THE LOIS SES DATABASE

Introduction

The SES database contains all the data collected during the SES field programme except for the underway and moored instrument data sets. This is a large and complex data set and this inevitably leads to some complexity in the structure of the database. However, tools have been provided to ease the task of getting what you need from the database.

The database is presented on the CD-ROM in both ASCII 'kit form' and as a complete implementation in Microsoft JET 3.0 and 3.5 formats. The JET formats are fully compatible with Microsoft *Access* 7.0 and 8.0 (*Office97*) respectively. An earlier JET format may be used with a later version of *Access* providing that the database file is copied from CD-ROM onto a disk that isn't read-only and that database objects are not redefined. JET format may also be read by many other packages, either directly or through Open Database Connectivity (ODBC) interfaces.

This document is designed to lead you to the data you require in the database. We will therefore start by telling you what is there in terms of types of data stored, followed by a description of the tools you may use to obtain the data. The information is presented in the following series of documents.

Database Contents

This provides an overview of the types of data held within the database and should be the starting point for any user of the database.

Database Data Documentation

Extensive documentation has been compiled on the protocols used to collect of the SES data set together with issues of data quality. This information is **vital**. The burden of deciding whether the data you extract is 'fit for purpose' for your application is placed on you, the user. **Ignore this documentation at your peril.**

Structure of the Database

The database is relational and therefore consists of a series of tables. Documentation is provided that summarises what is held in each table, defines the table fields and specifies the relationships between the tables.

Using the BODC Database Explorer

The BODC Database Explorer is a *Windows* application that provides easy access to the database contents.

Using Microsoft Access to Explore the Database

A copy of the database is included in two variants of Microsoft JET format. This document explains how users who have a copy of *Access* available may use this package to interrogate the database.

Using the 'Kit Form' Database

The 'kit form' database concept has been a feature of project data sets electronically published on CD-ROM by BODC. Essentially, it is a relational database dumped as a set of flat ASCII files, including all the key fields that link the tables together. The 'kit form' database is designed to accommodate those users who wish to import all or part of the database into their own relational database management system.

SES DATABASE CONTENTS

Introduction

We first need to clarify what is meant by the ‘SES 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 SES 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 over 150 Mbytes in a relatively compact format. This increases significantly 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 resulted 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, mooring deployments or corer deployments. Traverse events are events that occur along a significant distance and generally pertain to oceanographic hardware that was 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

Benthic Data

Particle Flux Data

Current Measurements

Production Data

Settling Velocity Tube Experiments

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

In addition to this primary event index, there are secondary indices containing 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
Instrument profiles	PRINDEX
SeaSoar profiles	SSINDX
Sediment trap data	STINDX
SVT experiments	SVTINDX
Tidal constituents	TIDAL_CONS_HEAD
Moored instruments	MOORINDX

Instrument type codes used in table CTDINDX and PRINDEX are defined in table CTDTYP and bottle type codes are stored in table BOTYPINDX. Parameter codes used in MOORINDX are defined in MOOR_PARAMS.

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 parameters 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. This is only useful 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 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.

The database contains several types of instrumental water-column profile. XBT, profiling radiometer, FLY probe and sound velocity profiles may be found in table PRDATA. This has a normalised structure like BOTDATA, but the nature of the data is simpler and cross-tabulated queries work. Consequently, these data may be interrogated using either *Access* or the Database Explorer. Additional optical data (CTD downwelling and upwelling 2-pi PAR data) are held in 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. BINCTD also contains SeaSoar profiles. These are individual legs of the SeaSoar oscillations stored in this manner to facilitate the integration of SeaSoar and CTD data.

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

Benthic Data

The benthic data include geological data (including biogeochemistry) and biological data. All benthic data may be found in either CORETOT or COREPROF. COREPROF contains along-core profile data. 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 does the job just as well.

Particle Flux Data

By particle flux data we really mean sediment trap data. These may all be found in table TRAPDATA, which contains a dozen 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 does the job just as well.

Current Measurements

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

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

Many of the moored instrument deployments during SES 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.

The SES moored ADCP data have been tidally analysed. The results of these analyses are stored in table TIDAL_CONS_DATA.

Production Data

Two types of production experiment were carried out during SES. Size-fractionated primary production was determined directly by on-deck incubation experiments. The data from these are stored in C14HDR and C14DAT.

The photosynthetic parameters alpha and Pmax were determined by P:I experiments. These are stored in table BOTDATA, which has a normalised structure 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.

Settling Velocity Tube Experiments

During SES, settling velocity tube (SVT) experiments were carried out to determine the sinking rates of suspended particulate matter and pigment-bearing particles. The results are held in tables SVTOTAL and SVSAMP. SVSAMP contains the particle concentration time series that form the basis of the experimental results. SVTOTAL contains the parameters derived from the resulting cumulative settling curves, such as settling rates.

Using the 'Kit-Form' Database

'Kit-Form' Database Concept

All relational databases are made up of tables. If such a database is broken down into its component tables, the result is a set of objects that have little or no value as discrete entities. However, these objects may be regarded as a kit of parts that may be assembled to produce a fully functional database. This is the fundamental concept behind the 'kit-form' database.

The kit consists of a series of files. Each file contains a table from the database as a totally portable ASCII file. The result is a dump of the database that may be easily incorporated into any relational database management system on any platform.

Files and Formats

The files of the 'kit-form' database may be found in the DBKIT directory. The names of the files generally conform to the name of the table in the database to which they correspond. The only exceptions are a small number of tables whose names were longer than 8 bytes. These have been truncated to give DOS-compatible filenames but may be unambiguously associated with table names in the database structural documentation. All files have the file extension CSV.

The file format is very simple and straightforward. The files are in standard ASCII code. Note that as the CD-ROM was assembled on a PC, the record terminators are 2-byte CRLF sequences, not the single-byte terminators found on UNIX systems.

The first record of each file contains the table field names separated by commas. The second and subsequent records contain the table data with one row per record. Obviously, the order of the fields in the data records corresponds to the order of the field names in the header.

In general, text fields have no delimiters except for the leading and trailing commas. However, the text fields from some tables included embedded commas. In these cases, the text fields are enclosed by double quote (") symbols.

This file format is sometimes known as 'comma separated value' or 'CSV' format.

Suggested Applications for the 'Kit-Form' Database

The 'kit-form' concept was conceived to provide a totally portable export mechanism for BODC's project databases. The 'target user' was someone who wished to import data in bulk into their own database, either preserving the BODC structures or reformatting some or all of the data into an alternative schema. This has been done with considerable success by a number of users from 'kit-form' databases on previous BODC electronic publications.

However, this is not the only way in which the 'kit-form' database files may be used. Part of the software interfaces for the North Sea Project and BOFS CD-ROMs actually used the 'kit-form' database files. It is therefore perfectly possible for users who do not have access to relational database management systems to develop conventional application programs against 'kit-form' database files. They are, after all, just a set of very ordinary data files.

One final suggestion. Spreadsheet users might find it useful to load up the EVENT.CSV file. This is an inventory of the database events. Most modern spreadsheet packages have limited database query capabilities. If these are applied to the data from EVENT, users can get a feel of properties of the data such as 'what came from where' and therefore determine whether the data set is of interest.

The BODC Database Explorer

The BODC Database Explorer is a *Windows* application that allows data from the database to be retrieved in a grid format that may be exported to other applications, such as spreadsheets. It has been tested successfully under *Windows95, Windows98* and *Windows NT 4.0*.

The program is designed to support one or more BODC CD-ROMs containing JET 3.0 (*Access* 7.0) databases providing the project-specific installation program has been run for each CD-ROM to be used. It includes full information on its use through an on-line help system, including functional descriptions of all the menu options and control buttons. However, a brief description of how to get started is included here.

When the program is launched through either the BODC entry in the Start menu, a shortcut or *Windows Explorer*, a splash screen is briefly displayed followed by the opening of the program control window. The following actions are then required to display data.

- Select the Open Project option from the File menu and choose the project appropriate to the CD-ROM currently loaded.

- Click on the Define option in the Data menu and choose the type of data required.

- Use the Selection dialog presented to choose the subset of stations you want and the header fields that you require on each row of the grid.

- Click on the Show option in the Data menu.

- Use the dialogs provided to select the parameters you wish to include in your grid. These cover three hierarchical levels that describe the parameters in increasing detail. If you are unsure which of the higher-level categories to include, then err on the side of inclusion rather then exclusion. You can always reject unwanted parameters at the more detailed levels that follow.

- The header parameters included in the grid may be modified, if required, by choosing the Index Fields option from the Data menu.

This is all you need to do to access the data. Control over how the data are presented is provided through both the menus and the toolbar buttons. Consult the on-line help or simply experiment to discover what these can do.