Using Darwin Core to make tracking data discoverable via the Atlas of Living Australia

Peggy Newman
Software Developer, ZoaTrack / Atlas of Living Australia
Biodiversity e-Infrastructure
- Australian region (broad)
- Aggregator - not curator or collector
- Open data for re-use
- Australian node of GBIF
- observations & specimens

**Contributors:**
museums
herbaria
research institutions
government
citizen science

**Users:**
science research
ecology research
education
environmental assessment
conservation management/planning

Where does a species occur?
What species occur in an area?
ALA and GBIF facilitate other countries to develop platforms using ALA e-infrastructure.

More are being implemented or are under discussion.
ZoaTrack: separate application supported by ALA specifically for animal movement data

- analysis tools & visualisations (e.g., home range estimation)
- cleansing/filtering tools
- data embargoes to support a research data life cycle
- automated sensor feeds (e.g., Argos, Globestar)
Why share data between ZoaTrack and ALA?

- **Discoverability**
  - ALA is the go-to for Australia’s biodiversity records
  - ALA is linked to GBIF and other Living Atlas platforms
- Tracking data is valuable, good quality, well described data that should be represented in the biodiversity record

How to share between such different databases?

- Multiple observations per individual
- Non standard data model
- Occurrence based datasets
- Simple Darwin Core
• ALA mainly uses Darwin Core - ingests flat files

• **Aggregate** each track into a single comprehensive record

<table>
<thead>
<tr>
<th>Darwin Core Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>basisOfRecord</td>
<td>MachineObservation</td>
</tr>
<tr>
<td>eventId</td>
<td>Identifier for the track (ie. sensor deployment on an animal). Foreign key to source system.</td>
</tr>
<tr>
<td>eventDate</td>
<td><strong>Date range</strong> for first to last detection</td>
</tr>
<tr>
<td>eventRemarks</td>
<td>Sensor description</td>
</tr>
<tr>
<td>footprintWkt</td>
<td><strong>Polygon</strong> representing track (home range, convex hull, alpha hull)</td>
</tr>
<tr>
<td>collectionId</td>
<td>URL link back to source system</td>
</tr>
<tr>
<td>decimalLatitude</td>
<td>Nominal occurrence (first detection)</td>
</tr>
<tr>
<td>decimalLongitude</td>
<td>Nominal occurrence (first detection)</td>
</tr>
<tr>
<td>measurement fields</td>
<td>Count of detections</td>
</tr>
</tbody>
</table>
Key Issues

• Discoverability – exposure - consistency
• General repository versus specialized
• Detailed Darwin Core record easily answers the two key questions
  • Where does this species/group occur?
  • What species occur in this area
• The difference between single occurrence records and tracks is made obvious to the users up front (no confounding)
• Ease of implementation
Next Steps

• TDWG involvement

• Upcoming Activities:
  • Darwin Core Hour discussion on Bio-logging and Camera Trap Data
  • TDWG 2018 Conference – workshop/symposium session on applying Darwin Core to Bio-logging data
WRAM & Swedish LifeWatch

**Wireless Remote Animal Monitoring (WRAM)**
The National Swedish Biotelemetry e-Infrastructure for sensor data from animals

**Swedish LifeWatch (SLW)**
The Swedish e-Infrastructure for Biodiversity & Ecosystem Research

**Task: WRAM delivers aggregated position data to SLW**

- WRAM currently contains e.g., 18.3M GPS records for 1,248 individual moose

- After aggregation 24,747 monthly ‘occurrence records’ with estimated position error can be delivered
Aggregating data

WRAM: individual moose movement track for each month
Aggregating data

Minimum Convex Polygone (MCP)
Aggregating data

Center of Gravity & radius of circle with MCP area = SLW occurrence data & approximate location error
Swedish LifeWatch - WRAM

The Analysis portal for biodiversity data

### Species observations provided by Swedish LifeWatch

<table>
<thead>
<tr>
<th>Data provider</th>
<th>Number of observations</th>
<th>Number of public observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species Observations System (A/portal). (Swedish Species Information Centre (A/Idatabanken))</td>
<td>54,354,163</td>
<td>50,300,330</td>
</tr>
<tr>
<td>Observation database of Radialsted species (Swedish Species Information Centre (A/Idatabanken))</td>
<td>725,905</td>
<td>0</td>
</tr>
<tr>
<td>DINA (The Swedish Museum of Natural History)</td>
<td>20,467</td>
<td>20,467</td>
</tr>
<tr>
<td>MVM (Environmental data MVM. SLU)</td>
<td>987,924</td>
<td>987,516</td>
</tr>
<tr>
<td>The National Register of Survey test-fishing (NORS) (Department of Aquatic Resources. SLU)</td>
<td>2,675,474</td>
<td>2,675,474</td>
</tr>
<tr>
<td>The Database for electrofishing in streams (SEBEL) (Department of Aquatic Resources. SLU)</td>
<td>297,719</td>
<td>394,470</td>
</tr>
</tbody>
</table>

At least one data provider must be selected.
Bird tracking - GPS tracking of Lesser Black-backed Gulls and Herring Gulls breeding at the southern North Sea coast

- 150 animals, ~8.5M records
- Each position is a Simple Darwin Core Occurrence record
  - Machine Observations
  - published to GBIF using IPT
  - Repetition, but volumes are not considered a problem
INBO / GBIF Example: Lossless Data

Darwin Core Mappings

Sample 100 Records