

The OMEX II Database

Introduction

The OMEX database contains all the data collected during the OMEX II field programme except for the underway and moored instrument data sets. Many different types of data are included, which 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 formats compatible with *Access95*, *Access97* and *Access2000*. The data 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, followed by a description of the tools you may use to obtain the data. The information is presented in the following 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 in the collection of the OMEX II data set together with issues of data quality. This information is **vital**. The quality of the OMEX II data is variable and the protocols used by the different partners vary significantly. 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 describes 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. Documentation is provided on what this program can do and how to use it.

Using Microsoft Access to Explore the Database

The database is included in 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.

OMEX II Database Contents

Introduction

We first need to clarify what the 'OMEX II Database' in the context of this document means. It is perfectly valid to describe the whole of the CD-ROM as the 'database', including this manual which includes material such as images that are data in their own right. However, a significant subset of the OMEX II data has been organised into an integrated structure within a relational database. It is these data that are referred to here.

The database data model is based on the concept of events, which are defined as actions that result in the generation of oceanographic data. Events range from the spectacular such as a mooring deployment to the trivial such as turning on a tap to collect a water sample.

The foundation of the database is therefore an inventory of events, including their fundamental attributes such as time and position. The oceanographic data are linked onto this framework of events through a series of secondary index tables that manage the one-to-many relationships between events and data.

The database makes use of alphanumeric codes to avoid replicating large text fields. These are defined internally through a series of code tables or dictionaries. The most significant of these is the BODC Parameter Dictionary that currently includes over 6,000 parameter code definitions.

Getting Hold of the Data

There are three ways of obtaining data from the database:

- Through the BODC Database Explorer program
- Through Microsoft Access
- From the ASCII 'kit-form' database files

The entire database is accessible from the 'kit-form' database files and these may be perused using standard software such as spreadsheets or even word processors. However, you should only really think about using the 'kit-form' database if you are not working with *Windows* or you want to get the data into an RDBMS other than *Access*.

For *Windows* users, the two tools available are best suited to different types of data:

Database Explorer Software

The Database Explorer is the recommended mechanism for obtaining the following types of data:

- CTD data
- Water bottle profiles
- Benthic data
- Sediment trap data
- Net haul data
- XBT, FLY probe and radiometer profiles
- Underway ADCP data
- Drifting buoy data
- Long incubation (24-hour) carbon, nitrogen and phosphorus production data (other production data are water bottle parameters)
- Phosphate uptake kinetics data
- CPR data
- Integrated water column profile data

Microsoft Access

The following types of data may only be obtained by opening the database file with an appropriate version of Microsoft Access. A comprehensive set of forms are provided, but users with the full (as opposed to run-time) version of Access are free to copy the data to hard disk and create queries and forms to suit their own requirements.

- Events
- CTD calibration details
- Parameter Dictionary
- Inventory of moored instrument data

Forms have also been included in the database to provide a secondary interface to the data types covered by the Database Explorer. However, most users will find these a less convenient data delivery mechanism for most data types.

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 conform to the name of the table in the database to which they correspond. 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.

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 than 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.