Database documentation: oyster

K. A. Mackay

NIWA Internal Report
2000

Revised on 30 November 2000

Contents

1 Database documentation series **3**

2 Foveaux Strait oyster sampling programme **3**

3 Data Structures **4**
- **3.1** Table Relationships 4
- **3.2** Database Design 6
  - **3.2.1** Oyster dredge surveys 7
  - **3.2.2** Survey on the impact of dredging on oysters 8

4 Table Summaries **9**

5 oyster Tables **10**
- **5.1** Table 1: t_survey 10
- **5.2** Table 1: t_vessel 10
- **5.3** Table 1: t_station 11
- **5.4** Table 1: t_damage 14
- **5.5** Table 1: t_treatment_codes 15
- **5.6** Table 1: t_damage_codes 15
- **5.7** Table 1: t_cluster_codes 15

List of Figures

1 Entity Relationship Diagram (ERD) for the **oyster** database **5**

1 Database documentation 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.

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 Foveaux Strait oyster database **oyster**, and is a 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 the relationships between the tables in **oyster**.

This document is intended as a guide for users and administrators of the **oyster** 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 Foveaux Strait oyster sampling programme

In 1985, the population of takeable (legal size >=58mm) Foveaux Strait oysters (*Tiostrea chilensis*) was estimated to have been 1140 million. An epidemic of the protistan parasite *Bonamia* sp. probably began in 1985 and was diagnosed from samples taken after high oyster mortalities in 1986. In response to this outbreak, a series of surveys was undertaken, starting in 1990, to monitor the distribution and intensity of the outbreak. These surveys showed that the infection and subsequent mortality spread through the population in the following years reducing the population to 771 million by July 1990, and 319 million by February 1992, by which time infection had reached the periphery of oyster distribution. In 1992, the population in the area surveyed in 1975 was less than 10% of that present in 1975 and recruitment was considered to be at risk. The fishery was partially closed to fishing in 1992 and fully closed in 1993 to allow the population to rebuild.

Changes in the distribution of *Bonamia* sp. in Foveaux Strait between 1990 and 1995 had indicated that the prevalence and intensity of infection had waned and that mortality of oysters in the future is unlikely to be as great as it has been in the immediate past. The size of the oyster population estimated in 1992 and 1993 suggested that the population was increasing though the increase was not statistically significant. The population has been surveyed several times since 1995 to monitor further changes.

In 1997, the survey also allowed for the opportunity to examine the effect of different dredge methods on the damage sustained by oysters at harvesting.

3 Data structures

3.1 Table relationships

This database contains several tables. The ERD for **oyster** (Figure 1) shows the logical structure[1] of the database and its entities (each entity is implemented as a database *table*) and relationships between these tables. Each table represents an object, event, or concept in the real world that has been 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]. This schema is valid regardless of the database system chosen, and it can remain correct even if the Database Management System (DBMS) is changed. Most of the tables in the **oyster** database also contain special attributes, called foreign keys[3].

Section 5 shows a listing of all the **oyster** tables as implemented by the EMPRESS DBMS. 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 format:

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

where the attribute(s) make up the primary key (the key attributes) and the index name is the primary key name. This prevents records with duplicate key values from being inserted into the table, e.g., a new survey being inserted with an existing survey number.

The **oyster** database is implemented as a relational database. That is, tables are linked to one another by their relationships. All relationships in **oyster** are of the type *one-to-many*[4]. This is shown in the ERD by connecting a single line (indicating 'many') from the child table (e.g., *t_station*) to the parent table (e.g., *t_survey*) with an arrow-head (indicating 'one') pointing to the parent. For example, consider the relationship between the tables *t_survey* (the parent table) and *t_station* (the child table). Any one survey in *t_survey* can have one or more stations in *t_station*, but any one station can only be a part of one survey.

[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 any attribute, or a combination of attributes, in a table that is a primary key of another table. Tables are linked together through foreign keys.
[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 survey in *t_survey* can have many stations in *t_station* but any one station can only come from one survey.

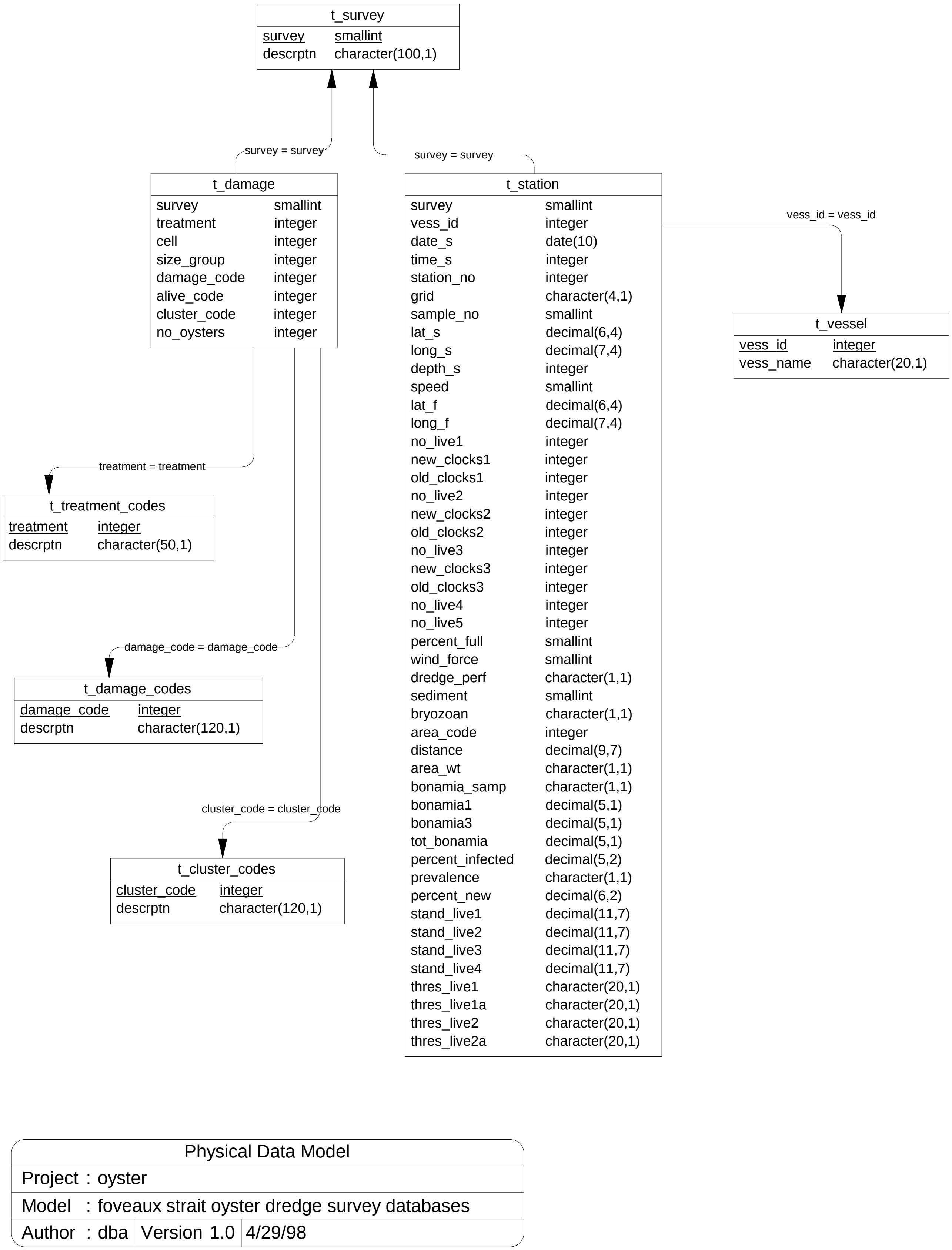

Figure 1: Entity Relationship Diagram (ERD) of the oyster database.

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

These relationships are enforced in the database by the use of 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:` `error message (attribute) INSERT parent table (attribute)`

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

`Referential:` `No such survey (survey) INSERT t_survey (survey)`

This means that the value of the attribute *survey* in the current record must already exist in the parent table *t_survey* or the record will be rejected and the error message "No such survey" will be displayed.

All tables are indexed. That is, attributes that are most likely to be used for searching, such as *survey*, have like values linked together to optimise search times. Such indices are shown in the table listings (Section 5) by the following syntax:

`Indices:` `NORMAL (2, 15) index_name ON (attribute{, attributes})`

Note that indices may be ***simple***, pointing to just one attribute, or ***composite***, pointing to more than one attribute. The numbers '…(2, 15)…' are EMPRESS default values relating to the amount of space allocated to index storage.

3.2 Database design

The genesis of **oyster** lies in the need for a central data repository for data collected by the Foveaux Strait oyster survey programme. The data were originally residing on several PCs and in a variety of formats while analyses were carried out.

Once the analyses were completed, the groomed data, usually in the form of spreadsheets, were made available for loading onto the EMPRESS RDBMS. These spreadsheets form the basis of the database table design and on a conceptual level an attempt has been made to make these tables as robust as possible for any further surveys that may eventuate.

[5] Also known as integrity checks.

3.2.1 Oyster dredge surveys

Conceptually, a dredge survey has two main entities: a survey entity (*t_survey*, Table 1) and a station entity (*t_station*, Table 3) with one survey comprising of many stations. However, during analysis, each survey's data was stored in it's own single spreadsheet, with each record in the spreadsheet broadly corresponding with a station (or dredge) site. Hence, these spreadsheets collectively represent the station entity and were loaded verbatim into the table *t_station*. The attribute *survey* was added to *t_station*, as a number to identify which survey the station was a part of. The table *t_survey* contains these codes and the survey title that they represent.

Different surveys employed a variety of sampling methodologies, from fixed grid stations to stratified random sampling. All these methodologies are accommodated in the station table resulting in a wide range of attributes, from *station_no* and *sample_no*, to *grid_no*. However, for the stratified surveys, no data defining strata was given for loading into the database, hence the absence of a strata entity.

Spreadsheet design mentality dictates that all attributes are represented in columns, with many of the columns being used to hold results of various calculations. However, standard database design theory dictates that tables should, where possible, conform to certain normalization rules. The purpose of these normalization rules is to produce a databases design that is highly flexible, allowing the schema to be extended when needed to account for new attributes, entities, and relationships. They can also reduce redundancy in the database and make sure the database is free from certain anomalies. Usually, the aim is to conform to the first three rules. That is, the database should be in a 3^{rd} Normal Form (3NF).

All oyster survey spreadsheets to date are in 1NF in that every attribute is single-valued for each record. However, to be 2NF, all the non-key attributes have to be fully dependent on the primary key; i.e., each attribute in a record has to be dependant on the attributes *survey* and *station_no*. Here, the spreadsheets failed because the vessel name is dependent on the vessel registration number, not the station primary key. To conform to 2NF, the vessel names are therefore stored in their own table *t_vessel* (Table 2) and are linked to *t_station* by the attribute *vess_id*.

To make *t_station* 3NF requires that every non-key attribute is not transitively dependent on the primary key. That is, there should be no attributes that are the result of calculations of other attributes in the table. The problems arise because any changes in the values of the other attributes will not automatically change the value of the transitively dependent attribute, resulting in data anomalies, unlike spreadsheets, which can do this using formula. Because the original spreadsheets were used for analyses, they have several attributes that fall into this category (such as the attribute *stand_live1* that is calculated from the attributes *no_live1* and *distance*). These transitively dependent attributes should be removed from *t_station*, but because **oyster** is used as a static data repository rather than a dynamic data analysis tool, these anomalies will not appear. Therefore, they remain in *t_station*.

The table *t_station* contains a number of attributes for absolute, percentage, and standardised abundance of oysters dredged at each station; e.g., *no_live1*, *stand_live2*, and *bonamia3*. A number to denote the shell size range they represent suffixes these attributes, with: 1 = takeable (legal size

>= 58mm); 2 = sublegal (< 58mm); 3 = pre-recruits (between 50 and 57mm); 4 = juveniles (< 50mm); 5 = combined pre-recruit and takeable (>= 50mm).

3.2.2 Survey on the impact of dredging on oysters

The exception to the survey/station relationship is the 1997 survey on the impact of dredging on Foveaux Strait oysters, which looked at differences of levels of damage sustained by oysters at harvesting by different dredging methods. Conceptually, this survey is a one-to-many relationship between the survey entity (*t_survey*) and the damage entity (*t_damage*, Table 4). Individual records in *t_damage* don't represent a station, but a cluster of oysters, where a cluster is defined as either one adult (legal size >= 58mm) oyster or a clump of juvenile spat formed around a single nucleus.

Much of the data in *t_damage* are in the form of numeric codes. Separate tables contain the definitive lists of these codes, and their descriptions. The codes are for the types of damage to oysters (*t_damage_codes*, Table 6), treatment (or method of dredging) of oysters at harvesting (*t_treatment_codes*, Table 5), and cluster types (*t_cluster_codes*, Table 7).

4 Table summaries

This database is arranged as a set of four main tables containing survey data, and three other tables describing the various codes that are used.

The following is a listing and brief outline of the tables contained within **oyster**:

1. **t_survey** : contains descriptions of the various Foveaux Strait oyster surveys undertaken and a numeric code to identify each survey.
2. **t_vessel** : contains vessel registration codes and vessel names.
3. **t_station** : contains station details including date, position, distance dredged, and depth, as well as catches of oysters by size class and levels of infection by *Bonamia* sp.
4. **t_damage** : contains details of damage to clusters (a single takeable oyster or a cluster of juveniles around a single nucleus) sustained by various dredging methods.

Also contained in the **oyster** database are master tables containing all codes and their descriptions. Codes used in **oyster** are checked against these master tables before being inserted. The following is a summary list of these master code tables:

5. **t_treatment_codes** : contains codes and descriptions of the different ways in which the oysters are treated (fishing methods) at harvesting.
6. **t_damage_codes** : contains codes and descriptions for the various levels of damage sustained to an oyster.
7. **t_cluster_codes** : contains codes and descriptions for the various types of clusters ranging from a single takeable oyster to a cluster of multiple generations of oysters around a single nucleus.

5 oyster Tables

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

5.1 Table 1: t_survey

Comment: Table of oyster survey codes and descriptions.

Attributes	Data Type	Comment
survey	smallint	Unique sequential numeric code for each survey.
descrptn	character(100,1)	Description of survey.

Creator: dba
Indices: UNIQUE survey_pk ON (survey)

5.2 Table 2: t_vessel

Comment: Registration numbers and names of vessels used during Foveaux Strait oyster surveys.

Attributes	Data Type	Comment
vess_id	integer	Vessel registration number.
vess_name	character(20,1)	Vessel name.

Creator: dba
Indices: UNIQUE vessel_pk ON (vess_id)

5.3 Table 3: t_station

Comment: Station and catch details for an oyster survey station.

Attributes	Data Type	Comment
survey	smallint	Numeric survey code (refer: t_survey).
vess_id	integer	Vessel registration number (refer: t_vessel).
date_s	date(5)	Date on start of dredge station.
time_s	integer	Time (24hr, NZST) on start of dredge station.
station_no	integer	Sequential station number.
grid	character(4,1)	Station grid reference.
sample_no	smallint	Sample number.
lat_s	decimal(6,4)	Latitude at start of dredge station (decimal degrees).
long_s	decimal(7,4)	Longitude at start of dredge station (decimal degrees).
depth_s	integer	Depth (m) at start of dredge station.
speed	smallint	Towing speed (knots).
lat_f	decimal(6,4)	Latitude at finish of dredge station (decimal degrees).
long_f	decimal(7,4)	Longitude at finish of dredge station (decimal degrees).
no_live1	integer	Number of live legal size oysters (>=58mm) in dredge.
new_clocks1	integer	Number of legal size (>=58mm) new clocks in dredge.
old_clocks1	integer	Number of legal size (>=58mm) old clocks in dredge.
no_live2	integer	Number of sublegal size (<58mm) new clocks in dredge.
new_clocks2	integer	Number of sublegal size (<58mm) new clocks in dredge.
old_clocks2	integer	Number of sublegal size (<58mm) old clocks in dredge.
no_live3	integer	Number of live oysters between 50mm and 57mm in dredge.

new_clocks3	integer	Number of new clocks between 50mm and 57mm in dredge.
old_clocks3	integer	Number of old clocks between 50mm and 57mm in dredge.
no_live4	integer	Number of live oysters < 50mm in dredge.
no_live5	integer	Number of live oysters >50mm in dredge.
percent_full	smallint	Percent fullness of the dredge.
wind_force	smallint	Wind force (Beaufort scale) during dredge.
dredge_perf	character(1,1)	Did the dredge fish well? (Y/N)
sediment	smallint	Sediment type (0=weed;2=shell; 3=shell/gravel;4=pea gravel;5=sand; 6=silt;7=sponges;8=bryozoan).
bryozoan	character(1,1)	Bryozoan substrate (Y/N)?
area_code	integer	Stratum area code.
distance	decimal(9,7)	Distance dredged (km).
area_wt	character(1,1)	Area weight (0=0km grid;1=1nm grid;2=3.96km grid).
Bonamia_samp	character(1,1)	*Bonamia* sp. sample taken?
Bonamia1	decimal(5,1)	Percent prevalence of *Bonamia* sp. in oysters >=58mm.
Bonamia3	decimal(5,1)	Percent prevalence of *Bonamia* sp. in oysters between 50mm and 57mm.
tot_Bonamia	decimal(5,1)	Percent prevalence of *Bonamia* sp. in all oysters.
percent_infected	decimal(5,2)	Percent of a 50 oyster sample infected with *Bonamia* sp.
prevalence	character(1,1)	Levels of prevalence of *Bonamia* sp. in all oysters (o=0;l=1-10;m=11-25;h=26+;*=no sample).
percent_new	decimal(6,2)	Percent of new clocks over new clocks+live(>50mm).
stand_live1	decimal(11,7)	Number of live oysters >= 58mm in a standard 0.37km dredge.
stand_live2	decimal(11,7)	Number of live oysters < 58mm in a standard 0.37km dredge.
stand_live3	decimal(11,7)	Number of live oysters between 50mm and 57mm in a standard 0.37km dredge.

stand_live4	decimal(11,7)	Number of live oysters between <= 50mm in a standard 0.37km dredge.
thres_live1	character(20,1)	
thres_live1a	character(20,1)	
thres_live2	character(20,1)	
thres_live2a	character(20,1)	

Creator:	dba
Referential:	No such survey (survey) INSERT t_survey (survey)
Indices:	NORMAL (2, 15) station_survey_ndx ON (survey) NORMAL (2, 15) station_vess_ndx ON (vess_id)

5.4 Table 4: t_damage

Comment: Impact of dredging on Foveaux Strait oysters as determined by damage to oysters.

Attributes	Data Type	Comment
survey	smallint	Numeric survey code (refer: t_survey).
treatment	integer	Code to describe the fishing method and the treatment of the oysters (refer: t_treatment_codes).
cell	integer	Cage cell number. 6 cells per cage.
size_group	integer	Oyster size group (1=Spat<10mm; 2=Pre-recruits 10-57mm;3=Takeable oysters>=58mm).
damage_code	integer	Code to describe damage to oyster as a result of treatment (refer: t_damage_codes).
alive_code	integer	Dead or alive code (0=Alive;1=Dead).
cluster_code	integer	A hierarchical code which describes the formation of clusters by tracing successive settlements (refer: t_cluster_codes).
no_oysters	integer	Usually takeable and pre-recruits were recorded as 1 oyster per record. Spat were assessed in groups and numbers of individuals given.

Creator: dba
Referential: No such survey (survey) INSERT t_survey (survey)
No such treatment (treatment) INSERT t_treatment_codes (treatment)
No such damage_code (damage_code) INSERT t_damage_codes (damage_code)
No such cluster_code (cluster_code) INSERT t_cluster_codes (cluster_code)

5.5 Table 5: t_treatment_codes

Comment: Table of oyster treatment codes and descriptions.

Attributes	Data Type	Comment
treatment	integer	Code to describe the fishing method and treatment of oysters at harvesting.
descrptn	character(50,1)	Description of treatment code.

Creator: dba
Indices: UNIQUE treatment_codes_pk ON (treatment)

5.6 Table 6: t_damage_codes

Comment: Table of oyster damage codes and their descriptions.

Attributes	Data Type	Comment
damage_code	integer	Code to describe damage to oyster as a result of treatment.
descrptn	character(120,1)	Description of damage.

Creator: dba
Indices: UNIQUE damage_codes_pk ON (damage_code)

5.7 Table 7: t_cluster_codes

Comment: Table of oyster cluster codes and their descriptions.

Attributes	Data Type	Comment
cluster_code	integer	A hierarchical code which describes the formation of clusters by tracing successive settlements.
descrptn	character(120,1)	Description of the cluster settlement.

Creator: dba
Indices: UNIQUE cluster_codes_pk ON (cluster_code)