Field to database to aggregator and beyond: documenting the flora of Melanesia

Shelley A. James
iDigBio, Florida Museum of Natural History
Field to database born digital – important?

- Making data accessible, discoverable, useable sooner
- Improve collections management efficiency
- Sustainability
- Expedition and collection tracking
- Publishing and citation
**Minimum data fields needed for biodiversity aggregators?**

<table>
<thead>
<tr>
<th>Darwin Core</th>
<th>Audubon Core</th>
<th>Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>• recordID</td>
<td>• recordID</td>
<td>• Institution</td>
</tr>
<tr>
<td>• occurrenceID (unique!)</td>
<td>• occurrenceID of specimen</td>
<td>• Collection</td>
</tr>
<tr>
<td>• scientificName</td>
<td>• URL</td>
<td>• Contact &amp; info</td>
</tr>
<tr>
<td>• eventDate</td>
<td>• Camera EXIF</td>
<td>• Description</td>
</tr>
<tr>
<td>• recordedBy (ORCID!)</td>
<td>• photographer</td>
<td>• URL</td>
</tr>
<tr>
<td>• Locality information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• catalogNumber</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• institutionID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• collectionID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Geological Context</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Collect data with a particular level of reuse in mind!*
Valevahalo (base) camp, Guadalcanal
**Locality (incl. state/region)**

<table>
<thead>
<tr>
<th>Lat.</th>
<th>N/S</th>
<th>Long.</th>
<th>E/W</th>
<th>Error</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Habitat (substrate/host/assoc. species)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Habit (bark, wood)**

**Habit (leaves)**

**Flowers/Flora**

**Fruits**

DNA Images Seed

**Notes/Local names**

Live coll. Alcohol Not pressed

Field det. Unicate

**Collector**

<table>
<thead>
<tr>
<th>Date</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SAJ</td>
</tr>
</tbody>
</table>

**Collection team**

<table>
<thead>
<tr>
<th>Date</th>
<th>20</th>
</tr>
</thead>
</table>
**Locality** (incl. state/region)

<table>
<thead>
<tr>
<th>Lat.</th>
<th>N/S</th>
<th>Long.</th>
<th>E/W</th>
<th>Error</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Habitat** (substrate/host/assoc. species)

- GPS
  - WGS84 (=GDA94)
  - Other

**Habit** (bark, wood)

**Habit** (leaves)

**Flowers/Sp</br

<table>
<thead>
<tr>
<th>DNA</th>
<th>Images</th>
<th>Seed</th>
<th>Live coll.</th>
<th>Alcohol</th>
<th>Not pressed</th>
<th>Unicoll</th>
</tr>
</thead>
</table>

**Date** 20

**No. SAJ**
The Scenario

Botanical collecting expeditions
> 250 flowering and fruiting taxa/specimens

3+ voucher specimens
1 tissue sample → products
Alcohol collection fruits, flowers
10+ images per collection (RAW + JPG)
Living collection
Notebook entry
Collection Tracking!

- Field number
- Expedition metadata
- Collection specimen numbers
- Tissue specimen number
- Field images, digitized herbarium vouchers
- Genbank and other downstream identifiers
- Other duplicate specimen numbering systems
- etc.
QR codes and collections data

<table>
<thead>
<tr>
<th>dwc:recordID</th>
<th>dwc:family</th>
<th>dwc:scientificName</th>
<th>dwc:recordID</th>
<th>dwc:family</th>
<th>dwc:scientificName</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAJ1765</td>
<td>Rutaceae</td>
<td>Murraya</td>
<td>81996B92C3264AF7</td>
<td>James, S.A.</td>
<td></td>
</tr>
<tr>
<td>SAJ1767</td>
<td>Rubiaceae</td>
<td>Ophichniza</td>
<td>B7D3959BH1209BA</td>
<td>James, S.A.</td>
<td></td>
</tr>
<tr>
<td>SAJ1768</td>
<td>Urticaceae</td>
<td>Urtica</td>
<td>F2F1F688C56726B8</td>
<td>James, S.A.</td>
<td></td>
</tr>
<tr>
<td>SAJ1769</td>
<td>Lamiaceae</td>
<td>Plectranthus scutellarioides</td>
<td>2685C18B4934D2C2</td>
<td>James, S.A.</td>
<td></td>
</tr>
<tr>
<td>SAJ1770</td>
<td>Gesneriaceae</td>
<td>Cyrtandra cf. filibriz</td>
<td>32D953C14883486B</td>
<td>James, S.A.</td>
<td></td>
</tr>
<tr>
<td>SAJ1771</td>
<td>Cunoniaceae</td>
<td>A4AD17D88384281</td>
<td>James, S.A.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAJ1772</td>
<td>Gentianaceae</td>
<td>Neurontis corymbosa</td>
<td>34CD02D7A34AE</td>
<td>Channel, S.</td>
<td></td>
</tr>
<tr>
<td>SAJ1773</td>
<td>Piperaceae</td>
<td>Piper</td>
<td>A1A4A096494A4B01</td>
<td>James, S.A.</td>
<td></td>
</tr>
<tr>
<td>SAJ1774</td>
<td>Orchidaceae</td>
<td>Bulbophyllum</td>
<td>599E57C1EDF1DF9F</td>
<td>Channel, S.</td>
<td></td>
</tr>
<tr>
<td>SAJ1775</td>
<td>Asteraceae</td>
<td>Erigeron</td>
<td>C8A37B166A6324F2D</td>
<td>James, S.A.</td>
<td></td>
</tr>
<tr>
<td>SAJ1776</td>
<td>Sapindaceae</td>
<td>Harpullia</td>
<td>7C9C55B0015449C</td>
<td>James, S.A.</td>
<td></td>
</tr>
<tr>
<td>SAJ1777</td>
<td>Cunoniaceae</td>
<td>Spreantherum</td>
<td>52B1372D4F04D05E</td>
<td>James, S.A.</td>
<td></td>
</tr>
<tr>
<td>SAJ1778</td>
<td>Actinidiaceae</td>
<td>Saurasia</td>
<td>8BC6A336C8364AC1</td>
<td>Channel, S.</td>
<td></td>
</tr>
<tr>
<td>SAJ1779</td>
<td>Melastomataceae</td>
<td>Medinilla</td>
<td>8BB85A6C9149C6</td>
<td>Channel, S.</td>
<td></td>
</tr>
</tbody>
</table>

916E0E076A344FA7BEC4084443F3CC66
dwc:eventID
Other useful tips

- Keep GPS track; Geotag images
- Photograph tag number before each specimen
- Don’t delay on transcribing/cleaning/integrating data!
More tricks and tips!

• Mobile technology and apps
• Tools for cleaning and standardizing data
  • Data collection apps
  • Georeferencing tools
  • Photo geotagging
  • Measurement tools (height, direction, distance)
  • Audio collection tools
  • OpenRefine
  • Google Earth, other visualization software
  • Taxonomic name etc. validation services
More tricks and tips!

• Mobile technology and apps
• Tools for cleaning and standardizing data
• Field information management system
• Develop a sustainable workflow
• Publish using identifiers!
• Reach out to iDigBio for information:
  – Field to Database Wiki
  – Glossary of Terms
Expedition identifiers & metadata
Excel spreadsheet generation
Controlled vocabulary
GUID generation
Data validation
DwC
More tricks and tips!

- Mobile technology and apps
- Tools for cleaning and standardizing data
- Field information management system
- Develop a sustainable workflow
- Publish & archive data using appropriate identifiers!
More tricks and tips!

• Mobile technology and apps
• Tools for cleaning and standardizing data
• Field information management system
• Develop a sustainable workflow
• Publish & archive data using identifiers!

• Reach out to iDigBio for information:
  – Field to Database Wiki
  – Glossary of Terms

www.iDigBio.org
Using biodiversity data to plan expeditions

- iDigBio
- GBIF
- Other biodiversity data sources
- Physical specimens
- Previous survey documentation
- Literature
www.iDigBio.org

I Dig Bio do you?

Shelley James, sjames@flmnh.ufl.edu