**LHI Charter Fishery - Database Quality Control**

**File**: Logbook\_DAT\_2015.mdb

Below are notes on initial & brief quality control assessment of LHI Logbook data. I focused on data tables for (1) Fishing Activity and (2) Catch. Initial focus was on missing values and potential outliers. Other issues may be identified in future during more detailed analysis.

We need to make decisions regarding what records to exclude from analysis. Some records have count data but no length data for key reporting species. Other discrepancies are probably data recording or entry errors and can probably be cleaned up quickly (e.g. high counts of very small sharks)

**Relevant Access Database Tables**

**1. Fishing Activity**

**Miscellaneous**

* 4425 FAC\_ID entries
* No FAC\_ID = 1 > unsure why not, removed?
* No activity for Vessels: (i) VSL\_ID **11**: no entry in Vessel table (ii) VSL\_ID: **15**, Barracuda II is present in Vessel table and not marked as Retired, therefore still active? Note: Barracuda II does not appear in 2012 Report. > unsure why no Vessel 11, Vessel table now updated to reflect currently active vs retired vessels

**Outliers**

* **No. of Hours**: 5 trips > 12 hrs : FAC\_ID: 595,596,597 (all Barracuda – VSL\_ID: 1, Note: 596 = 26hrs), 655 (15hrs) & 4390 (18hrs) [all pre-2013 except for 18hrs/4390] > one trip correct (overnight trip – written in FAC\_ID ‘notes’). All other trips >21hrs from vessels owned by same operator in 2005, where catch was entered for each month and hrs given as monthly total. Written in notes of each FAC\_ID and ‘DATA ISSUE’ checked.

**Missing Data** – No data entered (different to value 0 entered)

* **No. Hours**: 81 trips with no entry

VSL\_ID: *No of Trips* = [**3**:*11*, **6**:*28*, **12**:*7*, **13**:*23*, **14**:*2*, **16**:*4*, **18**:*2*, **19**:*1*, **21**:*3*]

[73 are pre-2013 report. 8 Appear to be 2013-2015 but no Date data entered] >hours not given in raw data. Written in notes of each FAC\_ID and ‘DATA ISSUE’ checked.

**Location**: 42 trips with no entry (34 of these trips also have no **No. Hours** data, and 6 of this set have no **Date** or **No. Hours**) [36 are pre-2013, 6 appear to be 2013-2015 but no date data entered] >location not given in raw data. Written in notes of each FAC\_ID and ‘DATA ISSUE’ checked.

* **Date**: 10 trips with no date entered (8 of which have no No. Hours entered either, and 6 of this set have no Location) [1 is pre-2013 as entered 2010, but other 9 are new (2013-2105)] > all records located and correct date entered
* **Date**: When is cut off for analysis? Most recent is 30/6/2016. Data has NOT been cut off at 2015. So 2016 is incomplete? **Discard 115 trips that have been entered for 2016**? – all 2016 trips have data entered for all fields. >all data now entered until the end of 2018

**2. Catch**

**Outliers**

**Length:** 6 CAT\_ID entries where Length = 0cm. [5 are pre-2013, yellowfin tuna record is 2013-2015] 3 of these were for Wahoo which is one of the species reported on in detail.

> lengths not recorded in raw data. Written in notes of each CAT\_ID and ‘DATA ISSUE’ checked. “0” deleted so length field blank.

* **Length**: 7 CAT\_ID entries for Galapagous whaler have length = 1cm. Notes say “No Length Given”. All these are for **Greenback** (VSL\_ID: 14) – see note below (this vessel suspected of underreporting shark bycatch in 2012 Report). All 7 CAT\_ID entries record a value for count, ranging from 2 – 8. So maybe Greenback is recording count but not length? Doesn’t give confidence that recording count accurately. [These actually pre-2013] >length value deleted so length field blank (given no lengths given in raw data). Written in notes of each CAT\_ID and ‘DATA ISSUE’ checked.
* **Length**: 2 CAT\_ID entries for Galapagous whaler have length = 2cm. No notes. Vessel on both occasion was **Fin** (VSL\_ID: 21). A new vessel since 2012 report. Haven’t checked to see how it has reported shark bycatch on other trips. [new, 2013-2015] >these records may have been fixed already? No longer present in database with SHK=2cm.
* **Length Frequency Distribution (By Species)**

Will need to check data entry for some small and large values recorded for key species.

*Galapagous whaler*

In previous report I think smallest reported size of galapagous whaler ~ 70cm. In current dataset there are **1,751** individuals with size listed **< 60cm**.

* From above 7 CAT\_ID entries for SHK of length = 1cm (total count = 38 sharks)[pre-2013] > see above
* From above 2 CAT\_ID entries for SHK of length = 2cm (total count = 3 sharks)[new 2013-2015] > see above
* 16 CAT\_ID entries for SHK length ranging from 10 – 30cm. Note. The count recorded for all these 16 CAT\_ID entries is high and ranges from 90 – 130. Is it possible that entries for Count were erroneously entered in Length field and vice versa (either in logbook or computer data entry)? ALL of 16 CAT\_ID entries are attributed to VSL\_ID: 16 (**Carina**) and are from trips that appear to have been entered close together (within the range FAC\_IDs: 4301 – 4340). These entries represent 1,710 sharks. To put it in context – the next highest count in CAT\_ID records is for SCAD (horse mackerel) at 72. [These are all new BUT some are 2016 records and therefore likely to be excluded – but should be fixed up for subsequent analyses] >re-entered correctly

*Kingfish*

95 individual kingfish have no length entered (see CAT\_ID details below in “Missing Data”). Otherwise the range of sizes is similar to 2012. > lengths not recorded in raw data. Written in notes of each CAT\_ID and ‘DATA ISSUE’ checked.

*Silver Trevally*

Length distribution similar to 2012, although there seem to be a few records of large trevally (>120cm): 2 of 125cm (CAT\_ID 1710) >unable to locate raw data for this entry, 1 at 180cm (CAT\_ID 15479) >incorrect, re-entered as 80cm, 1 at 155cm (CAT\_ID 17777), 1 at 150cm (CAT\_ID 21423) >correct as per raw data,1 at 140cm (CAT\_ID 18720) >unable to locate raw data for this entry. And 3 records of large silver trevally from the same trip (FAC\_ID 3929) of 120cm, 123cm & 125cm.> correct as per raw data [Mix of pre and post-2013...Note, just want to check that these large values are real] some are some aren’t as above

*Wahoo*

3 CAT\_ID (16641, 16698, 16750) where wahoo length is entered as 0 (Notes: no lengths recorded). Counts were recorded in each CAT\_ID and total = 20 individuals. Again, VSL\_ID 14 (**Greenback**) responsible for all. [pre-2013] no lengths given, written as ‘released’ due to being ‘surplus to requirements’

*Eastern Gar*

Length distribution similar range to 2012

*Doubleheader*

Length distribution similar range to 2012

*Bluefish*

Length distribution similar range to 2012

*Black Cod*

Length distribution similar range to 2012

* **Count**: 3 CAT\_ID entries for species that contained Count = 0.

CAT\_ID: 18291 / FAC\_ID: 3257. Unknown spp, count=0, length=0. Note says taken by shark (prob before landed) [Old, pre-2013] >correct as per raw data

CAT\_ID: 25768 / FAC\_ID: 4083. Yellowfin Tuna – it was retained but count=0, length = 0 [new, 2013-2015] >no YFT in raw data, entry corrected

CAT\_ID: 27056 / FAC\_ID: 4188. Shark. Not retained. Length=100 but count = 0 [new, 2013-2015] >raw data notes “no shark interaction” and SHK number = 0. Entry corrected (length removed).

* **Count.** Distribution of CAT\_ID count data. Most distributions show expected exponential decay with recorded frequency decreasing with count (catch) size. Most obvious exception is for **Galapagous whalers** – see above for detail on 16 CAT\_ID with outlier counts (they seem too high).

**Missing Data**

**Length**: 17 CAT\_ID entries without length data. > lengths not recorded in raw data. Written in notes of each CAT\_ID and ‘DATA ISSUE’ checked.

*Kingfish*

CAT\_ID: 18631 / FAC\_ID: 3200. Count = 28 and all Not Kept (all undersize or C&R vessel?) [pre-2013]

CAT\_ID:26501 / FAC\_ID: 4134. Count = 67. All kept (theoretically all legal size?) [new 2013-2015] >

*Galapagous Whaler*

12 CAT\_ID galapagous whaler entries without length data (all trips recorded catch & length for other spp, just not sharks). Corresponds to 12 separate trips (FAC\_IDs) and all by same vessel: **Greenback** (VSL\_ID: 14). Note: this was one of the vessels that reported virtually no galapagous whaler bycatch in previous report whilst still reporting relatively high kingfish catch (2012) and was therefore suspected of under-reporting shark bycatch. Count of galapagous whalers was reported for these 12 trips. [pre-2103, all late in 2012]

*Unknown*

2 CAT\_IDs of unknown fish without length data. Unknown fish must have looked appealing as they are marked as retained.

CAT\_ID: 22426 / FAC\_ID: 3695. Count = 1. Retained. [new, 2013-2015] >unknown species kept as specimen for identification

CAT\_ID: 4704 / FAC\_ID: 778. Count = **62**. Retained. 4 Kingfish were identified from trip (FAC\_ID 778), length was recorded and all were retained. [pre-2013, old 2005 record]

* **Count**: There were no CAT\_ID entries with missing count data.
* **Species**: No catch data missing Species entry and all abbreviations in catch seem correct