Image based Digitisation of Entomology Collections:

Leveraging volunteers can significantly increase digitisation capacity

Paul Flemons, Rhiannon Stephens, Leonie Prater, Michael Elliott
Collection Informatics, Australian Museum

nature culture **discover**

Australian museum

ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge

Entomology collections are large and inaccessible

- Entomology Collections are very large, but in most cases remain largely undigitised and therefore all but inaccessible
    - Australian Museum 11.4%,
    - Queensland Museum 9.4%,
    - Museum Victoria 9.4%,
    - Australian National Insect Collection 4.2%

(statistics courtesy Atlas of Living Australia)

Problem – lack of resources

- Resourcing large-scale digitising still beyond the budgets of most museums.
    - Funding bodies (in Australia at least), governments included, see digitising as a core activity and so are unwilling to fund the staff required to make it happen at the scale that is required to have an impact on the large undigitised collections held by many museums.

A new approach

Developed by the Australian Museum, with funding assistance provided by the Atlas of Living Australia.

- Combines:-
  - the benefits of image-based digitisation
  - high throughput
  - scaleability

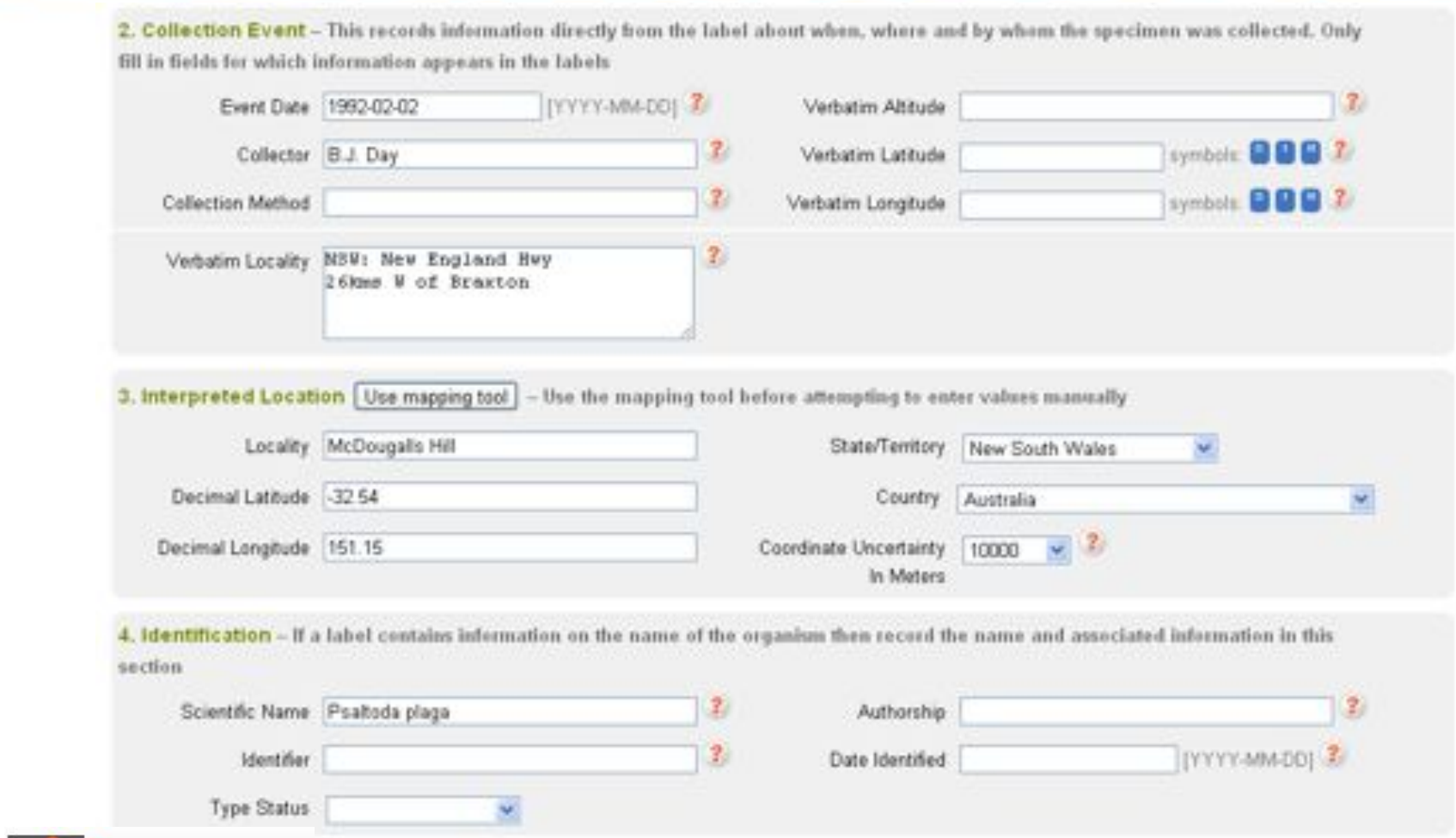
- Uses a team of some 60 volunteers and 4 digitising workstations operating 4 days a week, 5 hours a day
  - with sufficient space and equipment could scale up considerably more

- We estimate that with current setup we will digitise between 50000 and 75000 specimens per year depending on the groups being imaged.

Why Image based Digitising?

- Image based Digitising is the new databasing
  - Specimens are imaged and entered into the collection database along with their associated label data.

Traditional Digitising – or databasing



2. Collection Event – This records information directly from the label about when, where and by whom the specimen was collected. Only fill in fields for which information appears in the labels

Event Date: 1992-02-02 [YYYY-MM-DD]
Collector: B.J. Day
Collection Method:
Verbatim Locality: NSW: New England Hwy 26kms W of Braxton
Verbatim Altitude:
Verbatim Latitude: symbols
Verbatim Longitude: symbols

3. Interpreted Location Use mapping tool – Use the mapping tool before attempting to enter values manually

Locality: McDougalls Hill
Decimal Latitude: -32.54
Decimal Longitude: 151.15
State/Territory: New South Wales
Country: Australia
Coordinate Uncertainty In Meters: 10000

4. Identification – If a label contains information on the name of the organism then record the name and associated information in this section

Scientific Name: Psaltoda plaga
Identifier:
Type Status:
Authorship:
Date Identified: [YYYY-MM-DD]

ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge

Image based digitising – database record

2. Collection Event – This records information directly from the label about when, where and by whom the specimen was collected. Only fill in fields for which information appears in the labels

Event Date	1992-02-02 [YYYY-MM-DD]	Verbatim Altitude	
Collector	B.J. Day	Verbatim Latitude	symbols
Collection Method		Verbatim Longitude	symbols
Verbatim Locality	NSW: New England Hwy 26kms W of Braxton		

3. Interpreted Location Use mapping tool – Use the mapping tool before attempting to enter values manually

Locality	McDougalls Hill	State/Territory	New South Wales
Decimal Latitude	-32.54	Country	Australia
Decimal Longitude	151.15	Coordinate Uncertainty In Meters	10000

4. Identification – If a label contains information on the name of the organism then record the name and associated information in this section

Scientific Name	Psalloda plaga	Authorship	
Identifier		Date Identified	[YYYY-MM-DD]
Type Status			

ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge

Image based Digitising

Benefits include:

1. images are a readily accessible digital voucher of specimen and labels for verifying data

2. reduced need for specimen handling

3. having a virtual specimen in the event of collection loss or damage (eg fire, flood, earthquake), or when the specimen is on loan

4. enabling remote access to original label data for review by researchers

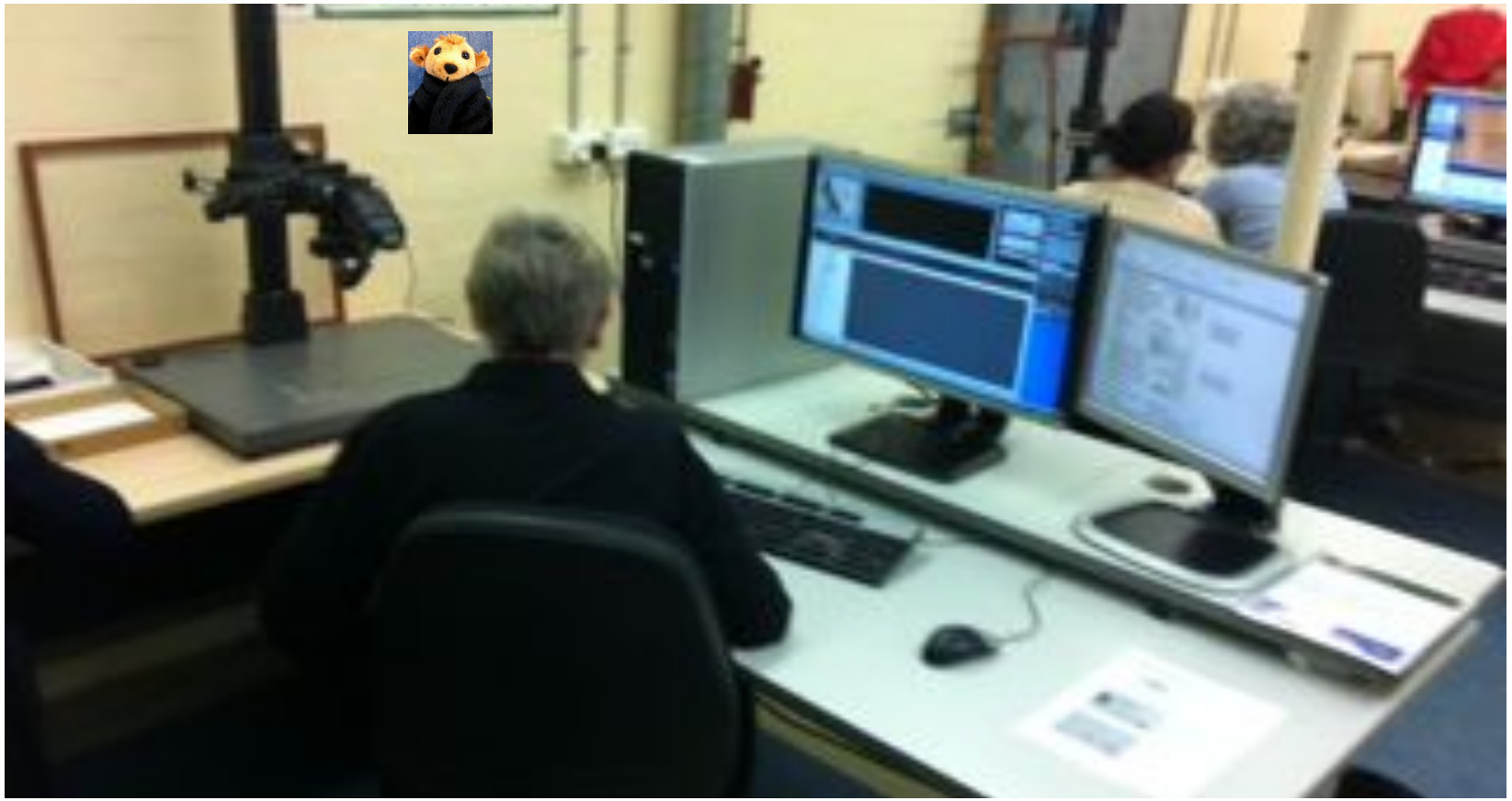
5. some limited potential for species identification from an image

6. enabling option for full data entry by “non-experts” :
    1. at time of image capture
    2. through crowdsourcing mechanisms

Image based Digitising - using Volunteers

ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge

Digitising Equipment



Digitising Process

ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge

Digitising Process

Digitising Process

Documentation

- Website - http://www.australianmuseum.net.au/Rapid-Digitisation-Project
- Manuals

- Videos

- http://www.australianmuseum.net.au/Video-Guide-Handling-Specimens
- http://www.australianmuseum.net.au/Video-Introduction-to-Handling-of-Specimens

Digitising Process - output

- Images
  - Purpose
    - Capture labels and specimens as a virtual record
  - Jpg
  - Size 5mb
- Metadata
  - Species, date, photographer, databaser, catalogue number, drawer no.
  - Captured into an Access database initially
  - Then imported into KE EMu

The Digitising Volunteers

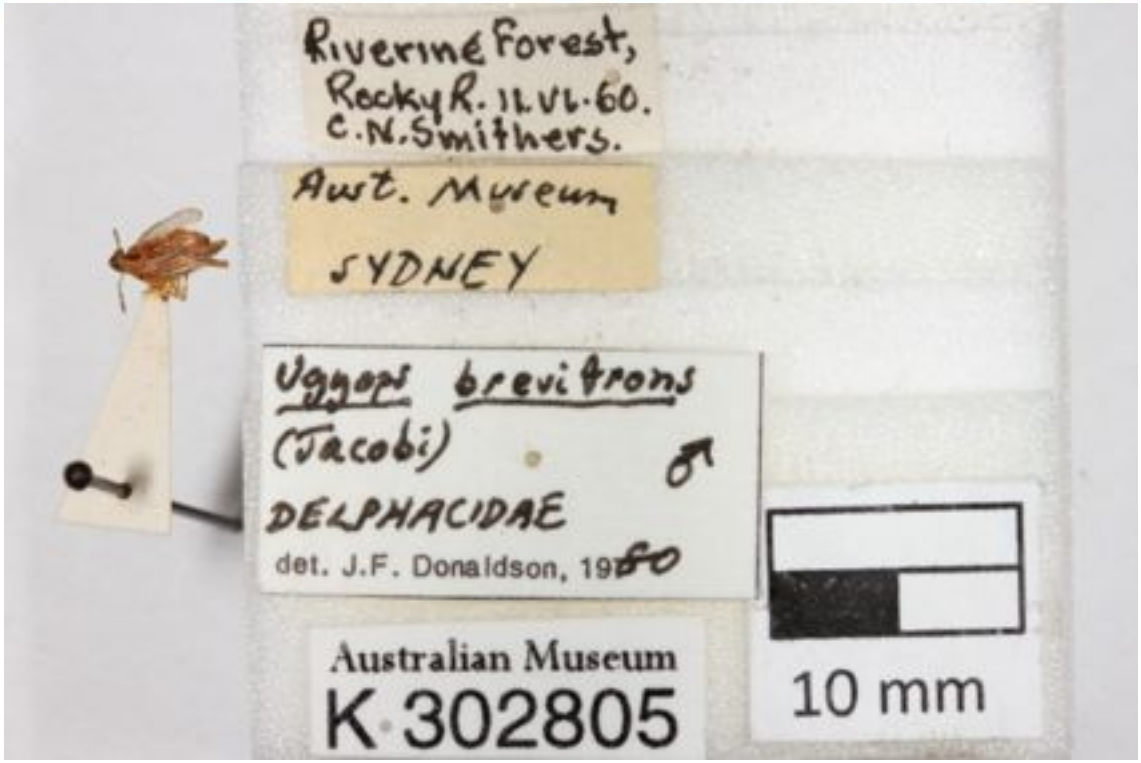
- Recruitment
  - Through traditional Museum networks
    - Members of the Museum society
    - Existing Museum volunteers
- Training
  - Custom designed training course
    - Orientation
    - Training Videos and Manuals
    - Hands on training
- Coordination and Supervision
  - Two part time staff share the tasks of recruiting, training, coordinating and supervising
    - Equivalent of 1.2 full time staff
    - 4 volunteer contact days
    - 1 non-contact day – specimen preparation, data management and documentation

The Digitising Volunteers

- Current Volunteer Team:
  - 60 volunteers
  - volunteer drop out rate has been minimal with most volunteers committing weekly, some fortnightly
  - a 2:1 ratio of female/male volunteers
  - age range: a third under 30; a third between 30-49 and a third over 50 yrs.
  - university students (10); full time workers (Saturday's); part time workers and retirees
- Input: 1.2 EFT staff
- Output: equivalent to between 3 and 4 EFT staff

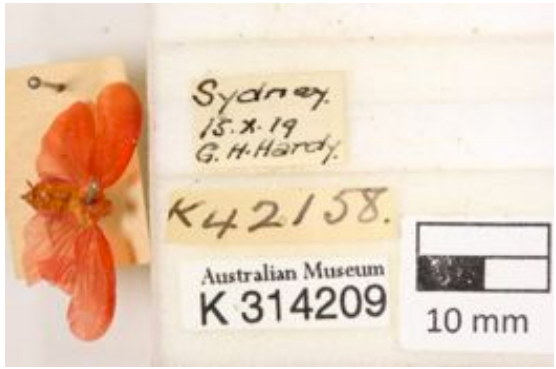
Digitising Results

Digitising Results



Digitising Issues

- Transcription of label information
- Supplying the “hungry beast” with specimens
- Errors
  - Duplicates
  - Poor images



Digitising Results

- September 2011
  - Specimens: 3473;
  - Taxa – hawk moths and leafhoppers
  - specimens damaged- 121(ie large nos due to brittle hawk moths already broken and damage easily when handled; glued tree hoppers falling off cards)

- Total project – May – September 2011
  - Total = 10,575
  - Taxa – cicadas, leafhoppers and treehoppers, hawk moths;
  - specimens damaged-249

- Projected estimates for year 2012 – using 4 workstations
  - 50000 to 750000
  - depending on taxa

Thank you

www.australianmuseum.com.au

nature culture **discover**

Australian museum

ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge