Digital images forever:
implementing an imaging system in a cultural institution

by

Scott Donkin

Computing Arts: Digital Resources for Research in the Humanities 2001

University of Sydney
26 - 28 September 2001
Sydney, Australia

Introduction

This paper deals with the decision-making and challenges that arose out of the implementation of a large-scale digital imaging project by the Powerhouse Museum. It discusses the use of digital imaging in the Powerhouse Museum as well as the development and use of its Image Management System (IMS).

Also covered are some of the procedures and techniques used by the Powerhouse for capturing images as well as the reasons behind such procedures.

There is an overview of some of the digital imaging equipment used by the Powerhouse and details of our experiences. Any mention of specific product brands or models should be taken as just an opinion based on experiences we have had and are not intended to be a recommendation. Suitability of a product for a job will vary according to individual needs.

Any monetary values are used as a guide only and are in Australian dollars (AUD$).

For the purposes of this paper, use of the terms 'digital image' and 'digital imaging' shall be limited to photographic imaging of two- or three-dimensional objects using digital equipment or scanning of conventional photographic materials, with regards to producing and maintaining a digital image archive. The terms are also used to relate to the methods and procedures used for capturing, manipulating and outputting images electronically for the purpose of archive, retrieval and sharing as well as the catalogue information that relate to the images that are captured and stored. Document scanning or document management is not covered.

What is Digital Imaging?

Quite simply, digital imaging is the capture or creation of an image in an electronic form. This includes the electronic manipulation, storage, output and management of the digital image using anything from a desktop computer though to dedicated graphics workstations and related servers.

Unlike their conventional photographic counterpart, digital images are binary data files stored on, for example, magnetic or optical media. Seemingly exact copies of a file can be made at will and shared with others all over the world almost instantly and much more economically.

 A digital file can be enhanced, manipulated or changed using image-editing software on a computer workstation. It can be output to a digital printer that yields a print which looks and feels exactly like a professionally produced photographic print – and as such, depending on the printing process, have its own permanence characteristics. Using the right equipment, digital image files can also be *written* onto conventional film for slide projection, though such slides if used for printed reproduction are of questionable quality.

A digital file can be transmitted over a network or via a microwave link, for example, and placed on a page in a newspaper or magazine and published on the same day.

Unlike conventional photographic materials, a digital file itself will not fade or deteriorate in time. Stored on appropriate media, an image file will theoretically last forever. In reality it is the stability of the media and the availability of technology required to 'read' it that becomes an issue.

One of the simplest ways to capture a digital image is with an electronic *digital camera*. Pocket-style versions can cost as little as a few hundred dollars with the professional studio set-ups running up to and beyond $100,000. Digital images created this way never need to be copied onto conventional film or paper material. They can quite adequately be displayed on computer screens and stored on a disk.

Images that have been shot conventionally and represented on film (positives or negatives) or paper prints can also be captured digitally using scanning equipment. Scanners are also available to everyone from the home-office user for a few hundred dollars through to the graphic arts pre-press units. The latter generally start in the high four-figure region and go up to over half-a-million dollars.

The phrase "you get what you pay for" certainly applies to digital imaging. The more you pay for a digital camera or scanner generally equates to better image quality and, particularly, the size of the printable image. A low-end digital camera will only offer results equivalent in size and quality to instant *Polaroid* material when printed as hard copy. Similarly, budget scanners offer low-grade results: certainly not suitable for reproduction on a printing press and enlargements not much bigger in size than the scanned original.

For an organisation interested in creating and maintaining faithful digital images of their collection, events and so forth, a complete home-office budget digital imaging set-up costing a few thousand dollars will not offer much in the way of image quality, speed or reliability. For the purpose of an archive, the idea is to make a faithful reproduction of the original object or scene. Budget digital imaging equipment does not offer these kinds of results.

Digital Imaging in Cultural Institutions

In broad terms, it can be said that most cultural institutions maintain collection objects as well as supporting photography and documentation of those objects. These images and archives are used for recording the status and condition of the collection as well as for such uses as research, educational and public relations.

Cultural institutions are in the business of maintaining their collections 'forever'. However, such collections must be accessible for them to have any worth. Regardless of the storage techniques used for these collections, it is the physical accessing of the individual objects that accelerates their degradation or, worse, risks damage due to handling mishaps.

The use of conventional photography has allowed a record of the current state of an object or scene to be captured for all time. Images of collection objects make them accessible, allowing staff and the public close examination of the object without risk to object condition.

Ironically the conventional photographic materials that are used as a record of these objects and scenes won't actually last forever. Such materials have a definite life. Depending on the materials and atmospheric conditions (Wilhelm et al, 1993; http://www.wilhelm-research.com) colour photographic transparencies can be expected to show noticeable colour shifts and losses of density after around 15 years or so. Properly processed and stored prints can last anywhere from 10 years for minilab prints through to 50-plus years for commercial colour prints and hundreds of years for Ilfochrome (Cibachrome) prints. These figures vary of course, but it is the constant handling and the standard of storage in most institution Photo Libraries[1] that lead to a greater rate of deterioration. To minimise fading and deterioration, film and paper-based images really need proper storage such as an environment of low temperature and humidity with no light and little or no manual handling. Of course, photographs stored in such an inaccessible archive using these techniques are of little use to anyone.

Therefore, much of the film and print material in most institutions' photo libraries that are used for research and reproduction will probably become a less than an accurate record within a matter of years due to continual handling as well as environmental effects.

With the advent of digital imaging, photographic images can be electronically captured and stored along with appropriate database records in a more compact, long-lasting and easily retrievable way. Theoretically this means a digital image of an object or scene will be an accurate record forever – providing they can be accessed using future technology.

Therefore, the methods and technology used to record and hold the digital images as well as the reliability of the catalogue data also have direct influence on the accuracy and future use of the digital files.

---

[1] The permanence of conventional photographic film or prints depends on many factors including the material itself and how it is handled and stored, as well as exposure to temperature, humidity and light and UV levels (Wilhelm et al, 1993; http://www.wilhelm-research.com). It is considered here that the Photo Libraries of most cultural institutions are designed for quick, easy and constant access and are not huge cold-storage archive facilities designed with archival permanence in mind.

Image Quality

Throughout this document I refer to the term *image quality*. In designing and implementing a digital imaging project for capturing and arching an object collection, one of the expected outcomes will be a system which provides images that are clean, sharp and a good representation of the original collection objects, scenes and so on.

The perception of quality is a highly subjective one. In terms of images, most people with little or no knowledge of imaging or photography still have a pretty good idea of whether an image or collection of images are of good or poor quality.

In that we can see that the Project team involved in an imaging system will want to provide images suitable for the purpose of the archive and their target audience.

Such images therefore should be sharp and reasonably colour faithful for them to be a good representation of the original. At the same time, images and output derived from them should be free of the dust spots that generally result from the capture process.

The Powerhouse Museum

The Powerhouse's Photo Library is home to around 650,000 film frames taken by its photographers – with around 1,000 new frames added each month. Many frames are actually bracketed exposures[2] of the same shot. Therefore there are probably about 300,000 *different* images of collection objects, events, exhibition launches, publicity shots etc.

Images held by the Photo Library are in constant demand. For some Museum staff, accessing these images can be time consuming and inconvenient. The Powerhouse's site is rather large, occupying two city blocks plus the Sydney Observatory several kilometres away. Staff wanting access to Photo Library images have had to either 'describe' to the Photo Librarian what they are looking for in an image or actually go to the Photo Library to manually search for images for their projects. Either way, this has not been an ideal situation. Also, each use of an image requires manual handling and poses a risk to the original material itself [3].

Due to the increasing demand for images – especially simultaneous requests for the same images, it became necessary to develop a system whereby staff could access images over the Museum's computer network.

In 1996 and after successfully bidding for a capital works grant, the Powerhouse embarked on a 5-year *Imaging Project* to develop its *Image Management System* and build an *Image Resource Centre* (IRC). The IRC - a digital-imaging department, complete with imaging hardware and workstations, would then be responsible for the capture and archive of images and related data.

---

[2] In order to guarantee at least one well-exposed image, a photographer will often *bracket exposures* by making slight changes to aperture or shutter settings on the camera or by adjusting light levels. Often this means taking one image at the determined exposure setting, plus one slightly 'lighter' and one slightly 'darker'.

[3] The PHM Photo Library mainly holds either black & white or colour images on film material (negative or transparency) in one of three film formats. All film material is 'original' meaning it is the chemically processed, original film from the camera. As such, no first generation or 'negative' exists. Handling this original material therefore poses risks to the image.

The Powerhouse Museum's Imaging Project

The goal of this project was to digitally capture images held in the Powerhouse's Photo Library and initially make them available to staff over the internal network via client software.

Images that were regarded as either important or reusable were initially chosen for scanning and archiving with related data. Eventually, these images would also be made available via the web to external customers and the public. To achieve this goal, suitable client and server software would be required, as well as the equipment to support it.

At the time of the Project's implementation, there was little available in the way of image archive and retrieval software; or at least none that suited the purposes of the Museum and addressed all the project requirements. It was therefore determined that client software be developed for searching and viewing images held on a server as well as other modules for capturing images and data as well as administration.

The Image Management System (IMS) software was created by a development company and linked to the image and data repository based around an Oracle database (www.oracle.com) and a FileNet document management system (www.filenet.com). The development of the IMS was based on specifications & functionality created by the Powerhouse's *Image Project Team*.

The development of any custom software is normally a huge undertaking and the Powerhouse's IMS was no exception. The time and effort costs were enormous with the development lasting several years: endless meetings, beta testing, bug reports, retesting, more meetings – the usual process of creating, evaluating and rolling out custom software.

The final product – the Powerhouse Museum's Imaging System was rolled out in mid-2000 and does what was intended. It delivers images and related data to staff for the purposes of research, promotion, education and more. The public also has access to the IMS by visiting the Museum and using a specially set up terminal for image searching and viewing.

These days there are some good packages available that handle image and data archiving that would probably suit an organisation the size and diversity of the Powerhouse; meaning if we had to do it all again there'd be little need to develop custom software.

Decisions

What to capture

The decision of what photos should be captured to an Image Management System will normally be governed by the size of the budget and the available time and resources.

The Powerhouse's Photo Library, as already stated, stores around 650,000 photographic images. It would be pointless attempting to scan each of the best exposed frames – possibly 300,000 or more images; as such a venture would take years and be very expensive.

Many of the Powerhouse's Photo Library images are of events such as exhibition launches, people visiting, educational programs, as well as general PR images and so forth. And while these images may have significant importance to those involved in the individual projects, the long-term use of many of these images is doubtful. Certainly in the Powerhouse's case the demand for such images would be intense for the two or so weeks after such an event then would rapidly decline to very little or no demand at all. Therefore, most of the images representing these categories were excluded from any immediate plans for capturing to the Powerhouse's IMS.

On the other hand, images of collection objects are very suitable for the Museum's Image Management System. Such digital images could be used and reused to provide instant access to the objects thereby negating the need to disturb the object itself.

Quite a proportion of the photos of collection objects in the Powerhouse's Photo Library consist of multiple views and / or detail shots of an individual object. In some cases, shots of individual *parts* [4] of objects are also taken. If all of the images relating to one object are captured, there would be considerable extra time and expense involved in acquiring all of these images into an IMS – particularly if hundreds of thousands of photos were involved. For the reasons of time and budget, it is sometimes necessary to exclude more than one or two alternate views of an object at the scanning stage. Of course, this practice can be cause for hot debate amongst curators, registrars and imaging folk alike.

It was therefore determined at the time of the Project implementation that representative views of collection objects and views of exhibition installations be scanned and saved as a priority; with other images scanned as they were requested. So staff requiring scanned images of, say, the dinner table set up at a function, for use in an upcoming brochure, would get their scans provided on disk by the Museum's Image Centre and a copy of the digital file would be saved to the IMS.

Apart from existing photography, other digital projects have been implemented and continue to date. The Conservation department has a full-time photographer taking digital camera pictures of objects before and after treatment. The Museum's Photography staff not only shoot highly creative imagery of collection objects, events and so on, but also identification images of objects as they are acquired.

---

[4] Where an object has a separable portion, such as the cover of a teapot, this constitutes a separate 'part' of the main object and so receives a sub- number when the object is registered.

Copyright issues

The Museum makes it practice for its photographers to photograph all its collection objects for 'future use'. But just because a staff photographer has taken the photographs, and therefore the Powerhouse owns the copyright of that *picture*, the Museum does not have the automatic right to reproduce the image.

Contemporary art objects make up at least part of the Powerhouse's collection. Copyright of the objects, in many cases (unless otherwise agreed), *might* remain with the artist. The action of photographing these objects is in fact a copying process and as such *could* be a violation of copyright. Of course, photographs of such objects *might* be OK under the *Copyright Act*[5] for reasons of research and education. However, saving the images to an image repository for future 'use' may not be a good idea in case those images are accidentally used for a purpose that may put the organisation in a difficult situation.

In saving images and related data to an imaging system, a means to block the export and printing of images (as well as 'screen grabbing') would need to be employed to limit use of copyright protected images to research (viewing on screen) only. At present, the Powerhouse's IMS does not address such issues but instead relies on the diligence of staff to check the copyright status of subject matter with the relevant curator and the Rights & Permissions (copyrights) Officer. This is obviously not an ideal situation. The Powerhouse's IMS does, however, limit the printing and / or exporting of all images according to the logon privileges set for the user. This tends to be more of a burden than a help as even legitimate users of certain images are forced to apply to the Photo Library or the Image Centre to get printable versions of the files. Obviously this undermines the whole purpose of the IMS: to provide instant access to images.

A second phase of the Project, to have begun after the initial five years, has still not been addressed. Part of this new phase was to enhance the IMS to provide limitations upon printing or saving images based on copyright status of individual files, regardless of logon privileges.

For this to be implemented there needs to be copyright data relating to each of the images. To date there has not been an ongoing process of recording copyright status of each of the photographic images held in the Photo Library. So even if limitations on IMS accessed images based on copyright status can be made, such status for each image will need to be researched and recorded. With the Photo Library staffed by just one librarian plus a copyright officer, each whose time is consumed by ongoing internal and external requests, there is little chance of much research into copyright status of even part of the entire image base happening at all. The problem is exacerbated by the fact that images of many objects would need to be referred to the relevant curator. I'd be surprised if similar problems weren't faced by other large institutions if only because of the complexities of the Copyright Act.

The Capture Process

From 1996 onwards, contractors and specialised staff based in the Image Resource Centre scanned selected images and saved them to CD, ready for upload when the IMS software was completed.

---

[5] Always make your own inquiries. Refer to a copyright lawyer or the Austalian Copyright Council for advice. www.copyright.org.au/

Images were saved to CD's, with a backup set of CD's sent off-site to a fireproof storage facility. Data relating to the images was saved to an interim *Microsoft Access* database and recorded information as original Photo Library ID numbers, a unique scan number (see scan project numbers, below), collection object numbers, names of people & places etc, and any relevant captions.

This interim Access database was not used to record the full details of collection objects. The Museum had already been using for some time the Argus Collection Information System with complete details for each object recorded. It would be pointless to double up on information by including it in the IMS, so the idea was to provide an API[6] link to Argus from the IMS via the object number. This was to happen once *Argus 2000* was released. However, due to some other issues and problems with the release date and stability of Argus 2000, this feature has yet to be added to the IMS. Again, this was to be a *Phase 2* implementation.

There were around 60,000 records of captured images held in the Access database. Once development of the Museum's Image Management System software had been completed, data in the interim Access database would be imported. However, problems relating to data integrity ensued. Notably, the external company commissioned to create the forms in the database had not provided for any input masks. That is, there was nothing stopping a user from entering something other than what was intended to be entered into a field.

For the first eighteen months of the Image Project, the scanning contractors worked unsupervised. They were hired by the Museum to do the scanning and data entry but the Image Centre Administrator's position hadn't been created until later 1997 so there were no methods in place to audit their work.

Once it was time to upload the data to the IMS, errors occurred left, right and centre. Mostly these were due to alpha characters present where numbers should be and so on. One operator seemed to have felt that anything scanned that needed review warranted a question mark character appended to the eight-character scan number. To compound this, the operator decided that if he was unsure of the project number of the scan, he'd embed the question mark within the scan number, after the three digital project code (see *File Naming*, below). All of this could have been avoided by programming the Access database interface to restrict data entry fields to accepting only the data intended for them.

Further problems were encountered when the IMS software was completed and image and data uploads were in progress. Where a record indicated that a transparency or negative belonged to a range (or sheet) of images, the records would show the first and last record reference separated by a hyphen. In some records, the operators had not followed protocol and were saving the ranges with other separators including commas and or spaces and double-spaces. The Image System had not been prepared for this and crashed as the uploads were being processed. As simple as the fixes for these errors sound, they were the cause for a lot of extra work and testing. In short, the problems could have all been eliminated by better supervision of staff and tighter database design.

Most of the 140,000 scanned images and data files that were created before the IMS rollout have now been uploaded to the IMS. However, the process is largely done by hand and requires close attention to detail so the final 30,000 images will probably take another couple of months to send to the IMS.

---

[6] Application Programmer Interface (API): some software programs have built-in 'interfaces' that allow developers of software to link to / launch features of these programs.

Of course, the Museum is continually acquiring new objects – and deaccessioning others. The object information for these items, recorded in *Argus*, is uploaded as a flat-field ASCII file to the IMS on a nightly basis. When images of these new objects are uploaded, the data is waiting for them. That said, the deaccessioned objects are still held in the IMS and accessible via searching the related Object number(s). A feature that was perhaps overlooked was the amendment of IMS records to reflect the deaccessioning of an object. Any image and its data needing to be removed from the IMS because of deaccessioning or because the image of the object has been otherwise requested to be removed, has to be done manually.

With the 5-year project now officially concluded, it is the Museum's responsibility to provide the continual funding required to sustain the IMS and the Image Centre. That is, the staff, the hardware and the recently completed 100Mbit fibre optic network throughout the buildings.

New photography is continually being produced by the photography department and added to the Photo Library. Even considering some of the original decisions on what to exclude from scanning projects – such as most PR images would not be added to the IMS because of the short lifespan of the images, there are still around 1,000 *tradition* photographic images to consider for scanning each month. Now that the project funds have all been spent, the Museum must rely on the Image Centre's 2-person operation to take care of the scanning of ongoing photography as well as the increasing demand for staff and external requests for purpose-made scanned images. There was talk some time ago of training the photographers to use the scanners to capture and catalogue their images to the IMS but this has still not happened. To date, the Imaging Centre and the Photography dept remain separate entities.

Then there are the digital camera images. Images captured using the professional digital equipment by the Museum's photographers plus the point-and-shoot images taken by staff. Although digital cameras replace the need for conventional image handling and processing, the files themselves still need a degree of electronic processing. This can be anything from a crop of the image or tonal adjustments through to noise reduction, colour correction and sharpening.

It is expected that, with the various digital imaging projects the Museum has running at any one time, around 25,000 scanned photos and digital camera images will be added to the IMS each year. The Powerhouse has the opportunity to capture up to 50% more images than this with the addition of just one more Image Centre staff member. However, with continued staffing cuts experienced by all government departments, the chances of this becoming a reality are slim.

As the Museum moves closer to complete digital image capture procedures, it would be hoped that this total would increase. However, high-end digital camera files do require considerable electronic processing, especially if colour matching (to an original) is required, so retraining of photography staff will be needed as well as automated processing of the digital files if increased throughput is expected.

The Image Resource Centre

Whilst the Powerhouse's Imaging Project was being developed, the "Image Resource Centre" (IRC), was created. A digital imaging centre equipped with state-of-the-art

imaging hardware, its purposes were to both capture images and data for the IMS and to provide services and self-services to staff for their projects.

During and since the completion of the Imaging Project, the IRC has evolved into an in-house service bureau offering high resolution scanning, manipulation and printing services for staff and external clients.

The Image Centre also provides self-services for staff for scanning and printing their own images for work projects. In addition, the IRC offers the loan of around fourteen small digital cameras to staff for quick shots for identification and research purposes. Most all of the images taken for these *other* projects are saved to the IMS.

The self-service use of small digital cameras for identification shots has been a boon for staff. Some may argue, though, that the low quality (compared to traditional photos) of pocket digital cameras - especially by inexperienced photographers, are not suitable for saving to an Image System. However, we must not lose sight of the fact that any and all images have some sort of value – if even only for identification. The IMS was set up first and foremost as a research tool. That is, many users simply need to view an image of a collection object on screen. Digital cameras are perfect for that.

Digital Cameras – staff become photographers

Shortly after taking my post as Image Centre Administrator, I noticed that there was a huge demand for the ten or so shared Polaroid cameras around the Museum. Every time an object was acquired, Polaroids were taken. Each time an object was considered for an upcoming exhibition, more Polaroids. And so on. The problem with instant Polaroid pictures is that they are not archival and the quality is variable. Most of all, the prints are self developing using an alkali gel which dries naturally – the prints are not washed. Therefore, residue of these chemicals can contaminate other things they come in contact with, such as files and, more importantly, people.

The cost of running a Polaroid camera is obvious. Around $2.00 per shot and a user will often shoot more than one shot to ensure they've got a good image.

In early 1998, I managed to convince the Museum management of the benefits of replacing the Polaroids with small digital cameras. We purchased 14 *Sony Mavicas*. These cameras take a floppy disk as media meaning that, when it was time to download, print and save their images, the user didn't need to fiddle with connecting cords or have any understanding of computer directories or file management. On the downside, straight from the camera the images we highly compressed – low quality, JPEGs. In any case, compared to Polaroid, the Mavica images were still superior and the benefits outweighed anything that could be achieved with the Polaroids.

The idea was that the cameras were loaned to staff for work project use. A dedicated workstation was set up with a Tektronix photo quality laser printer for processing and printing. Photoshop was configured to make proof sheets with captions (current versions of Photoshop now do much of this for you) and all of a sudden we had a digital camera facility for staff giving them better than Polaroid images, one, four or eight to an A4 page with captions, and output costs of around 12 cents per printed image. In short, the cameras were a runaway success.

All staff had to do was keep a simple record of each object they photographed, such as a registered object number, and hand that and the disk over to the Image Centre. The images would be then uploaded to the IMS for all future use and the disks would be recycled.

This all meant that the Museum was saving around $5000 a year in Polaroids and the IMS was getting a lot more images, particularly the ones that we had no other record of.

Well, it all worked in theory.

After the eighteen months, there were over 1,000 disks issued to staff for use in the Mavica cameras. This probably represented an average of 20 images each (Mavicas manage to squeeze up to 35 files on each disk). Therefore around 20,000 new images were waiting to be returned - perhaps 10,000 of them suitable for upload to the IMS.

However, people being people… we faced one main obstacle. The cameras made it so easy for staff to grab quick shots for a report or proposal that they would quickly shoot, download and print the images they needed and then put the disks in their desk drawers with the honest intention of writing up the data sheets 'later'.

Of the 1,000 disks issued in that time, about 750 were never returned by the Mavica users. Most of them would come to me apologising for not handing in the images and data sooner and promising to do it 'next week'. Well, you know the rest.

Obviously the system had its benefits in the time and materials savings from the alternate systems. However there had to be a way of allowing these people the ease of capturing and using digital camera images whilst enabling the Image Centre to get a copy of those images and the data.

Then I remembered that I myself had been guilty of similar actions a year or so earlier. I joined the Powerhouse Museum in 1996 as a photographer after more than 12 years of being a self-employed commercial photographer. In my spare time I would scan and catalogue the images I'd taken over the years but soon realised that it was a daunting task that needed some sort automation. Using my skills as a computer programmer, I created an application that would gather images and data as they were scanned.

With this in mind, I modified and enhanced my original home-made programme into a utility that would automatically download digital camera (or scanned) images and present them to the user, asking for information along the way.

The Mavica digital cameras were removed from service and given to the education and visitors section of the Museum for school kids to use. In their place I purchased 15 new Kodak DC290's. These cameras have a flash card for storing images and download is via a hot connectable (you can leave the camera and computer turned on) USB connection.

The staff, now using the new cameras and unable to pop a floppy disk into their own workstation were now forced to download their images in the Image Centre (which was the original idea anyway). Upon doing so, the application I created, called DCD (for *Digital Camera Downloader*) would begin the download of their images. Once complete, the images on the camera would be cleared and the camera readied for the next user (all automatically).

Now, the user would be presented with a list of all their images. Clicking on any one shows a reasonably large preview of the image. Now for the good part: once the image is displayed, the user is asked if they wish to 'keep' or 'delete' the image. Deleting it means they lose it forever (well, there is a failsafe mechanism, but they don't know that..). Choosing to *keep* an image means they must provide information.

Until they provide some information, the user cannot make any prints nor can they transfer their images to another workstation. They are then asked about the *type* of image it is: object, non-object, research, report etc. Depending on their answer, they then need to type in the information relating to that image type. As soon as they press the save button, the image is saved off ready for the IMS after verification by Image Centre staff, and a copy made available in their 'user folder'. Users can then do what they want with the image – lighten, darken etc and they won't be affecting the master file sent to the IMS.

DCD is configured so that a user can 'process' 50 images, supplying all data (usually around three or four fields) in under 15 minutes. This compares well with the alternatives of about an hour completing manual record sheets and downloading via copying or moving files from floppy disks.

Most importantly, all the data they supply is formatted into a comma delimited text file that can be easily read by humans and in a format that is ready for upload to the IMS.

With this new system I developed, the Powerhouse is able to capture an extra 10,000 or more images per year to the IMS whereas previously around 500 were being saved.

Of course, not everyone was happy about such a system. A few curators felt that they and they alone should have control of the images they take. However, the system was ultimately making life easier for them because they now only needed to check the IMS if *any* sort of image exists for a collection object before they go off for an hour or so shooting more images they don't need.

The Image Management System

Users and uses

The main role of the Powerhouse's IMS was to allow staff to search and retrieve images and related data quickly and easily. However, with staff of over 300 and a diverse range of needs for images, the plan was to satisfy as many needs as possible within the budget and scope of the project.

After consultation with a group representing the main image users of the Museum, as well as looking at digital image storage costs, it was determined that the IMS should deliver images capable of being printed at up to A4 size at 300dpi. These specifications would satisfy about 80 to 90% of the image requirements throughout the Museum. Images archived at this size - saved as JPEG format, are relatively easy to store and retrieve and can be printed at good quality on a desktop printer or viewed on screen.

It is important to understand the concept of image size and resolution with regards to output quality. A discussion on this subject continues below, but for now we can say that an image scanned at 300dpi at A4 (210 x 297mm) in the RGB colour space produces a file around 24Mb uncompressed. If the current image base of 140,000 images were saved as raw, uncompressed files at 24Mb each, storage requirements would be around 3.5 Terabytes (3,500Gb). Even based on today's well priced storage solutions, this kind of hard drive space is serious money. And this doesn't take into account the fact that any system administrator worth his weight will want to use a RAID solution which would at least double or triple that space requirement. In 1995-6 when the hardware was being purchased to provide server storage for the Museum's scanning projects, implementation of such storage space was not financially viable.

The Museum instead purchased two CD jukeboxes: one with a 200 disk capacity, the other with a 150 dock capacity plus the ability to burn CDs. Even though it was in the specification for the design of the Imaging System, the developers of the application modules had major problems in making the CD burning Jukebox actually burn CDs. The idea was that images would be scanned and saved to the IMS where thumbnails would automatically be produced and the master file was saved to a hard disk directory. When the contents of this directory had reached a set level, 620Mb, for example, the system would automatically create a CD. Unfortunately, the third-party software components that promised this function just didn't work. So it was decided to drop that part of the functional specification and have Image Centre staff burn CDs by hand. At least they could also verify a CD burn by putting the CD in another computer's CD drive.

Many colleagues have in the past criticised the Powerhouse's decision to save its digital images as JPEG files. It was also the subject of hot debate amongst the Project members. Unfortunately with the size of the project and the costs of storage it was the only way the project could have been sustained.

JPEG is a lossy compression format. Most of the Museum's images scanned to the above specifications were found to save out as a JPEG file at around 1.5 to 4Mb each. This represents a hefty 88% saving in storage space requirements. Now the 140,000 images would require, say 420Gb of storage (at an average of 3Gb each), not including our friend, the System Administrator's need to at least mirror this in a RAID solution.

These days the Powerhouse stores all IMS images on a RAID server (hard drive based). Hard drive storage allows for quick access to requested files. The fact that images are held

in JPEG format means that delivery over the network is also fast. It is only once the image is delivered to a client machine that the file is decompressed to its ~24Mb size. Therefore desktop systems have been configured with at least 64Mb of RAM.

With today's hard drive storage costs being around 5 to 10% of what they were five or so years ago, the Museum can now well afford to save captured images as uncompressed TIFF format files. The Image Centre has in fact been keeping scans of images determined to be highly reusable in TIFF format. Unfortunately, even with more affordable high capacity storage systems, there has not been funding made available to put these mechanisms into place. As at early 2001 the Museum had only invested in enough storage to cater for current storage needs – the 140,000 images it already had captured plus room for a few thousand more. Any files captured as TIFF have been saved by the Image Centre onto CD-R disks and a JPEG derivative made and uploaded to the IMS.

Even if there was a specification change now and future images were archived as uncompressed TIFFs, one question would remain. What would we do with the 140,000 existing JPEG files? Well, we leave them as they are. Every major imaging project is going to experience tough decisions based on technological changes – probably every five years or so.

Regardless of quality differences between an uncompressed TIFF file and its level 10 to 12 JPEG derivative, there are important reasons why JPEG images suit certain types of image repository / server setup. In the case of the Powerhouse's IMS, theses were some of the considerations:

- The images need to be delivered via a network. Our JPEGS at around 12% the size of the same image as a TIFF place far less demand on network infrastructure.
- Both versions of the image, when printed to even a high quality digital photo printer have barely - if any, noticeable differences.
- The same images sent to magazines and newspapers will end up being resaved to a higher compression factor (lower quality) JPEG. These organisations send their images, now imbedded in a page layout, over the wire / microwave link to their print works as a low to mid-quality JPEG file (an average of JPEG level 6).

Of course, the main reason why you would still want to save images as uncompressed TIFF format is for re-usability. As stated, it's hard to pick the differences between the same images saved as a high quality JPEG. That is, until you reopen the JPEG and then *re-save* it. Re-saving a JPEG image means that the application will recompress the image. JPEG is famous for its compression artefacts. Recompression will have disastrous effects upon the image quality of that same file.

Therefore, for an Imaging System delivering JPEG images, only the final saved version of such images should be made into JPEGS and only after all digital processing.

User Needs

During the pre-development phase of the IMS it was determined that staff ultimately using IMS images would generally have the following needs:

1. Search and identification: on screen only viewing of search results.

2. Screen: web pages, email and PowerPoint presentations.
3. Research and Report: insertion into word processor documents.
4. Reference prints: laser and digital photo prints - reference material, handouts.
5. High resolution reproduction: publicity, magazine, catalogues.
6. Display prints for exhibition: high end small prints for exhibition label graphics.
7. Large prints: posters and banners.

Most of the uses mentioned in the points above are being satisfied with images scanned at A4 @ 300dpi. It was conceived the IMS could serve each and every one of the users of the system – except (in the mid-90's), the last three. The cost of scanning equipment with optics and CCD sensors good enough for the demands of high resolution images as required by these last three items was out of reach. In 1996, $25,000 bought you a "professional" desktop A3, 800dpi scanner complete with "transparent media adapter". What you actually got was a scanner capable of scanning transparent or reflective originals up to A3 in size and output them between two to three times the original's size on an inkjet printer. This, of course, was a big thing in those days. We were using a UMAX scanner of with above specifications mainly for scanning 4x5" transparencies and any printed materials we had. The results were so-so. Certainly good enough for the purposes of the IMS but it couldn't be considered appropriate for magazine reproduction. The results in the latter case were poorly coloured and unsharp.

These days the same money will buy a film scanner capable of scanning 4x5" originals up to poster size, pin sharp and accurate in colour. The results perfect for prepress use.

In the beginnings of the Image Project it was decided that until the equipment was available within our budget, we'd send the originals out for scanning. In any case, even if somebody wanted a high resolution digital file, they would generally need to at least find the image using the IMS and possibly make a reference laser print before deciding on the final picture.

Images required at larger sizes - posters for example, would normally be scanned especially for the purpose. In any case, many digital bureaux still prefer to scan their own files because they have gone to great lengths to create an accurate colour management workflow – the setup that guarantees what they scan is what the print (a science in its own right). Of course, one of the advantages of having high resolution scanning equipment is to save on the costs of having scans done outside. Such scanning can cost anywhere from $25 to $75 per image. For anyone setting up their own high resolution imaging centre and intending to send self-scanned high resolution files to a digital print bureau, it is advisable to make contact with the bureau of choice and work together to establish a configuration that both parties can work with.

While the Powerhouse certainly has needs for larger or higher quality magazine work, these would normally represent a small proportion of the total imaging requests in the IRC. In any case, the total number of photos sent to an outside organisation for high resolution scanning has certainly justified the purchase of a high-res scanner.

The Image Centre's High Resolution Scanning Facilities

At the time of the Powerhouse's Imaging Project implementation, the equipment purchased for its Image Centre was of good enough quality to satisfy the 80 to 90% of uses. Images destined to be printed larger than A3 (beyond Powerhouse equipment capability) or for magazine reproduction were sent to bureaux.

More recently, the Powerhouse invested in newer, high resolution scanning equipment that offers reproduction quality scans as well as giving the Museum the ability to produce photo quality digital prints on an A3 high resolution digital printer. This alone has resulted in considerable savings for the Museum in the costs of producing professional standard photos from transparencies and negatives. The Powerhouse also offers this scanning and printing service to other organisations.

The only downside to this $50k scanner, a *Scitex Eversmart Pro*, is it is limited to scanning only 8 bits of information per colour channel. That is, 256 levels for each of the red, green and blue channels. By mixing various combinations of the three primages (RGB), we can get a palette of (2^8) x (2^8) x (2^8) = 16,777,216 colours. This is still regarded as "true colour" in the computer world.

At the time, the scanner was state-of-the-art. Of course, two years down the track and there are now better scanners available in the same (and lower) price ranges capable of capturing up to 16 bits of (or more) information per channel – 48 bits total. That is, two to the power of sixteen colours – 65,536 levels for each of the Red, Green and Blue Channels. This yields a palette of over 281 trillion possible colours. Of course, no screen or printing device is capable of printing this total palette representation all at once. In fact, nothing out there will even reproduce the 16.7 million colours of the 8-bit-per-channel system all at once. However, the advantages of such higher bit-depth systems is that a truer capture of an original is possible than the 24 bit systems (8 bits per channel).

Most importantly, better gradations of colour are possible. Imagine you have a photograph of a collection object featuring a long graduated blue colour. An 8 bit per channel scan will mean that this graduated tone will be limited to being represented by just 256 shades within the blue channel. Scanning at twice the bit depth, 16 bits per channel, allows for 65,536 different levels of that same blue. In any case, most high-end digital photo prints you see at the moment are produced from an 8 bit per channel file.

Photoshop users can archive 16 bit per channel image files for later use and have a better representation of the original. For viewing these images an application must be capable of opening them. Certainly a web page viewer and most image viewing applications won't open such files. For simply viewing on most people's computers, the first need converting to 8 bit per channel. But the process of converting means that a better 8 bit representation of the files is created.

Digital Imaging Hardware & Methods

This section discusses some of the methods of capture and storing images used by the Powerhouse Museum. It also looks at the type of hardware available for digital imaging use in an imaging project.

Scanning & Capturing Images

In line with the Powerhouse Museum's Imaging Project, the Museum captures and archives its digital images with the view of using them for screen display (research, web, PowerPoint presentation etc) or printed output – using one of several networked printers at A4 size.

After images are scanned or photographed with a digital camera, they are cropped to suit. In some cases the image may end up being a square format or perhaps a wide panorama, so not of A4 proportions. Images that are squarer than the A4 format are set to 210mm wide. Panoramic images are set to 297mm wide, so still fit inside an A4 page without need for readjustment.

Of course, pocket digital cameras are not able to produce files so large, so scaling is set to the capabilities of the camera. The currently used Kodak DC290s offer file sizes capable of printing high quality laser prints around 5x7" in size.

Dust spots are removed using the Photoshop cloning tool and/or using a product called *Scitex Final Touch*[7] that identifies suspect marks on an image that might be dust spots. Final Touch gives the user the option of cloning out all of these marks or just the ones they choose. Whilst the product is a good idea, it gives the user a false sense of security because it is all too tempting to press the 'global clean' button and let the software do its magic. Final Touch and those like it, such as Agfa's *Dusta La* Vista, remove specs of detail that the software assumes to be dust spots (up to around 2 pixels in diameter) and with it any detail that looks like but really isn't dust! So care must be taken not to wipe out precious detail. All-in-all such products save a lot of retouching time but really aren't something that should be used for cleaning images that are supposed to be accurate representations of photographs of collection objects because the risk of destroying detail is too great.

All scanned images require some sort of sharpening simply because of the nature of the process. The Image Centre does a minor sharpen with Photoshop's unsharp mask filter with settings determined according to the individual image. The name Unsharp Mask relates to the methods used some years ago by graphic arts labs and some photographers (myself included) for sharpening negatives using a mask that was made by a process of contact printing the original onto litho film.

To make a digital image appear sharper, the Photoshop USM tool increases the edge contrast of adjoining tonal variations in an image.

Most scanner software these days offers some sort of USM tool. As a rule, you should try to avoid using the scanner's unsharp masking (USM) because if you are unsatisfied with the result (sharpening is based on the type of image), you have to rescan. Using the

---

[7] Final Touch is a product that is sold only with a Scitex prepress scanner. Other scanner manufacturers make similar software sold only with their hardware, such as Agfa's *Dusta La Vista*.

Photoshop USM tool allows you the luxury of the undo / history commands to reverse anything you've done.

Again, the aim is to have archive images that are suitable for the Museum's general use and require no addition work. My own opinion is that images should see no sharpening at all before they are committed to archive. The act of sharpening using, for example, Photoshop's unsharp mask tool, permanently alters the state of the image. Once saved, it cannot be undone. That said, the fact that images destined for the Powerhouse's IMS are saved as JPEG files means that any sharpening performed on them after conversion to JPEG will only make more apparent the artefacts resultant from the compression routine.

Whilst working with images, versions are saved as either a TIFF or PSD file. Once work is complete, the files are saved as an A4-sized, 300dpi, RGB in JPEG format, using compression level 10 to 12 for Photoshop 5.5 and over (or level 8 for Photoshop versions prior to 5.5). This generates a file size of around 1.5 to 4Mb[8].

Where images may be required for future rework or possible publication (such as exhibition images that will be in hot demand), copies of the images are saved at either or both of A4 and A3 size at 300dpi, RGB. For these purposes, the files are kept as TIFF's and stored on CD in the Photo Library. When those images are requested for use, the images are copied to another CD and sent to the staff member requesting them.

Captured images of manuscripts or other objects with handwriting or printed text, as well as images of objects with fine detail, such as postage stamps should never be saved using a lossy file format such as JPEG. The very nature of the compression routine means that the captured fine detail in such items will be irreversibly destroyed and the detail will usually be too hard to read. Instead, either save the images as a TIFF file – though this means that an A4 RGB file at 300dpi will be around 24Mb in size – or, where colour accuracy is not an issue, save the images as a GIF. GIF files compress the colour palette and not the detail. A standard GIF file (these days, there a several save options) uses a palette of just 256 colours total, so a TIF file saved as GIF will much smaller than then original file size.

File naming

All scanned or digital camera images are processed in the above way and allocated file names in 8.3 format[9] that reflect the project under which they were created. For example, 35mm colour negatives scanned on our Nikon 35mm film scanner are allocated a different project number than medium format colour transparencies scanner on the high resolution Scitex scanner.

The format of our file names comprises a three character alphanumeric representing the project that created the image plus five trailing numbers as a sequential image number, plus the dot and extension, such as:

00A01234.jpg

---

[8] The ability of the JPEG compression routine to compress an image is based on the amount of detail in that image. A picture of a polar bear in a snowstorm will compress more, giving a smaller JPEG file, than a photo of close-up detail of a paisley shirt.
[9] Standard DOS file naming convention where 8.3 represents <filename in 8 letters>.<extension in three letters>

This would represent a JPEG image that was the 01235th image captured (we start counting from zero, then one and so on) under the "00A" project, where 00A might be for flatbed scans of 2-dimensional objects. Using this type of numbering system allows for 100,000 images per project.

The system used for the project component of the filename uses full-alphanumeric count, so numbered in this way:

> **000, 001, 002… 009 then 00A, 00B, 00C … 00Z then**
> **010, 011, 012… 019 then 01A…**

Each position therefore allows for 36 characters (0 to 9 plus A to Z), thereby giving us a possible 36^3 (46,656) projects. All up, this eight character numbering system for our file names allows us to allocate over 4.6 billion unique numbers to our images.

How much time to scan images?

The time needed to scan images is what I would consider the most important consideration when planning an imaging project.

How long will it take to scan images? Productivity is directly related to the availability of resources and their quality: the best throughput is achieved by good, full time operators using high-end equipment.

In the Powerhouse's case, much of the material already existed as colour transparencies and negatives stored in the Photo Library. As such, a good percentage of film had already seen a lot of manual handling.

It was conceived that with a high quality scanner connected to a fast computer, a competent operator should be able to get through around 50 images per day, including recording data for each. This is for film-based images, in the form of slides, single transparencies or film strips.

On the other hand, scans from new photography – film fresh from the processing lab and not having seen and manual handling, would take considerably less time. On the equipment installed to the Image Centre in the mid-to-late 90's, our operators were scanning around 100 images per day, including data entry.

Even though a piece of film looks good to the eye or even under a magnifying loupe, the fact that it has seen previous handling means that there is probably marks and residue that will require removing on screen after you scan it. Common are dust spots that have adhered with the help of humidity as well as marks, abrasions and finger prints and left-over residue from the oils used by a drum-scanner operator at a print house[10]. All up, some older images can require retouching on screen that requires considerable time.

---

[10] Drum scanners are the high-end graphic prepress units that have a transparent cylindrical *drum* that originals are taped to. Film originals are coated in a transparent oily fluid and encased in a sheet of plastic. The oils 'fills in' scratches and blemishes so they don't appear on the final scan. After scanning, the operator is supposed to thoroughly clean the film before returning them to their owner. In reality, the cleaning job is often not very thorough or, on the other hand, too thorough. Either way, original film material can be returned with an almost permanent residue or physical damage from attempts to clean them.

Normally, such dust cleaning is done using the *clone* tool in Photoshop. Matching areas are cloned or duplicated over the dust spots and marks. It's relatively easy but can become a challenge when cleaning spots on areas of fine details.

At the same time, colour balancing of an image – doing what we call "an appearance match" to the original, would be done by eye using one of several tools available in Photoshop for such a purpose.

For images destined to the Powerhouse's IIMS, the decision was made to do a maximum of six to eight minutes of work on each file with Photoshop. This includes all cropping and sizing as well as dust spotting and colour balancing. This would give us images that were reasonably clean and reasonably suitable for the appropriate uses of the IMS: screen viewing and digital photo and laser printing.

Of course, now that the Powerhouse's Image Centre is equipped for high resolution scanning services, meeting requests for scans suitable for prepress reproduction or large output means a lot more time needs to be spent on colour balancing and cleaning the files of dust spots. In my experience, times of over one hour of an operator's time are not uncommon when preparing digital files captured from film originals that are faded and / or damaged and required for prepress or high resolution output.

Choosing the hardware

In setting up an image capture program as part of an imaging project, a decision will need to be made on what to capture. Aside from the subject matter of the images, already discussed above, the hardware and methods used for capture will be determined what format of original is to be captured. For instance, if most of a collection is 35mm slides, then a standard flatbed scanner will be of little use.

Scanners

I'll limit the discussion to scanning for the purpose of an image archive and not for printing press output.

Scanners are devices that connect externally to a computer workstation for the purpose of capturing a digital image from a photographic or two-dimensional original, be it film or paper. Depending on the type of originals you have and the type of output you required – and your budget, there is a scanner model that will do the job.

Type	Suitability	Cost
Film – slide scanner	35 mm slides and film strips	$2,500 to $3,500
Film – medium format	35mm through to 6x9cm film	$4,000 to $25,000
Film – large format	35mm through to 4x5" film	$10,000 to $50,000
Flatbed - desktop	Reflective originals up to A4 or A3 in size (depending on model). Large format film originals can be scanned with the use of a 'transparency adaptor lid', though quality of scans is questionable.	$99 to $20,000
High Resolution Flatbed (A4 / A3 units)	Film originals from 35mm up to A4 / A3 in size. Reflective originals up to A4 / A3 in size. Ideal for printed publication.	$30,000 to $90,000
Drum Scanner -desktop	Flexible originals only. 35mm film originals up to A4 in size. Reflective originals that are in single sheet format up to A4. Ideal for printed publication.	$30,000 to $50,000
Drum Scanner – prepress units	Flexible originals only. 35mm film originals up to A3+ in size. Reflective originals that are in single sheet format up to A3+. Used mainly for printed publication.	$100,000 to $500,000+

As can be seen from the table above, there are scanners available from $99 to over $500,000. You don't get much for $99 and you'll need to be running a prepress film house night and day to justify the half-million-dollar-plus equipment.

So what do you look for in the way of scanners for your imaging project? If you have mainly 35mm film originals plus some flat art (photo prints, pages from books etc), then a 35mm film scanner and an A4 flatbed (or A3 if you need it) will be fine. Possible outlay will be about $3000 for the film scanner plus $1,000 – $2,000 for the A4 scanner. If all you have are medium and large format originals you may need to consider one of the larger film scanners. On the other hand, if none of the collection exists as photographs, a digital

camera might be the way to go. Whatever you choose, you'll need a computer to process and prepare the images for archive.

Don't be fooled by the advertising material that a sub-$25,000 flatbed scanner with a transparent media adapter (light source in the lid for scanning film) will be suitable for any serious scanning of film originals - particularly 35mm film originals. Most of these scanners have an optical resolution of 1200dpi. If you are scanning and outputting your files to a 300dpi colour printer, then your 35mm film frames will only be able to be enlarged 4x their size – a maximum of 140mm x 96mm (5.5" x 3.8"). How do we work that out? Divide your output resolution by your input resolution to get the enlargement (scaling) factor.

After deciding what type of scanner you require, you need to look at the brands and models available and their specifications.

First and foremost, buy a scanner that connects via SCSI – don't look at the parallel models – those that connect to a printer port. They are too slow. If you have more than one scanner and only one computer, SCSI connection will allow up to seven devices (sometimes more, depending on the SCSI card) to be connected simultaneously. Both scanners *should* work fine connected to the same machine (have your supplier demonstrate this to you) – though you can't use them both at the same time. The scanner not in use will need to be powered off for the other to work.

Buy a name-brand scanner that is well supported locally and comes with an on-site warranty. That is, don't end up with something that you need to lug across town or across the country (you wouldn't think about sending it by courier, would you?). Scanners, no matter their value, do not like being moved around too much. They contain lens assemblies and precision mechanisms that seem to get knocked around when moved[11].

Secondly, spend as much as you can on a scanner. This will be the device that creates your master digital images. If you buy a cheap printer and the results are ordinary, you can always reprint on a better printer. If you create a poor digital image with a cheap scanner, no amount of repair work in your image editing software, nor a decent printer, can improve the output quality.

Finally, find out how the cost of spare parts for the units – especially globes and light tubes. When we purchased our Scitex scanner, the salesman had a sudden bout of amnesia when I enquired about the replacement cost of the fluoro tube light sources. It turned out they were $140 each – and the Scitex takes three. In their defence, the tubes are specially made and have a clear window on one side, but really, that kind of money is ridiculous. On the other hand tubes for other high resolution scanners, such as the *Imacon FlexTight,* cost not much more than $30.

Manufacturers will quote a scanner's *resolution* (see resolution, later) as a figure in *dpi* or *dots-per-inch*. This is a measure of the amount of detail in the scanned image. Each of these *dots* is the finest point of detail that can be reproduced. Resolution is the most

---

[11] If you must move a scanner then make sure you 'lock the optics'. Most all scanners have facilities for locking down the moving parts – the lens assemblies and so forth. Locking them normally means turning a screw or a knob. Consult the manual. Some manufacturers will refuse to accept a scanner for repair if received with the optics unlocked. And whatever you do, make sure you unlock the optics before powering it up, otherwise, serious damage can occur!

important specification of the scanner and directly influences both the quality of the scanned image and the enlargement factor of your final output, such as a print.

Often the manufacturers will quote a scanner's the resolution as a range, such as 600dpi to 4800dpi. It is the <u>lowest</u> figure they quote that we are interested in. The lowest of the numbers is the *optical* or true resolution. The scanner software that drives the scanner from your computer will *interpolate* or fake the extra resolution. If a manufacturer quotes a figure like 600 x 1200dpi (normally for flatbed scanners), then it is the lower of the figures, again, that is the optical resolution. In these cases, they use the scanner's motors to step back and forth as it scans down the bed (surface), thereby doubling the amount of detail it acquires. It sounds better, but it isn't. For this type of scanner to capture at 1200dpi, the width of the scan capturing at only 600dpi will need to be interpolated using software to the 1200dpi.

So what do we look for in *optical* resolution? As much as possible! Take, for example, a desktop flatbed scanner. Most of even the cheapest scanners these days have a 600dpi optical resolution.

Without spending too much time going in to details – there are a lot of books and resources on the Internet if you're interested, the optical of your scanner will help you determine by how much you can enlarge and output (print or display the final image).

Desktop colour printers, such as laser printers (regardless of their quoted resolution), *require* an image scanned at 300dpi for a quality print. Many inkjet printers like the range from Epson yield great results from a scan at 240dpi. A monitor *requires* 72dpi (Mac) or 96dpi (PC). To determine the size of enlargement from an original before you scan it, simply divide the optical resolution of your scanner into the output device's *required* resolution.

The figures quoted here are industry standard. Scanning at higher values will blow out the file sizes until they are unmanageable. In any case, doubling input resolution with yield a file size four times bigger and sharpness might only be marginally improved. In fact, scanning much more than twice the recommended resolution for any output device can actually degrade image quality.

Therefore, you can use a 600dpi flatbed scanner to make a print of an original at twice its size on one of the standard desktop printers mentioned above (600 / 300 = 2).   Similarly, the same scanner is capable of enlarging an image up to around six times its size (600dpi / 96dpi = 6 approx).

A Word on SCSI

As I've mentioned, you should buy a scanner that connects via SCSI. Most SCSI scanners come with a small SCSI card. Don't discard this in favour of a super fast SCSI card, like an *Adaptec 2940 Ultra-Wide* or something. Although most scanners work best or *only* with an Adaptec card (some manufacturers will not offer support if their scanner is connected to any other brand!), high-speed SCSI cards can cause data overruns – errors and lockups of the scanner and / or computer.

If you do run two scanners from the one machine, you may also find that they won't like living together on the one card. Sometimes you can get by with making sure only the scanner in use is switched on (this is good practice anyway). Failing that, you might have

to resort to having two SCSI cards in your machine and run one scanner from each. Normally, each card will happily co-exist with the other, particularly if they are the same brand. Also, some scanners don't in fact like living with any other SCSI device (like Zip drives) on the same card – or prefer to be the first or even the last device in the chain. In any case, total cable length for scanners should not be more than one metre (not the usual 2 metres for SCSI chains!). At other times, you may have scanner lockups if the hard drive you are writing the scan to is a SCSI device and on the same chain as the scanner. Actually, it is good practice to have the scanner and hard drive live off separate SCSI cards anyway. Most of all make sure you buy a decent, shielded SCSI cable, and a good *active* terminator or you'll be chasing your tail looking for the cause of your problems!

Image Files

Colour Space

Colour can be created in many different ways. The colour images on a computer screen are made up by mixing different intensities of just three colours of light: Red, Green and Blue (RGB). For printing press colour, such as in a magazine, the colour is made up of mixing quantities of just four inks: Cyan, Magenta, Yellow and Black (CMYK). Desktop colour printers, too, use CMYK toners or inks for creating colour. Some desktop inkjet and wide format (floor standing, poster-style) printers use six or more ink colours for "photographic" results.

For best results when creating digital images with a scanner you need to know what output device will be used to print or display the image and capture the image at the right resolution and convert it to the colour space required by the device.

It becomes a little complicated here, because if you are creating images destined for a printing press you really need to contact the print shop to get colour specification of their printing methods (and these can change on a daily basis). You then set up that information in the colour preferences of the image editing software, such as Adobe Photoshop, which will code the correct information into your digital files so colours look right when printed at the print shop or prepress house.

Of course, if you are scanning images for archive and then *possible* later use in a magazine, then you really won't know what setting to use in your image editing program's CMYK colour preferences.

As you will see shortly, you should also scan a file to the required size and resolution for your output device as well. Again, this is not practical when creating an image archive. So for the purposes of an archive, choose a size and resolution that will suit most of your purposes – and save the images as RGB.

RGB and CMYK colour spaces are device dependant. That is, you need to tune each to a particular output method. However, it can be said that CMYK files are even more device dependant in that they need to be encoded for a particular printer or press house.

Some of the best reasons for archiving images as RGB for an Image Management System like the Powerhouse's are:

1. Most of uses for the images will require RGB images, so you will save time in not having to do conversions.
2. The file will be three-quarters the file size than if it was saved as CMYK (look at the formula for file sizes below and use a 4 in place of the 3), therefore storage of the image(s) will be cheaper.
3. All scanning equipment captures in RGB – right up to the half-million dollar units. They use *light* to capture the image of an original. This light is split up into its Red Green and Blue components. It is only after this that the scanner equipment and / or the software driving it does the conversion to CMYK where needed. So it makes sense to start with an RGB file and convert to CMYK when needed.
4. You will need to set the CMYK preferences to the settings required by the print shop (and these change daily), so converting *up* to CMYK from RGB will not

throw away any information, though there are certain colours in the RGB space that cannot be reproduced in the CMYK space[12].

When printing to desktop colour printers, the question of which colour space to use would often be posed. Most of these printers use CMYK toners or inks, so naturally most people think that converting a digital file to CMYK would be the right thing to do. Well, not necessarily.

Let's take an inkjet printer, for example. Forgetting the cheap units that use only CMY coloured inks, most good inkjets have at least CMYK colours, if not six (or more) colours – the common extra ones being *light* shades of cyan and magenta and, in the bigger printers, green. Most of these units (check the user documentation) expect an RGB file and convert to CMYK on the fly. Whilst they will work happily with a CMYK file (though colour reproduction will be affected), often they will convert it to RGB and then back to it's own flavour of CMYK before processing the image, meaning a lot of wasted time and possible some lost data along the way. Desktop colour laser printers are a similar story. Again, check user documentation but some do expect an RGB file and convert to their own CMYK space on the fly.

You may be thinking that sending RGB files to a printer and letting it convert to CMYK would not promote accurate colour: and you would be right. However, most desktop and network printers used for reference prints (as opposed to prepress proofing printers) do not have much in the way of colour calibration facilities. For this reason, it is best to experiment with colour output and your settings in the scanning equipment and image editing application you are using and find your best colour preference values for your needs. Surprisingly, good results can be had for those willing to work at colour management for this kind of equipment.

Colour Accuracy
The whole idea of a collection imaging system is to save accurate representations of collection objects. But what is an accurate representation? The Powerhouse's IMS saves images that have been appearance matched – with respect to colour, using the naked eye and a relatively good quality monitor.

To go into a discussion of colour would consume a lot of paper – and there are some very good books on the subject.

Let me just point out that in referring to colour accuracy, I am talking about human perception of what is a good or bad representation of something in a reproduction, based on memory and experiences as well as the human ability to make comparative judgements, the conditions under which they make these comparisons and their emotional state.

For an imaging workflow, accurate colour matching is only achievable on a system that has been tuned using rather expensive colourimeter equipment. The subject of colour management and colourimetry is long and technical. For our purposes we can say that accurate colour matching in an imaging workflow – that is, getting printed or displayed output to match the original is near impossible without the right equipment and the know-how.

---

[12] The more base colours used in a colour space, the more possible colour combinations that are achievable. Converting a file back and forth between colour spaces, especially going from a 'bigger' colour (like CMYK) space to a smaller one (like RGB), will result in a loss of colour information.

Imagine you scan a transparency and display it on a monitor. Does the image file you are see look like it's a match to the original transparency? If so, then does this mean the scanned image is an accurate representation of the original? If you were to say yes, then you might very well be wrong.

Suppose the scanner used to capture the image was creating captured images a little too green and then the monitor you've used to display the image was operating a little too magenta (opposite of green). By sheer luck, the two have worked together to cancel out each other's inaccuracy to give an image that appears to be right. In fact, the image file itself is not a match to the original transparency because the scanning equipment has not been properly calibrated. Viewed on calibrated equipment, the file would indeed look too green – as scanned.

The point here is that just because an image looks good when output (to monitor or paper), doesn't mean that your system is tuned. Often, it's sheer luck.

Simply investing in good quality scanners, monitors and graphics cards doesn't mean you will get a good end result. Each piece of equipment in the chain of events has its own idea about how all colours should look.

Colourimetry is the science of measuring and predicting colour based on the differences between human colour perception and equipment used to capture and output colour.

For a properly colour managed workflow from scanner through image editing workstation to printed output, you need to measure and record each device's response to a standard colour target. That is, a piece of photographic material that depicts a series of over 200 colour patches of known value. Standard colour targets are made by AGFA or Kodak. The Kodak IT8 targets are the most commonly available and come in two forms: as a transparency (for scanning transparent media) or a print (for reflective work). They seem to be hard to buy in this country at times probably because demand isn't all that high and then because the have a definite shelf life: they're photographic images and they fade.

The idea is that you place the IT8 target in your scanner and scan it. Software is then used to read the values of the known colour patches and it tables the differences between what the scanner *sees* and how the patches should reproduce. The resultant *profile* is embedded into each image file created by the scanner. From here on, each time the image is opened by ICC compliant software, the embedded profile is read and the image colours adjusted to compensate for the colour shift resultant from each device's shortcomings.

Of course, the application software – such as Photoshop, will need to have its colour management options set up correctly and output profiles for each device you intend to use made available to it. In addition, your operating system will need to support colour management. Macintosh and Windows both support colour management, though Windows 95 and Windows NT4 do not.

From here on, if you intend to output the file to monitors, for web use for example, you can then embed a profile suitable for monitors (such as sRGB) in the file so that the colours represented reproduce best on most people's monitors. By embedding a new profile, the old is removed and the differences between it and the new profile – sRGB in this case, are applied.

Of course, if you are going to be doing any serious image editing, you will not only need a decent quality monitor, such as Trinitron-based Sony's or a LaCIE, but you will also need to colour profile it. Phototoshop offers reasonable colour adjustments for those without the budget for professional calibration equipment. However, in a fully colour managed imaging environment you will want to invest in a good colourimeter and take readings from each monitor in your workflow. From these readings, your colourimeter software will build a monitor profile that can be used by Photoshop and your operating system to display images they was they should look.

Oh… and the cost of a *decent* colourimeter? Around $18,000.

So in here I have a point to make. Most institutions setting up an Imaging System need to make a big decision. Do they invest a big slice of their hardware money in colour measurement tools or do they do what many other institutions do and that is to use the Photoshop colour settings adjustments and do their own workflow tweaking – manually? Having been there myself, I can tell you that it's hard to convince management into believing that a hand-held device costing twice as much as an imaging workstation is going to make such a big difference.

A lot of it depends on your situation. If you are an Art Gallery, you really should tie down your system with proper colour management equipment and procedures – and spend the money to get a good colourimeter. However, if you are a car museum, then perhaps getting exact colour reproduction might not be of paramount importance.

Buying the equipment is only part of the story. The ability to utilise such precision instruments takes a lot of practice and a solid understanding of colour theory. On the other hand, there are companies out there who will come to you, for a fee of course, and do all the calibrations and profile building for you. This would be a very sensible road to take, though calibration should really be done on a weekly basis, so you need to factor this in. However, you should ask around for recommendations of who is best as I'm not sure that all of these guys, with their new toys, really understand what they are doing – or rather, bother to understand what it is that you, the customer is trying to achieve.

It is important to note that regardless of how well calibrated your workflow, if you don't have well photographed images to work with, then you are wasting your time and money trying to reproduce your object collection faithfully.

Printers

In choosing a desktop colour printer it is important to realise that most brands and models, do not offer archival permanence. The inks, toners and the papers they use allow for images that last anywhere from a few months (!) to ten or so years. Developments of archival inks by some manufacturers, notably Epson, are seeing a claimed permanence of hundreds of years[13] when using the right paper.

Desktop inkjet printers are the most notorious for a short life span. However, desktop inkjet printers, while offering very good through to outstanding photographic results, normally give you a print that noticeably fades after just a few months when displayed.

Normally the inks used in these printers are water-based, so prints are easily damaged when exposed to water splashes or humidity. Some of these printers, though, do have water fast inks. The Canon printers in fact use inks that allow you to submerge the print in water without any problems. However, the prints still fade in a relatively short amount of time. Inkjet prints can be laminated to protect them from water, humidity and even UV but this only adds to the already high cost of printing them and doesn't do much to increase their life with respect to fading – as often fading is a combination the effects of light (UV) and chemicals (paper and inks interaction with each other and the atmosphere).

The biggest problem I see with inkjet printers, especially in an organisation where a lot of prints are likely to be made, is the cost of running them. While the A4 photo quality units are as little as $600 to buy, when used to print full-colour photos on A4 paper, the prints cost as much as $4.00 or more each. That said, small organisations may have no other choice but to buy such printers. Certainly the photographic results from these units is exceptional and in any case, the fact that images are being made from 'permanent' digital files means that an exact duplicate of a print that has faded can be made again.

These 'photo quality inkjet printers' as the manufacturers call them, offer continuous tone colour. In other words, they have no discernible 'dots' and so look like a regular photograph. Actually, there a spots of ink there, it's just that the new units print with such a fine dot that the output looks continuous. Really, they look fabulous. However, to create the apparent continuous tone, the printer lays down a lot of ink. The ink itself costs a lot of money. For some Epson models, for example, cartridge sets (a full colour cartridge and a black cartridge) cost about $75.00 - good enough for about 25 full A4 prints. Add that to the cost of custom gloss paper, at around $1 per sheet and you are up for a lot more than the 'so many cents per page at 5% coverage' that the advertising blurb will lead you to believe.

Now you might say that $4 for an A4 'photo' is cheap compared to a pro photo lab, where you can spend around $30 for the same print on conventional paper. However, given an inkjet print will start to show noticeable fading after a few months to a year when on display, as opposed to a regular photo that can be displayed for around 18 years before fading, you start to realise just how expensive the inkjets are!

Don't be fooled by the manufacturer's claims of "7 cents per page at 5% ink coverage". Those figures are more accurate if you're printing a one colour, plain text, word processor document. However they do not apply to photo prints where there are no regular blocks of the page left uncovered by ink.

---

[13] Refer to Wilhelm Research (www.wilhelm-research.com). Note that most desktop inkjet printers and most all laser printers do not allow for use of archival inks or toners.

For printing full colour images on an inkjet printer, each of the semi-transparent colour inks will cover the page at about 70% - that's 70% per colour for each of the six or more colours in the cartridge. The black ink coverage runs at about 20% in the same photo. Do the sums using these figures based on 7 cents @ 5% coverage per ink and you soon realise the cost of a photo print runs well into the 'dollars'.

One other thing to consider about desktop inkjet printers is speed. They can take between 5 and 10 minutes per A4 full colour photo. Considering most of these printers cannot be placed on a network with their own print queue, the computer connected to the printer will be tied up for the duration of the print job.

Large-format inkjet printers

These are the floor standing, large-format printers capable of printing posters that are a metre or more wide and as long as the roll of paper you feed into them.

The better known models such as Hewlett Packard and Roland use a standard water-based ink requiring you to either laminate, frame or otherwise coat the print for protection from water.

Most of these units require connection to a dedicated computer with specialised hardware inside called a Raster Image Processor (RIP). These process the image data and convert to poster format very quickly. The cost of the card alone can be thousands or more likely, tens of thousands of dollars. Software RIPs are available these days and do a reasonably good job. These too are quite costly.

As with the desktop models, the same caveats apply to the large format inkjets. The print sizes and quality, when viewed at an appropriate distance[14] are very nice indeed. The cost of printing the images, the computer hardware required, as well as the time taken to print a poster (sometimes measured in hours) makes this form of photo printing suitable only for those who can justify the costs of hardware, materials and time.

Depending on the model printer and the inks used in them, some prints from these wide format printers are archivally permanent and are said to last for over a hundred or more years.

Laser printers

If carefully chosen for use as a photo quality printer, certain models of desktop and floor standing colour laser printers can yield very good results at reasonable speed and cost.

As with other medium or large organisations, the Powerhouse's IT department prefers HP printers for general workgroup (non-photo) printing and have had good reliability from them.

For photo printing, though, my personal preference is for the Tektronix range of colour laser printers. The results on plain paper are astounding. As such, I've purchased many of Tektronix printers for the Image Centre and for other departments directly involved in capturing images for the IMS.

---

[14] In photography terms, a photo should be viewed from no less than two to three times the length of the diagonal measurement of the paper it is printed upon. For example, an A4 print should be viewed at arm's length.

Regardless of the Powerhouse's or my own opinions, remember, your own needs are different than ours. If you were looking for a good photo printer, I'd recommend you look at everything available in your price range. Study colour output quality on the various papers that each printer would accept[15].

When selecting a colour laser printer, you should be concerned with the cost of the consumables as well as the warranty and on-site support – particularly the speed of response to service calls. All laser printers break down when you need them the most, so the quality and timeliness of service is worth its weight in gold.

A photo quality laser printer, like an inkjet, does take considerable time to print an image. It is not uncommon to have an A4 image @ 300dpi take five or more minutes on a photo quality laser printer. Sure that's more than a single page from your word processor but photos contain a lot more complex data and hence the file sizes are huge in comparison. However, unlike most inkjets, photo quality colour laser printers can be networked, so print queues can be set up and multiple users can keep on sending files. Even if the printer does not have a network card option, you can often connect it to your network via a print server – such as a simple and inexpensive Hewlett Packard *JetDirect*.

The one thing to realise about photo quality laser printers is that the initial investment is high – around $4000 for a good A4 unit and double that and more for an A3 unit. So too is the cost of consumables, with toners running at about $500 per colour, about $300 for black, and the other 'consumable parts' such as imaging units at up to $1200 each, and fuser units at several hundred dollars. However, most all the consumables go a very long way, with the most expensive parts lasting a year of more even in a printer that sees a lot of use. In all, the cost of printing an A4 image on a photo quality laser printer is generally around $1 to $2. Expect the prints from most photo quality laser printers to last between seven and twelve years[16]. Therefore, for reference prints that need to be filed away, photo quality laser prints are a better alternative to the more expensive and quickly fading desktop inkjet prints.

Dye Sublimation Printers

These printers use a special ribbon, the width of the paper being printed to, for the *transfer* of images to the paper. The term *sublimation* refers to the process of turning a solid to a vapour without actually going through a liquid stage. The printer uses the special cellophane-like ribbon to carry and transfer the images to the carrier paper – either a gloss or matte surface.   The print process takes around 3 minutes per A4 image and a bit longer for A3. Considering it uses the same amount of ribbon regardless of whether you are printing a full colour image or a page of text or even if you accidentally print a blank white page, output costs are the same each time you make a print. Dye sublimation printers cost about $4 to $6 per A4 page and double that for A3. The consumable parts are generally only paper and ribbon. That is, unlike laser printers, there are no expensive imaging units and the like that need changing once in every while.

---

[15] Most colour laser printers have a manual feed door to allow single sheets of paper, such as letterheads and thick paper to be fed into the machine. Be very wary about the thickness or 'weight' of paper you feed into your expensive, photo quality printer. You'd be surprised at how easy it is to jam a printer with paper not much thicker than photocopy bond. And jams of this kind inside these units are not covered by warranty. Follow the manufacturer's recommendations closely.

[16] Most well known printers have test results published by Wilhelm Research (www.wilhelm-research.com).

You can buy ribbons especially for printing A4 images, so the cost per print is lower than the above, however, if you intend to print A3 at all, then do not buy the A4 rolls. The reasons for this are simple: the ribbon is extremely fragile and they do not like dust. Once rolled onto a carrier assembly, the roll stays there until it is fully transferred to a roller on the other side of the assembly and fully used up. This means that if you remove the ribbon and assembly mid-roll, you need somewhere clean, dark and dry to store it. In short, don't waste your time.

The print results from dye-sublimation printers are exceptional. Although they have a slight waxy look to them, they easily pass for a regular photo print. However, note that the prints from *some* models fade - rather quickly. The newer models, suck as those by Kodak yield prints that are supposed to last many years – but this is dependent on the models using UV coated materials. Similar to conventionally made photographic prints and even a lot of alternate digital prints, stored in the dark, dye-sublimation prints suffer also *dark fading*[17]. So storing them away from light, although prolonging their life, will not stop fading.

Depending on the brand, features and output size of a dye-sublimation printer, don't expect much change from $12,000 for an A4 / A3 unit.

The Fuji Pictrography
These printers are worthy of a mention, even if the initial outlay for hardware is $30,000 for an A4 unit and $40,000 for the A3-capable unit.

This floor standing continuous-tone printer resembles more a washing machine than a digital photo printer and is, surprisingly, an RGB device. Another interesting point is that it can only be connected to a SCSI port on a PC or Mac. Network options are not. The only way to connect to these printers from another computer is via third-party software and *printer sharing* – meaning the computer it is connected to must be left switched on.

The Pictrography (yes, that's how it's spelled) uses colour lasers (red, green and blue) to 'cut' an image to a paper negative or *donor*. The Donor material is then wet down and placed in contact with the gloss or matte receiver. A dye-transfer process ensues. After a short time, the two are peeled apart with the receiver, now a final print, dried and expelled from the printer (with the used donor sent to a waste bin) all in about 2 minutes and all automatically.

These printers are arguably the best digital photo printers I've seen. I purchased one for the Powerhouse's Image Resource Centre a couple of years ago and the cost savings over conventional colour prints are incredible, not to mention the time savings. The Pictrography, though, would be worthless without the high quality scans afforded by our high resolution scanner. The image quality would be ordinary if scans were produced on a low-end desktop scanner.

The most impressive side of these printers is that they can both pass as a normal commercial photograph and last as long. The running costs are $7 (A4) and $14 (A3) per print. Compare that with a print from a lab that costs $30 and $60 respectively and takes three days turnaround.

---

[17] This is the fading caused by the instability of dyes over time and not as a result of exposure to light or UV. (Wilhelm et al, 1993; http://www.wilhelm-research.com).

The Pictrography is very easy to run and maintenance free, requiring only a refill of it's small removable water reservoir bottle once per day and a water filter change ($16) each two months.

The printer is controlled via a "TWAIN plug in" - from within Photoshop, for example. Colour control is via a set of colour sliders allowing brightness, contrast, and RGB. Having been a photographer much of my life, I likened it to printing colour photos in the darkroom – without the chemical fumes.

The standard Fuji driver for this unit is good enough for superb photographic prints of good colour clarity. However, a third-party product called *Raster Plus* allows for colour management and network-ability not afforded by the Fuji driver. In addition, this product allows you to queue images meaning you don't have to wait around 30 seconds after sending one image to the printer before sending the next.

Which printer to buy for an imaging project?

If you have a choice, give desktop inkjets a big miss purely for cost and speed reasons and the fact that the prints fade very quickly. Although the hardware costs are less than $1000, prints that last less than a year and cost $4 each is false economy for an organisation needing to do a lot of photo printing. However, many organisations simply don't have a choice. Budgets are such that they cannot afford a capital outlay into the thousands of dollars to print A4 images and in any case, a $600 inkjet printer gives, arguably, better results than anything other printing technology under $5000.

Whilst photo quality laser printers have an initial high outlay cost, the long term running costs are much lower than most other forms of printing that are otherwise suitable for an imaging project and the quality of prints are superb depending on the model.

If you have a lot of use for extremely high quality photo prints and or archival quality prints, then a dye-sublimation printer, such as the Kodak 8660 or the Fuji Pictrography are worth a look, though for good results you'll need high quality scans to feed them.

Resolution

There are several uses of the term *resolution* that ultimately lead to confusion.   Quite literally, the term refers to an ability to resolve or reproduce fine detail. For digital images captured by a scanner or a digital camera, resolution is a measurement of the number of pixels (short for picture elements) in the vertical and horizontal directions. Each pixel is the finest point of information in the captured image. After you capture the image, the digital file has a line of code within to say that "this file is set to output at 300dpi". When you send that image to a printer, for example, the file is read and output with 300 points of information per linear inch, both vertically and horizontally. In this case, 300dpi (dots-per-inch).

When you print an image, the term resolution relates to the number of spots of printed information to the linear inch and directly affects the picture quality of the final image.

The output size, the resolution and the colour space of the image determines the file size of a digital file.

If you scan an image to a size of 10 x 8 inches @ 300dpi, in RGB, your file size will be as follows:

File-size in bytes =   (height x res) x (width x res) x (colour depth)

 - where colour depth is 3 for RGB and 4 for CMYK (printing press)

In this case…

File-size   = (10 x 300) x (8 x 300) x 3
                 = 21,600,000 bytes
                 = 20.6 Mb   (remember, 1Mb = 1,048,576 bytes)

If we scan the same image at half the resolution – 150dpi, we get a file size of 5.25Mb or a quarter of the size.

As a side note, an image scanned at A4 @ 300dpi in RGB will become around 24Mb. The same file scanned to A3 will become 96Mb. Doubling the output size with quadruple the file size because we will now have an image with double the quantity of dots horizontally as well as double the dots vertically.

For best results, it is important to know the required output size of a scan and its resolution before we start a job. As a rule of thumb, it is not a good idea to re-size a scan more than 10 to 20 percent of its size. Let's look at why.

If we have an A4 scan at 300dpi in RGB and need to print the image at A5 (half the size of an A4), we would need to reduce its size using an imaging application like *Adobe Photoshop*. If we were to simply squeeze the size down to A5, the resolution itself would double. In other words, squeezing the size of an image down would simply bunch up the dots in the image closer together. So in this case, the image would become A5 @ 600dpi.

Now we have too much resolution[18] for our printer that requires only 300dpi. So now we have to tell the imaging software to *resample* the image to 300dpi. Great, now our image is at the required A5 @ 300dpi. But… what did the software do to the image to change it from 600dpi to 300dpi? It <u>threw away</u> every second dot in the image in each of the vertical and horizontal directions. That is, three-quarters of the information in the image has been discarded (remember, it's been resized from a 24Mb file to a 6Mb file).

Imagine now that this scaled-down image is a detailed image of hair. Originally, the tones are fine, graduated ones. The detail is fine and in some cases, one row of dots might represent a single hair. After the image has been scaled down and every second row of data has been thrown away, some of this fine detail will be discarded. Now what we have is an image where the fine detail becomes coarse and the tonal gradations become stepped. In actual fact, a good image processing application like Adobe's Photoshop will do some serious calculations when resampling images and a lot of the fine detail will not be lost but in any case, some of it will still be interpolated meaning that a lot of the image will have faked-in dots of information.

This is the sole reason why it is important to know what size and resolution is required to output an image and scan to those specifications.

By the same token, if the original A4 @ 300dpi image was enlarged to A3, the image editing software doing the resizing would first need to stretch the image to double the size, resulting in a file that is A3 @ 150dpi (the *distance* between the dots having been doubled). To make the image 300dpi again it would have to insert a dot between every other dot. That is, the software interpolates the image – it inserts dots that have been coloured using an average of the values of surrounding dots. The result will be an A3 @ 300dpi image but three-quarters of the data has been faked in (the file has gone from 24Mb to 96Mb). To the eye, the image will look unsharp and the colour inaccurate.

What about high resolution?
Throughout this text I've used the term *high resolution*. This phrase can indeed refer to a file that is encoded to thousands of DPI. However, digital image files output *thousands of dpi* are of little use to anyone for the reasons mentioned previously.

In fact, generally high resolution refers to a large file size suitable for large size output. With most small scanners capable of 600 to 1200dpi input resolution, the phrase high resolution is often used to refer to scans from equipment rated at higher optical capabilities. The phrase is also used in its literal sense to refer to a scan of exceptional detail and sharpness.

High resolution scanners often have optical capabilities in the magnitude of 3000dpi plus. Therefore a scan of a 4x5" transparency scaled at its same size (100%) at 3000dpi will yield a file size of about 527Mb. Now, we've already said that most image files need to be about 300dpi for output on most digital photo printing equipment, so theoretically, this file is way too big. However, we have also seen that the scanner merely encodes the scaling and resolution settings into the file and it is really the number of dots of information – related directly to the file's size that determines how big the image file can be output on a given device.

---

[18] Despite popular belief, doubling the resolution value will not give you a better, sharper image. In many cases your image quality will appear less sharp when printed in certain equipment!

So, we've seen earlier that each time we double the output size of an image, we halve its resolution. So for our 3000dpi file, this equates to:

$$3000dpi \ @ \ 4x5" \ = \ 300dpi \ @ \ 40" \ x \ 50"$$

If you consider the file size calculations demonstrated earlier, you would also see that the same 300dpi file at 3000dpi resized as above will still store in memory or on disk at 527Mb.

Most importantly, this means that this enormous file will print perfectly on a 300dpi printer without interpolation. Simply, when you resize an image this way – reducing output resolution while increasing output size proportionally, the file size remains the same. And if the file size remains the same it means that no resampling has occurred – ie no adding or removing information from the file.

File formats

When we scan a picture, or capture one with a digital camera, the image file is initially a *raw* file. It is a binary file made up of noughts and ones that describe the location of the red, green and blue pixels in the image.

When the file is saved, we choose an image file format so that it can be written to disk and read again later. There are some 50 or more different image file formats including native formats used by scanner makers and image editing software for accessing custom features while their software is in use.

Each format has a particular use or feature that is suitable for a range of applications. However, in choosing a file format for storing and transmitting images it is important to choose one that is universal - one that can be opened later on (perhaps in years to come) or by the recipient of the images; and one that suits the purposes of your work.

Some file types offer a compression routine that allows for smaller file sizes when stored on disk. Some compression routines are lossy, meaning they sacrifice file integrity for file size. Others are lossless allowing you to open and resave images without any degradation. Both types of compression have their merits and generally the lossy types offer a greater rate of compression.

The whole idea of using a compression format, such as JPEG is to allow for better storage and quicker transmission times over intranets and the Internet. If you do have a few JPEG compressed image files, prepare them for sending via email, for example, by placing them in one directory and then compressing that entire directory using, for example, the Zip compression utility. Your overall file size will actually get bigger.

On the other hand, if you do have TIFF image files you cannot afford to compress with JPEG, one option is to *Zip* these files. The resultant file size will only be a little bit smaller than the original simply because the famous Zip compression routine is not designed for image files – but it does compress them a little. Of such small file size savings can be the difference between a file attachment via email making it to its destination and not getting there at all.[19]

The most commonly used image formats at the Powerhouse for images in the Powerhouse's Image Centre are TIFF, JPEG, GIF, and PSD. The first three are used for saving images to the archive. The last one is an Adobe proprietary format used for saving intermediate files relating to an image being processed.

It is important to note here that PC operating systems require filename to have a three letter extension so it can identify the file type and open the file using the application assigned to working with such files. Mac systems do not need the extension though they will still open files with the three letter extension.

With that in mind, you should be aware that Macintosh systems will not as a default add the extension to a filename as it is created or resaved. Such files then cannot be opened by simply double clicking on them in a PC environment. In fact, when you select *file → open*

---

[19] Large size emails sent to Internet addresses will often *bounce* because some system that the email has to pass through on its way to its final destination has a limit on the file size of emails that it will handle. This limit can be as little as 1Mb. These days, though, the norm is somewhere between 2Mb and 5Mb.

in Photoshop, the dialogue box will only show, as a default, known image file types[20]. As such, when creating and resaving all file types in a Mac environment, be sure to add the appropriate file extension.

TIFF files

Tagged Image File Format (TIFF) files use the extension[21] " .tif ". TIFF is a free to use, non-proprietary format. The files are high resolution, uncompressed bitmaps[22] that are the most commonly used by all major image editing applications and able to be read by most operating systems. TIFF files are useful when an uncompressed, faithful copy of an image is required. However, the file sizes are big: as already stated, an A4, RGB image at 300dpi when saved as a TIFF file will occupy over 24Mb of hard drive space. Unlike JPEGs, however, a TIFF image can be opened, altered and resaved without degradation of the file due to the action of re-saving (compression).

TIFF files with LZW compression

TIFF files can actually be compressed using the LZW routine (available under Photoshop), resulting in non-lossy (quality is not lost in favour of file size) compression. Like uncompressed TIFFs, these files can also be opened, altered and resaved – even resaved with the LZW compression turned off, without any degradation due to the saving (compression). However, some image editing and image viewing applications cannot open LZW compressed TIFF images – depending on the application, an error might be reported or the image displays as garbage. For this reason it is not a safe bet to encode a TIFF file with LZW compression because there is less chance that you will be able to open it in years to come.

Most importantly, there is no easy way of telling whether a TIFF image has been saved with LZW compression or not by looking at the file in inside a directory or once opened on screen. The (DOS) extension will still be saved as ".tif". If you have TIFF files that you suspect are saved with LZW compression, there is a trick to determining is they were in fact saved with LZW.

Open the file in Photoshop and then select *save as* from the *file* menu. If, during its last save, the LZW option was selected, Photoshop will show the checkbox for LZW as being *ticked* at the bottom of the *save as* dialogue box. Note also that Photoshop will report the file size for the image as a larger size than that reported by Windows or the Mac operating systems. Of course, both these methods are manual and time consuming.

JPEG files

This is a *lossy* file format - data is altered and lost in favour of making a smaller file. JPEG is actually a compression format and not an image file format and, like TIFF, is a non-proprietary format meaning is it free to use. It is one of the most widely used formats of all and offers large compression ratios, but at a price. Depending on the amount of compression used to compress a file for storage, JPEG will have a direct influence on the quality of the image when viewed.

---

[20] Under Photoshop on a PC system, if you know the file type of an image file that is missing its extension you can use the *open as* option to open such files as their filetype.
[21] Some computer platforms, especially Windows, rely on a three-letter 'extension' after a 'dot' to indication to the operating system what type of file is being accessed. This allows the operating system to determine what to do with the file. Common extensions are .doc for Microsoft Word documents, .tif for TIFF images, .jpg for JPEG images and so on.
[22] Bitmapped files are those with data described as a pixel map: a grid style map that uses coordinates to indicate what coloured pixels go where.

An average image scanned at 300dpi @ A4 (210x297mm), using the RGB colour space, will occupy around 24Mb of memory on a computer and the same space on a disk when saved. That is, regardless of any compression for storage, once the image is opened on screen, it is loaded into memory. To store such files uncompressed means that only around 26 images will fit on a CD (a standard CD will hold 650Mb of data). Compressed using JPEG compression level 8/10 under Photoshop 5.0 or level 10/12 using Photoshop 5.5 and later, the file will compress to around 1.5 to 4Mb on disk. At an average of, say, 3Mb each, this would allow around 216 images to be stored.

Although the format is a lossy one, the quality of a level 10/12 (Photoshop 5.5) image is almost as good as the original and depending on the amount of fine detail in the original, may be quite undetectable.   The same image compressed using level 3/12 compression may result in an extremely small file but very poor in image quality, showing lots of artefacts or odd pixels throughout the image. Most often, artefacts appear around transition areas of detail.

At the same time, care should be taken when applying JPEG compression to images containing fine detail because this detail can be irretrievably damaged once the file is saved as a JPEG. In particular, you shouldn't use JPEG as the format when saving images of text, such as documents. The JPEG compression will normally degrade the text to a point where it is almost unreadable. For text images, you should save the files as either TIFF or GIF. The former yielding much larger files, the latter not offering colour accuracy.

For most image archives, particularly when lots of images are involved, meaning storage costs are high, JPEG is the best solution. One word of caution though. Don't reopen JPEGs with the intention of resizing, retouching or altering. If a JPEG file is resaved with changes, the entire file is re-compressed, resulting in a severe degradation of the file. In short, do all sizing, retouching and manipulation and only then save as JPEG format when done. If you must open a JPEG image for rework, then save this new version as an uncompressed image such as a TIFF file.

GIF files

Graphics Interchange Format (".gif "). This is a proprietary 8 bit format originally used for images for transmission over Compuserve, a well-known US Internet provider. Images converted to GIF are down-sampled from 24 bit colour (2^{24} = 16.7 million colours) to 8 bit colour (2^8 = 256 colours). The palette used for the images generally gives a rather good interpretation of the original image with a file size reduced down to one third the original. While GIF files do not exhibit artefacts like JPEGs, tones that are graduated, like skies etc, will look banded or *posterised*. GIF files are more forgiving of rework and resaving than are JPEGs.

One thing about GIF files. Compuserve owns the license to GIFs. In the past few years, Compuserve granted that use of the GIF format would be free except if you are intending to *create* GIF files, in which case you must pay a license fee. Purchasers of Adobe Photoshop are licensed automatically to create GIF files upon buying the software. Developers of imaging software must pay a license fee for using routines for creating GIF images in their applications. When the Powerhouse developed its Image Management System software, a license fee was paid for the GIF routines.

Photoshop PSD files

Photoshop (PSD) format is Adobe's proprietary image format. This is what Adobe has to say about it: "it is the only format supporting all available image modes in Photoshop (Bitmap, Grayscale, Duotone, Indexed Color, RGB, CMYK, Lab, and Multichannel), guides, alpha channels, spot channels, and layers (including adjustment layers, type layers, and layer effects)".

Simply, if you have a lot of image manipulation to do, particularly image processing such as correcting for noise resultant form using a low-end digital camera or scanner, then like any good operator, you should regularly save your work. At times you will find that Photoshop will only allow you to save images as a PSD file because, during the course of your image processing work, you will have converted your image to, say, Lab mode.

While PSD format allows for more flexibility during the use of Photoshop, it is still a proprietary format and, as such, is not supported by many image viewers or editing applications. Either way, it is not the format to be used for archiving images simply because it has less of a chance of being supported in years to come.

Media

CD Storage

At the moment, the only media to consider for image storage are CDs. In that I mean that it is the most universal, high capacity, removable storage media available that is not affected by issues of propriety. The disks themselves have a relatively long life and, physical damage aside have little risk of data loss[23]. However, if we are to store disks for a considerable amount of time, then disks that are made for archiving should be used. The Kodak Ultima Gold CDs, for example, are claimed by Kodak to 'last' for at least 160 years. Let's face it, how many of us are going to be around to sue them if the disks don't make it?

CD disks can be written to using a CD recorder, a device that looks for all intents and purposes like a regular CD drive, internal or external. Using specialised software, you can copy any of the files you can write to a hard drive.

The capacity of normal CDs is 650Mb – about 26 A4 images @ 300dpi, RGB files, saved as TIFF - or up to 1000 or more of the same images when saved as JPEGs with high compression (low quality – suitable for web use).

There are CDs with higher capacity – around 700Mb, but you must have a CR writer capable of writing to the disk and you need a drive capable of reading it. Not worth the trouble!

CDs can be bought for as little as 40 cents each these days. If you are using the disks for temporary storage of non-important files, then these are probably fine. For anything else, look for name-brand disks such as Kodak or Verbatum. In particular, invest in disks with a good quality dye (the data layer) such as a said-to-be-stable AZO-Metal.

You can buy CD blanks with a special non-printed back surface that is specifically made for printing onto using a purpose-build inkjet CD printer (such as the Primera Signature III, imported here by Verbatum). If you do buy these disks, again, buy a good brand. The fact that the back of the CD has a very opaque coating on it means that the manufacturers can cheat – by using a very thin dye in the CD. This means poor quality disks.

The dyes used in CDs should be opaque when held to the light. If you can see the filament of a light globe through it, chances are, it's a cheap disk. I have seen reports from people who have used such disks who have said that after a few months there disks were not readable.

When saving or *burning* files to CD, there are several file format options to choose from. Each have their merits but most will make those disks of files readable by fewer operating systems. The Powerhouse creates all CDs under the ISO-9660 standard which means that most all operating systems will be able to read their contents. In addition, the files are saved as Joiliet which supports long file names.

---

[23] The data written to CDs is done using a laser to 'cut' holes in the metal of the disk. In time, reactions in the metal and the substrate material cause these holes to change and migrate, causing loss of data and unreadable disks.

DVD

DVD disks are the latest incarnation of CD disks and hold many gigabytes of data. However, although the technology has been in use for a few years now, it is still young and the mechanisms required for writing data to them are still relatively expensive.

At the same time, there are three major formats of DVD, each with their own benefits and each with varying or no compatibility with the other. For instance, DVD-RAM is a read-write format and stores a massive 4.7Gb of data per side but must be used in a caddy (the disk resides in a cartridge) and therefore cannot be used in anything but a DVD unit that accepts a caddy. DVD-RW and DVD+RW are both single-sided, read-write technologies and each stores 4.7Gb of data but groups of manufacturers are still squabbling over who has the best format and whose DVD readers will read disks created by somebody DVD writers.

The present choices in DVD make them a risk for long term archival storage, however it is being tipped by many that DVD will replace everything from CDs through to videotape - and in the not too distant future.

The one comforting point, of course, is that whichever form of DVD survives, the manufacturer's will no doubt make the units backwards compatible with current CD-R – CD-RW technology simply because of the enormous popularity of CDs and the 20 years they've been around. How long such backward compatibility lasts, though, is anyone's guess.

Planning for the Future

In designing methods to capture images and store image files and data it is of paramount importance to consider the future. There is not a lot of point spending a great deal of money on an imaging project if in a few years time the mechanisms required to read the storage media no longer exists. So too, the file formats themselves will need to be readable by whatever application software is available.

If you think that you won't need to worry about these problems within your lifetime, try taking a Beta video cassette out of the cupboard and finding something to play it on. The Sony Betamax video recorders were tipped to be *the* video format – and that was under 20 years ago!

Look at 5¼" floppy disks. I know I still have some of these from an Apple machine I had as late as 12 years ago. I have little chance of accessing that data today.

Although DVD disks are out of reach for many at the moment, in a couple of years they'll probably be common place and sounding the death knell for the 650Mb CD. I wonder how long it will be after then that the reverse compatibility that some drives have with CD disks is faded out.

The Powerhouse currently saves its images to long-life 650Mb Kodak Ultima Gold CDs. On each CD is an accompanying ASCII (plain text) file with records of each image on that disk.

The fact that these disks supposedly last around 160 years is not important. It is the overall *stability* compared to the next brand we are interested in. In reality, the disks will probably have a useable life of only 10 to 20 years before images and data need to be transferred to newer technology just to guarantee reusability. Therefore every so often, an archive of hundreds of thousands of images and data will probably need to be copied to new media and the files migrated to newer Image Management Systems.

The Powerhouse's image project began six years ago and with the exception of the change over from CD Jukebox to a RAID hard drive system to serve the images and the upgrade of computers, there has been little change to the original Project specifications with regard to image creation and storage specifications.

However, in the six years since 1995, there has been considerable technological change in the way of computer hardware.

For the same money we were spending on hard drive storage six years ago, some 30 to 40 times the storage space can now be had. Average network bandwidth has increased – the Powerhouse itself has changed over to a full 100Mbit fibre optic system from an old 2Mbit system. The average PC workstations were running Pentium 90 MHz CPU's then. Today they average well over 1300 MHz. The average RAM configuration for the same computers has gone from 16Mb to 256Mb.

The design of the Museum's Imaging System certainly considered the technology of the day and at the same time it did foresee many of the changes mentioned above. However, images designed for computers and networks of six years ago: JPEG files delivered over restricted bandwidth and stored on large but relatively slow storage mechanisms, are today no challenge.

Based on the speed of technological advances, it should be that every five years the administrators of all image archives stop and take a long look at their methods and technology versus what is on the horizon. There is every reason to believe that the next big thing in storage, DVD – already upon us (format bickering aside), will replace CD-R technology in the near future. And while the manufacturers are sure to offer backward compatibility so the new DVD hardware reads CD-R disks, the question is: *for how long?* Most certainly there will be opportunity to migrate data and images to these new technologies but these opportunities may not last long enough to migrate a big archive for those who don't plan carefully and regularly. In any case, the job will not be anywhere near as simple as a straight copy from one media type to another.

The choices of what to scan and how to scan are determined by your available resources and your situation. Every institution will have different needs – a unique user base and set of deliverables. The decisions made by the Powerhouse in planning its Imaging Project were right for its situation – at that time. Now, six years later, its imaging client base has evolved, technology has changed and the internet has emerged from its *geek* status. New plans and decisions need to be formulated to suit these changes and at the same time, an acceptance of what has been accomplished in the past and how to best utilise what we have created.

The Powerhouse's imaging archive will continue to grow – the scanning and archiving will be perpetual. We cannot afford to go back and scan images over again simply because there is now better technology available to do so. We must continue and also accept the methods we have used in the past while continuing to embrace new methods and better technology.

The Powerhouse Museum's Imaging System, as it stands today, does what was intended. It delivers images and related data to staff and public for the purposes of research, promotion, education and more. The Museum's Image Centre was set up to scan, manipulate, print and archive images. This Centre has evolved into a high-resolution digital imaging service and continues to create and upload images to the Imaging System.
However, without a plan for constant review and update, even archives of beautifully scanned images and comprehensive data are worthless if those archives cannot be reused because the mechanisms for accessing the media - or the applications that they run on, have become obsolete. The Museum's next step, therefore, should be continual re-evaluation.

Scott Donkin
September 2001

Biography

Scott Donkin was with the Powerhouse Museum for between 1996 and 2001.

As *Image Centre Administrator* he was responsible for imaging services, which include high-resolution photographic scanning, manipulation and printing for the 300+ members of staff plus external customers. Scott also administered the Image Management System, overseeing the capture of digital images and data and dealt with long-term storage and retrieval. He has thorough knowledge of digital camera and scanning equipment and provides professional advice for other institutions on a regularly basis.

In May 2001 Scott left the Powerhouse to join digital image archiving and colour management company, *Digital Art Directory* in Sydney. As their production manager, he looks after digital image scanning and archiving services for cultural and commercial clients as well as high-end digital reproductions of art works.

Scott's qualifications include Digital Colour Prepress, Information Technology, Object Oriented Programming, Photography and Scientific Photography. Prior to 1996 he was a commercial photographer for 12 years, specialising in product and architectural photography as well as infrared and ultraviolet imaging of questioned paintings.

His interests include photography, computer programming and his daughter, Emily.

Feel free to email him at scott_donkin@iprimus.com.au

Bibliography and Recommended Reading

Books

Wilhelm, H., Brower, B.
"The Permanence and Care of Color Photographs: Traditional and Digital Color Prints, Color Negatives, Slides, and Motion Pictures". 1993.
Preseveration Publishing Company. ISBN 0-911515-00-3 & ISBN 0-911515-01-1

Blatner, D., Fleishman, G., Roth, S.
"Real World Scanning and Halftones: the definitive guide to scanning and halftones from the desktop". 2ⁿᵈ Ed. 1998
Peachpit Press. ISBN 0-201-69683-5

Digital Prepress series by Agfa Educational Publishing:
  "Digital Color Prepress":
- "An Introduction to Digital Photo Imaging"
- Vol 1: "An Introduction to Digital Color Prepress"
- Vol 2: "A Guide to Color Separation"
- Vol 3: "Working with Prepress and Printing Suppliers"
- Vol 4: "An Introduction to Digital Scanning"
- "A Guide to Digital Photography"
- "An Introduction to Digital Color Printing"
- "The Agfa Guide to Digital Color Prepress" – CD-ROM (Mac)

Periodicals

- Australian Personal Computer Magazine.
  - www.apcmag.com
- PEI (Photo Electronic Imaging) Magazine.
  - www.peimag.com
- Photo District News.
  - www.pdn-pix.com

Internet

http://www.getty.edu/research/institute/standards/introimages/index.html - Getty Museum's Imaging Program
www.scantips.com.au – all round guide to scanning
www.amol.org.au – Australia Museums Online, gateway to cultural institutions large and small. Online collections available. There are some very good links for digital imaging.
www.adobe.com – makers of Abobe Photoshop. Some great tips here.
www.kodak.com – a very large site. Look hard for all sorts of info regarding imaging, photography, digital cameras, scanners, white papers.
www.agfa.com – Agfa, major competitor to Kodak offering similar information. Agfa make some very good scanners.
www.color.com – the International Color Consortium, the first and last place to go for colour information.
http://www.adobe.co.uk/print/prodzone/howcmworks.html - colour management info.