

A New Zealand sea lion demographics database

POP 2011-01

Finlay Thompson
Richard Mansfield



CSP TWG - 13 November 2012

Introduction

A database has been developed to store all the New Zealand sea lion tagging and sightings data.

The database is publically accessible on a website.

Today I will report on how it was created and demonstrate how it can be used.

<http://data.dragonfly.co.nz/nzsl-demographics/>

Overview of data preparation

The source data consists of various spreadsheets provided by CSP.

They were prepared by Louise Chilvers, and collect all the sightings and tagging data from 1980 to 2012.

From spreadsheets an relational database was created.

Grooming rules were applied to the data as it was loaded.

Data model

There are three separate data structures:

- **Sea lions** are deduced from the data, and are assigned a unique identifier.
- **Tags** have colour, shape, and number. Other identifiers also exist: brands and chips.
- **Sightings** bring together the identifier (tag) with a particular sea lion.

Data model

The **sea lion** record holds a unique label

- `label` - a unique label to reference each sea lion. It has the form of a date and tag number: YYYYMMDD-TTTT

Data model

The **tag** records include the following fields:

- `tag` - the number printed on the tag
- `colour` - a code indicating the colour
- `shape` - Coffin or Round
- `first_tagged_date` - Date of the tagging event
- `first_tagged_season` - Season of tagging event
- `first_location` - Location of tagging event
- `adult` - Boolean indicating if the animal was tagged as an adult

Data model

The **sighting** information can be divided into identification information and observational information.

Identification information includes:

- `date` - date of sighting
- `location` - two levels, corresponding to beach and island.
- `nature` - code indicating type of sighting
- `tag` - tag color, shape, and number
- `brand` - If a brand was sighted
- `chip` - NC if not checked, or the chip number if checked
- `n_tags` - code indicating how many and which tags were seen
- `comments` - comments regarding the sighting

Data model

The **sighting**'s observational information includes:

- `sex` - male or female
- `age_class` - a code indicating the age of animal
- `status` - code indicating current demographic status
- `pup` - sea lion label of associated pup
- `behaviour_comments` - comments regarding the animals behaviour

Grooming rules

There are 96 640 sightings records, 18 459 tags, and 18 295 sea lions in the groomed data.

30 grooming rules were applied, with most relating to standardising the various data fields.

- Location information - 96 505, eg: 'E' to 'ENDERBY'
- Behaviour comments - 3303 records, mostly removing initial white space
- pup tag colour - 2833 recrds, eg: 'Blue03' to 'B03'

Grooming rules

More interesting grooming rules were applied to 279 sightings records

These updates can be seen associated to the effected records.

For example:

- 47 'tagging' events were updated because there was an earlier 'tagging' event in the database for that tag.
- 12 records were updated because the tag shape was unlikely.

Grooming rules

422 sighting records had identified errors but were not corrected. They are indicated in the database interface.

These sightings will require further investigation.

The errors include 33 chip ids that were reported on more than one sea lion.

<http://data.dragonfly.co.nz/nzsl-demographics/explore/sighting?q=chip:00-0210-A06E>

Searching and filtering

We have used a filtering interface for exploring the data.

The following filters are currently defined:

- After and before a date
- sealion label
- tag, including with wild cards
- chip number
- location
- island
- the event type

More filters types can be added as required.

Downloading as CSV

The results of each data search can be downloaded as a CSV file.

Only 100 sightings are displayed on the screen, but all records are included in the CSV file.

With no filters applied, the download will include all 96 640 records.

Next steps

Editing records, to allow the fixing of outstanding data problems. The editing will be completely tracked.

Streamlined data entry of new data. The data entry will capture the observations from notebooks faithfully, then allow corrections to be made where necessary to reconcile with the existing data.