Rapid Digitisation Project: Final Report

Australian Museum

2011

ATLAS OF **LIVING**
AUSTRALIA
sharing biodiversity knowledge

Contents

1   Executive Summary

In 2010, the Atlas of Living Australia (ALA) approved funding to both the Australian Museum (AM) and the South Australian Museum (SAMA) to collaborate on a project to explore and develop means for engaging volunteers to assist with digitising natural history collections. For its component of the project, AM received $85 000 for employing a digitising project officer for six months and for purchasing two imaging workstations.

The AM focused on methods for image-based digitising of dry pinned entomology collections. The AM team developed a team of volunteers, training materials and processes and procedures. This document, and a website which includes training and procedure videos and manuals (http://www.australianmuseum.net.au/Rapid-Digitisation-Project/), document the products and outcomes of the project.

Digitising workstations were established to minimise cost whilst maximising productivity and ease of use. Costing in the vicinity of $8 000 each they represent very good value for money. Each workstation requires two volunteers for efficient functioning, though a single volunteer can use them.

Productivity of the workstations varied depending on the species group being digitised—fragile groups take longer—and because digitising rates vary among the volunteers. The average hourly rate for digitising pinned entomological specimens (cicadas, leafhoppers, moths, beetles, flies) varied between 15 to 20 per workstation per hour.

The project began with just a few volunteers and one workstation for each of specimens and collection registers. By the end of the allotted six month project there were four specimen workstations and one register workstation being operated four days a week, five hours a day, by a team of 52 volunteers. This increase in workstations was made possible by a grant in early 2011 from the ALA for imaging equipment purchase.

In August 2011 the ALA approved a $90 000 supplement which, in conjunction with two months of funding contributed by AM ,will support the project through to October 2012. This extra time will be beneficial in enabling the identification of digitising processes and procedures for other collections in terms of taxa, eg malacology, and media, eg collections in liquid preservatives.

2   Introduction

The Australian Museum (AM) has been exploring opportunities for engaging volunteers in image-based specimen digitisation since 2007. Previous studies demonstrated that utilising volunteers for imaging specimens and their labels was feasible and compared favourably with traditional text-only data entry techniques.

AM approached the Atlas of Living Australia (ALA) in 2010 with a funding proposal to develop a volunteer-based digitisation program. The ALA agreed to fund a six month trial, to which the AM contributed an additional two months funding to take the trial project through to end of October 2011. This document represents the final report of that trial project.

2.1   Scope

The scope of this project was to explore and develop methods and technologies for engaging volunteers in digitisation and registration of museum specimens. The project focused on the entomology collection, in part because it is a big collection that is largely undigitised, yet it lends itself to a methodical volunteer-based digitising process.

2.2   Requirements

Output:

- Maximum 5MB file size
- Clearly readable text on labels
- Clear, focused image of specimen
- Include attachment of registration number
- Partial record of metadata including: species name, registration number.

Process:

- Simple standardised process that could be easily replicated and implemented on multiple workstations by volunteers
- Process comparable in speed to direct data entry by volunteers
- Specimen safety—minimal breakages
- Harmonious working relationship with collection staff.

Equipment:

- Relatively low cost imaging and computing equipment.

2.3   The case for image based digitisation

There was a time when databasing (entry of text only data) of collection holdings was the preferred way of digitising a collection for collection management and data access. This is now being challenged by the digitising of collections where specimens are imaged and their associated label data entered as complementary data. Image digitising is the new databasing.

The advent of this approach has come from the realisation that having an image of the specimen and its associated labels has strong collection data management benefits including:

- A readily accessible digital voucher of each specimen and its labels for verification and reference

- Reduced need for specimen handling

- A virtual specimen in the event of collection loss or damage, eg fire, flood, earthquake, or for when the specimen is on loan

- Remote access to original label data for review by researchers

- A capacity for using handwriting to help identify a collector in the absence of a collector name

- A limited potential for species identification from an image

- Enabling the use of 'non-experts' in data entry with the benefit of knowing data quality and enabling dubious data can be checked without having to physically visit a specimen in the collection.

3 Equipment

3.1 Establishing the Digitisation Laboratory

AM provided a large room in which the Digitising Laboratory was established. The room was appointed with new double power outlets and network cabling (in the vicinity of $5 000) and secure access. AM also provides photocopying, IT support and office supplies as ongoing in-kind support to the project.

Additional costs associated with the Lab include:

- for volunteers: meeting table, lunchroom, lockers for personal belongings, fridge, food supplies (tea, coffee, milk, biscuits)

- for training material: photocopying for manuals (induction and training).

Once the Laboratory was set up the equipment was purchased.

3.2 Specimen workstations

Equipment list for a single specimen workstation. This workstation is used for imaging individual specimens and their labels.

Equipment	Details	Cost (ex GST)
Furniture	2 tables and 2 chairs	$895
	Non-slip mat	$35.87
Computer	Desktop computer, 24 inch monitor, 19 inch monitor	$3045
	Powerboard	$29.80
Software	Microsoft Office with Access	Institutional licence
	Canon Capture	Comes with camera
Imaging equipment	Canon 550D with 60mm lens	$1344
	Metz Mecablitz microslave flash unit	$471.82
	Kaiser RSX Camera Stand	$1363
	Rechargeable Batteries	$20.60
	Battery Charger	$69.92
Specimen handling and curation	Trolley to transport drawers to and from collection	$297.60
	Stages: cardboard boxes with a foam insert to hold the specimen while imaging	$12 each, need 2
	Labels blocks: a foam block to hold labels	$47.30/m^2 sheet
	Entomological forceps for handling pins	$3.70 each
	Fine curved forceps for removing and handling labels	$31.50 each
	Pins: stainless steel entomological pins used for pre-pinning K registration labels and for pinning hospital drawer labels into unit trays	$7.15 / packet
	Metal pinning block to return labels to pins	$37.95
	Microscope slide to hold down folded labels	Supplied by collection
	'Hospital' drawer for damaged specimens	Supplied by collection
	Unit Trays to hold damaged specimens in 'hospital drawer'	Supplied by stores
	Scissors for cutting registration labels	Supplied by collection
	Registration labels: each specimen imaged will receive a new K registration label	Supplied by collection
Total cost		**$7724.21**

3.3   Archive/register workstation

Equipment list for a single register workstation which is used to photograph collection registers, page by page.

| Equipment	Details	Cost (ex GST)
Furniture	2 tables and 2 chairs	$895
	Non-slip mat	$35.87
Computer	Desktop computer, 24 inch monitor, 19 inch monitor	$3045
	Powerboard	$29.80
Software	Microsoft Office with Access	Institutional licence
	Canon Capture	Comes with camera
Imaging Equipment	Canon 550D with 60mm lens	$1344
	Freestanding cold fluorescent lights	$845
	Kaiser RSX camera stand	$1363
Register Handling	Trolley to transport registers to and from Archive Collection	$297.60
	Bookrest (cradle) and pillow to support the register whilst imaging	Supplied by Archives
	Gloves, cotton or latex, to handle a register with red rot	Supplied by stores
	Black cloth to cover the background behind the register in the image	Supplied by Archives
	Glass weights to hold done pages flat to improve image quality	$10
	Brushes to brush any dirt from the inside of the spine	$8.50
	2B pencil, eraser and pencil sharpener to lightly number register pages that have not already been numbered	Supplied by Archives
Total Cost		**$7873.77**

4 Process/workflow

4.1 Selecting and supplying specimens to be digitised

AM collection staff chose curated drawers of specimens that were to be made available for imaging. Criteria were that the taxonomy and names for specimens in those drawers were as up-to-date as possible, and were unambiguous in relation to the specimens being databased. Specimens needed to be accessible within drawers, eg not cramped where labels and specimens would be damaged, relatively robust, or mounted to avoid damage when being handled.

The figure below shows the decision making process that determines which specimens, organised by drawers, were to be digitised through the imaging process described in this report.

It is a large task to ensure the supply of specimens for digitising. With four workstations operating four days per week, the rate at which specimens can be digitised puts a considerable burden on collection staff. Curated drawers need to be allocated for up to a month in advance to ensure that volunteers will not run out of specimens to digitise. Any digitising strategy that involves high volumes of specimens needs to consider the impact of this on other collection management work, or to adequately fund additional collection staff to prepare specimens.

Prioritization framework for Digitizing Australian Museum Entomology Collection Material
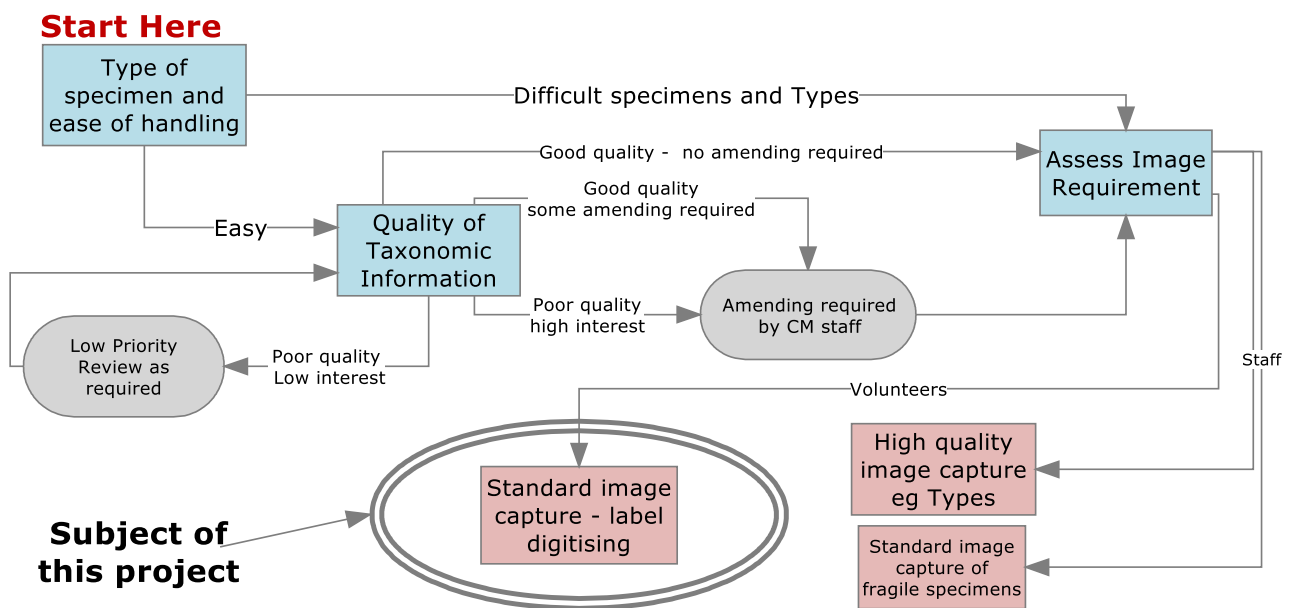
4.2 Specimen workstation

4.2.1 Preparation of drawers

The drawers of specimens require preparation before the digitisation officer removes them to the Digitisation Lab.

The preparation of the drawers is a two step process:

1. Curation: Performed by the collection staff and involves removing type specimens, ensuring each specimen is labelled adequately with its taxonomic name, and ensuring specimens are not overcrowded making it difficult for volunteers to handle. (This can have significant resource implications for collection staff see section 9).

2. Numbering drawers: Performed by the Digitisation Officer in order to track the location of drawers for replacement after processing. Small numbered labels are placed on the drawer knob. Note: Blue and red stickers are not used in this step as these colours represent type specimens in the collection.

4.2.2   Retrieve specimens

Drawers are removed from the collection by the digitisation officer in the order that they are numbered. Drawers are transported on a trolley from the collection to the Digitisation Lab.

4.2.3   Prepare specimens

At each workstation there are two volunteers: one volunteer handles the specimen (the specimen handler), the other volunteer images the specimen (the digitiser).

For more details of the process refer to the training guide: A Guide to the Handling and Digitising of Specimens (http://www.australianmuseum.net.au/document/Specimen-Training-Compressed).

Species labels are generally found below each specimen but that is not always the case. Digitisers are expected to spend time to assess how many species there are in the drawer they are about to work on and note the approximate numbers of specimens in each species. This should help avoid errors when entering data into the database. For example, if a specimen already has a new registration label, it has already been databased and therefore there is no need to image that specimen and data. Digitisers check that all family, subfamily, genus and species names in the drawer are entered into the database before starting to image that drawer.

A specimen is prepared for imaging using the following steps:

1. The specimen handler puts the stage in front of them and will always work within the stage to minimise potential specimen damage.

2. The specimen is removed from the drawer using fingers or forceps, bracing fingers on the edge of the drawer or unit tray to assist in pulling the pin out of the foam in a controlled fashion.

3. The position of the specimen within the drawer is noted using a stage labelled pin.

4. The specimen labels are removed from the pin using forceps.

5. The labels are placed on the rungs of the label block with the label closest to the specimen on the top rung and subsequent labels on the rungs below.
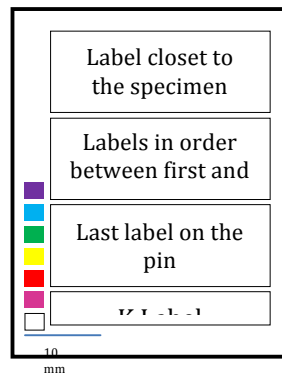
6. Place a new K registration number on the rung below the lowest label that was set on the pin. The K label is specific to AM.

7. Position the specimen on the stage with the specimen's head in the same orientation as the label block, ie with its head pointed toward the top of the stage.

8. Position the specimen and the label block to be at the same height. The label block holding the labels is only one layer of foam. To lift up the labels to be in line with the specimen use one, two or three foam layers as below. By having the specimen and labels on the same plane, they will both be in focus.
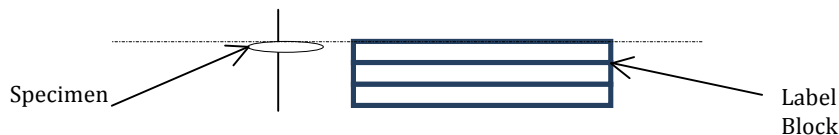
The specimen is now ready to be imaged.

4.2.4   Image specimens

1. Position the stage under the camera.

2. Ensure that the specimen and label are in the camera frame using the camera's remote live view finder. The specimen and label will be centred in the image and take up as much of the field of view as possible.

3. Once satisfied with the final image, the stage is passed back to the specimen handler for reassembling the label stack with the specimen.

4. Label information is entered into the database.

4.2.5   Record label data

A database has been created to help with good data entry and management practices.

The fields of the database are completed to correspond with that of the specimen image.

The image file name should match the new registration number.

Once the database fields have been checked and are correct, the digitiser is ready to receive another specimen for imaging.
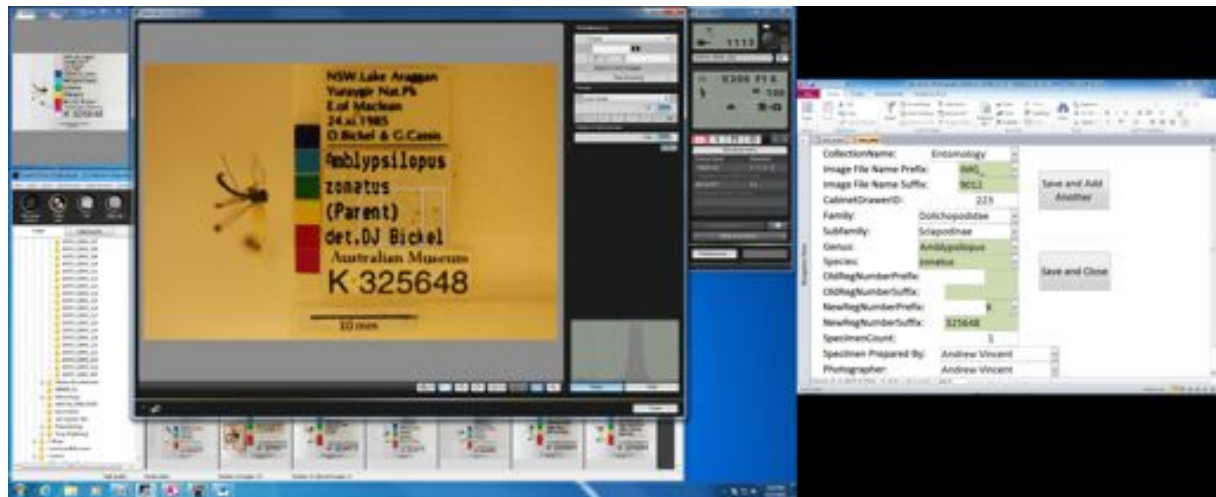
4.2.6   Replace specimen

After a specimen has been imaged it can be replaced in its drawer. The labels are moved to a foam 'pinning' block alongside the label block and are pinned in the order that they were removed from the specimen. The new K registration label becomes the last label to be attached.

The stage labelled pin is removed and the specimen is replaced in the same pin hole that it was originally in.

Damaged specimens, damaged labels and specimens without labels are placed in a 'hospital' drawer to be returned to collection staff for assessment and repairs.

Once each drawer has been imaged the information entered in the database is reviewed.

The drawer is now finished and can be returned to the collection.
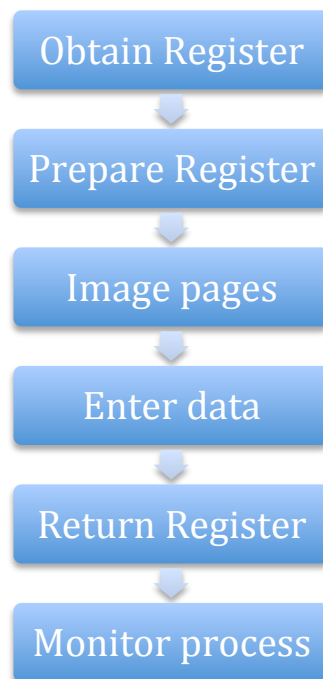
4.2.7   Return drawer

The drawer is returned to the collection by trolley, to the correct place in the cabinet using the numbered labels on the front of the drawer. The collection manager is notified when the hospital drawer is full.

4.2.8   Monitor process

The following information is recorded to monitor the project's outputs, staffing and volunteers:

- Number of specimens digitised per day by volunteer
- Number of drawers digitised per day
- Number of volunteer hours per day
- Number of damaged specimens
- Number of specimens damaged beyond repair
- Number of collection staff hours per day

4.3   Archive/register workstation

```
┌─────────────────────┐
│   Obtain Register   │
└─────────────────────┘
          ▼
┌─────────────────────┐
│  Prepare Register   │
└─────────────────────┘
          ▼
┌─────────────────────┐
│    Image pages      │
└─────────────────────┘
          ▼
┌─────────────────────┐
│     Enter data      │
└─────────────────────┘
          ▼
┌─────────────────────┐
│   Return Register   │
└─────────────────────┘
          ▼
┌─────────────────────┐
│   Monitor process   │
└─────────────────────┘
```

4.3.1   Obtain Register

A register is a paper record of museum objects held by the museum. In the 1980s, with the introduction of databases, paper records were no longer created.

Process

Obtain a register from the Archives Collection or from the individual collection using a trolley. Each register that is received into the Digitisation Lab is recorded in a spreadsheet. Each record includes the date the register was received, when digitising was finished and when it was returned to the collection.

4.3.2   Prepare register

There is a three step process to capture the image of the register:

1. The register handler prepares the register for imaging.

2. Imaging the register and data entry.

3. Preparing the next page for imaging.

Procedure

Each workstation is operated by two volunteers. One will handle and turn the pages of the register (the register handler) and the other volunteer, the digitiser, will image and database the pages.

The register is placed on the cradle (using gloves if red rot, ie deterioration of leather cover is apparent. The digitiser and the digitisation officer make an assessment of the register. If the register needs the pages to be numbered the digitiser will do so before imaging begins. For more details on the process refer to A Guide to Handling and Digitising Archival Material – Registers
(http://www.australianmuseum.net.au/document/Archive-Training-Compressed).

4.3.3   Image pages

The register handler prepares the register for imaging making sure that it is centred and flat in the camera view.
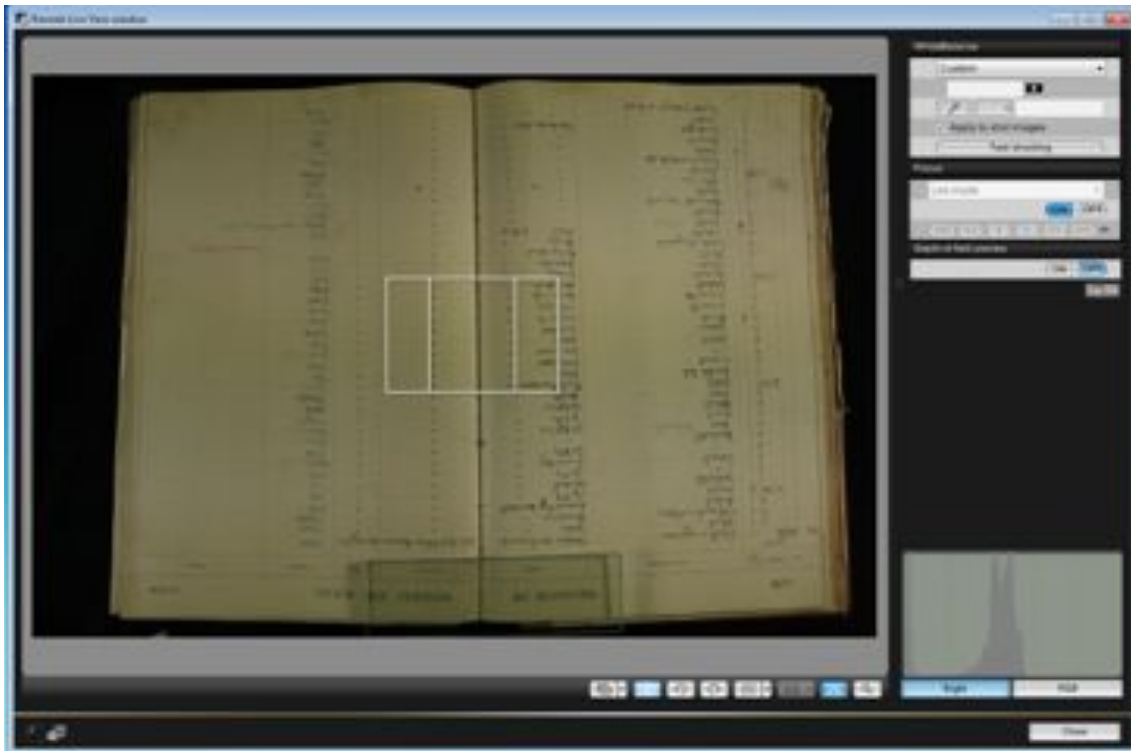
Use the camera's remote live viewfinder to ensure that the register is aligned and in the camera frame, and that the black cloth is covering the bookrest and the background.

The digitiser captures an image of each double page of the register and checks both the quality of the image and clarity of the writing in the image.

4.3.4   Enter data

A database has been created to help with good data entry and management practices.

The digitiser enters the register page information into the database for each image and the register handler provides a quality check on data entered. Each field on the database is checked to confirm it matches that on the register image screen. Once the digitiser is satisfied with the final image the register handler can prepare the next page for imaging.

4.3.5   Return register

Once the register has been digitised it can be returned to the Archive Collection or the collection in which it is stored.

4.3.6   Monitor process

The following information is recorded to monitor the projects outputs, staffing and volunteers:

• Number of archival material single pages digitised per hour by volunteer

• Number of Register volumes per day

- Number of volunteer hours per day
- Number of damaged archival materials
- Number of archival materials damaged beyond repair

5 Volunteer recruitment and management: Best Practice for museums

5.1 Recruitment, coordination and supervision of volunteers

The development and management of a team of well trained, productive volunteers dedicated to the digitisation of museum collections, whether large or small, requires the same basic approach.

A volunteer coordinator is essential. In practice, this role and its responsibilities may be spread across one or more people or positions. Ideally, however, we feel that a single position, which in the case of this project was shared between two individuals, is likely to produce the best outcome for a museum.

A volunteer coordinator is responsible for recruiting, training, coordinating and supervising volunteers. First and foremost they must have excellent people management skills and they must also have a good understanding of the technical processes involved in the digitising process. The extent to which they need to be technically proficient is dependent on the availability of other sources of technical expertise in the institution.

The major steps in the creation of the volunteer team are as follows:

Recruitment

Rostering

Induction

Training

Review

Ongoing Support

5.2 Recruitment

AM members were targeted because it was thought that they would have an interest in being involved in an AM project. An expression of interest email was sent out to the AM members and a very high level of interest was received. Based on this level of interest it was felt that no further effort would be required to obtain a suitable number of volunteers. Potential volunteers were asked to identify their preferred days, and their availability, to volunteer on a weekly or fortnightly basis.

5.3   Rostering

Potential volunteers were prioritised on their day preferences according to their response time to the expression of interest.

5.4   Induction

Small groups of volunteers were inducted during a morning induction session. An information package was provided including a volunteer handbook, *Explore* magazine, *Nature and Culture* booklet, ALA fact sheet and ALA postcards and volunteer induction forms.

There was an introduction to the working area and other volunteers, followed by a tour of the public exhibition within the museum.

5.5   Training

Following the completion of their induction session, volunteers were invited to attend a one day training session. Training was provided for both archives and specimens in the form of short videos in both handling and imaging registers and handling and imaging specimens. The videos were accompanied by training manuals, followed by hands-on practice with experienced volunteers.

5.6   Review

Digitiser volunteers undertake a six-week introductory period. At the end of this period the volunteer will complete a self-assessment review of their practice and the project.

5.7   Ongoing support

Digitiser volunteers receive ongoing practise support from peers as well as the digitisation officer.

6   Cost/benefit

6.1   Inputs

6.1.1   Inputs: setup (once off only)

Stage 1 (two workstations funded through ALA Rapid Digitisation Project)

- Digitisation Lab or space. In the case of AM this involved dedicating a room for the digitising workstations. However, smaller scale projects might involve establishing single digitising workstations in the vicinity of the supervising staff member, much as many museums already have volunteers databasing collections.

- Equipment
  - Cameras, computers, furniture, miscellaneous lab consumables
  - Two workstations: $16 000 (one for specimens, one for archives)

- Permanent staff inputs
  - Training of digitisation officers
  - Setting up lab
  - Purchasing equipment
  - Managing collection
  - Setting up database and entry interface

6.1.2   Inputs: ongoing (estimated)

- Digitisation Lab or space
  - Cleaning (covered by existing museum arrangements)
  - Electricity and network maintenance (should be covered by existing museum budgets)
  - Equipment maintenance (should be included in existing museum arrangements)

- Staff hours
  - Digitisation Officers
    - 21 hours each per week, 42 hours per week total
  - Collection staff (for one specimen digitising workstation)
    - Before digitizing: curation of specimens
      - 8 hours per week
    - During digitising
      - 3-4 hours per week (fortnightly meeting plus fixing specimens)
  - Manager—2 hours per week
  - Technical Support—2 hours per week

6.2 Outputs

6.2.1 Fully equipped digitising workstations

Each workstation consists of:

- Two desks

- A powerful desktop computer

- Camera, lens and light source

- Copy stand for vertical photography

6.2.2 Well trained volunteer workforce

Using the training approach and materials developed during this project enables an institution to develop a robust , versatile and flexible volunteer workforce. Such a team might be as small as a few volunteers up to as many as 50 or 60 or more.

AM, by developing a team of 50 or so volunteers is able to operate a five workstation Digitisation Laboratory continuously four days a week, and provide reserves to cover any volunteers who are away as a result of sickness or holidays or wish to have a break.

Volunteers will continue to be trained as needed to ensure the availability of enough volunteers to keep the lab functioning at full capacity four days a week.

Suitably skilled volunteers will be identified and trained in the use of both the digital microscope and high resolution large specimen workstation

6.2.3 Specimen digitising

Outcome

Each specimen digitised results in a record in EMu with taxon, registration number and includes an image of the specimen and associated labels in the EMu multimedia module.
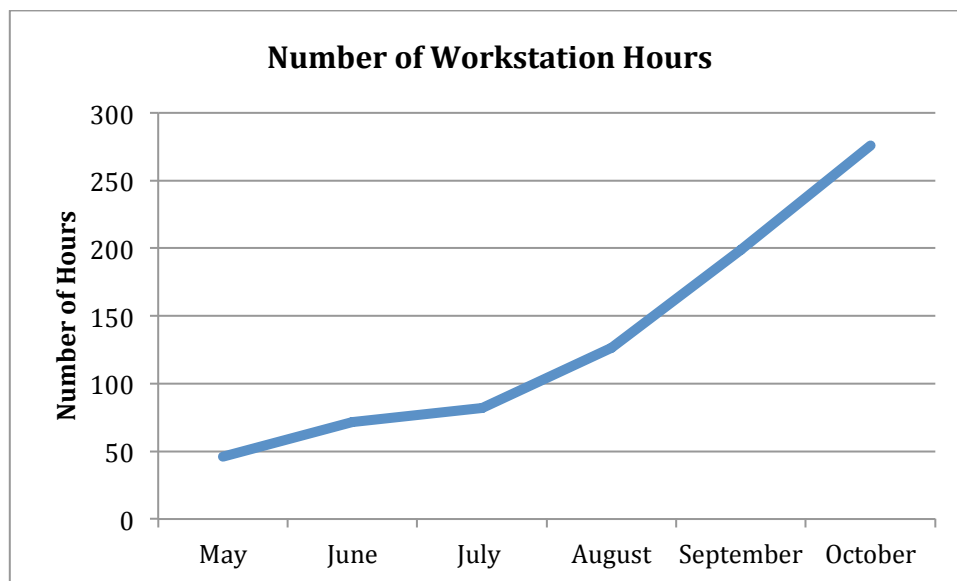
Outputs (one to four workstations)

The following graphs show various statistics over the period of the project.

As expected as the number of workstation hours per month (b.) increased so did the number of specimens digitized per month (a.).
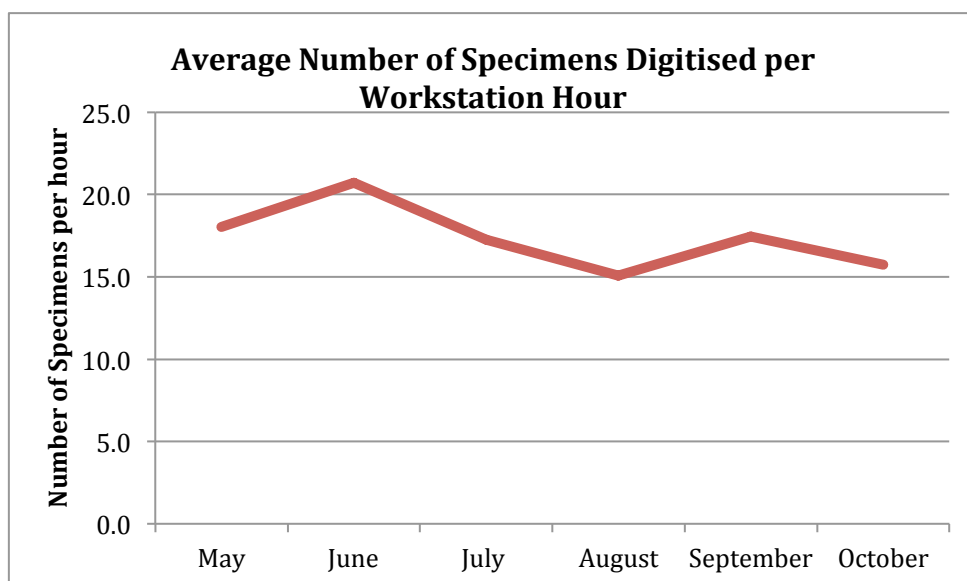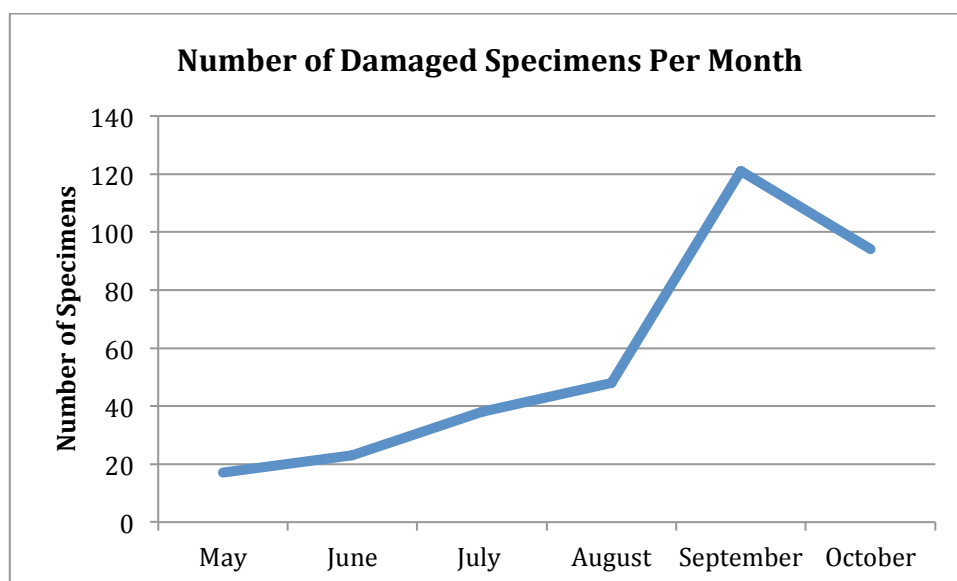
a.

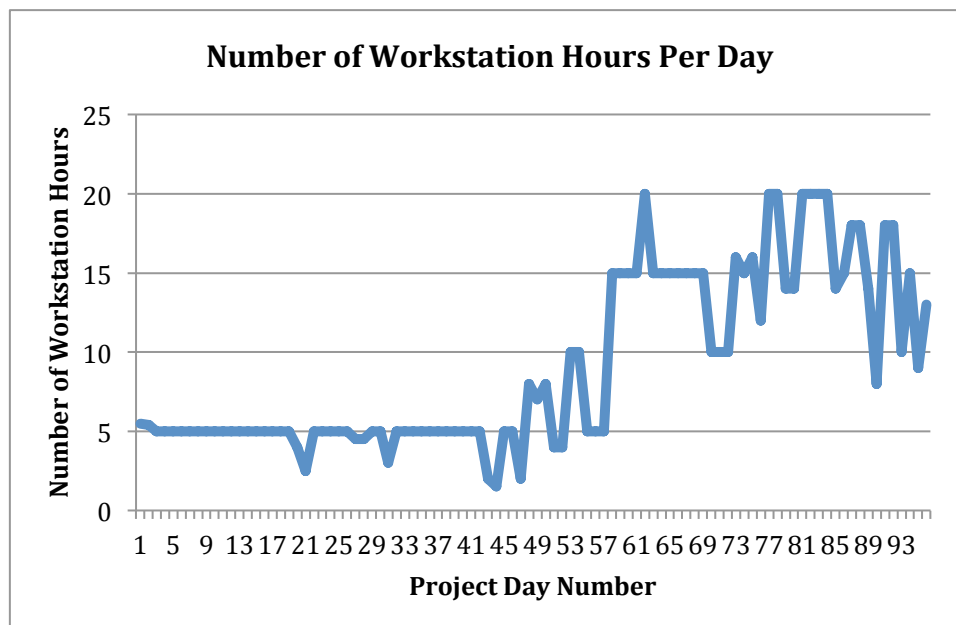**Number of Specimens**

b.

Number of Workstation Hours
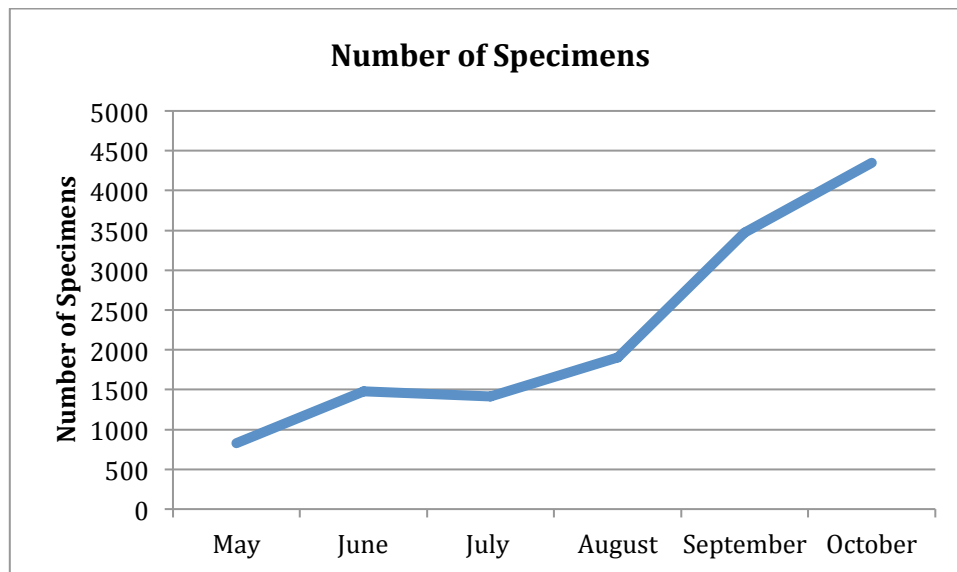
c. This graph shows that the number of specimens digitised per workstation hour has been fairly consistent throughout the project. The relatively large changes in daily productivity (see f below) are averaged out over a month and the productivity is fairly constant. The longer term variation is possibly due to the group being worked on. Cicadas (May, June, July) were easy to work with and so may have been a little faster to digitise.
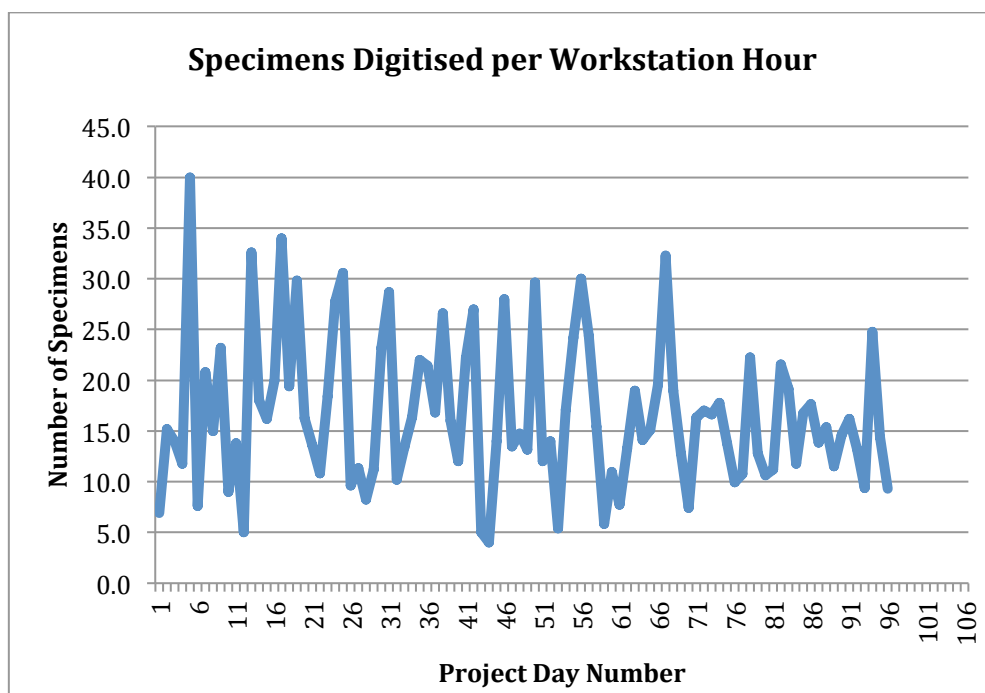
Average Number of Specimens Digitised per Workstation Hour

(Chart showing Number of Specimens per hour on the y-axis from 0.0 to 25.0, and months May through October on the x-axis. The red line starts at about 18 in May, rises to about 21 in June, decreases to about 17.5 in July, drops to about 15 in August, rises to about 17.5 in September, and falls to about 15.8 in October.)

d. As expected the number of damaged specimens per month increased as the number of specimens digitised per month increased. The peak in September is higher than might have been expected because the fragility of the group that was worked on that month: Sphingidae moths. All damaged specimens need to be dealt with by collection staff, so spikes in damage mean spikes in collection staff time to remedy.

Number of Damaged Specimens Per Month

(Chart showing Number of Specimens on the y-axis from 0 to 140, and months May through October on the x-axis. The blue line starts at about 17 in May, rises to about 23 in June, increases to about 38 in July, rises to about 48 in August, sharply increases to about 121 in September, and drops to about 94 in October.)

e. The number of workstation hours per day increased as the project took on more workstations. Even so it can be seen that there was still variation in the workstation hours. This is due to volunteer attendance being more likely to be affected by external factors than is the case with paid staff.

Number of Workstation Hours Per Day

Number of Workstation Hours (y-axis)

Project Day Number (x-axis)

f. There is great variability in the number of specimens digitised per workstation hour. See g below for more explanation.

Specimens Digitised per Workstation Hour

Number of Specimens (y-axis)

Project Day Number (x-axis)

g. This graph shows that there is great variability in the number of specimens digitised per day and that it is not simply a factor of the number of workstation hours per day. This suggests that other factors such as volunteer competency and diligence and the taxonomic group being worked on influence digitising rate also.
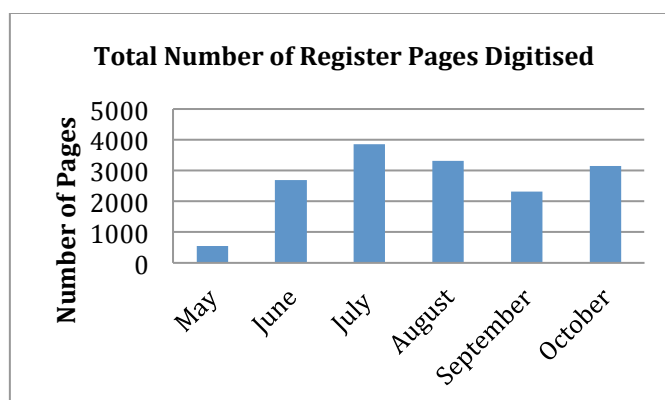
Number of Specimens Digitised per day as a function of WorkStation Hours
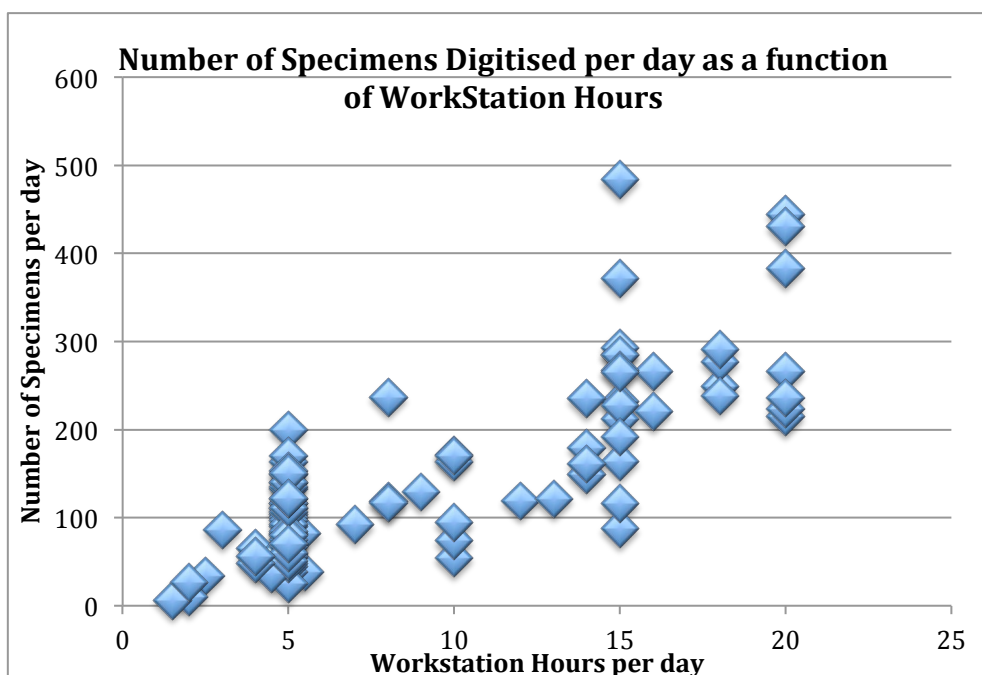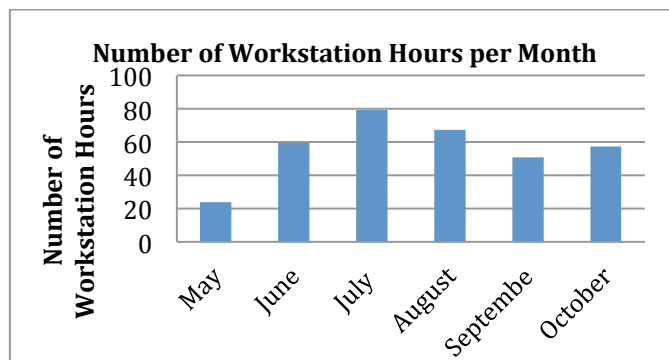
6.2.4 Archive digitising

Outputs (one workstation)

The variability in the register pages digitised per month (a.) can be seen to be primarily a function of the number of workstation hours per month (b.) and is less affected by volunteer and content variation. The rate of digitising of registers (c.) has remained fairly constant once the project was up and running for a while.
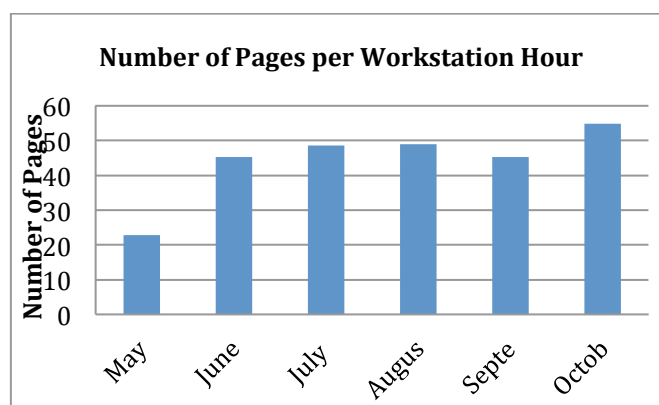
a.

Total Number of Register Pages Digitised

b.

Number of Workstation Hours per Month

	Bar chart of Number of Workstation Hours (y-axis, 0 to 100) by month (May, June, July, August, Septembe, October)

c.

Number of Pages per Workstation Hour

	Bar chart of Number of Pages (y-axis, 0 to 60) by month (May, June, July, Augus, Septe, Octob)

7 Funding exploration/outcomes and models

The processes and procedures developed during this project for digitising specimens and their labels are designed to be completely scaleable and adaptable to the resources available in an institution. This approach was taken because it is well understood that the funds available at any particular institution will vary greatly. With this in mind we will attempt to outline the range of options available using the processes and procedures identified in this report.

7.1 Potential sources of funds

In April, 2011 a funding options strategy was developed to explore different funding options for the project. In the first instance, the strategy was informed by discussions with key AM staff in fundraising and grant submission. These staff advised on possible charitable trust websites, corporate sponsors (both Australian and international), museum foundation funding rounds and internal museum funding protocols and processes.

A thorough review of trusts listed on Philanthropy Australia highlighted the following significant criteria limitations and grant exclusions imposed by many of the trusts:

- a strong social welfare emphasis

- interstate geographic exclusions

- emphasis on seeding grants rather than larger capital funds or recurrent funding

- priority given to community organisations rather than those organisations which have the capacity to obtain significant corporate sponsorship or community support

- unsolicited grants not considered

- exclude ongoing core operational costs.

Although it is understood that AM has had limited success with funding submissions to a range of trusts, it was felt that it would be worth approaching the WV Scott Trust as a possible source of funds This however will only be likely to provide a year's funding at best, leaving the source of the long term funding required still unidentified.

In the case of securing a corporate sponsor for the project, it is considered highly unlikely in the short term. Corporate sponsors have significant resourcing expectations in return for their sponsorship. At this stage, the project is not resourced to manage any such expectations. However, longer term it may be worth considering corporate sponsors as an option as the project will have more demonstrable outcomes and a stronger branding image.

Collaborative partnerships with other organisations were explored with the Centre for Volunteering and with Conservation Volunteers Australia. At this stage, there is no apparent partnership benefit to the museum as organisations such as these provide access to a source of volunteers, which AM already has through its members program; they provide a collaborative coordination role but do not have a funding stream.

Other potential funding options that could be explored are local sources such as Sydney City Council.

Our investigation of funding options came to the following conclusion on likely sources of funding for digitising projects:

- It is far easier to get funds for buying equipment and building infrastructure than it is for 'bums on seats'.

- With this in mind, some institutions may find it worthwhile seeking funds for equipment purchase and then allocating some existing internal resources to setting up the equipment and coordinating the volunteers in a manner that is amenable to their available staff resources.

- Short term projects of one or two years may be funded through trusts, particularly those related directly to the institutions activities, eg The Australian Museum Trust. Other trusts such as WV Scott Trust may be worth trying. Such short term projects should focus clearly on delivering a specific content such as a charismatic or high profile collection in its entirety.

- Long term digitising is only likely to be funded by large Commonwealth or State Government funding initiatives.

- In the absence of either of these sources of funding it is dependent on the institution itself to determine its priorities in terms of digitisation and focus what resources it can in pursuing those priorities.

7.2   Project expenses

7.2.1   Personnel

It is estimated that around $100 000 annually is required to fund the two digitisation officer positions. These people coordinate the volunteers, eg 2x3 days per week, and manage and coordinate the digitisation work undertaken by the volunteer workforce (approximately 50 volunteers). Further funding of these positions is integral to the success and continuation of the project.

7.2.2   Equipment

The equipment required to set up one workstation is clearly defined in section 3. The cost of setting up a workstation is somewhat flexible in that many institutions will already have the necessary equipment for specimen handling and curation or for archive handling and also the necessary furniture. This can considerably reduce the costs of setting up a workstation, to around $5 000, which covers the purchasing of the imaging equipment and computer software and hardware.

8 Conclusions and recommendations

8.1 Best practice

The processes and procedures detailed in this report constitute best practice for the predefined goal of image based digitising of individual specimens and the associated labels. We have produced documentation and videos that detail the handling of specimens and registers and a handbook for volunteers involved in the digitizing project.

The two most important components of this best practice are:

- the dedicated role of the digitising officer who recruits, trains and coordinates the volunteers, liaises with collection staff and implements the technical processes

- the curation of material prior to digitisation by collection staff which makes the digitising process is as effective as possible in terms of consistent identification of digitised specimens, ease of handling and selection of appropriate specimens (including removal of types).

8.2 Simple low cost options

Where institutions are unable to implement best practice because of resourcing constraints the processes and procedures detailed here can be scaled to suit available resources. For example, a single workstation could be established for around $5 000 and a small team of volunteers (two to ten) trained and coordinated by an existing staff member if the workstation was located in close proximity to the staff member.

8.3 Equipment selection

The selection of equipment for the imaging workstations was primarily focused on ensuring that the process was as simple and productive as possible. Image quality was seen as important but of less importance. Our target in terms of the images we were to capture and store was set at around 5MB, this being what we considered sufficient for the task and this was reflected in the equipment we have chosen.

Computer

A fast but standard specification was chosen to get the best balance between price and performance. Two screens were seen as important and so we settled for one large one for viewing the images, as image capture is controlled through the computer, and another screen for operating the database for data input.

Copy stands

Good quality copy stands are essential as they provide stability for the camera and a sound platform upon which the specimens can be imaged. Kaiser makes excellent stands.

Cameras

Any number of cameras could have been chosen and would have been suitable to the task. We chose a Canon 550D as we felt it was sufficiently good to do the job and represented very good value for money. We felt there was no need for a more expensive camera nor a higher resolution camera because we wanted to keep the image size to a manageable 5MB jpeg.

8.4 Issues

8.4.1 Equipment

Copy stands

The equipment used in the digitising workstations is easily obtainable, has a range of options in terms of replacements or alternatives and is relatively inexpensive. The one item that is crucial to the whole process, has few if any alternatives and has proven difficult to source in suitable numbers is the copy stand upon which the camera is mounted. The Kaiser brand was selected because of its quality build and in truth because there are few if any alternatives. Unfortunately the delivery of the Kaiser copy stands has been extraordinarily slow and at times threatened to jeopardize the success of our six month project. This is clearly a threat to any project, such as ours, that is relatively short term and dependent on specific items of equipment. Identifying an alternative more reliable source of copy stands or having plenty of lead time would be beneficial in such short term projects.

8.4.2 Curation of specimens by drawer in preparation for digitising

A factor that should not be glossed over is the potential resource impact of preparing drawers of specimens for digitising. This curation involves removing types (where it has been decided that types are not to be imaged by volunteers, as is the case in the AM project), ensuring specimens within the drawers are labelled adequately with taxonomic names and that the specimens are not overcrowded and thus difficult for volunteers to handle. These are tasks that need to be carried out by collection staff or, where appropriate, experienced volunteers.

Institutions need to ensure that adequate resources and lead time are made available to allow collection staff to curate drawers well ahead of scheduled digitising for those drawers. In addition, any application for funding of digitising projects needs to factor in resources required for the curation of material to be digitised. See section 6.1.2 for estimates of time required for curation.

It is difficult to estimate how much time is required specifically for the curation of specimens for digitising because curation is a normal part of collection management. What is clear though is that a dedicated rapid digitising project shifts the priorities of collection staff onto curation of specimens that may otherwise not have been in their workplans. Unless effectively resourced this can lead to conflict over work priorities in the collection. All insect collections contain a mixture of groups which range from well identified (usually because of associated input from skilled researchers) to completely unidentified or incorrectly identified. There may be well-identified groups that are not suitable for digitisation because of their physical state or there may be groups which would be ideal to digitise which are not well identified. The presence of well-identified collection material represents an investment that should not be taken for granted.

The table below gives some idea of the difficulty in estimating time required for curation. These are figures provided by David Britton, Collection Manager at AM.

Group	No. of drawers	No. of specimens	Curation time estimate
Notodontidae (moths)	26	~ 1000	7 days, included some identification (4 drawers)
Sphingidae (hawk moths)	50	1916	8-15 days, included some identification (6 drawers)
Cicadas	82	4386	25 days
Scarabaeidae (beetles)	5	2204	15 days including identification
Noctuidae (moths)	10	809	4 days
Leafhoppers etc	41	3385	20 days

9 Appendix: The data entry database

9.1 Database Setup

Volunteers capture the images of the specimens using Canon cameras controlled from workstations using Canon camera control software. Immediately after the image capture volunteers record data about the image into an Access front end database (see figure in section 4.2.5). The information recorded by the volunteers includes the specimen registration number, any old registration numbers, the taxonomic details and the image file name. Each of the front end workstations is connected to the same Access backend, resulting in four workstations and a single administration front end all depositing or retrieving data from the same source.

The administration front end uses a combination of VBA and queries to generate import files for EMu and export metadata files for ALA. Each operates at the press of a button that asks for information from the user such as the number of records to be processed or any supplementary information required to complete the files, eg the Internal Record Number of any associated taxonomy records for the EMu import file, before opening the resulting data for the user to examine and export the file. There are a number of ways the current process can be improved, eg for bulk importing and IRN matching to EMu, but those changes have yet to implemented.
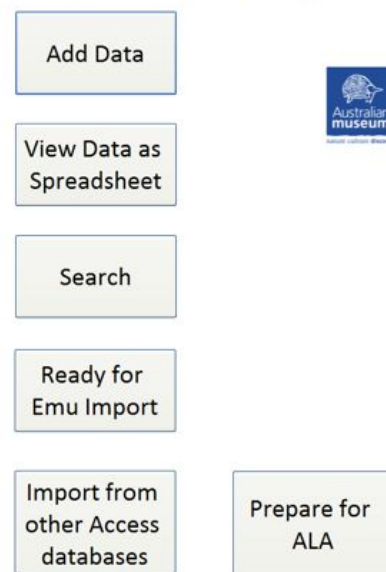
Apart from the actual time required by the volunteers to input the data, the generation of import files is quite fast. Indeed, the limiting factor in importing the records to EMu is EMu itself. It takes only minutes to generate approximately 7 000 record import file, but three working days (at approximately 1 000 records per import) for EMu to create the catalogue and attached multimedia records. On the few occasions that data have been imported into EMu without attached multimedia records, the imports have completed within minutes.

9.2   How the database works

The digitisation project uses specially designed databases that are so complex that nobody knows how they work. The computer has a sort of wibbly-wobbly bit that people wriggle around and somehow it does stuff. Underneath the hood the database is powered by four lizards called Ken, that run around on specially designed boots with springs on the bottom that allow them to slam dunk bits of data into the database. It has great colours and a font large enough to be seen from the moon. Every 15 minutes it tells the time in Spanish.

To get the data into EMu we first hold the emu down and sit on its neck. Then a small drop of single malt whisky is tipped onto its head and the first aria from Rigoletto is sung while a yabby pump is used in a gentle and deceptively coaxingly manner to get the tables in. So far we have used several antique French monoquocue Louis 16th dressing tables and these seem to work unfeasibly badly. The emu is then shaved and asked to take out a ridiculously large mortgage.

Then everyone smiles and gives high fives all around.

— —Michael Elliott