OMEX I DATABASE CONTENTS

Introduction

We first of all need to clarify what is meant by the 'OMEX I Database' in the context of this document. It is perfectly valid to describe the whole of the CD-ROM as the 'database', including the manual which includes material such as images that are data in their own right.

However, a significant subset of the OMEX I data has been organised into an integrated structure within a relational database. It is to these data that this document refers.

The database contains a vast amount of data. In terms of sheer volume it occupies some 100 Mbytes in a relatively compact format. This doubles when the overheads associated with storage in a relational database management system are added. However, it is not the volume of the database that is its most impressive statistic, but the diversity of the oceanographic data it contains.

The data model underlying the database is based on the concept of events. An event is defined as an action that results in the generation of oceanographic data. Events can be many things ranging from the spectacular such as a mooring deployment to the trivial such as turning on a tap to collect a sample from the non-toxic supply.

Broadly speaking, events are of two types. Point events are events that may be considered to relate to a single position, like CTD profiles, vertical net hauls or corer deployments. Traverse events are events that occur along a significant distance and generally pertain to oceanographic hardware that is towed by a ship.

However, it should be noted that the division between point and traverse events is not clear cut depending upon the scale at which the events are viewed. During a CTD cast, the ship is never completely still and at high resolution the CTD event could be considered as a traverse event. Likewise, an instrument tow on a tight grid within a small box could be considered a point event.

Within the database, event types are assigned on the basis of common sense with a view to making the interpretation of the data as easy as possible. This inevitably results in events being classified as point events wherever possible.

For the purposes of this document, the data within the database are grouped into broad categories designed with the way in which data are viewed by the oceanographic community very much in mind.

The following groupings are used:

Event Inventory

Data Dictionaries

Water Column Profile Data

Pelagic Biology Data

Benthic Data

Particle Flux Data

Current Measurements

Rate Measurements

Atmospheric Chemistry Data

As with any classification, there are grey areas where groups overlap or where the groupings based on oceanographic criteria do not map cleanly with optimum data storage structures. Care has been taken to identify these problem areas in the text.

The objective of this document is to provide an overview of the different types of data held in the database and to provide guidance on where to look for them.

Event Inventory

The Event Inventory is probably the most important component of the database. In addition to its inventory function, it stores data on event attributes, such as space and time co-ordinates. Obviously, without this information the data in the database would be useless.

The bulk of the inventory information is held in the EVENT table which contains times, positions and other ancillary information. This is supported by table EVENT_COMM which provides storage for plain language comments and G_CODE which defines mnemonics used in table EVENT. The table CRSINDX provides additional information on the cruises associated with the events.

In addition to this primary index of events, there is also a series of secondary indices that contain additional information that is specific to a particular type of oceanographic data. This information could have been stored in EVENT but it would have resulted in an unmanageable number of columns. However, these data are often of equal importance to the data stored in EVENT. For example sample depths are stored in table BOTTLE and without these, sample data would be useless.

The following secondary indices are present in the database.

ADCP data	ADCPINDX
Water and air samples	BOTTLE
Benthic data	COREINDX, CORESAMP
CTD data	CTDINDX, CTDCAL
Pelagic biology	NETINDX
Radiometer profiles	PRINDX
SeaSoar data	SSINDX
Sediment trap data	STINDX
XBT data	XBTINDX

Instrument type codes used in table CTDINDX are defined in table CTDTYP.

Data Dictionaries

The major data tables in the database use coded fields to store information. The most important of these fields are those identifying the parameter measured and the data originator.

The coding convention used for parameters is defined by a group of tables known as the Parameter Dictionary, namely ZUSG, ZUPM and ZUNT. The bulk of the code definition is stored in ZUSG, including the parameter name and the protocol used to measure it. The units in which the parameter is stored may be found by obtaining a coded field from table ZUPM that is defined in table ZUNT.

There is another table in the Parameter Dictionary called ZUCT but this is only of use to those wishing to assign parameter codes to data (i.e. BODC personnel).

Data originators are identified by simple numeric codes that are defined in table ORGCODE.

Water Column Profile Data

There are several different types of water column profile data held in the database. Broadly, these fall into two different types namely sample profiles and instrumental profiles.

Sample profiles, often known as 'bottle data', may be found in table BOTDATA. Many different parameters are stored here and as a general rule if any type of measurement made on a discrete water sample is required then this is the place to look.

BOTDATA has a highly normalised structure with a record for each parameter measured on each sample by a given originator. The possibility (and actuality) that more than one originator might measure the same parameter on the same sample gives rise to a nasty complication (a straightforward cross-tabulated query fails) when retrieving the data. Because of this, it is recommended that retrieval of the data from BOTDATA should only be attempted using the Database Explorer software.

A small amount of integrated profile data (size-fractionated chlorophylls from a couple of cruises) may be found in table INTBOT.

The database contains several types of instrumental water column profile. XBT profiles may be found in table XBT. Profiling radiometer data may be found in table PRPROF. However, there are a small number of CTD casts in the database which have downwelling and upwelling irradiance channels included. These are not included in PRPROF and must be obtained from table BINCTD.

Table BINCTD is predominantly concerned with the storage of CTD profiles. The data set includes all the parameters measured by the CTD package except for nephelometer data (but transmissometer data are included). Nephelometer data are held separately in table NEPH. This subdivision is due to a design weakness in the BODC system that is easy to identify with hindsight but somewhat more difficult to fix.

However, BINCTD also contains SeaSoar profiles. The concept of a SeaSoar profile may seem a little strange to those used to handling SeaSoar data as continuous time/depth series. However, it has been found extremely convenient to reduce these series to profiles by gridding and then taking each column of the grid as a profile. Not only does the data volume reduce to more manageable proportions but the result readily integrates with CTD data. Hence the common storage in table BINCTD.

Profiles of marine snow abundance and particle volume may be found in table MSP.

Pelagic Biology Data

Pelagic biology data may be found in several places in the database. Broadly speaking, there are three types of data that may be found in three different places.

Possibly the most extensive pelagic biology data set in the OMEX database is the Continuous Plankton Recorder data which provides a time series from 1964 to 1995. This may be found in tables CPR_COLOUR, CPR_ZOO and CPR_PHYTO.

Secondly, there is the pelagic biology data set that has been obtained through the deployment of 'nets', namely the Longhurst-Hardy Plankton Recorder and the Rectangular Mid-Water Trawl. The taxon-specific data from these are held in tables LHPR and RMT. Total population displacement volume and biomass may be found in table NETINDX.

Thirdly, significant quantities of pelagic biology data were measured on water samples, such as phytoplankton and zooplankton taxa enumeration and microzooplankton data. These are stored in table BOTDATA as water column parameters. This table should generally only be interrogated using the Database Explorer software. BOTDATA also holds large amounts of biogeochemical parameters that may be of interest.

Pelagic biologists may also be interested to know that the underway ADCP data, held in table ADCP, include a signal amplitude channel that may be used to estimate the distribution of biomass.

Benthic Data

The benthic data may be considered as geological data (including biogeochemistry) and biological data. The geological data may be found in one of two tables, namely CORETOT and COREPROF. COREPROF contains the more usual type of benthic geological data that take the form of profiles along the length of cores. Its content is both extensive and diverse and includes several hundred different parameters. CORETOT contains parameters that pertain to the whole of the core, such as fluxes and overall accumulation rates, or data from grabs. It includes a relatively small amount of data.

Both of these tables have the same fully normalised structure as table BOTDATA. Because this includes records for each parameter measured on a sample by each originator, conventional retrievals will fail if more than one originator has measured the same parameter on the same sample. However, unlike BOTDATA this is a possibility and not believed to be an actuality. Therefore conventional cross-tabulated retrievals may be attempted but most users will find it much easier to use the Database Explorer software that will do the job just as well.

The biological benthic data are held in a series of bespoke tables as insufficient time was available within the OMEX data management project to generate the large number of parameter codes necessary for the data to be included in COREPROF. The following tables will be of interest to those concerned with benthic biology.

Live and dead foraminifera	FORAMS
Meiofauna	MEIOHDR and MEIODAT
Macrofauna	MFHEAD and MFDAT
Megafauna	MEGAHEAD and MEGADAT

Particle Flux Data

By particle flux data we really mean sediment trap data. These may all be found in table TRAPDATA which contains over 50 different parameters.

This table has the same fully normalised structure as table BOTDATA. Because this includes records for each parameter measured on a sample by each originator, conventional retrievals will fail if more than one originator has measured the same parameter on the same sample. However, unlike BOTDATA this is a possibility and not believed to be an actuality. Therefore conventional cross-tabulated retrievals may be attempted but most users will find it much easier to use the Database Explorer software that will do the job just as well.

Current Measurements

The database includes Lagrangian current measurements in the form of drifting buoy or sediment trap tracks. These may be found in table ARGOS.

The second form of current data in the database are underway ADCP data which may be found in table ADCP.

Most of the moored instrument deployments during OMEX were concerned with the measurement of currents. Whilst the data are not held in the relational database, an inventory of the data series may be found in table MOORINDX. Mnemonics used in this table are defined in table MOOR_PARAMS.

Rate Measurements

The rate measurements in the database are measurements of uptake kinetics by the particulate phase from the dissolved phase. These consist of what are broadly termed production experiments together with some measurements of trace metal uptake kinetics.

Wherever possible, rate measurements have been treated as water sample parameters and are stored in table BOTDATA. Thymidine and leucine incorporation data will be found here as will 'per hour' carbon, nitrogen and phosphorus uptake data and parameterised (alpha and pmax) P:I data.

BOTDATA has a normalised structure which is complicated by the fact that different originators have measured the same parameter on the same sample. It is therefore recommended that it be interrogated using the Database Explorer software.

In a number of cases, such as 24 hour incubation experiments, time series experiments and full P:I profiles, it has proved impossible to store the data in BOTDATA with sufficient information for them to be of use. In these cases, bespoke table structures have been used.

The following tables are involved:

Trace metal uptake kinetics	MTALHDR and MTALDAT
^{14}C production experiments	C14HDR and C14DAT
^{15}N uptake experiments	N15HDR and N15DAT
$^{32}P/^{33}P$ uptake experiments	P33HDR, P33DARK and P33DAT

Atmospheric Chemistry Data

Atmospheric chemistry data are sample measurements that have a lot in common with water sample data. The only difference is that the sample analysed is water and not air. Consequently, they are stored together with the water sample data in table BOTDATA. The two types of data may be distinguished by the gear codes in the linked records in table EVENT or by the fact that air samples have negative 'depths' in table BOTTLE.

BOTDATA has a normalised structure which is complicated by the fact that different originators have measured the same parameter on the same sample. It is therefore recommended that it be interrogated using the Database Explorer software.