Soup for crowds: a new source of data on insect richness, diversity and abundance

Paul Flemons1 and Beth Mantle2

1. Australian Museum, Sydney, Australia
2. Australian National Insect Collection, CSIRO Ecosystem Sciences, Canberra, Australia
Patterns immediately obvious
Richness diversity abundance specific time and location.
Comparisons across time and space
Measures of richness and diversity are routinely used in research and planning.
Crowdsourcing could provide a means of extracting measures of richness, abundance and diversity from these “insect soups.”
How could we do that?
Step 2: Build a website where volunteers capture the patterns

- Group “like” insects
- Tag
- Extract image
- Count

Store in database

<table>
<thead>
<tr>
<th>Select</th>
<th>Symbol</th>
<th>Image</th>
<th>Class (Key)</th>
<th>Order (Key)</th>
<th>Sub-Order (Keys for each Order)</th>
<th>Family (Keys for each Orders)</th>
<th>Genus (Keys for each Family)</th>
<th>Count</th>
<th>Body</th>
<th>Wings</th>
<th>Legs</th>
<th>Colour</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Insecta</td>
<td>Coleoptera</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>Black</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Insecta</td>
<td>Diptera</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>4</td>
<td></td>
<td>3cm</td>
<td></td>
<td>Metallic green</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Insecta</td>
<td>Coleoptera</td>
<td>unknown</td>
<td></td>
<td></td>
<td>3</td>
<td>2cm</td>
<td>clear with brown pattern</td>
<td>Thin, long</td>
<td>3cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Insecta</td>
<td>Diptera</td>
<td>unknown</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>2cm</td>
<td></td>
<td>Metallic green</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>Black</td>
</tr>
</tbody>
</table>

Scale mm’s

• Group “like” insects
• Tag
• Extract image
• Count

Store in database

<table>
<thead>
<tr>
<th>Select</th>
<th>Symbol</th>
<th>Image</th>
<th>Class (Key)</th>
<th>Order (Key)</th>
<th>Sub-Order (Keys for each Order)</th>
<th>Family (Keys for each Orders)</th>
<th>Genus (Keys for each Family)</th>
<th>Count</th>
<th>Body</th>
<th>Wings</th>
<th>Legs</th>
<th>Colour</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Insecta</td>
<td>Coleoptera</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>Black</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Insecta</td>
<td>Diptera</td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td>4</td>
<td></td>
<td>3cm</td>
<td></td>
<td>Metallic green</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Insecta</td>
<td>Coleoptera</td>
<td>unknown</td>
<td></td>
<td></td>
<td>3</td>
<td>2cm</td>
<td>clear with brown pattern</td>
<td>Thin, long</td>
<td>3cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Insecta</td>
<td>Diptera</td>
<td>unknown</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>2cm</td>
<td></td>
<td>Metallic green</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>unknown</td>
<td>unknown</td>
<td>unknown</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>Black</td>
</tr>
</tbody>
</table>

Add Species
Virtual collections of insects can be created from the extracted images
Data Mining and Discovery

Metadata tags could be used to mine like images to explore a whole range of questions
Exciting possibilities for researchers and enthusiasts alike:

- discovering new species?
- range extensions?
- new distribution information?
Other Applications

other systematic and taxonomy related related applications?
Some Issues with Image Capture

Soup imaging workflow
making it efficient
Dirty soups
Crowded soups
Some other issues

Image tagging and management
controlled vocabulary for tagging?

Collection Management
loans
removal of insects from soups

Virtual Collection Management
how might that work
updating images as they have species identified
Set aside our prejudices about what can and cannot be done with “insect soups” and crowdsourcing and build communities that can help us mine this rich source of data.
Bulk sample
ANIC 3669

Melaleuca TAS
nr Bathurst Harbour
43°25'S 146°10'E
18-21 February 1991
A Calder & W Dressler
sweeping Leptospermum
flowers
Thank you

www.australianmuseum.net.au