Database documentation: dive

K. A. Mackay

NIWA Internal Report
2001

Revised on 30 January 2001

Table of Contents

List of Figures

1 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 diver sampling database **dive**, and is part of the database documentation series produced by NIWA. It supersedes the previous documentation by Mackay (1998)[1] on this database.

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 **dive** 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 Dive Sampling Programmes

Dive surveys cover a very wide range of survey designs. The main types of dive surveys covered by this database involve the collection of fish lengths using a variety of survey methods. The main types of survey method represented in this database are:

1. **Random dive surveys**. The area of interest is divided into sites (often 200m by 200m squares) and allocated a number. Sites are then selected at random. At each selected site, all specimens of the target species in the site are counted and a sample taken to the surface for measuring individual lengths.

2. **Fish aggregation surveys.** Divers count and estimate sizes of individual target species by eye. Dives are either done on: known aggregations of the target species, e.g., crayfish; random sites; or transects. In some cases, all specimens of the target species within a defined area are caught and brought to the surface for length measurements.

[1] K.A. MACKAY, K. 1998 Database documentation. dive. *NIWA Internal Report No. 36*. 12p.

3. **Transect dive Surveys**. Divers follow a transect line for a set time or distance. For finfish target species, counts and length estimates are made by eye. For other target species, divers may search along a transect using a square quadrat end-on-end. In such cases, counts of abundance of the target species are made, and include a sample taken to surface for measuring. Transects may be pre-defined or random. During random transect dive surveys, at randomly selected sites within a defined area, divers descend to a randomly chosen depth. Beginning in a randomly chosen direction, divers measure and record the occurrence of the target species within a quadrat.

To date, species involved in these surveys include blackfoot paua (*Haliotis iris*), Yellowfoot paua (*H. australis*), kina (*Evechinus chloroticus*), red rock lobsters (*Jasus edwardsii*), and blue cod (*Parapercis colias*).

3 Data Structures

3.1 Table Relationships

This database contains several tables. The ERD for **dive** (Figure 1) shows the logical structure[2] 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[3].

Note that Figure 1 shows the main tables only. Note that most tables contain foreign keys[4]. These foreign keys define the relationships between the tables in **dive**.

The **dive** 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 **dive**: one-to-many[5]. 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., *t_lgth*, to the parent table; e.g., *t_dive*, with an arrow-head (indicating 'one') pointing to the parent. For example, consider the relationship between the tables *t_dive* (the parent table) and *t_lgth* (the child table). Any one dive in *t_dive* can have one or more length records in *t_lgth*, but any one length record can only come from one dive. Note that the word 'many' applies to the possible number of records another is associated with. For a

[2] Also known as a database *schema.*
[3] A primary key is an attribute or a combination of attributes that contains a unique value to identify that record.
[4] A foreign key is an attribute or a combination of attributes that is a primary key in another table.
[5] 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 dive in *t_dive* can have one or more length records in *t_lgth*, but any one length record can only come from one dive.

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.

Note that the one-to-many relationships can be either mandatory or optional[6]. The optional relationship, denoted in the ERD by the symbol "O" at one or both ends of the relationship line, means that a record does not have to have any associated records. Conversely, the mandatory relationship denoted in the ERD by a bar symbol across the relationship line, means that a record has to have at least one associated record. For example, if we consider again the one-to-many relationship between the tables *t_dive* and *t_lgth*, which has a mandatory 'one' and an optional 'many'. This means that one dive record can have zero or more (many) length records recorded, but one length record in the effort table must have one, and only one, associated record in the dive table.

These links are enforced by referential constraints[7]. 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_dive*:

```
Referential:      Invalid target species code (target_species) INSERT
                  rdb : curr_spp (code)
```

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

```
*** User Error: insert constraint "Invalid species code" violation
```

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

[6] Known as the *cardinality* of a relationship – the constraint on relationships so that the possible extensions of the relation correspond to real-world associations.
[7] Also known as integrity checks.

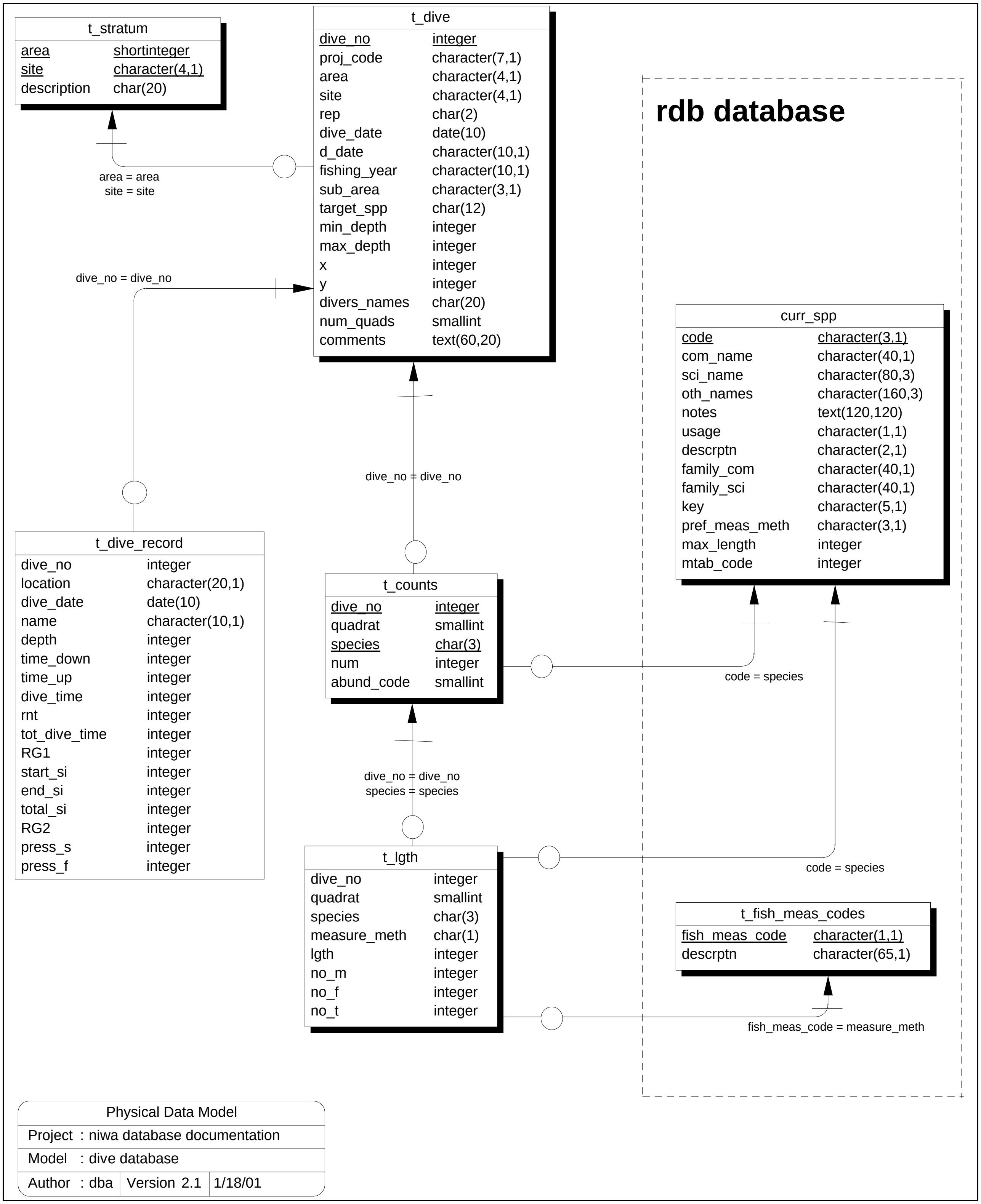

Figure 1: Entity Relationship Diagram (ERD) for the dive database

Section 5 lists all the **dive** 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 dive number. Note that the table *t_lgth* has no primary 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 core entity of the **dive** database is a single dive by a single or buddy-pair of divers. Details of each dive are recorded in the table *t_dive* (Table 1). Each dive is allocated a *dive_no,* which is a unique number to identify that dive. Each dive record must have the *area* code filled in to provide, at the very least, a geo-spatial reference. This may be anything from a general region (e.g., the Marlborough Sounds) to a particular geographic point (e.g., the wharf off Halfmoon Bay). The attribute *proj_code* is used to identify different dive survey data sets. Each dive site within an area and/or stratum is given a number, denoted by the attribute *site.* Dive sites may be repeatedly dived on. These repeats are given a number, denoted by the attribute *rep*. The date of the dive is stored in the attribute *dive_date*. However, for a large amount of the historical data (pre 1998), the exact date is not know. Therefore, the attribute *d_date* is a character field to stored text describing the dive date, such as "Jun-97". An additional date attribute, *fishing_year*, also stores the fishing year of the survey, in the format *yy1-yy2*; e.g., the 1997-98 fishing year is stored as "97-98". One dive can also involve more than one diver. The diver(s) names are recorded in the attribute *divers_names.*

Usually, the survey design is that the area is divided into sub-regions, such as strata or grids. Thus, each dive may be in a stratum within an area, as recorded by the attributes *area* and *stratum*. These attributes are a foreign key to the table *t_stratum* (Table 2) contains the descriptions of these strata. This is directly comparable to the table *t_grid_ref* in the **kina** database[8].

[8] Mackay, K. and Fisher, D. 1993: Marine Research database documentation. 7. Kina. *MAF Fisheries Greta Point Internal Report No. 213.* 19p.

Some dive surveys using metal quadrats that are repeatedly laid down along a transect line. All occurrences of the target species within the quadrat are counted and recorded by the divers. These quadrat counts are stored in the table *t_counts* (Table 3). Often divers will use an abundance code to describe the numbers of species, rather than an absolute number. Each record in *t_counts* is therefore the number or abundance for each species counted or measured within a single quadrat. If no quadrats are used in a dive, then *t_counts* represented a summary of the length information recorded in the *t_lgth* table. In such instances, the value of the quadrat attribute is set to a default of 1.

All length data recorded by divers are stored in the table *t_lgth* (Table 4). Each record on *t_lgth* corresponds to the measurement of an individual animal, rather than a length frequency. This structure allows *t_lgth* to be easily expanded to include other biological attributes including greenweight, gonad state, etc. For species such as shellfish; e.g., paua and kina, these lengths can be recorded after the dive by measurement of the catch. For other species, such as finfish, measurements are usually the result of estimates of fish length by the divers.

Dive details for each diver that participated in the dive can be recorded in the table *t_dive_record* (Table 5). This is information recorded for the divers personal dive logbook and includes such details as start, finish and total dive times, start and finish SCUBA tank air pressures, residual nitrogen time, start, finish and total surface intervals, and repeat groups as calculated from the standard PADI dive tables. This table is not normalized because several attributes are functionally dependant on other attributes within the table, not just the primary key. Specifically, the *total_dive_time* and *total_si* are simply the result of summing other attributes, namely *dive_down*, *dive_up* and *dive_time*, *rnt* respectively. Similarly, the PADI repeat groups *RG1* and *RG2* are dependant on these dive time and any surface intervals. The decision has been made to retain these attributes within the table, reflecting what the divers themselves entered in their dive logs. The calculations can be done later as a means of crosschecking the divers entries.

4 Table Summaries

The following is a listing and brief outline of the tables contained in **dive**:

1. **t_dive :** contains details for each dive taken, including area, date, and target species
2. **t_stratum :** contains details of individual dive sites or strata used during dive surveys.
3. **t_counts:** contains counts of species by dive number (and quadrat if used).
4. **t_lgth :** contains details for individual lengths of species as measured by divers.
5. **t_dive_record :** contains the dive details for each diver that participated in the dive, including total dive time, start and finish air pressure, start and finish surface intervals, and residual nitrogen time.

5 dive Tables

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

5.1 Table 1: t_dive

Comment: Table of individual dive information.

Attributes	Data Type	Null?	Comment
dive_no	integer	No	Unique sequential dive number
proj_code	character(7,1)		Project code of dive
area	character(4,1)	No	Area code. Maybe a QMA or a Research code (refer rdb:area_codes).
site	integer		Site number or stratum code.
rep	integer		Repeat dive number for a site
dive_date	date(4)		Date of dive
d_date	character(10,1)		Date of dive (char field)
fishing_year	character(10,1)		Fishing year e.g., 96-97
sub_area	character(3,1)		Sub-area code
target_spp	character(3,1)		3 char code for target species. Refer rdb:curr_spp
min_depth	integer		Minimum depth (m) of the dive.
max_depth	integer		Maximum depth (m) of the dive.
x	integer		Width (m) of the transect or site.
y	integer		Length (m) of the transect or site.
num_quads	smallint		Number of quadrats used during the dive.
divers_names	character(50,1)		Name(s) of the divers(s)
comments	text(60,20,20,1)		General comments for dive

Creator: dba

Referential: Invalid target species code (target_spp) INSERT rdb:curr_spp (code)

Indices:
UNIQUE dive_pk ON (dive_no)
NORMAL (2, 15) dive_area_ndx ON (area)
NORMAL (2, 15) dive_fishing_year_ndx ON (fishing_year)

5.2 Table 2: t_stratum

Comment: This table contains the details of individual dive sites or strata used in dive surveys.

Attributes	Data Type	Null?	Comment
area	character(4,1)	No	Area code. Maybe a QMA or a Research code (refer rdb:area_codes).
site	integer	No	Site or stratum code.
description	character(20,1)		Brief description of the site or stratum.

Creator: dba
Indices: NORMAL (2, 15) stratum_area_ndx ON (area)

5.3 Table 3: t_counts

Comment: Table of number of each species counted by dive_no (and quadrat if used).

Attributes	Data Type	Null?	Comment
dive_no	integer	No	Unique sequential dive number
quadrat	smallint	No	Sequential number for each quadrat searched during a dive. Default=1.
species	character(3,1)	No	3 char species code. Refer rdb:curr_spp
num	longinteger		Number of species counted by divers.
abund_code	smallint range 1-4		Code for the number of juveniles seen: "1"=1;"2"=2-4;"3"=5-16;"4"=>17

Creator: dba

Referential: Invalid dive number (dive_no) INSERT t_dive (dive_no)
Invalid species code (species) INSERT rdb:curr_spp (code)

Indices: NORMAL (2, 15) counts_dive_no_ndx ON (dive_no)
NORMAL (2, 15) counts_species_ndx on (species)

5.4 Table 4: t_lgth

Comment: Table of length measurements of a species measured during a dive.

Attributes	Data Type	Null?	Comment
dive_no	integer	No	Unique sequential dive number
quadrat	smallint		Sequential number for each quadrat searched during a dive.
species	character(3,1)		3 char species code. Refer rdb:curr_spp
measure_meth	character(1,1)		Code of method used to measure fish length, refer rdb:t_fish_meas_codes
lgth	integer		length (mm for shellfish, cm for finfish)
no_m	integer		Number of males counted at this length.
no_f	integer		Number of females counted at this length.
no_t	integer		Total number this species counted at this length.

Creator: dba
Referential: Invalid dive number (dive_no) INSERT t_dive (dive_no)
Invalid species code (species) INSERT rdb:curr_spp (code)
Invalid fish measurement code (measure_meth) INSERT rdb:t_fish_meas_codes (fish_meas_code)
Indices: NORMAL (2, 15) lgth_dive_no_ndx ON (dive_no)

5.5 Table 5: t_dive_record

Comment: Table of the dive details for each diver that participated in the dive, including total dive time, start and finish air pressure, start and finish surface intervals, and residual nitrogen time. Note that these are based on the PADI Dive Tables which are no longer used.

Attributes	Data Type	Null?	Comment
dive_no	integer	No	Unique sequential dive number.
location	character(20,1)		General location of the dive site.
dive_date	date(10)		Date of the dive.
name	character(10,1)	No	Diver's name.
depth	integer		Maximum depth of the dive (m).
time_down	integer		Time (NZDT, hhmm) at the start of the dive.
time_up	integer		Time (NZDT, hhmm) at the end of the dive.
dive_time	integer		Dive time (minutes) as calculated by time_up – time_down.
rnt	integer		Residual nitrogen time (minutes).
tot_dive_time	integer		Total time (minutes) of the dive as calculated by dive_time + rnt.
RG1	character(1,1)		Repeat group at beginning of surface interval (from PADI dive tables).
start_si	integer		Time (NZDT, hhmm) of the start of the surface interval.
end_si	integer		Time (NZDT, hhmm) of the end of the surface interval.
total_si	integer		Total time (minutes) of the surface interval.
RG2	character(1,1)		Repeat group at end of surface interval (from PADI dive tables).
press_s	integer		Air pressure at the start of the dive.
press_f	integer		Air pressure at the start of the dive.

Creator: dba

6 rdb business rules

6.1 Introduction to business rules

The following are a list of business rules applying to the **dive** database. A business rule is a written statement specifying what the information system (i.e., any system that is designed to handle market sampling data) must do or how it must be structured.

There are three recognised types of business rules:

Fact	Certainty or an existence in the information system.
Formula	Calculation employed in the information system.
Validation	Constraint on a value in the information system.

Fact rules are shown on the ERD by the cardinality (e.g., one-to-many) of table relationships. Formula and Validation rules are implemented by referential constraints, range checks, and algorithms both in the database and during validation.

Validation rules may be part of the preloading checks on the data as opposed to constraints or checks imposed by the database. These rules sometimes state that a value should be within a certain range. All such rules containing the word 'should' are conducted by preloading software. The use of the word 'should' in relation to these validation checks means that a warning message is generated when a value falls outside this range and the data are then checked further in relation to this value.

6.2 Summary of rules

Dive table (t_dive)

dive_no	Must have a value entered and be a unique sequential integer greater than zero.
proj_code	Can be any combination of up to 7 alphanumeric characters. Should be a valid project code
area	Must be a valid area code as listed in the *area_codes* table of the **rdb** database.
site	Must be an integer greater than zero and should be within the reasonable range to 1 to 20.
	Multiple column check on *area* and *site*: The combination of *area* and *code* must be a valid combination as listed in the *t_stratum* table.
rep	Must be an integer greater than zero and should be within the reasonable range to 1 to 10.
dive_date	Must be a valid date and should be after 1 January 1992.
d_date	Can be any combination of alphanumeric characters but should contain names of months (or part thereof) and/or 2- or 4-digit years.
fishing_year	Should contain a pair of 2- or 4-digit years separated by one or more characters.
sub_area	Can contain any combination of 3-character alphanumeric codes.
target_spp	Must contain a valid species code as listed in the *curr_spp* table in the **rdb** database and should be one of the species codes listed in the Appendix.
min_depth	Must be an integer greater than or equal to zero and should be within the reasonable range of 1 to 10.
max_depth	Must be an integer greater than or equal to zero and should be within the reasonable range of 5 to 40.
x	Must be an integer greater than or equal to zero and should be within the reasonable range of 0 to 10.

Dive table (t_dive) cont….

y	Must be an integer greater than zero and should be within the reasonable range of 1 to 10.
num_quads	Must be an integer greater than or equal to zero and should be within the reasonable range of 0 to 10.
divers_names	Can have any combination of up to 50 ASCII characters.

Dive strata table (t_stratum)

area	Must have a value entered and be a valid area code as listed in the *area_codes* table of the **rdb** database.
site	Must have a value entered and be an integer greater than zero and should be within the reasonable range to 1 to 10.
description	Can have any combination of up to 20 ASCII characters.

Specimen counts table (t_counts)

dive_no Must have a value entered and be a valid dive number as listed in the *t_dive* table.

quadrat Must be equal to 1 or an sequential integer greater than or equal to 1. Should be within the reasonable range of 1 to 10.

species Must have a value entered and contain a valid species code as listed in the *curr_spp* table in the **rdb** database and should be one of the species codes listed in the Appendix.

Multiple column check on *dive_no*, *quadrat* and *species*:
The combination of *dive_no*, *quadrat* and *species* must unique for each record.

num Must be an integer greater than or equal to zero and should be within the reasonable range of 0 to 20.

Multiple column check on *species* and *num*:
The number in *num* should be less than the reasonable maximum for the *species*. Maximum species numbers are listed in the Appendix.

abund_code Must be a valid abundance code as listed in the Appendix.

Multiple column check on *num* and *abund_code*:
The num and abund_code attributes are mutually exclusive. If a value existed in one, then the other must be null.

Specimen length frequency table (t_lgth)

Multiple column check on *dive_no, quadrat* and *species*:
The combination of *dive_no*, *quadrat,* and *species* must exist in the *t_counts* table.

measure_meth Must be a valid fish measurement method code as listed in the *t_fish_meas_codes* in the **rdb** database.

lgth Must be and integer greater than zero and within the reasonable range of 5 to 300.

Multiple column check on species and *lgth*:
The fish length should be less than the maximum length for the species. Reasonable maximum species lengths are listed in the Appendix.

no_m The number of male fish must be an integer greater than or equal to zero.

no_f The number of female fish must be an integer greater than or equal to zero.

no_t The total number of fish must be an integer greater than zero.

Multiple column check on number of males, females and total fish:
The total number of fish must be greater than or equal to the number of males plus the number of females.

Multiple column check on number of total fish and species:
The total number of fish should be less than the reasonable maximum for the *species*. Maximum species numbers are listed in the Appendix.

Multiple column check on number of total fish and *num*:
The total number of fish must be less than or equal to *num*, the number of the species counted in the *t_counts* table.

Dive log record table (t_dive_record)

dive_no Must be a valid dive number as listed in the *t_dive* table.

location Can have any combination of up to 20 ASCII characters.

dive_date Must be a valid date and should be after 1 January 1992.

name Must have a value entered and can any combination of up to 20 ASCII characters. Should be one of names listed in the attribute *divers_names* in the *t_dive* table.

depth Must be an integer greater than or equal to zero and should be within the reasonable range of 5 to 40.

time_down Start time of the dive must be a valid 24-hour time and fall within the range of 0 – 2359. Should be within daylight hours, which ranges from 0600 to 1800.

time_up Finish time of the dive must be a valid 24-hour time and fall within the range of 0 – 2359. Should be within daylight hours, which ranges from 0600 to 1800.

Multiple columns checks on *time_down* and *time_up*:
The finish time of the dive must be after the start time of the dive.

dive_time Must be an integer greater than zero and should be within the reasonable range of 15 – 80.

Multiple columns checks on dive time and the start and finish dive times:
Dive time must be the difference, in minutes, between the start and the finish times of the dive.

rnt Must be an integer greater than zero and should be within the range of 3 to 350.

tot_dive_time Must be a integer greater than zero.

Multiple columns checks on dive time, residual nitrogen time and total dive time:
Total time must be equal to dive time plus residual nitrogen time.

RG1 Must be a 1-character code within the range of A to N.

Dive table (t_dive) cont….

start_si — Start time of the surface interval must be a valid 24-hour time and fall within the range of 0 – 2359. Should be within daylight hours, which ranges from 0600 to 1800.

.

end_si — Finish time of the surface interval must be a valid 24-hour time and fall within the range of 0 – 2359. Should be within daylight hours, which ranges from 0600 to 1800.

Multiple columns checks on start and finish surface interval times:
The finish time of the surface interval must be after the start time of the surface interval.

total_si — Must be an integer greater than zero and should be within the reasonable range of 10 – 720

Multiple columns checks on total surface interval and the start and finish surface interval times:
Total surface interval must be the difference, in minutes, between the start and the finish times of the surface interval.

RG2 — Must be a 1-character code within the range of A to N.

Multiple columns checks on Repetitive Group 1 and Repetitive Group 2:
Repetitive Group 2 must be less than Repetitive Group 1.

press_s — Must be an integer greater than zero and should be within the reasonable range of 1500 to 3000.

press_f — Must be an integer greater than zero and should be within the reasonable range of 100 to 1500.

Multiple columns checks on start and finish air pressures:
Finish air pressure must be less than the start air pressure.

7 Acknowledgments

The author would like to thank Peter Gerring and Rob Stewart for their technical input, and Dave Banks for his review and editorial comment for this document.

Appendix

Codes of the species most likely to be targeted and measured by dive sampling.

Code	*Common Name*	*Scientific Name*
BCO	Blue cod	*Parapercis colias*
CRA	Red rock lobster	*Jasus edwardsii*
PAA	Yellowfoot paua	*Haliotis australis*
PAI	Blackfoot paua	*Haliotis iris*
PAU	Paua (generic)	*Haliotis* spp.
SUR	Kina	*Evechinus chloroticus*

Maximum specimen counts by species

Species	*Maximum number*
BCO	40
PAU	3000
SUR	2500

Species abundance codes.

Code	*Number of species*
1	1
2	2-4
3	5-16
4	>17