

Database documentation: pseine

K. A. Mackay

NIWA Internal Report
2001

Last revised in May 2003

Contents

1	Introduction to the Database Document series	3
2	Purse seine fishery catch effort data	3
3	Data Structures.....	4
3.1	Table relationships.....	4
3.2	Database design	7
4	Table Summaries	8
5	pseine Tables	9
5.1	Table 1: effort	9
5.2	Table 2: catch.....	11
5.3	Table 3: landing	11
5.4	Table 4: catch_eff	12
7	References.....	14

List of Figures

Figure 1: Entity Relationship Diagram (ERD) for the pseine database.....	5
Figure 2: ERD of the catch_eff table.....	7

Revision History

Version	Change	Date	Responsible
1.1	First release	30 January 2001	K. A. Mackay
1.2	Renamed attribute vessel_id to vessel_key in the tables effort, catch_eff and landing. Data in vessel_key updated.	15 May 2003	Brian Sanders

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 purse seine fishery catch effort database **pseine**, 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 **pseine** 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 Purse seine fishery catch effort data

Kahawai (*Arripis trutta*) is the most frequently targeted fish for the purse seine fishery. Kahawai is a pelagic schooling species and the fishery is mainly of value to the purse seine fleet. The kahawai purse seine fishery cannot be understood without taking into account the other species, which the vessels target. For approximately five months of the year (December to May) the northern fleet, based in Tauranga, tends to target exclusively for skipjack tuna (*Katsuwonus pelamis*) with very little bycatch. When skipjack are not available the fleet fish for a mix of species including kahawai, jack mackerels (*Trachurus* spp.), and blue mackerel (*Scomber australasicus*). These are caught 'on demand' as export orders are received (to reduce product storage costs). The southern fleet, based in Nelson, fished exclusively for the mackerels and kahawai when fishing in southern waters. Since the mackerels and Kahawai school together in some areas, attempts to target for either of the mackerels will often result in a bycatch of kahawai.

Apart from the main targeted species, minor amounts of other species regularly caught in the purse seine fishery include: barracouta (*Thyrsites atun*), trevally (*Pseudocaranx dentex*), blue warehou (*Seriolella brama*), spiny dogfish (*Squalus acanthias*), and kingfish (*Seriola lalandi lalandi*).

The purse seine fishery operates mainly in the Bay of Plenty and on the east Northland coast. A small purse seine fishery operates between the Clarence River mouth and the Kaikoura Peninsula, with at least 70% of the catch being the Peruvian Jack Mackerel (*Trachurus murphyi*). Minor amounts of purse seining also occur along the west coast of both islands in the main jack mackerel fishing ground (Horn, 1991a).

All commercial fishing catch and effort data for the purse seine fishery are recorded for the Quota Management System and are stored in the Ministry of Fisheries databases. The **pseine** database holds a research version of these data. These data are used to calculate biological reference points such the Maximum Sustainable Yield (MSY) in terms of Current Annual Yield (CAY) or Maximum Constant Yield (MCY).

The data in the **pseine** database are a copy of all purse seine fishery catch and effort data from the Ministry of Fisheries databases from 1989 to 1993 and are summarised in Jones (1995).

3 Data Structures

3.1 Table relationships

This database contains several tables. The ERD for **pseine** (Figure 1) shows the logical structure¹ of the database and its 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 key².

Note that Figure 1 shows the main tables only. Note that most tables contain foreign keys³. These foreign keys define the relationships between the tables in **pseine**.

The **pseine** 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 **pseine**: one-to-many⁴. These relationships can be seen in ERDs by connecting a single line (indicating “many”) from the child table; e.g., *catch*, to the parent table; e.g., *effort*, with an arrowhead (indicating “one”) pointing to the parent. Note that the word ‘many’ applies to the possible number of records in one table that one record in another table is associated with. For a given instance, there might be zero, one, two, or more associated records, but if it is ever possible to have more than one, we use the word ‘many’ to describe the association.

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 *effort* and its child table *catch*. The symbol “O” by the child *catch* means that a unit effort can have zero or many catch records, while the bar by the parent *effort* means that for every catch there must be a matching effort record. So it is possible for a unit of effort to produce no catch.

¹ Also known as a database *schema*.

² A primary key is an attribute or a combination of attributes that contains a 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.

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 unit of effort in *effort* can have many catches in *catch* but one catch can only come from one effort.

Physical Data Model		
Project	: pseine	
Model	: Purse Seine catch/effort database	
Author	: spn	Version 1.2 5/15/03

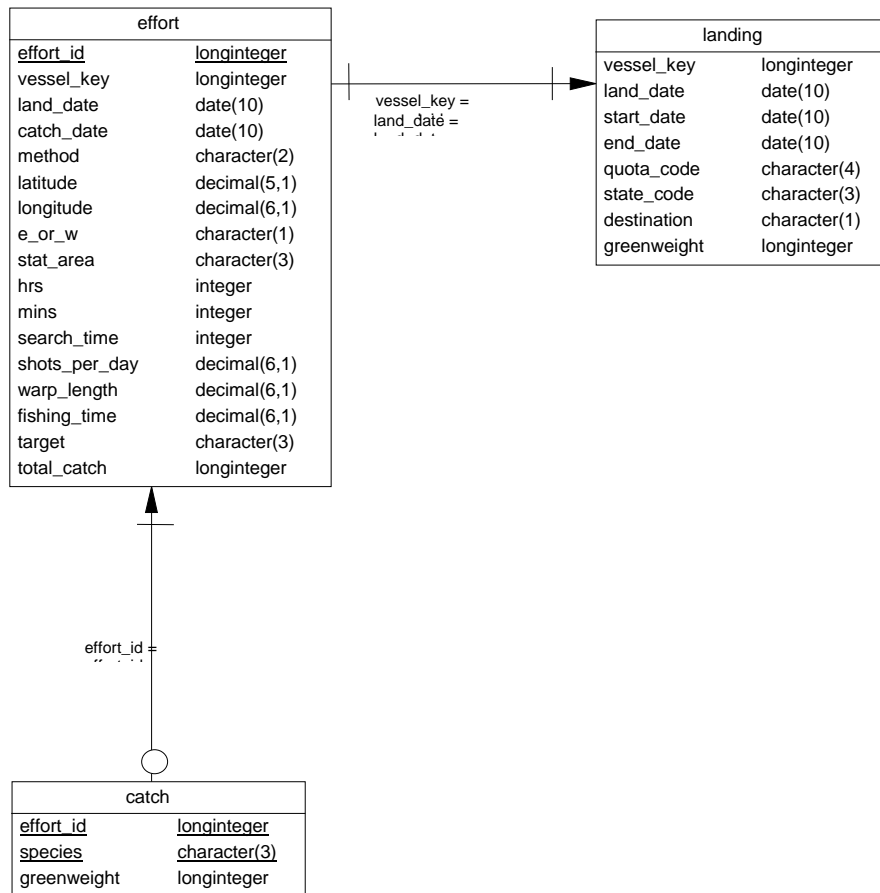


Figure 1: Entity Relationship Diagram (ERD) for the pseine database

Note that the word ‘many’ applies to the possible number of records another is associated with. For a given instance, there might be zero, one, two, or more associated records, but if it is ever possible to have more than one, we use the word ‘many’ to describe the association.

These links are enforced by referential constraints⁵. 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.

⁵ Also known as integrity checks.

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 *catch*:

```
Referential:      No such effort id (effort_id) INSERT effort (effort_id)
```

This means that the value of the attribute *effort_id* in the current record must already exist in the parent table *effort* or the record will be rejected and the following message will be displayed:

```
*** User Error: insert constraint "No such effort id" violation
```

For tables residing in external databases, the parent table name will be prefixed by the name of the database.

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

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

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 effort identifier.

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

Purse seine fishery data is source from the Ministry of Fisheries Catch and Effort database. Raw data are loaded into 3 main tables where further data grooming takes place. These tables are based on three basic entities which result from the Catch Effort Landing Return (CELR) form that fishers are required to complete at the end of each days fishing; i.e., fishing effort, catch, and landing.

The first of these tables is *effort* (Table 1). An effort record is the details of a number of purse seine shots taken within a 24-hour period. Each record is identified by a primary key *effort_id*. Fishing vessels are identified with a numeric code as provided by Ministry of Fisheries. Details stored within effort include: date, position or statistical area code, total search time during the day, the number of purse seine shots per day, total time spent fishing, target species, and total catch weight of all species combined.

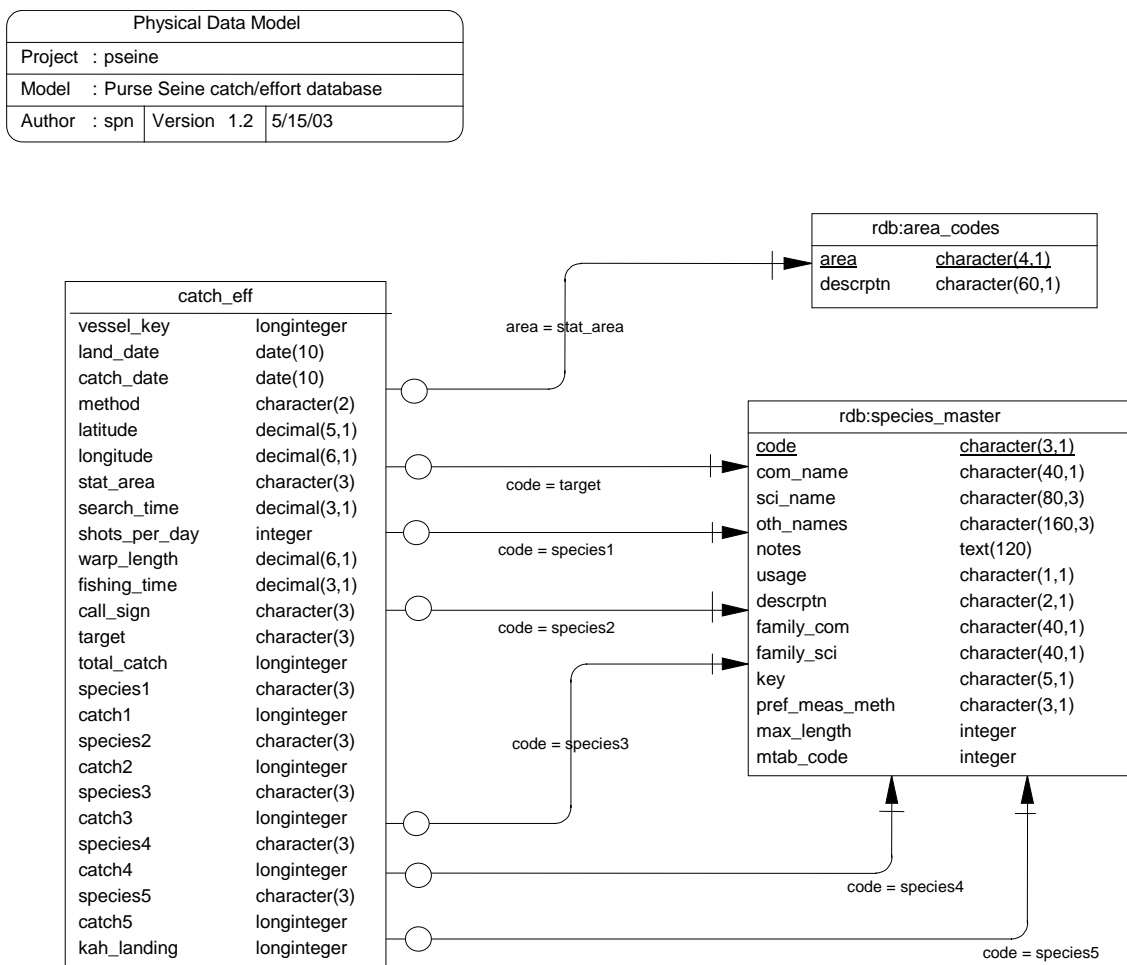


Figure 2: ERD of the *catch_eff* table.

Next, the table *catch* (Table 2), simply records the greenweight of each species caught by each effort record (as identified by *effort_id*).

At the end of a fishing trip, the catch is landing to a Licensed Fish Receiver. These LFR landing records are stored in the table *landing* (Table 3). Each landing record is identified by the vessel id and landing date. Each landing record contains the greenweight from a fishing trip for each quota code (a combination of species code and QMA code) and processed state.

It should be reiterated that these data in these tables are raw data loaded directly from the Ministry of Fisheries database. As such many of them contain inconsistencies, invalid fields, and duplicate records. These three tables are useful in that they allow researchers to go back to the raw data for a repeat analysis or a reinterpretation of values.

The final table *catch_eff* (Table 4) is a de-normalised amalgam of the first three tables (*effort*, *catch*, and *landing*) and contain cleaned and groomed data. As only the table *catch_eff* stores cleaned data, it is the only table to have relationships to the code tables *area_codes* and *species_master* in the **rdb** database (Figure 2). One *catch_eff* record represents one unit of effort (as per the *effort* table). Each record is identified by a combination of vessel identifier and catch date (not an arbitrary effort identifier as in the *effort* table). Details recorded include: catch date, position or statistical area code, total search time during the day, the number of purse seine shots per day, total time spend fishing during the day, target species, and total catch weight. For each day of purse seine effort, the top five species by catch weight have been calculated from the catch table and are stored into the attributes *species1* to *species5* and *catch1* to *catch5*. As a result of kahawai being the main focus of this database, the actual kahawai catch landed for each unit of effort is recorded in the attribute *kah_landing*.

4 Table Summaries

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

1. **effort** : contains raw effort data from the Ministry of Fisheries Catch and Effort database; i.e., a 24-hour period of purse seine shots.
2. **catch** : contains raw catch data from the Ministry of Fisheries Catch and Effort database for each unit effort.
3. **landing** : contains daily processing details of each species. Contains processing data and calculated greenweight
4. **catch_eff**: contains cleaned catch and effort data recorded by commercial purse seine vessels.

5 p seine Tables

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

5.1 Table 1: effort

Comment: Table of raw data from the Ministry of Fisheries Catch and Effort database of a unit of effort; i.e., a 24-hour period of purse seine shots.

Attributes	Data Type	Null?	Comment
effort_id	longinteger	No	Unique sequential number assigned to each effort record as imported from source.
vessel_key	longinteger		Mfish identification number of the fishing vessel.
land_date	date(5)		Landing date for trip.
catch_date	date(5)		Date for each day fishing.
method	character(2,1)		Fishing method code: PS = Purse seine.
latitude	decimal(5,1)		Latitude at start of fishing. In DDMM.m format where 43 degrees 34.6 minutes is stored as 4334.6
longitude	decimal(6,1)		Longitude at start of fishing. In DDDMM.m format where 173 degrees 45.6 minutes is stored as 17345.6
e_or_w	character(1,1)		Meridian of starting position for a purse seine shot - E for East and W for West.
stat_area	character(3,1)		Statistical area at start of fishing, 3 digit code.
hrs	integer		Search time in units of hours.
mins	integer		Search time in units of minutes.
search_time	integer		Total time spent searching in 24-hour period (in nearest whole hours).
shots_per_day	decimal(6,1)		Total number of shots/sets in 24 hour period.
warp_length	decimal(6,1)		Length of net and all warps used in shot (metres).
fishing_time	decimal(6,1)		Total time spent with fishing gear in the water in a 24-hour period (in decimal hours).

effort (cont...)

Attributes	Data Type	Null?	Comment
target	character(3,1)		3-character code for the target species.
total_catch	longinteger		Total estimated greenweight (kg) for catch in a 24-hour period.

Creator: dba

Indices:
UNIQUE BTREE ON (effort_id)
NORMAL (2, 15) BTREE ON (vessel_id)
NORMAL (2, 15) BTREE ON (land_date)
NORMAL (2, 15) BTREE ON (catch_date)
NORMAL (2, 15) BTREE ON (target)

5.2 Table 2: catch

Comment: Table of raw catch data from the Ministry of Fisheries Catch and Effort database for each unit effort.

Attributes	Data Type	Null?	Comment
effort_id	longinteger	No	Unique sequential number assigned to each effort record as imported from source.
species	character(3,1)	No	3-character species code.
greenweight	longinteger		Unique sequential number assigned to each effort record as imported from source

Creator: dba
Referential: No such effort id (effort_id) INSERT effort (effort_id)
Indices: NORMAL (2, 15) BTREE ON (effort_id)
NORMAL (2, 15) BTREE ON (species)

5.3 Table 3: landing

Comment: Table of raw landing data from the Ministry of Fisheries Catch and Effort database for each unit effort.

Attributes	Data Type	Null?	Comment
vessel_key	longinteger		Mfish identification number of the fishing vessel.
land_date	date(5)		
start_date	date(5)		
end_date	date(5)		
quota_code	character(4,1)		
state_code	character(3,1)		
destination	character(1,1)		
greenweight	longinteger		

Creator: dba
Indices: NORMAL (2, 15) BTREE ON (vessel_id)
NORMAL (2, 15) BTREE ON (land_date)

5.4 Table 4: catch_eff

Comment: Table to store cleaned catch and effort data recorded by commercial purse seine vessels.

Attributes	Data Type	Null?	Comment
vessel_key	longinteger		Mfish identification number of the fishing vessel.
land_date	date(5)		Landing date for trip.
catch_date	date(5)		Date for each day fishing.
method	character(2,1)		Fishing method code (PS = Purse seine).
latitude	decimal(5,1)		Latitude at start of fishing. In DDMM.m format where 43 degrees 34.6 minutes is stored as 4334.6
longitude	decimal(6,1)		Longitude at start of fishing. In DDDMM.m format where 173 degrees 45.6 minutes is stored as 17345.6
stat_area	character(3,1)		Statistical area at start of fishing, 3 digit code. Refer rdb:area_codes.
search_time	decimal(3,1)		Total time spent searching in 24-hour period (in decimal hours).
shots_per_day	integer		Total number of shots/sets in 24 hour period.
warp_length	decimal(6,1)		Length of net and all warps used in shot (metres).
fishing_time	decimal(3,1)		Total time spent with fishing gear in the water in 24-hour period (in decimal hours).
call_sign	character(3,1)		Call sign of spotter plane used for searching.
target	character(3,1)		3-character code for the target species. Refer rdb:species_master.
total_catch	longinteger		Total estimated greenweight (kg) for catch in a 24-hour period.
species1	character(3,1)		Species of greatest quantity in catch, code from rdb:species_master.
catch1	longinteger		Estimated greenweight (kg) for species1.
species2	character(3,1)		Species of second greatest quantity in catch, code as for species1.
catch2	longinteger		Estimated greenweight (kg) for species2.

catch_eff (cont...)

Attributes	Data Type	Null?	Comment
species3	character(3,1)		Species of third greatest quantity in catch, code as for species1.
catch3	longinteger		Estimated greenweight (kg) for species3.
species4	character(3,1)		Species of fourth greatest quantity in catch, code as for species1.
catch4	longinteger		Estimated greenweight (kg) for species4.
species5	character(3,1)		Species of fifth greatest quantity in catch, code as for species1.
catch5	longinteger		Estimated greenweight (kg) for species5.
kah_landing	longinteger		Actual greenweight (kg) of kahawai landed.

Creator: dba

Indices:
NORMAL (2, 15) BTREE ON (vessel_id)
NORMAL (2, 15) BTREE ON (latitude)
NORMAL (2, 15) BTREE ON (longitude)
NORMAL (2, 15) BTREE ON (target)

7 References

1. Annala, J.H., Sullivan, K.J., and O'Brien, C.J. (*comps.*): 1999. Report from the Fishery Assessment Plenary, April 1999: stock assessments and yield estimates. pp. 35-38.
2. Horn, P.L. 1991a: Assessment of jack mackerel stocks off the central west coast, New Zealand, for the 1990-91 fishing year. *New Zealand Fisheries Assessment Research Document 91/6*. 14p.
3. Jones, J.B. 1995: Kahawai: Information presented at the 1994 stock assessment. *New Zealand Fisheries Assessment Research Document 95/9*. 24p.