Not designed as an archival format; initially designed to limit file size and allow for quicker internet access
  - When edited and resaved more data is lost because it is being compressed again
  - Has 12 levels of compression
  - Performs best on photographs and paintings of realistic scenes with smooth variations of color and tone.
  - When used with text or on maps, blurring may occur

- PNG – Uses lossless compression
    - What WDL uses for access images
    - Uses a more efficient compression algorithm than GIF, is open source and supports true color images.
    - Was designed for transferring images on the Internet
    - Does not support non-RGB color spaces
    - Is supported by all internet browsers

- PDF – Uses lossy compression
    - Used by WDL for downloading complete item
    - Good for documents [esp. books]
    - Originally proprietary, but as of 2008 it has became open source
    - Often used for born digital documents

Compression

- LZW – Lossless
    - Best choice for compression of TIFs
    - Reduces file size by almost 60%

- JPEG2000 – previously discussed

- ZIP – Lossless
    - Must be unzipped
    - Reduces file size by almost half
    - Works as a compression type as well as a container for files
    - Allows for a number of different compression types

Storage

CD – holds 700MB

- May not be practical for storing large amounts of data
- Stability depends on the type of dye used; (phthalocyanine) green dye is good, gold discs are best
- Burning speed can affect quality (best to burn discs at 4x or 8x)
- Should be stored in jewel cases
- Claim to last 15-100 years, but can fail within a few years, especially if not stored in a cool environment 41-72 degrees Fahrenheit (5-22 degrees Celsius)
- Need to be regularly tested to check file integrity (checksums can be helpful)

DVD – holds 4.7GB

- Can hold over 6 times what a CD does
- Not as stable or long lasting as CDs
- Should be stored and checked regularly like CDs

Hard Drive – Sizes keep increasing, but 2TB drives are reasonably priced [$100-$200]

- Can be expected to last at least 5 years, but may last much longer
- Internal or external
- Safest and most cost effective method for storing files
- Solid state drives [uses flash memory like a thumb drive] [can be expensive]
    - More stable
    - Faster than traditional HD
    - Should retain data for up to 10 years while powered down, but it is recommended that it be turned on and checked after at least 5 years

Tape – various sizes and formats

- Work better as a backup than for general storage

Cloud – limitless

- Works better as a backup than for general storage
- Data is stored on multiple virtual servers which are generally hosted by third parties
- Could get expensive over time [monthly charges]
- Not the best choice if internet access is poor

<u>Backups</u>

- All of the above can function as a backup
- Good to have an offsite back up in case of catastrophe
- Hard drive is the most common and affordable
- RAID array
    - A RAID array consists of a number of drives which collectively act as a single storage system and can tolerate the failure of a drive without losing data.
    - These drives operate independently of each other.
    - The number of disk drives on the server or even a workstation PC set up as a storage device, will help determine how the RAID array on the equipment will be formatted.

Issues

- Capacity
    - ensure you have enough HD space
    - regularly buy additional storage as the collection grows

- Transfer errors – use checksums to protect against this

- Bit rot - refers to the decay of physical storage media or the breakdown of the material onto which the data is stored

- Sustainability, obsolescence, data migration
    - None of these storage solutions last forever
    - Some may even become obsolete in the future
    - It is a good practice to regularly migrate data onto newer storage device

Quality Review / Post-Processing

- Do a visual inspection of images [if it's a large number of files, do a sample review] to detect any digital issues (which I'll go into further detail later) or missing pages
- It's a good idea to have different people doing the scanning and the review

Use image validation software, like JHOVE

Color Correction and Sharpening are just some of the post processing tasks that can be performed. Cropping and de-skewing are the easiest.

Color Correction can be done using the curves tool [modifies contrast as well as color and brightness of the image] or histogram equalization [increases the global contrast of an image].

- These can be complicated and may require specialized training

Problems

Noise - Unidentifiable marks picked up in the course of scanning or data transfer that do not correspond to the original. Usually caused by the sensor and circuitry of a scanner or digital camera. Smaller sensors are more prone to noise issues; using a digital back helps prevent noise. Some programs have noise reduction filters.

Artifacting – Visual digital effects introduced into an image during scanning that do not correspond to the original image. Some types of artifacts include pixilation, dotted or straight lines, or regularly repeated patterns. These can be caused by hardware or software malfunctions [conversion from analog to digital], compression, dust, scratches, or streaks on the scanning surface or camera lens.

Vignetting - Is a reduction of an image's brightness or saturation at the periphery compared to the image center. Can be caused by improper lighting, using the incorrect aperture setting on the camera, or capturing the image at an angle. This is not a very common problem in digitization, more common in standard photography.

Chromatic Aberration - Is caused by the failure of the camera lens to focus all colors to the same convergence point. It is seen as "fringes" of color along boundaries that separate dark and bright parts of the image. Many high end cameras have lens made of low dispersion glass to help prevent this.

Over-sharpening – Some cameras and scanners automatically sharpen images, although this is not recommended and should be disabled if possible. Over-sharpening usually occurs in post-processing. Can create sharpening halos and increase the visibility of jagged lines, noise, and other image artifacts.

Depth of field - Is the distance between the nearest and farthest objects in a scene that appear acceptably sharp in an image. DoF issues can occur when digitization with books and artifacts that do not lay flat or are three-dimensional. With a book, some words may appear sharp while others are blurred.

Color Reproduction – When the color of the digitized image does not match the color of the original. Can be caused by poor lighting or improperly calibrated equipment. If a color bar is used, this problem can be somewhat easily fixed. Without a color bar, and when you don’t have access to the original, any attempt at correction will be a guess.

Lens distortion – Two main types: barrel [pictured] and pincushion. Wide-angle lenses and wide-range zoom lenses often suffer particularly badly from this. Rare in digitization.

Clipping – Overexposure/Underexposure - Is the result of capturing an image where the intensity in a certain area falls outside the minimum and maximum intensity which can be represented. It is one of the most common problems and can occur when operators don't understand the controls in the scanning software and inappropriate settings are selected. It is easily seen when looking at the histogram of the image. Can be reduced, but not fixed.

Other Considerations

Copyright – make sure materials being digitized are not copyright protected [check local laws]

Funding – Sources of funding, one time allotment or on going funding. Initial funding may be used for equipment, but ongoing funding will be needed for staff and maintenance.

Staffing – Ideally a digitization team should have a director, metadata specialist, technical support staff, a curator, an imaging specialist, and several digitizers. Staff should be well versed in the digitization process and have knowledge of how equipment is operated.