Database documentation: hmetal

K. A. Mackay

NIWA Internal Report
2001

Revised on 30 January 2001

Contents

List of Figures

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 heavy metal sampling programme database **hmetal**, 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 **hmetal** 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 Heavy Metal Sampling Programme

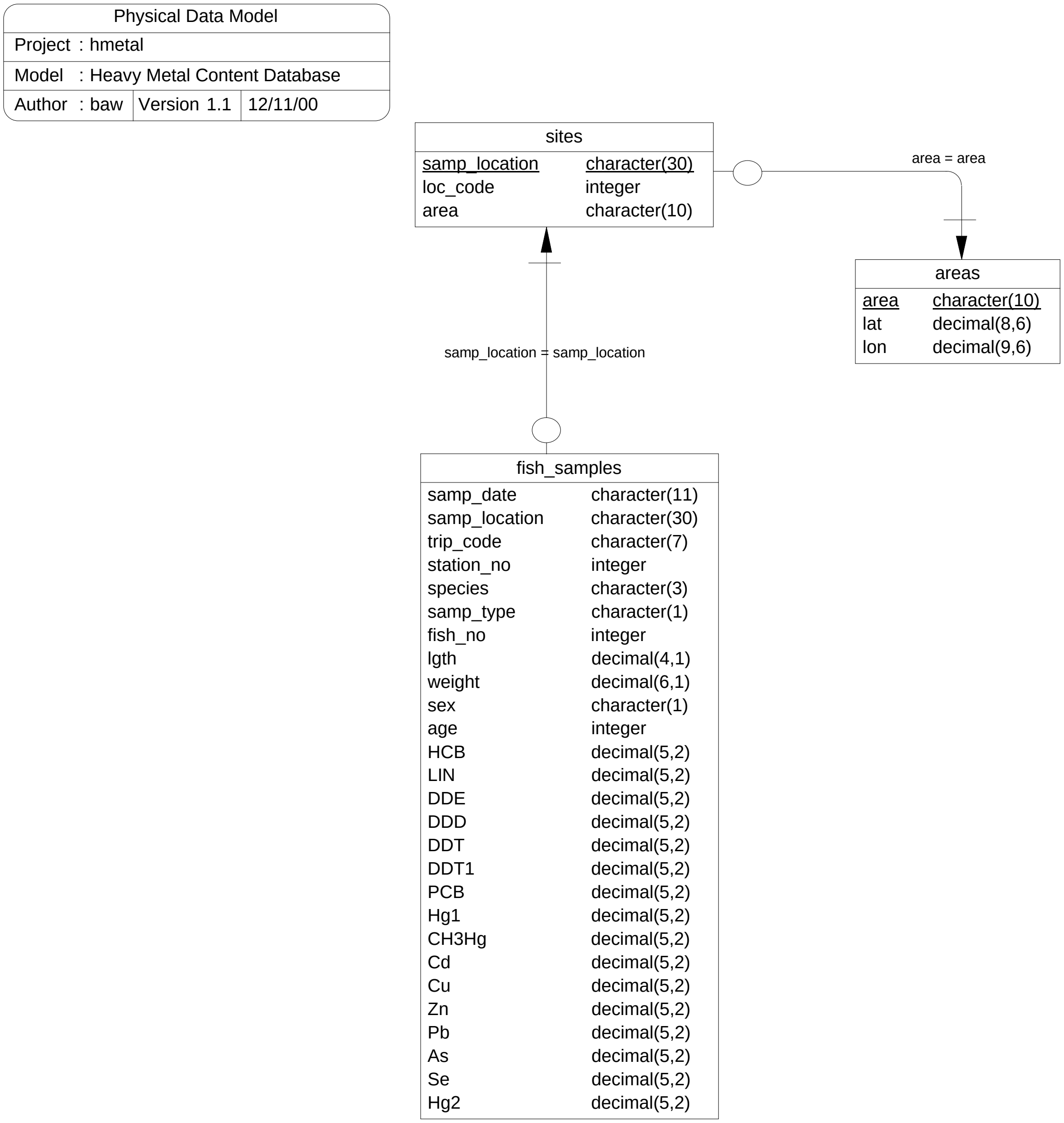
Before 1984, MAF Fisheries co-ordinated the collection of tissue samples for heavy metal analysis of commercial fish and shellfish species. This was to provide heavy metal level guidelines for health requirements and for the MAF export certification system. By 1984 most commercial species had been analysed for heavy metals and the formal heavy metal sampling programme was wound down. Total heavy metal levels in several species of commercially important fish have been summarised by Fenaughty et al. (1988).

In 1990, a further forty-eight black cardinalfish (*Epigonus telescopus*) were sampled from a major commercial fishery off the east coast of the North Island to determine their flesh mercury (Hg) levels. The results from this study have been published in Tracey (1990).

Muscle tissue was taken from the left dorsal side of fish specimens, directly posterior to the operculum, except with skate (*Raja* spp.), where flesh was taken from the wings, stargazer (*Kathetostama giganteum*), where flesh was taken from the posterior body, and southern bluefin tuna (*Thunnus maccoyi*), from which both tail and belly flesh were taken. An additional subcutaneous muscle sample was taken from some orange roughy (*Hoplostethus atlanticus*). Mantle, digestive gland, and gonad tissue were sampled from squids; the whole organism was homogenised for shellfish analysis.

Total mercury was analysed by the method of Robertson et al. (1975), with minor modifications (van den Broek et al. 1981). Later samples were generally acid-digested then analysed for Hg using cold-vapour atomic absorption spectrometry along with a reference sample of a known Hg value (Louie et al. 1985). The analytical techniques used for determining cadmium, copper, zinc, and lead concentrations were those of Gorsuch (1959) and Brooks and Rumsey (1974). Selenium levels were analysed by the method of Watkinson (1966). Organochlorine insecticide levels were examined by the technique of Solly and Harrison (1972).

Lengths and weights were not recorded for all samples. In some instances only partial lengths and weights (i.e., of trunked and gutted fish) were recorded, particularly with factory samples. Lengths were measured by the appropriate measurement method (see Appendix).



Physical Data Model		
Project : hmetal		
Model : Heavy Metal Content Database		
Author : baw	Version 1.1	12/11/00

sites	
<u>samp_location</u>	<u>character(30)</u>
loc_code	integer
area	character(10)

areas	
<u>area</u>	<u>character(10)</u>
lat	decimal(8,6)
lon	decimal(9,6)

fish_samples	
samp_date	character(11)
samp_location	character(30)
trip_code	character(7)
station_no	integer
species	character(3)
samp_type	character(1)
fish_no	integer
lgth	decimal(4,1)
weight	decimal(6,1)
sex	character(1)
age	integer
HCB	decimal(5,2)
LIN	decimal(5,2)
DDE	decimal(5,2)
DDD	decimal(5,2)
DDT	decimal(5,2)
DDT1	decimal(5,2)
PCB	decimal(5,2)
Hg1	decimal(5,2)
CH3Hg	decimal(5,2)
Cd	decimal(5,2)
Cu	decimal(5,2)
Zn	decimal(5,2)
Pb	decimal(5,2)
As	decimal(5,2)
Se	decimal(5,2)
Hg2	decimal(5,2)

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

3 Data Structures

3.1 Table relationships

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

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

The **hmetal** 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 **hmetal**: one-to-many[4]. These relationships can be seen in ERDs by connecting a single line (indicating 'many') from the child table; e.g., *fish_samples*, to the parent table; e.g., *sites,* 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 *sites* and it's child table *fish_samples.* The symbol "O" by the child *fish_samples* means that a catch site can have zero or many sample records, while the bar by the parent *sites* means that for every sample there must be a matching catch site record.

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

---

[1] Also known as a database *schema.*

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

[3] A foreign key is an attribute or a combination of attributes that is a primary key in another table.

[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 site in *sites* can have many fish samples in *fish_samples* but one fish sample can only come from one catch site.

[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 *fish_samples*:

```
Referential:      Invalid sample location (samp_location) INSERT
                  sites (samp_location)
```

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

```
*** User Error: insert constraint ‘Invalid sample location’ 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 **hmetal** tables as implemented by the Empress RDBMS. As can be seen in the listing of the tables, a table’s primary key has an unique index on it. Primary keys are generally listed using the following format:

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

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 area code.

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

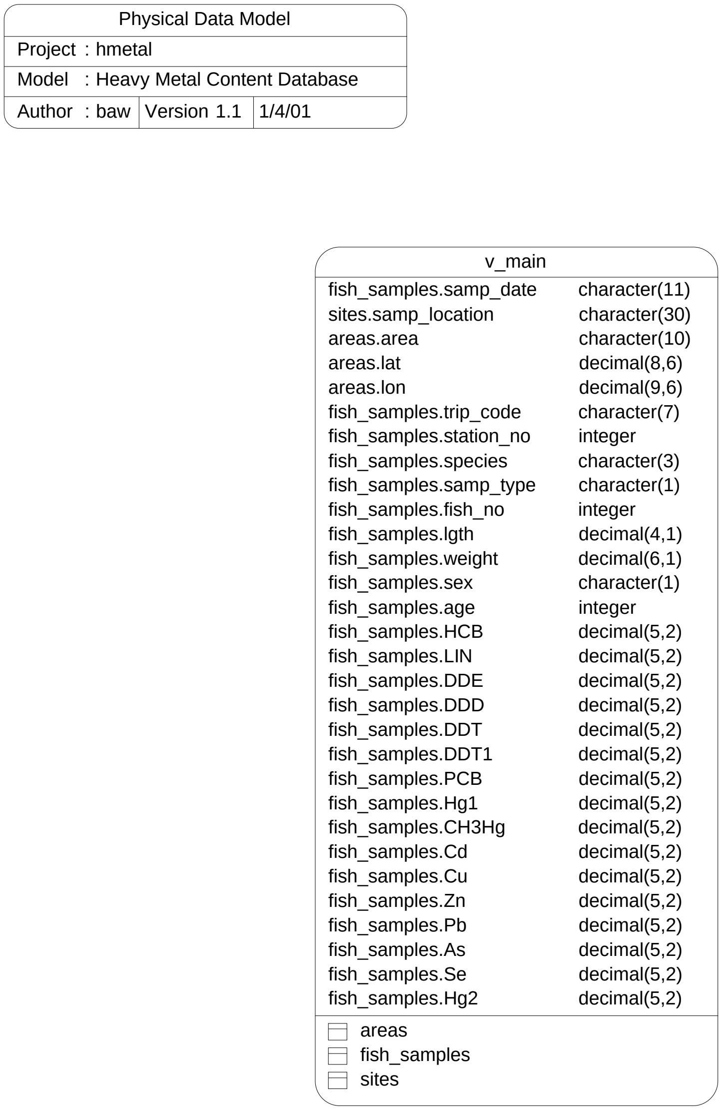
The **hmetal** database is constructed around two main entities: a sample site and a fish sample. A third entity has been added to record a spatial component to these data.

The first table *sites* (Table 1) contains area codes and location details of sites from which fish samples were taken for analysis.

In order to allow spatial analysis of the data, a third table of positions *areas* (Table 2) was created. This table assigns each of the sample sites with a latitude and longitude based on the value of the *area* attribute.

Individual fish samples are recorded in the *fish_samples* table (Table 3). Details recorded include sample date and location, species code, fish length, weight, sex, and amounts of various organochlorine insecticides and heavy metals such as DDT, zinc, arsenic, lead, and mercury. All amounts were recorded in as total mg per kg^{-1} muscle tissue. Fish samples are linked to the *sites* table by the attribute *samp_location*.

A view *v_main* (Figure 2) allows users to perform a denormalised multi-table join on all these tables to select complete records, which include details fish samples, their sample sites, and positions.



Physical Data Model		
Project : hmetal		
Model : Heavy Metal Content Database		
Author : baw	Version 1.1	1/4/01

v_main	
fish_samples.samp_date	character(11)
sites.samp_location	character(30)
areas.area	character(10)
areas.lat	decimal(8,6)
areas.lon	decimal(9,6)
fish_samples.trip_code	character(7)
fish_samples.station_no	integer
fish_samples.species	character(3)
fish_samples.samp_type	character(1)
fish_samples.fish_no	integer
fish_samples.lgth	decimal(4,1)
fish_samples.weight	decimal(6,1)
fish_samples.sex	character(1)
fish_samples.age	integer
fish_samples.HCB	decimal(5,2)
fish_samples.LIN	decimal(5,2)
fish_samples.DDE	decimal(5,2)
fish_samples.DDD	decimal(5,2)
fish_samples.DDT	decimal(5,2)
fish_samples.DDT1	decimal(5,2)
fish_samples.PCB	decimal(5,2)
fish_samples.Hg1	decimal(5,2)
fish_samples.CH3Hg	decimal(5,2)
fish_samples.Cd	decimal(5,2)
fish_samples.Cu	decimal(5,2)
fish_samples.Zn	decimal(5,2)
fish_samples.Pb	decimal(5,2)
fish_samples.As	decimal(5,2)
fish_samples.Se	decimal(5,2)
fish_samples.Hg2	decimal(5,2)

- areas
- fish_samples
- sites

Figure 2: ERD of the view v_main

4 Table Summaries

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

1. **sites :** contains heavy metal sample site location details including location name and area code.
2. **areas :** contains all area codes used in the sites table and their spatial coordinates expressed as latitude and longitude.
3. **fish_samples:** contains details for individual fish including species, fish length, weight, sex, and amounts of various organochlorine insecticides and heavy metals such as DDT, zinc, arsenic, lead, and mercury.
4. **v_main :** is a view that provides a denormalised view of all the tables in the **hmetal** database.

5 hmetal Tables

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

5.1 Table 1: sites

Comment: Table of heavy metal sample location site names and their area codes.

Attributes	Data Type	Null?	Comment
samp_location	character(30,1)	No	Brief description of the sample site location.
loc_code	integer	No	Unique site number.
area	character(10,1)		Area code. Used to link to latitude and longitude positions in the areas table.

Creator: dba
Referential: Invalid area code (area) INSERT areas (area)
Indices: NORMAL (2, 15) BTREE ON (area)
UNIQUE BTREE sites_PK ON (samp_location)

5.2 Table 2: areas

Comment: Table of all area codes used in the sites table and their spatial coordinates expressed as latitude and longitude.

Attributes	Data Type	Null?	Comment
area	character(10,1)		Area code
lat	decimal(8,6)		Latitude (decimal degrees; 90 to –90; -ve for southern hemisphere)
lon	decimal(9,6)		Longitude (decimal degrees; 0 to 360)

Creator: dba
Indices: NORMAL (2, 15) BTREE ON (lat)
NORMAL (2, 15) BTREE ON (lon)
UNIQUE BTREE areas_PK ON (area)

5.3 Table 3: fish_samples

Comment: Table of individual fish details including species code, fish length, weight, sex, and amounts of various organochlorine insecticides and heavy metals such as DDT, zinc, arsenic, lead, and mercury.

Attributes	**Data Type**	**Null?**	**Comment**
samp_date	character(11,1)	No	Date that the fish was sampled. Generally recorded in DD MON YYYY format but when dates were uncertain, just the month or even just the year was entered.
samp_location	character(30,1)	No	Location from which the sampled fish was caught. Refer to the sites table.
trip_code	character(7,1)	No	7-character trip code from which the sampled fish was caught. Some trip codes may link to the trawl database. Trips form unknown sources are prefixed by "xxx".
station_no	integer		Station number of a trip from which the sampled fish was caught. Can be used to link to the trawl database.
species	character(3,1)	No	3-character species code. Refer to the curr_spp table.
samp_type	character(1,1) smatch “[BDFGKLMPTW]”	No	1-character sample type code: B=belly; D=digestive gland; F=flesh; G=gonad; M=mantle; T=tentacle or tail; W=whole.
fish_no	integer		Individual fish number. May relate to records in trawl:t_fish_bio.
lgth	decimal(4,1)		Fish length (cm).
weight	decimal(6,1)		Fish weight (grams).
sex	character(1,1) smatch “[FMU]”		1-character code for the sex of the samples fish: M=male; F=female; U=unsexed.
age	integer		Age (years) of the sampled fish.
HCB	decimal(5,2)		Amount of Hexachlorobenzene measured in the sample (mg.kg-1).
LIN	decimal(5,2)		Amount of Lindane measured in the sample ($mg.kg^{-1}$).
DDE	decimal(5,2)		Amount of DDE measured in the sample ($mg.kg^{-1}$).

fish_samples (cont…)

Attributes	**Data Type**	**Null?**	**Comment**
DDD	decimal(5,2)		Amount of DDD measured in the sample (mg.kg^{-1}).
DDT	decimal(5,2)		Amount of DDT measured in the sample (mg.kg^{-1}).
DDT1	decimal(5,2)		Amount of total DDT (i.e., DDT, DDD, and DDE) measured in the sample (mg.kg^{-1}).
PCB	decimal(5,2)		Amount of PCB measured in the sample (mg.kg^{-1}).
Hg1	decimal(5,2)		Amount of Mercury (1st sample) measured in the sample (mg.kg^{-1}).
CH3Hg	decimal(5,2)		Amount of CH_3Hg measured in the sample (mg.kg^{-1}).
Cd	decimal(5,2)		Amount of Cadmium measured in the sample (mg.kg^{-1}).
Cu	decimal(5,2)		Amount of Copper measured in the sample (mg.kg^{-1}).
Zn	decimal(5,2)		Amount of Zinc measured in the sample (mg.kg^{-1}).
Pb	decimal(5,2)		Amount of Lead measured in the sample (mg.kg^{-1}).
As	decimal(5,2)		Amount of Arsenic measured in the sample (mg.kg^{-1}).
Se	decimal(5,2)		Amount of Selenium measured in the sample (mg.kg^{-1}).
Hg2	decimal(5,2)		Amount of Mercury (2nd sample) measured in the sample (mg.kg^{-1}).

Creator: dba
Referential: Invalid sample location (samp_location) INSERT sites (samp_location)
Invalid species code (species) INSERT curr_spp (code)
Indices: NORMAL (2, 15) BTREE ON (samp_location)

The following listing is a view on the union of the tables *sties*, *areas*, and *fish_samples*.

5.3.1 v_main

Comment: View on union of of all the tables to provide a denormalised view of the hmetal database.

View:

```
select attr 'samp_date', attr 'samp_location', attr 'area', attr
'lat', attr 'lon', attr 'trip_code', attr 'station_no', attr
'species', attr 'samp_type', attr 'fish_no', attr 'lgth', attr
'weight', attr 'sex', attr 'age', attr 'HCB', attr 'LIN', attr
'DDE', attr 'DDD', attr 'DDT', attr 'DDT1', attr 'PCB', attr 'Hg1',
attr 'CH3Hg', attr 'Cd', attr 'Cu', attr 'Zn', attr 'Pb', attr 'Se',
attr 'Hg2' from 'fish_samples' 'f', 'areas' 'a', 'sites' 's' where
((attr 'f'.'samp_location' = attr 's'.'samp_location' and attr
's'.'area' = attr 'a'.'area'))
```

Attributes	**Data Type**	**Null?**
samp_date	character(11,1)	No
samp_location	character(30,1)	No
area	character(10,1)	
lat	decimal(8,6)	
lon	decimal(9,6)	
trip_code	character(7,1)	No
station_no	integer	
species	character(3,1)	No
samp_type	character(1,1)	No
fish_no	integer	
lgth	decimal(4,1)	
weight	decimal(6,1)	
sex	character(1,1)	
age	integer	
HCB	decimal(5,2)	
LIN	decimal(5,2)	
DDE	decimal(5,2)	
DDD	decimal(5,2)	
DDT	decimal(5,2)	
DDT1	decimal(5,2)	
PCB	decimal(5,2)	

Attributes	**Data Type**	**Null?**
Hg1	decimal(5,2)	
CH3Hg	decimal(5,2)	
Cd	decimal(5,2)	
Cu	decimal(5,2)	
Zn	decimal(5,2)	
Pb	decimal(5,2)	
Se	decimal(5,2)	
Hg2	decimal(5,2)	

Creator: dba

6 References

Brooks, R.R., and Rumsey, D. 1974: Heavy metals in some New Zealand commercial sea fishes. *New Zealand Journal of Marine and Freshwater Research. 8*. pp 155-166.

Fenaughty, C.M., Tracey, D.M., and Lock, J.W. 1988: Heavy metal and organochlorine concentrations in New Zealand aquatic fish, crustaceans, and molluscs. *New Zealand Fisheries Data Report 34.* 44p.

Gorsuch, T.T. 1959: Radiochemical investigations on the recovery for analysis of trace elements in organic and biological materials. *The Analyst. 84*. pp 135-173.

Louie, H.W., Go, D., Fedezina, M., Judd, K., and Dalins, J. 1985: Digestion of food samples to total mercury determination. *Journal of Associated Official Analytical Chemistry*. *68*. pp 891-893.

Robertson, T., Waugh, G.D., and Mol, J.C.M. 1975: Mercury levels in some New Zealand snapper, *Chrysophrys auratus*. *New Zealand Journal of Marine and Freshwater Research. 9*. pp 265-272.

Solly, S.R.B., and Harrison, D.L. 1972: DDT in some New Zealand marine and freshwater fauna. *New Zealand Journal of Marine and Freshwater Research. 6*. pp 456-462.

Tracey, D.M. 1993: Mercury levels in black cardinalfish (*Epigonus telescopus*). *New Zealand Journal of Marine and Freshwater Research. 27*. pp 177-181.

Van den Broek, W.L.F., and Tracey, D.M. 1981: Concentration and distribution of mercury in flesh of orange roughy (*Hoplostethus atlanticus*). *Journal of Marine and Freshwater Research. 15*. pp 137-146.

Watkinson, J.H. 1966: Flourometric determination of selenium in biological material with 2,3-diaminonapthaleme. *Analytical Chemistry*. *38*. pp 92-97.

Appendix

Measurement conventions for species length as recorded in the **hmetal** database. Species codes are enclosed in parenthesis “( )”.

Fork length	Barracouta (BAR), bluenose (BNS), brown trout (BTR), alfonsino (BYX), blue mackerel (EMA), ribaldo (RIB), anchovy (ANC), pilchard (PIL), jack mackerel (JMA), kahawai (KAH), kingfish (KIN), orange perch (OPE), parore (PAR), rainbow trout (RTR), quinnat salmon (SAM), southern blue whiting (SBW), gemfish (SKI), porae (POR), skipjack (SKJ), snapper (SNA), silverside (SSI), southern bluefin tuna (STN), silver warehou (SWA), tarakihi (TAR), trevally (TRE), common warehou (WAR), white warehou (WWA), black cardinalfish (EPT)
Total length	Basss/grouper (BAS), blue cod (BCO), brill (BRI), New Zealand sole (ESO), sand flounder (SFL), black flounder (BFL), black oreo (BOE), conger eel (CON), gurnard (GUR), hake (HAK), hapuku (HAP), hoki (HOK), javelin fish (JAV), john dory (JDO), lookdown dory (LDO), leatherjacket (LEA), long-finned eel (LFE), ling (LIN), red cod (RCO), short-finned eel (SFE), sea perch (SPE), smooth oreo (SSO), stargazer (STA), seal shark (BSH), Owston’s spiny dogfish (CYO), elephant fish (ELE), Lucifer dogfish (ETL), ghost shark (GSH), rig (SPO), pale ghost shark (GSP), mako shark (MAK), northern spiny dogfish (NSD), school shark (SCH), shovelnose spiny dogfish (SND), spiny dogfish (SPD), witch (WIT)
Standard length	Orange roughy (ORH)
Rear of eye to tail fork	Striped marlin (STM)
Wing width	Skate (SKA, RSK)
Mantle length	Arrow squid (ASQ), giant squid (GSQ)
Carapace length	Rock lobster (CRA)
Carapace width	Paddle crab (PAD)
Longest dimension of shell	Bluff oyster (OYS), pacific oyster (POY), paua (PAU), blue mussel (MSB), green-lipped mussel (MSG), scallop (SCA)