

Database Documentation for

Ministry of Fisheries Marine Biodiversity & Biosecurity Database

biods

Fred Wei

NIWA Fisheries Data Management
Database Documentation Series

Last modified in May 2006

Version Control

Version	Status	Changed By	Reason	Date
0.1	draft	Fred Wei	initial version	12 Sep 2003
0.2	draft	Fred Wei	revised based on D Fisher's review	30 Sep 2003
1.0	Internal review	Fred Wei	revised based on K Mackay's review	6 Oct 2003
1.1		Fred Wei	Added Constraint column in table summaries.	7 Oct 2003
1.2		Fred Wei	Added more fields in tables based on 28 th Nov meeting.	1 Dec 2003
1.3		Fred Wei	Revised based on Don's comments and MFish's biosecurity database workshop	18 Dec 2003
1.4		Fred Wei	Revised based on real data and Graham Fenwick's feedback. Split Species into Taxon and Taxon Detail, added Habitat, Organisation and Allocation tables.	8 Oct 2004
1.5		Fred Wei	Added four new tables: t_effort_detail, t_sample_freq, t_sample_biol, t_sample_stage to cope with db trawl CE data.	11 Aug 2005
1.6		Don Robertson	Some text clarification added in relation to MAF BNZ marine biosecurity research data.	9 May 2006
		Fred Wei	Minor correction made to ERD survey & effort tables join.	10 May 2006

Contents

1. Database documentation series	4
2. Biodiversity and biosecurity data.....	4
2.1 Data Source.....	4
3. Data structures.....	5
3.1 Conceptual Data Model.....	5
3.2 Database description	7
4. Table summaries	9
5. BIODS Tables	11
5.1 Meta	11
5.2 Survey	12
5.3 Taxon_Classification	13
5.4 Taxon	13
5.5 Taxon_Detail	15
5.6 Site	16
5.7 Habitat.....	17
5.8 Effort	17
5.9 Effort_Detail	18
5.10 Sample.....	19
5.11 Sample_Freq.....	20
5.12 Sample_Stage.....	21
5.13 Sample_Biol.....	21
5.14 Organisation	23
5.15 Allocation	23
5.16 Archive.....	24
5.17 Method.....	25
6. Data Loading and Validation	26
6.1 Introduction to business rules	26
6.2 Business Rules	26
7. Acknowledgements	28
8. References	28

List of Figures

Figure 1: Entity Relationship Diagram (ERD) of the biodiversity database.	6
--	---

1. Database documentation series

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

The Ministry of Fisheries data sets incorporate 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 since 1995, and data collected by NIWA and other research providers for the Ministry of Fisheries. More recently (since 2000) research data from biosecurity initially funded by MFish then MAF Biosecurity New Zealand have also been included.

This document provides an introduction to the marine biodiversity and biosecurity database, and is a part of the database documentation series produced by NIWA.

All documents in this series include a summary of 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 and business rules. The ERD graphically shows how all the tables linked together.

This document is intended as a guide for users and administrators of the marine biodiversity and biosecurity database. This database is named as **biods**.

Access to this database is restricted to nominated personnel as specified in the Schedule 6 of the current 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.

Management of MAF Biosecurity New Zealand marine biosecurity research data are also managed in **biods** by MFish with MFish subcontracting this service to under contract to NIWA.

2. Biodiversity and biosecurity data

2.1 Data Source

2.1.1 Biodiversity

In June 2000, the New Zealand Government launched a comprehensive five-year [Biodiversity Strategy](#) which included programmes to increase our knowledge of marine biodiversity and knowledge of invasive species for maintaining biosecurity. These programmes support the strategy's goal to halt the decline in New Zealand's biodiversity and the government's strategic goal of protecting and enhancing the environment.

This marine biodiversity and biosecurity database is designed to hold any data collected within the framework of the MFish Marine Biodiversity and Biosecurity programmes and subsequent research projects funded by MAF Biosecurity New Zealand.

2.1.2 Biosecurity

Until October 2004, the Ministry of Fisheries was the main government agency with the job of managing risks to our marine environment, from exotic species that could become established in New

Zealand waters through the discharge of ballast water, as fouling on vessel hulls or by other means. Since 2004 MFish has been commissioned by Biosecurity New Zealand to contract out biosecurity research projects on its behalf.

The marine biodiversity and biosecurity database is also designed to hold data about the presence/absence and spatial distribution of all species collected on these projects including marine non-indigenous species from marine biosecurity research projects.

3. Data structures

3.1 Conceptual Data Model

The data model for the marine biodiversity and biosecurity database is based on the raw spreadsheet data format from several biodiversity and biosecurity survey projects, and is general enough to capture a broader range of biodiversity and biosecurity data. It is expandable to accommodate new requirements.

The entities *Taxon*, *Site*, *Sample* and *Method* are ideal candidates for inheritance data structures. But, because most currently used Database Management Systems (DBMS) cannot implement this feature, self-referenced structure is used instead.

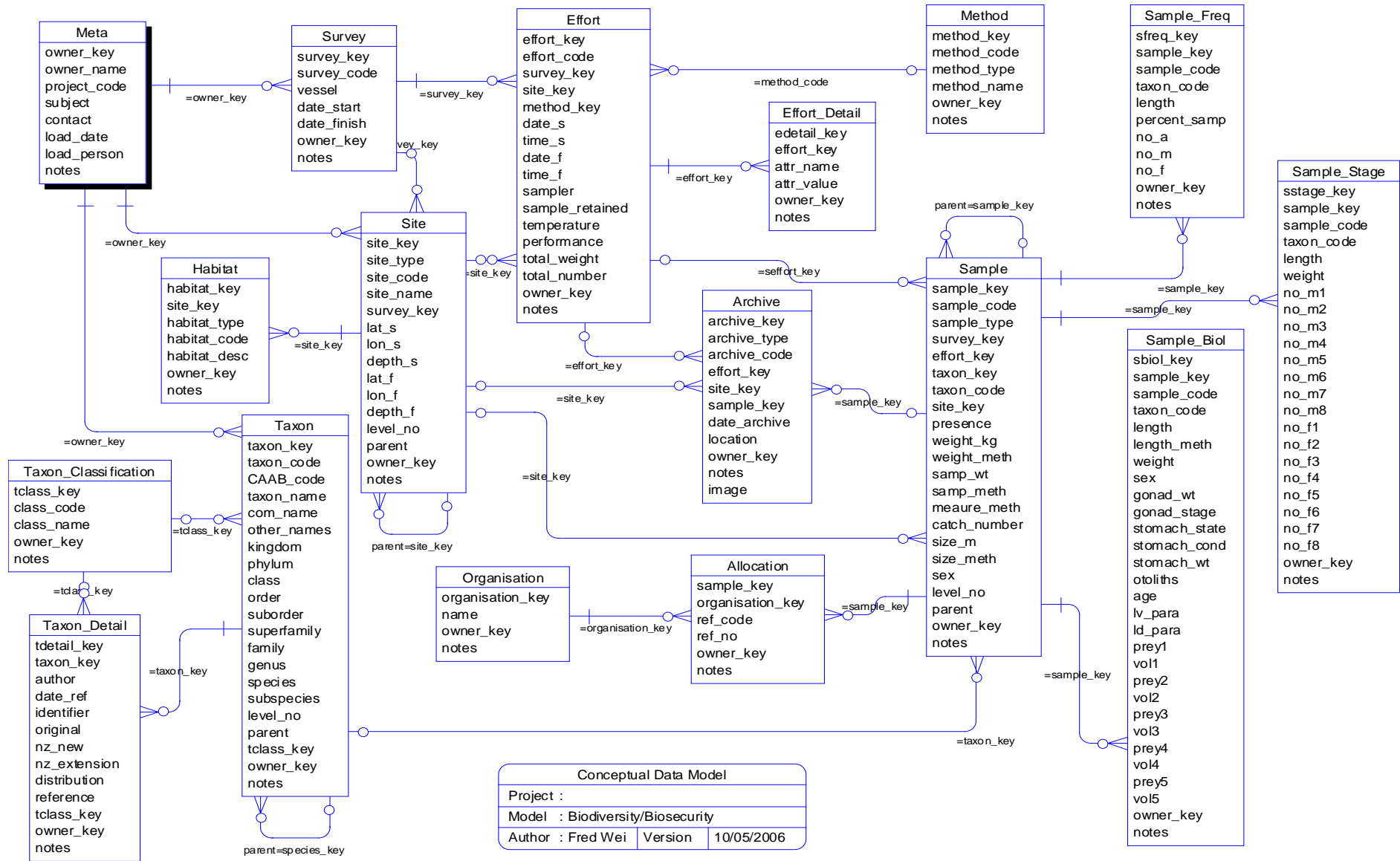


Figure 1: Entity Relationship Diagram (ERD) of the biodiversity database.

3.2 Database description

The ERD for **biods** (Figure 1) shows the logical structure of the database and its entities (each entity is implemented as a database *table*) and relationships between these. All of the table's attributes are shown in the ERD. The underlined attributes represent the table's primary key¹. This schema is valid regardless of the database system chosen, and it can remain correct even if the DBMS is changed.

Each table represents an object, event, or concept in the real world that has been selected to be represented in the database. Each *attribute* of a table is a defining property or quality of the table.

Most of the tables in the **biods** database also contain special attributes, called foreign keys².

Section 4 shows a listing of all the **biods** tables as implemented by relational DBMS. 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 format:

<i>key_name</i>	BTREE	Primary key
-----------------	-------	-------------

These keys prevent records with duplicate key values from being inserted into the table.

The **biods** database is implemented as a relational database. That is, 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. All relationships in **biods** are of the type *one-to-many*³. This is shown in the ERD by connecting a single line from the parent table; e.g., **Survey**, to the child table; e.g., **Effort**, with crows foot (indicating 'many') pointing to the child.

Note that the one-to-many relationships can be either mandatory or optional. 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

These relationships are enforced in the database by the use of Foreign Key Constraint. This kind of constraint does not allow orphans to exist in any table; i.e., where a child record exists without a related parent record. This may happen when:

- i. a parent record is deleted;
- ii. the parent record is altered so the relationship is lost;
- iii. a child record is entered without a parent record.

¹ A primary key is an attribute or a combination of attributes that contains a unique value to identify that record.

² 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.

³ 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 *Survey* can have many sampling efforts but any one sampling effort can only come from one survey.

All constraints in **biods** prevent these from occurring. Constraints are shown in the table listings by the following example:

survey_key	BTREE	Foreign key, refer to Survey(survey_key)
------------	-------	--

This means that the value of the attribute *survey_key* in the current record must already exist in the parent table **Survey** or the record will be rejected and an error message will be displayed.

All tables in the database are indexed. 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 example:

species	BTREE	Normal
---------	-------	--------

4. Table summaries

The **biods** database contains 12 tables corresponding to the 12 entities shown in the concept model ERD. All table names are the same as entity names in lower case and prefixed with t_, eg. entity Meta will be table t_meta in biods etc.

The following is a listing and brief outline of the main tables and the relationships between them in the **biods** database.

- **Meta :**
Contains meta information and setup ownership of the data. With meta records in this table, some of the data maintaining work becomes very easy.
- **Survey :**
Contains survey/trip/voyage information for sampling activities in which biods data are collected.
- **Taxon :**
Contains taxonomy information. Notice this table is self-referenced, so it can accommodate any level of taxonomic information by assigning an appropriate level number.
- **Taxon Detail :**
Contains detailed taxonomy information. One taxonomic record can have multiple records in this table.
- **Site :**
Contains site/location information for sampling activities in which biods data are collected. The relationship between **Survey** and **Site** is optional, meaning a site record may be without a survey to refer to. This reflects the reality in that sites information can be reused by more surveys.
This table is also self-referenced, so it can accommodate any combination of location relationships, e.g. port-berth-pile-quadrat, port-berth-sled-replicate, beach-stratum-transect-quadrat-replicate, etc. The hierarchy structure definition should be described in the *notes* column of the top level site record.
- **Habitat :**
Contains habitat information related to a site.
- **Effort :**
Contains sampling effort information for surveys. It can also accommodate total catch/samples resulting from each sampling effort.
- **Sample :**
Contains sample information for sampling efforts. The relationship to the **Site** table may be redundant, but in some cases it's necessary, and also improves query efficiency. Samples caught from a sampling effort can be divided into several groups based on a super species classification (like genus), these groups will be further classified based on different species. To reflect this flexible data structure, this table is also a self-referenced, the hierarchy structure definition should be described in the *notes* column of the top level sample record.
- **Archive :**
Contains specimens/files/photos catalogue information for sampling, sites, or samples.
- **Method:**
Contains method information for sampling effort.
- **Organisation :**
Contains information related to organizations, persons who have some associations with the samples

- **Allocation**
Contains sample allocation information. A sample can be allocated to multiple organizations, an organization can accept multiple samples.

5. Biods Tables

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

5.1 Meta

Contains meta information and setup ownership of the data in the **biods** database.

Attribute Name	Data Type	Constraint	Description
owner_key	integer	Primary key	Primary key for a data set owner
owner_name	character(32)		Name for the data set owner
project_code	character(32)		Project code associated with the data set
subject	character(32)		Any subject info associated with the data set
contact	character(32)		Contact persons for the data set
load_date	date		Date when the data set is loaded
load_person	character(32)		Person who loads the data set
version_desc	character(64)		Any major change description since initial loading for the data set
notes	character(256)		Any additional description for the data set

Indexed Attribute	Index type	Description
owner_key	BTREE	Primary key

5.2 Survey

Contains survey/trip/voyage information for sampling activities in which **biods** data are collected.

Attribute Name	Data Type	Constraint	Description
survey_key	integer	Primary key	Primary key for a survey
survey_code	character(32)		Artificial code created for the survey
vessel	character(32)		Vessel used for the survey, could be an ID number, code or name
date_start	date		Survey starting date
date_finish	date	later than date_start	Survey finishing date
owner_key	character(32)	Foreign key to (Meta), not null	Foreign key linking to the meta record for this survey
notes	character(256)		Any additional description for the survey

Indexed Attribute	Index type	Description
survey_key	BTREE	Primary key

5.3 Taxon_Classification

Contains taxonomic classification information.

Attribute Name	Data Type	Constraint	Description
tclass_key	integer	Primary key	Primary key of a classification
class_code	character(32)		alphanumeric code for a classification
class_name	character(64)		name for a classification
owner_key	integer	not null	Defines data ownership
notes	character(2048)		any additional information

5.4 Taxon

Contains taxonomic information for the data collected in the **biods** database. The level_no should be set correctly to reflect the taxonomic level, eg. 9 for species, 8 for genus.

Attribute Name	Data Type	Constraint	Description
taxon_key	integer	Primary key	Primary key of a species
taxon_code	character(32)		Artificial code created for the species
CAAB_code	character(8)		8-character species code: 2 digits to describe the phylum/group; 3 digits for the family within the phylum/group; 3 digits for the genus/species within the family
taxon_name	character(64)		Scientific taxonomic name
com_name	character(64)		Common taxonomic name
other_names	character(256)		Other names
kingdom	character(64)		kingdom name of this taxon
phylum	character(64)		phylum name of this taxon
class	character(64)		class name of this taxon
order	character(64)		order name of this taxon
suborder	character(64)		suborder name of this taxon
superfamily	character(64)		superfamily name of this taxon
family	character(64)		family name of this taxon
genus	character(64)		genus name of this taxon
species	character(64)		species name of this taxon
subspecies	character(64)		subspecies name of this taxon
level_no	smallint	between 1 and 10	Numbers 1 to 10 are assigned to each taxonomic level, <i>kingdom</i> =1, <i>species</i> = 9. Because the self-referenced structure of this table, this level number helps making a query

parent	integer	Foreign key to (Taxon)	Foreign key referencing to a parent record in this table. The parent's level is one level higher than children's, this is not a mandatory attribute, meaning it could have no parent
tclass_key	integer	Foreign key to (Taxon_Classification)	Foreign key linking to the classification record for this species
owner_key	character(32)	Foreign key to (Meta), not null	Foreign key linking to the meta record for this species
notes	character(1024)		Any additional description for the taxon, a hierarchy structure should be described in level 1 record.

Indexed Attribute	Index type	Description
taxon_key	BTREE	Primary key
taxon_code	BTREE	For improving query efficiency

5.5 Taxon_Detail

Contains detailed taxonomy information for the data collected in the **biods** database. One record in Taxon table can contain multiple detailed records in this table.

Attribute Name	Data Type	Constraint	Description
tdetail_key	integer	Primary key	Primary key of a taxon detail record
taxon_key	character(32)		Foreign key link to taxon
author	character(32)		Scientist who first published the species
date_ref	date		Any significant event date associated with the species
identifier	character(32)		Scientist who identified the species
original	character(32)		In which area/country the species is originated
nz_new	boolean		If the species is first found in New Zealand
nz_extension	character(512)		Distribution extension in New Zealand of the species
distribution	character(1024)		Overall distribution of this species
reference	character(1024)		Reference info associated with the species
tclass_key	integer	Foreign key to (Taxon_Classification)	Foreign key linking to the classification record for this species
owner_key	integer	not null	defines data ownership
notes	character(2048)		Any additional taxonomic detail information

Indexed Attribute	Index type	Description
tdetail_key	BTREE	Primary key

5.6 Site

Contains site/location information for sampling activities in which biodiversity data are collected.

Attribute Name	Data Type	Constraint	Description
site_key	Integer	Primary key	Primary key for a site
site_type	character(16)		Type of a site, e.g. port, pile, replication, etc.
site_code	character(32)		Artificial code created for the site
site_name	character(32)		Name of the site
survey_key	integer	Foreign key to (Survey)	Foreign key to the parent survey, this relationship is optional, the site may be independent of a survey
lat_s	decimal(8,3)		Decimalized degree of starting latitude
lon_s	decimal(8,3)		Decimalized degree of starting longitude
depth_s	decimal(8,3)		Starting depth in meter
lat_f	decimal(8,3)		Decimalized degree of finishing latitude
lon_f	decimal(8,3)		Decimalized degree of finishing longitude
depth_f	decimal(8,3)		Finishing depth in meter
level_no	smallint	between 1 and 10	Numbers assigned to each site level. Usually number 1 is given to the highest site level for a survey, e.g. port, beach. Because the self-referenced structure of this table, this level number helps making a query
parent	integer	Foreign key to (Site)	Foreign key referencing to a parent record in this table. The parent's hierarchy level is one level higher than children's, this is not a mandatory attribute, meaning it could have no parent
owner_key	character(32)	Foreign key to (Meta), not null	Foreign key linking to the meta record for this site
notes	character(256)		Any additional description for the site, a hierarchy structure should be described in level 1 record

Indexed Attribute	Index type	Description
site_key	BTREE	Primary key

5.7 Habitat

Contains habitat information related to a site.

Attribute Name	Data Type	Constraint	Description
habitat_key	integer	Primary key	Primary key for a site
site_key	integer	Foreign key to (Site)	Foreign key to the site where the habitat record associated to
habitat_type	character(32)		Type of the habitat related to this site
habitat_code	character(32)		Artificial code created for the habitat
habitat_desc	character(128)		Any descriptive text
owner_key	integer	not null	Data ownership key
notes	character(256)		Any additional description

Indexed Attribute	Index type	Description
habitat_key	BTREE	Primary key

5.8 Effort

Contains sampling effort information for surveys.

Attribute Name	Data Type	Constraint	Description
effort_key	integer	Primary key	Primary key for a sampling effort
effort_code	character(32)		Artificial code created for the effort record.
survey_key	character(32)	Foreign key to (Survey)	Foreign key to the parent survey
method_key	integer	Foreign key to (Method)	Foreign key to the method used
site_key	integer	Foreign key to (Site)	Foreign key to the site where the sampling effort occurs
date_s	date		Date when the sampling effort started
time_s	time		Time when the sampling effort started
date_f	date		Date when the sampling effort finished
time_f	time		Time when the sampling effort finished
sampler	character(32)		Person who collected the sample
sample_retained	boolean		Whether sample has been retained
temperature	decimal(5,2)		Celsius degree of temperature when this effort happens
Performance	character(32)		Performance text for the effort
total_weight	decimal(8,2)		Total sample weight in kg, redundant for easy query, could be left empty
total_number	integer		Total samples in number, redundant for easy query, could be left empty
owner_key	integer	not null	Data ownership key
notes	character(256)		Any additional description

Indexed Attribute	Index type	Description
effort_key	BTREE	Primary key

5.9 Effort_Detail

Contains sampling effort detailed information for surveys.

Attribute Name	Data Type	Constraint	Description
edetail_key	integer	Primary key	Primary key for a sampling effort detail
effort_key			Foreign key to the parent effort
attr_name	character(32)		Any detail attribute name
attr_value	character(32)	Foreign key to (Survey)	Value for the attribute
owner_key	integer	not null	Data ownership key
notes	character(256)		Any additional description

Indexed Attribute	Index type	Description
effort_key	BTREE	Primary key

5.10 Sample

Contains sample information for sampling efforts.

Attribute Name	Data Type	Constraint	Description
sample_key	integer	Primary key	Primary key for a sample or samples
sample_code	character(32)		Artificial code created for the sample/sample lot
sample_type	character(32)		Classification of sample
survey_key	integer		Duplicate link to Survey record to help query
effort_key	integer	Foreign key to (Effort)	Foreign key to the parent sampling effort
taxon_key	integer	Foreign key to (Taxon)	Foreign key to the species where the sample/samples belong to. If the samples are only classified, say to genus level, then this key will point to a genus level record
taxon_code	character(32)		Duplicate species code to help query
site_key	integer	Foreign key to (Site)	Foreign key to the site where the sample/samples are collected. This is a redundant relationship for improving query efficiency
presence	boolean		Presence of the taxon at the site.
identifier	character(32)		Who identified the sample
date_ref	date		Reference date attached to the sample
Absence	boolean		If <i>true</i> meaning that the species on the site does not found
weight_kg	decimal(10,4)		Sample weight in kg
weight_meth	character(8)		Method code for how the weight is measured
samp_wt	decimal(10,4)		sample weight (vs the total catch weight weight_kg)
samp_meth	character(1)		method code for sampling, defined in rdb
measure_meth	character(1)		measurement method code defined in rdb
catch_number	character(32)		Number of samples
size_m	decimal(6,3)		Sample size in meters
size_meth	character(8)		Method code for how the sample is measure in size
sex	character(1)	Value in M, F or U	M for male, F for female, U for unknown
level_no	smallint	between 1 and 10	Numbers assigned to each sample level. Usually number 1 is given to the highest sample level, e.g. a sample lot. Because the self-referenced structure of this table, this level number helps

			making a query
parent	integer	Foreign key to (Sample)	Foreign key referencing to a parent record in this table. The parent's hierarchy level is one level higher than children's, this is not a mandatory attribute, meaning it could have no parent
owner_key	integer	not null	Data ownership key
notes	character(256)		Any additional description for the sample, a hierarchy structure should be described in level 1 record

Indexed Attribute	Index type	Description
sample_key	BTREE	Primary key

5.11 Sample_Freq

Contains sample's length frequency information, the same structure as trawl database.

Attribute Name	Data Type	Constraint	Description
sfreq_key	integer	Primary key	Artificial integer as primary key
sample_key	integer	Foreign key to (Sample)	Foreign key to the parent sample summary
sample_code	character(32)		Artificial code created for the sample/sample lot
taxon_code	character(32)		Duplicate species code to help query
length	integer		Sample's length in cm
percent_samp	integer		Duplicate link to Survey record to help query
no_a	integer		Number of all sex samples
no_m	integer		Number of male samples
no_f	integer		Number of female samples
owner_key	integer	not null	Data ownership key
notes	character(256)		Any additional description

5.12 Sample_Stage

Contains sample's gonad stage information, the same structure as trawl database.

Attribute Name	Data Type	Constraint	Description
sstage_key	integer	Primary key	Artificial integer as primary key
sample_key	integer	Foreign key to (Sample)	Foreign key to the parent sample summary
sample_code	character(32)		Artificial code created for the sample/sample lot
taxon_code	character(32)		Duplicate species code to help query
length	integer		Sample's length in cm
weight	Decimal(7,1)		Sample's weight in gram
no_m1	integer		Number of male samples in stage 1
no_m2	integer		Number of male samples in stage 2
no_m3	integer		Number of male samples in stage 3
no_m4	integer		Number of male samples in stage 4
no_m5	integer		Number of male samples in stage 5
no_m6	integer		Number of male samples in stage 6
no_m7	integer		Number of male samples in stage 7
no_m8	integer		Number of male samples in stage 8
no_f1	integer		Number of female samples in stage 1
no_f2	integer		Number of female samples in stage 2
no_f3	integer		Number of female samples in stage 3
no_f4	integer		Number of female samples in stage 4
no_f5	integer		Number of female samples in stage 5
no_f6	integer		Number of female samples in stage 6
no_f7	integer		Number of female samples in stage 7
no_f8	integer		Number of female samples in stage 8
owner_key	integer	not null	Data ownership key
notes	character(256)		Any additional description

5.13 Sample_Biol

Contains sample's Biological data (gonad staging, stomach contents etc), the same structure as trawl database.

Attribute Name	Data Type	Constraint	Description
sbiol_key	integer	Primary key	Artificial integer as primary key
sample_key	integer	Foreign key to (Sample)	Foreign key to the parent sample summary
sample_code	character(32)		Artificial code created for the sample/sample lot
taxon_code	character(32)		Duplicate species code to help query
length	decimal(4,1)		Sample's length in cm
length_meth	character(1)		Code of method used to measure sample's length

weight	decimal(7,1)		sample weight in gram
sex	char(1)		1=male, 2=female, 3=immature or unable to determine
gonad_wt	decimal(5,1)		weight of sample gonad
gonad_stage	character(1)		1 digit code for the stage of gonad development
stomach_state	character(1)		Code used to describe the state of the stomach
stomach_cond	character(1)		Code used to describe the condition of the stomach contents
stomach_wt	integer		weight of sample stomach in gram
otoliths	character(1)		Whether otoliths taken from this sample (Y = yes; N =no)
age	character(2)		Age read from otoliths - 2-digit age or b=broken otolith, u=unreadable otolith
lv_para	integer		Parasite count on left ventral muscle tissue
ld_para	integer		Parasite count on left dorsal muscle tissue
prey1	character(3)		code for 1st species found in stomach, may also be MINITAB code
vol1	smallint		percentage volume of 1st species to total stomach content
prey2	character(3)		code for 2nd species found in stomach, may also be MINITAB code
vol2	smallint		percentage volume of 2nd species to total stomach content
prey3	character(3)		code for 3rd species found in stomach, may also be MINITAB code
vol3	smallint		percentage volume of 3rd species to total stomach content
prey4	character(3)		code for 4th species found in stomach, may also be MINITAB code
vol4	smallint		percentage volume of 4th species to total stomach content
prey5	character(3)		code for 5th species found in stomach, may also be MINITAB code
vol5	smallint		percentage volume of 5th species to total stomach content
owner_key	integer	not null	Data ownership key
notes	character(512)		Any additional description

5.14 Organisation

Contains reference information of subject bodies like organization, persons etc.

Attribute Name	Data Type	Constraint	Description
organisation_key	integer	Primary key	Primary key for an organization record
name	character(32)		Name of an organisation
owner_key	integer	not null	Data ownership key
notes	character(256)		Any additional description

Indexed Attribute	Index type	Description
organisation_key	BTREE	Primary key

5.15 Allocation

Contains allocation information of samples to organizations or persons etc.

Attribute Name	Data Type	Constraint	Description
sample_key	integer	Primary key	Point to the primary key of a sample or sample lot
organisation_key	integer	Primary key	Point to the primary key of an organization
ref_code	character(32)		Reference code of this allocation
ref_no	integer		Reference number of this allocation
owner_key	integer	not null	Data ownership key
notes	character(256)		Any additional description

Indexed Attribute	Index type	Description
sample_key	BTREE	Primary key
organisation_key	BTREE	Primary key

5.16 Archive

Contains specimens/files/photos catalogue information for sampling, sites, or samples.

Attribute Name	Data Type	Constraint	Description
archive_key	integer	Primary key	Primary key for an archive
archive_type	character(32)		Type of an archive, file, specimen, photo, etc
archive_code	character(16)		Artificial code created for the archive
effort_key	integer	Foreign key to (Effort)	Foreign key to the sampling effort this archive relates to
site_key	integer	Foreign key to (Site)	Foreign key to the site this archive relates to
sample_key	integer	Foreign key to (Sample)	Foreign key to the sample this archive relates to
date_archive	decimal(8,3)		Date of archiving
location	character(64)		Where this archive is located, can also be used to point to the file path of this archive
owner_key	integer	not null	Data ownership key
notes	character(1024)		Any additional description for the archive
image	bytea		Binary type, used to store inline image file.

Indexed Attribute	Index type	Description
archive_key	BTREE	Primary key

5.17 Method

Contains method information for sampling effort.

Attribute Name	Data Type	Constraint	Description
method_key	integer	Primary key	Primary key for a method
method_code	character(64)	unique	Code of the method
method_type	character(64)		Type of the method
method_name	character(64)		Name of the method
owner_key	integer	not null	Data ownership key
notes	character(256)		Any additional description for the method

Indexed Attribute	Index type	Description
method_key	BTREE	Primary key
method_code	BTREE	unique constraint

6. Data Loading and Validation

There is no data loading scripts created yet. A general tool to load data from MS Excel spread sheets or plain text files into the database is considered suitable for this purpose.

6.1 Introduction to business rules

The following are a list of business rules applying to the **biods** database. A business rule is a written statement specifying what the information system must do or how it must be structured. In this instance the information system is any system that is designed to handle biodiversity data.

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 process.

6.2 Business Rules

Customized validation rules are yet to be defined

Data model integrity is ensured by database built-in constraints. If the term ‘must’ is used then a database constraint is enforced. If the term ‘should’ is used then the business rule may not be enforced with a database constraint, and is most likely enforced on the application layer.

6.2.1 Meta

- *owner_key*: must have a value and the value must be unique.

6.2.2 Survey

- *survey_key*: must have a value and the value must be unique.
- *owner_key*: must have a value and the value must exist in *owner_key* of table **Meta**.

6.2.3 Taxon

- *taxon_key*. must have a value and the value must be unique.
- *owner_key*: must have a value and the value must exist in *owner_key* of table **Meta**.
- *parent*: may have a value and the value must exist in *species_key* of table **Species**.
- *level_no*: must have a value, and the value must between 1 and 10.
- the combinations of taxonomic names and level number should be unique, e.g. (taxon, level_no).

6.2.4 Taxon Detail

- *detail_key*. must have a value and the value must be unique.

6.2.5 Site

- *site_key*: must have a value and the value must be unique.
- *owner_key*: may have a value and the value must exist in *owner_key* of table **Meta**.
- *survey_key*: may have a value and the value must exist *survey_key* of table **Survey**.
- *parent*: may have a value and the value must exist in *site_key* of table **Site**.

- *level_no*: must have a value, and the value must be between 1 and 10.
- *site_code*: must be unique.

6.2.6 Effort

- *effort_key*: must have a value and the value must be unique
- *survey_key*: must have a value and the value must exist in *survey_key* of table **Survey**.
- *site_key*: must have a value and the value must exist in *site_key* table **Site**.
- *method_code*: must have a value and the value must exist in *method_code* of table **Method**.

6.2.7 Sample

- *sample_key*: must have a value and the value must be unique
- *effort_key*: must have a value and the value must exist in *effort_key* of table **Effort**.
- *site_key*: may have a value and the value must exist in *site_key* table **Site**.
- *taxon_key*: may have a value and the value must exist in *taxon_key* of table **Taxon**.
- *parent*: may have a value and the value must exist in *sample_key* of table **Sample**.
- *level_no*: must have a value, and the value must be between 1 and 10.
- *sex*: may have a value, and the value must be in M, F and U.

6.2.8 Organisation

- *organisation_key*: must have a value and the value must be unique

6.2.9 Allocation

- *sample_key*: must have a value and the value must exist in *sample_key* of table **Sample**.
- *organisation_key*: must have a value and the value must exist in *organization_key* of table **Organisation**, the combination of (*sample_key* and *organization_key*) must be unique.

6.2.10 Archive

- *archive_key*: must have a value and the value must be unique
- *effort_key*: may have a value and the value must exist in *effort_key* of table **Effort**.
- *site_key*: may have a value and the value must exist in *site_key* table **Site**.
- *sample_key*: may have a value and the value must exist in *sample_key* of table **Sample**.

6.2.11 Method

- *method_code*: must have a value and the value must be unique
- *parent*: may have a value and the value must exist in *method_code* of table **Method**.
- *level_no*: must have a value, and the value must be between 1 and 10.

7. Acknowledgements

The author would like to thank Don Robertson and David Fisher for reviewing and editorial comment, Kevin MacKay for his technical comment and for writing section 2, Graeme Inglis for the initial port survey data model and port survey sampling data and Peter Marriott for his knowledge on field work.

8. References

Ng, S. 1992: Standards for setting up databases and their applications. *MAF Fisheries Greta Point Internal Report No. 180*. 31p.