Database documentation: fish_ce

K. A. Mackay

NIWA Fisheries Data Management
Database Document Series

Updated May 2007

Contents

List of Figures

Version Control

Version	Status	Changed By	Reason	Date
1.0	release	Kevin Mackay	First release version	Sep 2001
1.1	release	Fred Wei	Added t_meta table to track datasets and some attributes in tables.	Jan 2004
1.2	release	Fred Wei	Added efflength in table t_fishing_event.	8 May 2007

1 Introduction to the Database Document series

The National Institute of Water and Atmospheric Research (NIWA) currently carries out the role of Data Manager and Custodian for the fisheries research data owned by the Ministry of Fisheries (MFish).

The Ministry of Fisheries data set incorporates historic research data, data collected more recently by MAF Fisheries prior to the split in 1995 of policy to the Ministry of Fisheries and research to NIWA, and currently data collected by NIWA and other agencies for the Ministry of Fisheries.

This document is a brief introduction to the New Zealand Catch Effort (Research interpretation) database **fish_ce**, and is part of the database documentation series produced by NIWA.

All documents in this series include an introduction to the database design, a description of the main data structures accompanied by an Entity Relationship Diagram (ERD), and a listing of all the main tables. The ERD graphically shows how all the tables link together and their relationship with other databases.

This document is intended as a guide for users and administrators of the **fish_ce** database.

Access to this database is restricted to specific nominated personnel as specified in the current Schedule 6 of the Data Management contract between the Ministry of Fisheries and NIWA. Any requests for data should in the first instance be directed to the Ministry of Fisheries.

2 New Zealand Catch Effort data

The New Zealand Catch Effort system stores catch, effort, landings, production, and environment information provided to the Ministry of Fisheries by commercial fishers. Of these data:

- Catch data are rough estimates of the catch (kg of each species) made by fishers as they fish.
- Effort data summarise the amount of effort that a fisher/vessel put into catching fish, specify what method was used and what species was targeted.
- Landings data summarise either the actual quantity of fish landed at a wharf (or transferred to another vessel at sea). Landings data are considered more accurate than estimated catch data.
- Production data summarise the estimated quantity of fish processed onboard a vessel during a day. This is more accurate than estimated catch but less accurate than actual landings.
- Environment data summarise the depth of the sea in which a vessel was fiashing, and the sea and weather conditions at the time of fishing.

The information received from fishers is recorded on one of five forms:

1. **CELR** – Catch Effort Landing Return. Records estimated catch, effort and actual landings for approximately 30 different fishing methods. This form is very generic. Fishers superimpose one of seven cardboard templates over the form to tell them what information is entered in each field. Fishers that fill in a CELR do not fill in any other type of form. One form is used for each trip.

2. **TCEPR** – Trawl Catch Effort Processing Return. Records estimated catch, effort, processing and environment data for deep-sea trawlers. No landing data are recorded so fishers must also fill in a CLR. One form is used for each days fishing.
3. **TLCER** – Tuna Longlining catch Effort Return. Records estimated catch, effort, processing and environment data for surface longliners targeting tuna. No landing data are recorded so fishers must also fill in a CLR. One form is used for each days fishing.
4. **SJCER** – Squid Jigging Catch Effort Return. Records estimated catch, effort, processing and environment data for squid jiggers. No landing data are recorded so fishers must also fill in a CLR. One form is used for each days fishing.
5. **CLR** – Catch Landing Return. Records actual landings for a vessel. Only filled in if a fisher also filled in TCEPR, TLCER, or SJCER forms. One form is used for each trip.

From these five forms, the Ministry of Fisheries populates it's **catcheff** database, and from this selected users can access these data via the **warehou** database (essentially a filtered copy of **catcheff**).

Fisheries data supplied to Research Providers are extracted from **warehou** on an ad-hoc basis in the form of discrete extracts of data, tailored to each Research Providers need. This generates multiple copies of Catch Effort data to which the individual Research Providers can edit and modify to suit their various research objectives.

The **fish_ce** database is designed to store these different data extracts in a central, managed RDBMS, as well as to provide the ability to pre-emptively hold data from a variety of fisheries that Research Provider may need to access in future times.

3 Data Structures

3.1 Table relationships

This database contains several tables. The ERD for **fish_ce** (Figure 1) shows the logical structure[1] of the database and it's entities (each entity is implemented as a database *table)* and relationships between these tables and tables in other databases. This schema is valid regardless of the database system chosen, and it can remain correct even if the Database Management System (DBMS) is changed. Each table represents an object, event, or concept in the real world that is selected to be represented in the database. Each *attribute* of a table is a defining property or quality of the table. All of the table's attributes are shown in the ERD. The underlined attributes represent the table's primary key[2].

Note that Figure 1 shows the main tables only. Note that most tables contain foreign keys[3]. These foreign keys define the relationships between the tables in **fish_ce**.

[1] Also known as a database *schema.*
[2] A primary key is an attribute or a combination of attributes that contains an unique value to identify that record.
[3] A foreign key is an attribute or a combination of attributes that is a primary key in another table.

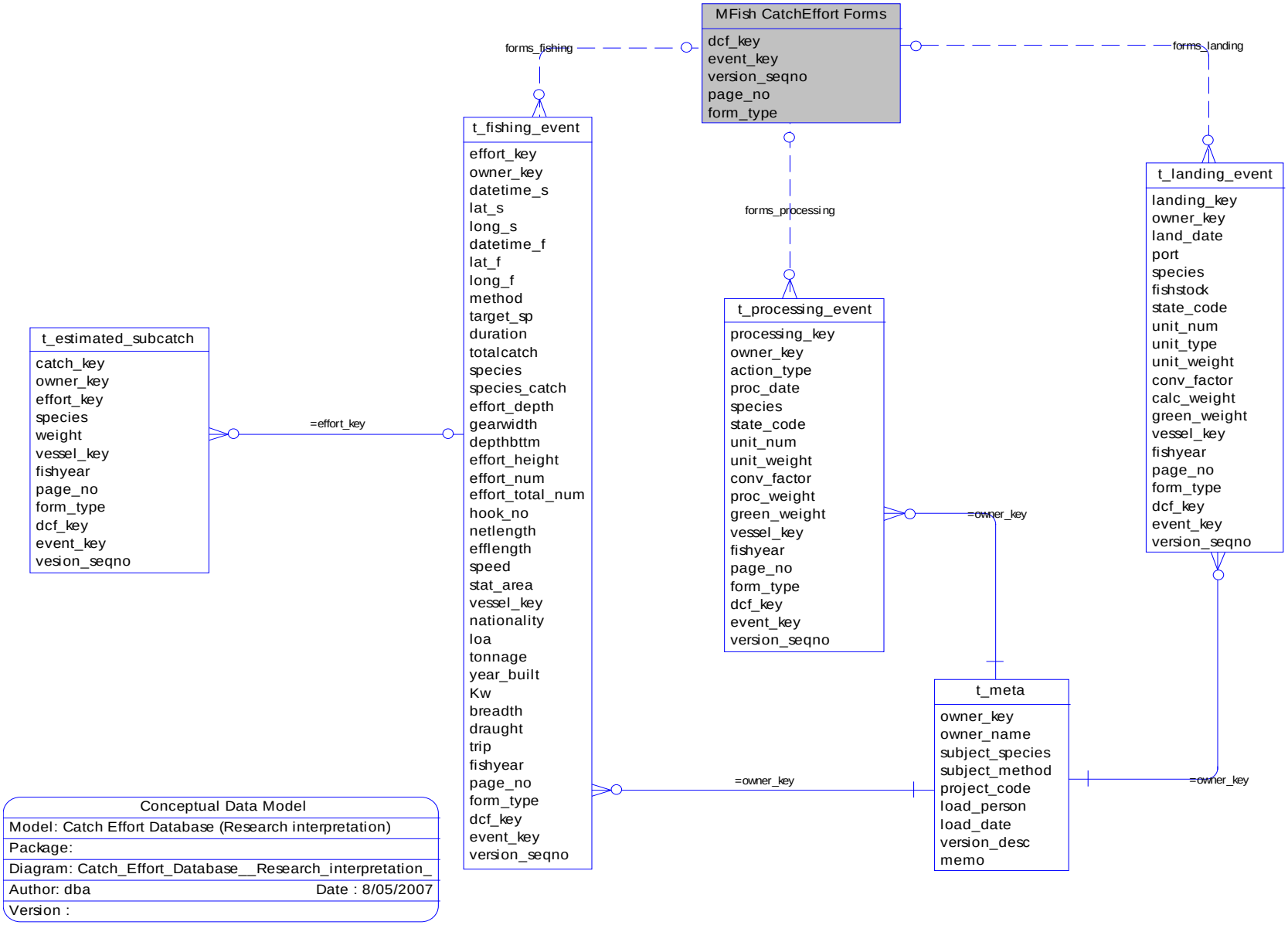

Figure 1: Entity Relationship Diagram (ERD) for the fish_ce database

The **fish_ce** database is implemented as a relational database; i.e., each table is a special case of the mathematical construct known as a *relation* and hence elementary relation theory is used to deal with the data within tables and the relationships between them. There are three types of relationships possible between tables, but only one exists in **fish_ce**: one-to-many[4]. These relationships can be seen in ERDs by connecting a single line (indicating "many") from the child table; e.g., *t_estimated_subcatch*, to the parent table; e.g., *t_fishing_event*, with an arrowhead (indicating "one") pointing to the parent.

Every relationship has a mandatory or optional aspect to it. If a relationship is mandatory, then it has to occur at least once, while an optional relationship might not occur at all. For example, in Figure 1, consider that relationship between the table *t_fishing_event* and it's child table *t_estimated_subcatch.* The symbol "O" by the child *t_estimated_subcatch* means that a fishing event can have zero or many catch records, while the bar by the parent *t_fishing_event* means that for every catch there must be a matching fishing event record.

These links are enforced by referential constraints[5]. Constraints do not allow *orphans* to exist in any table; i.e., where a child record exists without a related parent record. This may happen when: a parent record is deleted; the parent record is altered so the relationship is lost; or a child record is entered without a parent record.

Constraints are shown in the table listings by the following format:

```
Referential:      constraint name (attribute[, attribute])  |INSERT|
                                                            |DELETE|
                  parent table (attribute[, attribute])
```

Note that the typographical convention for the above format is that square brackets "[ ]" may contain more than one item or none at all. Items stacked between vertical lines | | are options of which one must be chosen.

For example, consider the following constraint found in the table *t_estimated_subcatch*:

```
Referential:      Invalid fishing event (effort_key) INSERT
                  t_fishing_event (effort_key)
```

This means that the value of the combination of attributes *effort_key* in the current record must already exist in the parent table *t_fishing_event* or the record will be rejected and the following message will be displayed:

```
*** User Error: insert constraint "Invalid fishing event" violation
```

Section 5 lists all the **fish_ce** tables as implemented by the Empress RDBMS. As can be seen in the listing of the tables, a table's primary key has an unique index on it. Primary keys are generally listed using the following format:

```
Indices:          UNIQUE index_name ON (attribute[, attribute])
```

[4] A one-to-many relationship is where one record (the *parent*) in a table relates to one or many records (the *child*) in another table; e.g., one fishing event in *t_fishing_event* can have many catches in *t_estimated_subcatch* but one catch can only come from one fishing event.
[5] Also known as integrity checks.

where attribute(s) make up the primary key and the index name is the primary key name. These prevent records with duplicate keys from being inserted into the tables; e.g., a record with an existing event key.

The database listing (Tables 1-5) show that the tables also have indices on many attributes. That is, attributes that are most likely to be used as a searching key have like values linked together to speed up searches. These indices are listed using the following format:

Indices: NORMAL (2, 15) index_name ON (*attribute*[, *attribute*])

Note that indices may be simple, pointing to one attribute or composite pointing to more than one attribute. The numbers "…(2, 15)…" in the syntax are Empress RDBMS default values relating to the amount of space allocated for the index.

3.2 Database design

The **fish_ce** database is built around the premise that all fishing trips are based on a series of events. Where an event is a specific occurrence at a particular position on earth and at a certain time to a vessel or fisher. The MFish Catch Effort system recognises four types of events: a fishing event (when a trawl, pot set, longline set, jig, etc. is made); a processing event (when an amount of fish is processed over a certain time period); an environmental event (a weather or sea condition measurement); and a trip event. These three event type are specialisations of the generalise entity "events". This can be modelled as a "GENSPEC" structure. Generalisation and specialisation are pictured in Figure 2 using a triangle containing the words "IS A" to connect the components to each other and to the higher-level entity. Each event, regardless of event type, is identified by an *event_key* attribute.

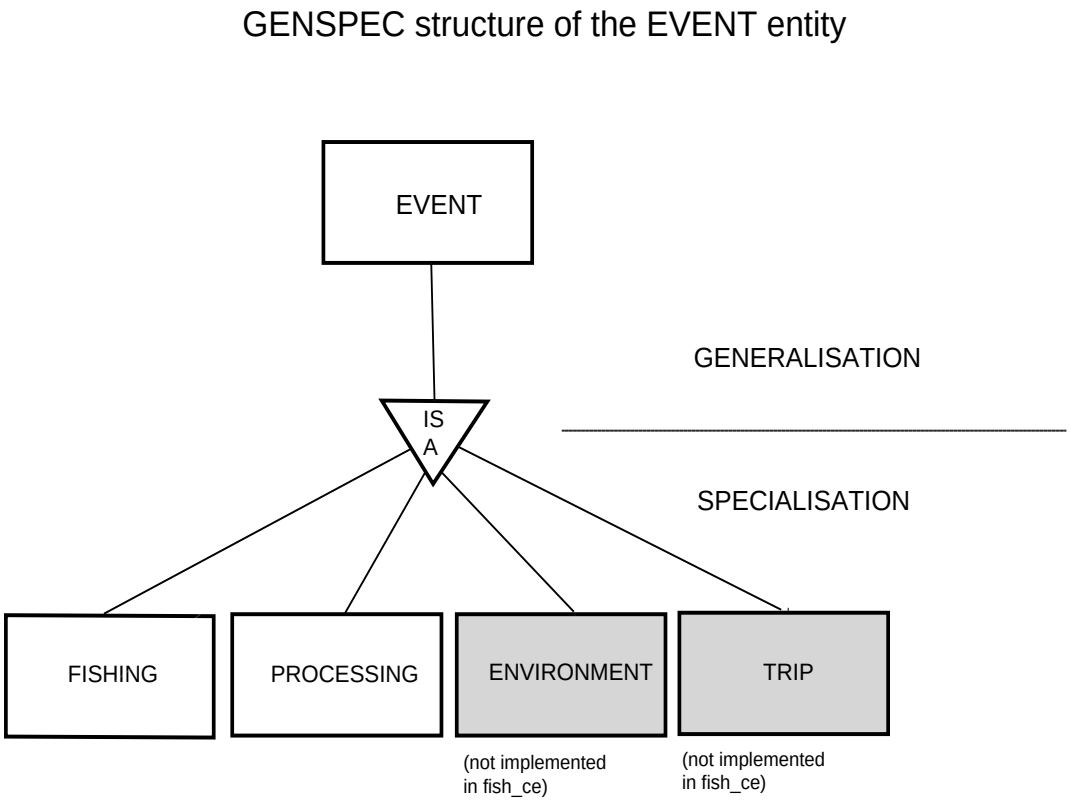

Figure 2: GENSPEC diagram for the "events" entity

In the Mfish **warehou** database, this GENSPEC structure is implemented as tables: one each for event, fishing event, environmental event, and trip event; fish processing events are further broken down into four separate tables for processed catch, landings, tuna individual catch, and squid tallies.

Extracts for data for Research Providers generally require the merging of data from both the generalisation and specialisation components of **warehou**. The tables in **fish_ce**, therefore, reflect this data merging and only represent the most common events required by Research Providers: the fishing event and the fish processing event (both fish processing at-sea and landings).

Full documentation of the **warehou** database is given in Duckworth (1997).

Each record in all **fish_ce** tables comes from a paper form. It is very important for Research Providers to be able to link different events from the same form together. Each form has it's own page number, but this is not unique between different form types. MFish have created the attribute *dcf_key* to store a system generated number to uniquely identify each form. This attribute, combined with *version_seqno* is used to link different events on the same form.

The first table is *t_fishing_event* (Table 1). This is equivalent to linking the **warehou** tables *event*, *fishing_event* tables (by the attributes *event_key* and *version_seqno*), and the vessel table (by the attribute *vessel_key*). Details stored include: dates, times, and location of the fishing event, the fishing method used; the vessel key; and various physical parameters about the gear used (e.g., trawl width, no of hooks used, length of set net, etc.). Because *t_fishing_event* has to handle all possible fishing type, the meaning of the various effort related attributes changes depending on the form type and fishing method used. Definitions for the effort related attribute are given in Appendix 1 for each form type and fishing method.

Details of the catch by species, as estimated by the fisher, are stored in the *t_estimated_subcatch* table (Table 2). Details include species code an estimate weight.

Fish processing at-sea details are stored in the table *t_processing_event* (Table 3). This is generally the daily processing summary in the TCEPR form, but also includes tuna counts and squid catches. Details include species code, processed state, number of units of processed fish, and the weight of each unit of processed fish.

Landing details are stored in the table *t_landing_event* (Table 4). These details come from the CLR and CELR forms. Details include species code, landed state, and landed weight.

4 Table Summaries

The following is a listing of the tables contained in the **fish_ce** database:

1. **t_fishing_event :** contains temporal and spatial data about the effort and overall catch for each fishing event.
2. **t_estimated_subcatch:** contains estimated green weights for each species caught.
3. **t_processing_event :** contains daily processing details of each species. Contains processing data and calculated green weight
4. **t_landing_event :** contains details about a landing or transhipment by species and processed state.
5. **t_meta :** contains dataset ownership information.

5 fish_ce Tables

The following are listings of the tables in the **fish_ce** database, including attribute names, data types (and any range restrictions), and comments.

5.1 Table 1: t_fishing_events

Comment: Fishing Event details.

Attributes	Data Type	Null?	Comment
effort_key	longinteger	No	Unique number to identify the record
owner_key	longinteger	No	Foreign key to reference t_meta table
datetime_s	time(0)		Start fishing date and time
lat_s	decimal(4,1)		Decimalised latitude of start of fishing truncated to 1/10th of a degree
long_s	decimal(4,1)		Decimalised longitude of start of fishing truncated to 1/10th of a degree
datetime_f	time(0)		Finish fishing date and time
lat_f	decimal(4,1)		Decimalised latitude of finish of fishing truncated to 1/10th of a degree
long_f	decimal(4,1)		Decimalised longitude of finish of fishing truncated to 1/10th of a degree
method	character(3,1)		Primary fishing method code
target_sp	character(3,1)		Target species code
duration	decimal(4,1)		Duration of fishing event (usage varies with form type)
totalcatch	longinteger		Total weight (kg) of catch for this fishing event as estimated at the time
species	character(3,1)		Species code for subject species
species_catch	decimal(8,2)		Total catch weight(kg) for the subject species
effort_depth	longinteger		Depth (m) of effort (ground rope in TCEPR forms only)
gearwidth	decimal(6,2)		Gear width (m)

Attributes	Data Type	Null?	Comment
depthbttm	longinteger		Depth (m) of sea bottom
effort_height	decimal(5,1)		Effort_height (usage varies with fishing type)
effort_num	longinteger		Effort number (usage varies with fishing type)
effort_total_num	longinteger		Effort total number (usage varies with fishing type)
hook_no	longinteger		Hook number (may be total hooks hauled per day OR maximum number of hooks used at any one time).
netlength	longinteger		Total length of nets hauled that day(m)
efflength	longinteger		Length of line in metres and is derived by multiplying the number of hooks by the hook spacing
speed	decimal(5,2)		Vessel speed (knots) during effort
stat_area	character(5,1)		Statistical area in which fishing started
vessel_key	longinteger		MFish generated number identifying the vessel fishing
nationality	character(5,1)		Vessel flag nationality
loa	decimal(6,3)		Vessel length (m) overall
tonnage	decimal(12,6)		Vessel tonnage
year_built	integer		Year vessel built
Kw	decimal(7,3)		Vessel engine power (Kw)
breadth	decimal(6,3)		Vessel breadth (m)
draught	decimal(6,3)		Vessel draught (m)
trip	longinteger		A system generated number allocated to each of the events that took place for one vessel between its trip start and end dates
fishyear	character(7,1) match “19??/??|20??/??”		Formatted fishing year (e.g., 1 Oct 1996 to 30 Sep 1997 = 1996/97)
page_no	longinteger		The ID number printed on each form. Not unique (like dcf_key) because there are 5 types of forms

Attributes	Data Type	Null?	Comment
form_type	character(3,1)		MFish form type code: CEL=CELR; TCP=TCEPR; CLR=CLR; SJC=SJCER; TUN=TLCER match “CEL|TCP|CLR|SJC|TUN”
dcf_key	longinteger		System generated number identifying a single CELR, TCEPR, CLR, SJCER, or TLCER form
event_key	longinteger		Unique fishing event number
version_seqno	integer		Version number of fishing event

Creator: dba
Referential: Invalid dataset owner (owner_key) INSERT t_meta (owner_key)
Indices: UNIQUE BTREE ON (effort_key)
NORMAL (2, 15) BTREE ON (vessel_key)UNIQUE BTREE

5.2 Table 2: t_estimated_subcatch

Comment: Estimated subcatch details by species.

Attributes	Data Type	Null?	Comment
catch_key	longinteger	No	Unique number to identify the record
effort_key	longinteger		Foreign key to reference t_fishing_event table
owner_key	longinteger	No	Foreign key to reference t_meta table
species	character(3,1)		Three letter code identifying the species caught
weight	longinteger		Estimated weight (kg) caught of the species
vessel_key	longinteger		MFish generated number identifying the vessel fishing
fishyear	character(7,1) match “19??/??|20??/??”		Formatted fishing year (e.g., 1 Oct 1996 to 30 Sep 1997 = 1996/97)
page_no	longinteger		The ID number printed on each form. Not unique (like dcf_key) because there are 5 types of forms
form_type	character(3,1) smatch “CEL|TCP|CLR|SJC|TUN”		MFish form type code: CEL=CELR; TCP=TCEPR; CLR=CLR; SJC=SJCER; TUN=TLCER
dcf_key	longinteger		System generated number identifying a single CELR, TCEPR, CLR, SJCER, or TLCER form
event_key	longinteger	No	Unique fishing event number
version_seqno	integer	No	Version number of fishing event

Creator: dba
Referential: Invalid dataset owner (owner_key) INSERT t_meta (owner_key)
Invalid fishing event (effort_key) INSERT t_fishing_event (effort_key)
Indices: NORMAL (2, 15) BTREE ESTCATCH_SPECIES_NDX ON (species)
NORMAL (2, 15) BTREE ESTCATCH_FISHYEAR_NDX ON (fishyear)
UNIQUE BTREE ON (catch_key)

5.3 Table 3: t_processing_event

Comment: Fish processing event details.

Attributes	Data Type	Null?	Comment
processing_key	longinteger	No	Unique number to identify the record
owner_key	longinteger	No	Foreign key to reference t_meta table
action_type	character(3,1) smatch “PRO|OFF”		General nature of processing event: PRO=processing; OFF=offal production
proc_date	date(5)		The start date for processing
species	character(3,1)		Three letter code identifying the species caught
state_code	character(3,1)		Processed fish state code
unit_num	longinteger		Number of containers or litres of oil produced
unit_weight	decimal(15,6)		Average weight (kg) of each container
conv_factor	decimal(6,4)		Conversion factor
proc_weight	decimal(15,6)		Processed weight (kg) (processed weight X conversion factor = green weight)
green_weight	decimal(15,6)		Calculated green weight (kg) of the fish
vessel_key	longinteger		MFish generated number identifying the vessel fishing
fishyear	character(7,1) match “19??/??|20??/??”		Formatted fishing year (e.g., 1 Oct 1996 to 30 Sep 1997 = 1996/97)
page_no	longinteger		The ID number printed on each form. Not unique (like dcf_key) because there are 5 types of forms
form_type	character(3,1) smatch “CEL|TCP|CLR|SJC|TUN”		MFish form type code: CEL=CELR; TCP=TCEPR; CLR=CLR; SJC=SJCER; TUN=TLCER
dcf_key	longinteger		System generated number identifying a single CELR, TCEPR, CLR, SJCER, or TLCER form
event_key	longinteger	No	Unique fish processing event number
version_seqno	integer	No	Version number of fish processing event

Creator: dba

Referential: Invalid dataset owner (owner_key) INSERT t_meta (owner_key)

Indices: NORMAL (2, 15) BTREE PROCESSING_SPECIES_NDX ON (species)
NORMAL (2, 15) BTREE PROCESSING_FISHYEAR_NDX ON (fishyear)
UNIQUE BTREE PK_PROCESSING ON (processing_key)

5.4 Table 4: t_landing_event

Comment: Fish landing event details.

Attributes	Data Type	Null?	Comment
landing_key	longinteger	No	Unique number to identify the record
owner_key	longinteger	No	Foreign key to reference t_meta table
land_date	date(5)		The start date for the landing
port	character(40,1)		Port of landing or callsign of transhipment
species	character(3,1)		Three letter code identifying the species caught
fishstock	character(5,1)		Fishstock code
state_code	character(3,1)		Processed fish state code
unit_num	longinteger		Number of containers or litres of oil produced
unit_type	character(3,1)		Type of packaging; e.g., container, box, sack, single fish, etc.
unit_weight	decimal(15,6)		Average weight (kg) of each container
conv_factor	decimal(6,4)		Conversion factor
calc_weight	decimal(15,6)		Calculated weight (kg) ((unit weight X number of units) X conversion factor = green weight)
green_weight	decimal(15,6)		Calculated green weight (kg) of the fish
vessel_key	longinteger		MFish generated number identifying the vessel fishing
fishyear	character(7,1) match “19??/??|20??/??”		Formatted fishing year (e.g., 1 Oct 1996 to 30 Sep 1997 = 1996/97)
page_no	longinteger		The ID number printed on each form. Not unique (like dcf_key) because there are 5 types of forms

form_type	character(3,1) smatch “CEL|TCP|CLR|SJC|TUN”		MFish form type code: CEL=CELR; TCP=TCEPR; CLR=CLR; SJC=SJCER; TUN=TLCER
dcf_key	longinteger		System generated number identifying a single CELR, TCEPR, CLR, SJCER, or TLCER form
event_key	longinteger	No	Unique fish processing event number
version_seqno	integer	No	Version number of fish processing event

Creator: dba
Referential: Invalid dataset owner (owner_key) INSERT t_meta (owner_key)
Indices: NORMAL (2, 15) BTREE LANDING_SPECIES_NDX ON (species)
NORMAL (2, 15) BTREE LANDING_FISHYEAR_NDX ON (fishyear)
UNIQUE BTREE PK_LANDING_EVENT ON (landing_key)

6 References

Duckworth, K. 1997. WAREHOU Database Documentation Base Views and Fields (Adapted from CATCHEFF database documentation Part 2 – Base views and fields). Version 2.0. Ministry of Fisheries Report. 41 p.

7 Appendix 1 – Reference codes

Fishing Method Codes

PS	Purse Seining
DS	Danish Seining - Single
DPS	Danish Seining - Pair
L	Lampara
BS	Beach Seining/Drag Netting
RN	Ring Net
H	Hand Gathering
DI	Diving
HL	Hand Lining
T	Trolling
PL	Pole and Line
BT	Bottom Trawl – Single
BPT	Bottom Trawl – Pair
MW	Midwater Trawl – Single
MPT	Midwater Trawl – Pair
D	Dredging
CP	Cod Potting
RLP	Rock Lobster Potting
EP	Eel Potting
FP	Fish Traps
BLL	Bottom Longlining
SLL	Surface Longlining
DL	Drop/Dahn Lines
TL	Trot Lines
SN	Set Netting
DN	Inshore Drift Netting
FN	Fyke Netting
PSN	Pair Set Netting

Types of Containers

BIN	Bin
BOX	Box
BAS	Basket
TRA	Tray
BLO	Block
CAG	Cage
BAG	Bag
CAR	Carton
SAC	Sack
STR	String

Landed State

GRE	Green (or whole)
GUT	Gutted
HGU	Headed and Gutted
DRE	Dressed
FIL	Fillets: skin on
SKF	Fillets: skin off
USK	Fillets: skin off untrimmed
SUR	Fillets: skin on trimmed
SUR	Surimi
TSK	Fillets: skin off trimmed
TRF	Fillets: skin off trimmed
DSC	Dressed – straight cut (Stargazer)
DVC	Dressed – V cut (Stargazer)
MEA	Fish Meal
SCT	Tailed (Scampi)
RLT	Tail (Rock Lobster)
TEN	Tentacles
FIN	Fins
LIV	Livers
MKF	Hoki Mince SKF
MGU	Hoki Mince HGU
HGT	Headed, Gutted and Tailed
HGF	Headed, Gutted and Finned
GGU	Gilled and Gutted
SHU	Shucked and Shelled
ROE	Roe
HDS	Heads
HET	Heads and Tentacles
FIT	Fish Tails
SHF	Shark Fins
MBS	Hoki Mince By-product SKF
MBH	Hoki Mince By-product HGU
MEB	Fish Meal By-product
FLP	Flaps
BEA	Beak and Mouth
LIB	Livers By-product
CHK	Cheeks
LUG	Lugs and Collars
SWB	Sounds or Swim Bladders
WIN	Squid Wings
OIL	Oil
TNB	Tentacles By-product
GBP	Gut by-product

Multiple definitions of effort related fields in *t_fishing_event* by form type and fishing method

Form type and method		***duration***	***effort_height***	***effort_num***	***effort_total_num***	***gearwidth***	***total_hook_num***
CELR	BT, BPT, MW & MPT	Time that gear was at target depth	Headline height (m)	Number of tows in the day		Wind Spread (m)	
	D	Time bet. start of first shoot and finish of last		Number of shots in the day		Dredge width (m)	
	SN & DN	Time from start of setting first net until end of hauling last				Mesh size (mm)	
	RLP, CP, EP, FP & FN			Number of pots/traps/nets in water at midnight	Number of pot/trap lifts in the day		
	SLL, BLL, DL & TL			Number of sets hauled in the day			Number of hooks hauled in the day
	HL, T & PL	Total catching time		Maximum number of lines used at 1 time			Maximum number of hooks used at 1 time
	PS, DS, L, BS & RN			Number sets/shots in the day			
	H, DI	Total person hours spent gathering/ diving		Number of people gathering or diving			
TCEPR			Headline height (m)			Wing spread (m)	
SJCER				Number of single reels in use			
TLCER							Number of hooks