THE OMEX I UNDERWAY DATA SET

Introduction

During a significant number of OMEX I cruises, the ship's computers automatically logged a range of sensors sampling surface sea water, together with meteorological instruments. The files described in this section, termed the underway data set, contain these data merged on a common time base with sampling intervals ranging from 30 seconds to 10 minutes (1 minute for most cruises).

The data files may be found in the UNDERWAY directory on the CD-ROM. This contains one file per cruise leg hat are named using the convention:

cruise_leg_mnemonic.BMM

The data are stored in BODC's Binary Merge Format. This is a binary format that may not be listed or printed. However, a software interface, the BODC Underway Explorer, is provided for *Windows95* users and a format specification is given to allow users of other operating systems to develop applications to access the data.

Each data file is accompanied by a data document that is included in this *Acrobat* manual.

A fundamental principle of Binary Merge Format and BODC's management of underway data files is that each data value is assigned a single character quality control flag. This provides the only quality control mechanism: any suspect data values, including total garbage, are labelled by a quality control flag set to 'S'. Problem data are not deleted. Consequently, **these flags must not be ignored.**

The underway section of this manual includes the following information:

Data Set Contents

A summary table is provided listing the cruise legs for which data are available, including their start and end dates and the parameters measured.

Database Data Documentation

A data document has been prepared for each cruise leg. This describes the instrumentation, the data processing and calibration protocols employed and any problems with the data noted by either the originators or BODC. 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.**

Using the BODC Underway Explorer

The BODC Underway Explorer is a *Windows95* application that allows data from the underway files to be presented as time series plots and listed in a data grid that may be exported to other applications. The program also provides an indication of the spatial context of the data through a map of the cruise track overlaid on a coastline and bathymetric contours.

Binary Merge Format Specification

This section provides a technical specification of the Binary Merge Format used for the data files. It provides sufficient information for users to be able to write their own applications for handling data in this format.

OMEX I Underway Data Set Contents

The OMEX I underway data set contains data from the following cruise legs. Note that the hieroglyphics in the 'Channels' section are Binary Merge Format parameter codes that may be found in the format specification.

Cruise leg:	Belgica BG9309
Dates:	19/04/1993 13:54 to 06/05/1993 07:20
Sampling:	60 seconds
Channels:	ABCFJOZaY1QM{EH
Cruise leg:	Belgica BG9322A
Dates:	21/09/1993 19:42 to 29/09/1993 15:53
Sampling:	60 seconds
Channels:	ABCFTWYZEHM{Q1Oa
Cruise leg:	Belgica BG9322B
Dates:	03/10/1993 08:26 to 06/10/1993 07:16
Sampling:	60 seconds
Channels:	ABCFYZEHMDG{Q1Oa
Cruise leg:	Belgica BG9412
Dates:	20/04/1994 07:20 to 05/05/1994 13:25
Sampling:	60 seconds
Channels:	ABCDFJOZaY1MQEHTW{G
Cruise leg:	Belgica BG94ZB
Dates:	11/04/1994 11:45 to 14/04/1994 03:55
Sampling:	300 seconds
Channels:	ABQDGFCM{EH
Cruise leg:	Belgica BG9506
Dates:	03/03/1995 09:13 to 17/03/1995 08:35
Sampling:	60 seconds
Channels:	ABCFJOYZ1DaGM{EHQ
Cruise leg:	Belgica BG9521
Dates:	11/09/1995 10:36 to 19/09/1995 09:30
Sampling:	60 seconds
Channels:	ABCFJOYZ1DaGM{EHQ
Cruise leg:	Belgica BG9522
Dates:	22/09/1995 07:14 to 29/09/1995 07:14
Sampling:	60 seconds
Channels:	ABCFGJOYZ1DaM{EHQ

Cruise leg: Charles Darwin CD83
Dates: 13/12/1993 10:00 to 13/01/1994 08:20
Sampling: 60 seconds
Channels: ABKJCF?IL!

Cruise leg: Charles Darwin CD84
Dates: 18/01/1994 12:48 to 01/02/1994 18:30
Sampling: 30 seconds
Channels: ABKCF?IJL!

Cruise leg: Charles Darwin CD85
Dates: 11/04/1994 08:04 to 07/05/1994 06:56
Sampling: 60 seconds
Channels: ABFJKC?IL!

Cruise leg: Charles Darwin CD86
Dates: 20/05/1994 00:16 to 12/06/1994 08:14
Sampling: 60 seconds
Channels: AB#()KJ

Cruise leg: Charles Darwin CD94
Dates: 03/06/1995 05:45 to 20/06/1995 07:53
Sampling: 60 seconds
Channels: ABCFIJK?L!

Cruise leg: Charles Darwin CD97
Dates: 03/11/1995 00:00 to 06/11/1995 09:38
Sampling: 60 seconds
Channels: ABCIKL?!F

Cruise leg: Discovery DI216
Dates: 26/08/1995 17:00 to 12/09/1995 06:52
Sampling: 60 seconds
Channels: ABKYZd1JCF?I!taO

Cruise leg: Discovery DI217
Dates: 27/09/1995 07:00 to 21/10/1995 16:30
Sampling: 60 seconds
Channels: ABKJd1YZaC?FO!tI

Cruise leg: Meteor M27_1
Dates: 31/12/1993 15:07 to 17/01/1994 07:13
Sampling: 60 seconds
Channels: ABJYZ1OoCFacd

Cruise leg: Meteor M30_1
Dates: 06/09/1994 14:42 to 20/09/1994 07:06
Sampling: 60 seconds
Channels: ABJYZ1OCFoadcw

Cruise leg:	Poseidon PS200_7
Dates:	25/06/1993 14:01 to 01/07/1993 18:13
Sampling:	120 seconds
Channels:	ABCFK#(

Cruise leg:	Poseidon PS211
Dates:	31/08/1995 12:58 to 11/09/1995 10:53
Sampling:	60 seconds
Channels:	ABCFYZ1abEu

Cruise leg:	Valdivia VLD137
Dates:	27/06/1993 08:20 to 11/07/1993 06:20
Sampling:	600 seconds
Channels:	ABYZabcO1K

Binary Merge Format Specification

Binary Merge Format is a binary format for the compact storage of high volume time series data. The format was initially developed for use on an IBM main-frame and subsequently adapted for use on UNIX workstations and PCs.

The file structure comprises a single header record followed by the datacycles. All the data on the CD-ROM in Binary Merge Format have a regular time channel with a sampling interval of between 30 seconds and 10 minutes.

The structure of the header record is:

Cruise identifier	-	12-byte character
Pointer to first data record	-	4-byte integer
Pointer to last data record	-	4-byte integer
Number of data channels excluding date and time (always present)	-	4-byte integer
Processing status mask	-	4-byte integer
Data source indicator	-	4-byte integer
Project indicator word	-	4-byte integer
Padding	-	set to binary zero
Channel identifiers	-	1 byte per flagged channel

The cruise identifier is of the form cruise_mnemonic/yy where yy is the year in which the data were collected. It is stored in ASCII character code.

The processing status mask indicates the data processing operations to which the data have been subjected. The principle of a bit mask is that each bit in the word is given a specialised meaning. In the description of the meanings of each bit below, the description is true when the bit is set on. The bit numbering convention used is 1 (most significant) through 32 (least significant).

The bit meanings are:

1 - Thermosalinograph salinity calibrated
2 - Thermosalinograh temperature calibrated
3 - Navigation checked and gaps filled by interpolation
4 - Unassigned
5 - Unassigned
6 - Unassigned
7 - Unassigned
8 - Transmissometer converted from voltage to attenuance

9 - Unassigned
10 - Unassigned
11 - Unassigned
12 - Unassigned
13 - Unassigned
14 - Unassigned
15 - Unassigned
16 - Phosphate baseline correction applied
17 - Phosphate calibrated
18 - Nitrate calibrated
19 - Nitrite calibrated
20 - Silicate calibrated
21 - Silicate additional drift correction applied
22 - Ammonia calibrated
23 - File has been workstation screened
24 - Irradiance channels calibrated
25 - Nitrate baseline corrected
26 - Nitrite baseline corrected
27 - Silicate baseline corrected
28 - Ammonia baseline corrected
29 - Urea baseline corrected
30 - Urea channel calibrated
31 - Unassigned
32 - Unassigned

The data source and project indicator words have no relevance to the data stored on the CD-ROM. They will always be set to zero and one respectively.

The padding words are included to ensure that the header contains the same number of bytes as the data records which follow. Consequently, the number of words of padding depends upon the number of data channels (it is in fact the number of data channels minus 7).

The channel identifiers are single characters, encoded in ASCII, which specify the channels (other than date and time which are always present) in the file. The order of the identifiers in the header specifies the order of the data channels in the data records.

The channel identifiers are defined as follows:

A = Latitude (deg +ve N)
B = Longitude (deg +ve E)
C = Temperature (°C)
D = Raw fluorescence from Turner Designs through-flow fluorometer
E = pCO_2 (μatm)
F = Salinity (PSU)
G = Chlorophyll from Turner Designs (mg/m^3)
H = TCO_2 (μmol/kg)
I = Optical attenuance (per m)

J = Bathymetric depth (m)
K = Distance run (km)
L = Photosynthetically available irradiance (W/m^2)
M = pH (pH units)
N = Temperature of the pH determination (°C)
O = Solar radiance (W/m^2)
P = Ammonia (µM)
Q = Dissolved oxygen at in-situ temperature and salinity (µM)
R = Oxygen saturation (%)
S = Density anomaly (kg/m^3)
T = Nitrate + nitrite (µM)
U = Nitrite (µM)
V = Phosphate (µM)
W = Silicate (µM)
X = Null channel
Y = Absolute wind speed (knots)
Z = Absolute wind direction (degrees from which the wind blows)
1 = Barometric pressure (mb)
2 = Dry bulb air temperature from port bridge sensor (°C)
3 = Wet bulb air temperature from port bridge sensor (°C)
4 = Dry bulb air temperature from starboard bridge sensor (°C)
5 = Wet bulb air temperature from starboard bridge sensor (°C)
6 = Dry bulb air temperature from starboard mast sensor (°C)
7 = Wet bulb air temperature from starboard mast sensor (°C)
8 = Significant wave height, H_s (m)
9 = Urea (µM)
/ = Long wave radiation (W/m^2)
+ = Port solar radiance (W/m^2)
- = Starboard solar radiance (W/m^2)
* = Port photosynthetically available irradiance (W/m^2)
: = Starboard photosynthetically available irradiance (W/m^2)
? = Raw signal from Chelsea Instruments Aquatracka fluorometer (V)
! = Chlorophyll from Aquatracka (mg/m^3)
_ = Attenuance calibrated in terms of calcite (mg Ca/m^3)
{ = Potentiometric alkalinity (µEq/kg)
> = Dry bulb air temperature from port mast sensor (°C)
< = Wet bulb air temperature from port mast sensor (°C)
a = Combined dry bulb air temperature (°C)
b = Combined wet bulb air temperature (°C)
c = Dew point (°C)
d = Relative humidity (%)
e = Downwelling long wave radiation (W/m^2)
f = Upwelling long wave radiation (W/m^2)
g = Absolute port wind speed (knots)
h = Absolute port wind direction (degrees from which the wind blows)
i = Absolute starboard wind speed (knots)
j = Absolute starboard wind direction (degrees from which the wind blows)
k = Port dew point (°C)
l = Starboard dew point (°C)

m = Port relative humidity (%)
n = Starboard relative humidity (%)
o = UV radiation (W/m^2)
p = Masthead photosynthetically available irradiance (W/m^2)
q = Vertical wind velocity (knots +ve upwards)
r = Port photosynthetically available radiance (W/m^2)
s = Starboard photosynthetically available radiance (W/m^2)
t = Photosynthetically available radiance (W/m^2)
u = Atmospheric pCO_2 (μatm)
v = Low detection ammonium (nM)
w = Atmospheric particle count (per cm^3)

At the right hand end of the header record are up to 3 blank padding bytes to ensure that the record length is a multiple of 4 bytes (to allow it to be specified in terms of words). The same number of padding bytes is also added to each datacycle record.

Each datacycle contains the date (word 1), time (word 2), the data values (words 3 to number of channels plus 2) and their flags. Date is stored in binary integer form as a 'Loch day number', defined as the number of days elapsed since the start of the Gregorian calendar. Time is stored in IEEE binary floating point representation (as used on UNIX systems) as a day fraction (06:00 = 0.25, 12:00 = 0.5 etc.).

The data values are stored as IEEE binary floating point numbers in the order prescribed by the channel identifiers in the header. At the rightmost end of the record are the data quality control flags occupying one byte each. The flag definitions used are as follows:

B - Bad data
G - Good data
I - Interpolated data
N - Null data
S - Suspect data
U - Data outside range of calibration

The main problem awaiting those who wish to access the Binary Merge files without the assistance of the software interface provided is the conversion of 'Loch day numbers' into calendar dates. IEEE floating point structure is rapidly establishing itself as a de facto standard and therefore should not prove to be a problem.

The following subroutine listings, one in FORTRAN and one in Pascal, convert a Loch day number into year, month and day.

```
    SUBROUTINE CMAADY(IDY,IDATE)
C#S   ****************************************************************
C      TITLE  S CMAADY   VR  1.0   AUTHOR  MDBS/SGL   DATE  79FEB01
C
C      S/R calculates date given the number of (complete) days since
C      1760.01.01  (= 0 days elapsed). Not valid for 21'st century
C#E   ****************************************************************
C
C  ARGUMENTS
C  ---------
C
C   IDY      -     No. of complete days elapsed
CO  IDATE    -     3 element array containing 1) year, 2) month
C                  and 3) day of month
C
      DIMENSION IDATE(3)
      INTEGER MONTH(12)/0,31,59,90,120,151,181,212,243,273,304,334/
      IC  =  0
C
      IDYC0 =  IDY - 51133
      ICEN  =  19
  10    IF(IDYC0.GT.0) GO TO 20
        ICEN  =  ICEN - 1
        IDYC0  =  IDYC0 + 36524
        GO TO 10
  20  IDYC  =  IDYC0 - 365
      IF(IDYC.GT.0) GO TO 30
      IDYC = IDYC0
      IDATE(1)  =  ICEN*100
      GO TO 40
C
C  NOW DETERMINE NO OF FULL LEAP YEAR CYCLES PRESENT
C
   30 NLPYR  =  (IDYC-1)/1461
      IYR  =  4*NLPYR
      IDYC  =  IDYC - NLPYR*1461
      IYRX  =  (IDYC-1)/365
      IF(IYRX.EQ.4) IYRX = 3
      IDYC  =  IDYC -  IYRX*365
      IDATE(1)  =  IYR + IYRX + ICEN*100 + 1
      IF(IYRX.EQ.3) IC = 1
C
C  NOW GET MONTH AND DAY
C
   40 L  =  13
      DO  50  I =1,12
        L  =  L - 1
        IF(L.EQ.2) IC = 0
        MN  =  MONTH(L) + IC
        IF(IDYC.GT.MN) GO TO 60
   50   CONTINUE
C
   60 IDATE(2) = L
      IDATE(3) = IDYC - MN
C
      RETURN
      END
```

```
Type
  IntArr = Record
             Y : Word;
             M,D : Byte
           End;
Const
  MonthSum : Array[1..12] of Word=(0,31,59,90,120,151,181,212,243,273,304,334);
Procedure Cmaady(Var Idy : LongInt; Var tDate : IntArr);
{     S/R calculates date given the number of (complete) days since  }
{     1760.01.01  (= 0 days elapsed). Not valid for 21'st century    }
Var
  iC,iCen,nLpYr,iYr : LongInt;
  mn,i,L,iYrx : LongInt;
  iDyc,iDyc0 : LongInt;
Begin
  IC := 0;
  IDYC0 :=  IDY - 51133;
  ICEN :=  19;
  While IDYC0 <= 0 Do
  Begin
    ICEN := ICEN - 1;
    IDYC0 := IDYC0 + 36524
  End;
  IDYC := IDYC0 - 365;
  IF IDYC<=0 Then
  Begin
    IDYC:= IDYC0;
    tDate.Y:= ICEN*100
  End
  Else
  Begin
    {  NOW DETERMINE NO OF FULL LEAP YEAR CYCLES PRESENT }
    NLPYR := Trunc((IDYC-1)/1461);
    IYR  :=  4*NLPYR;
    IDYC  :=  IDYC - NLPYR*1461;
    IYRX  :=  Trunc((iDYC-1)/365);
    IF IYRX=4 Then
      IYRX := 3;
    IDYC  :=  IDYC - IYRX*365;
    tDate.Y :=  IYR + IYRX + ICEN*100 + 1;
    IF IYRX=3 Then
      IC := 1
  End;
  { NOW GET MONTH AND DAY }
  L := 13;
  I:=1;
  While I<13 Do
  Begin
    L := L-1;
    IF L=2 Then
      IC:= 0;
    MN := MonthSum[L] + IC;
    IF IDYC>MN Then
      I:=13
    Else
      Inc(I)
  End;
  tDate.M:= L;
  tDate.D:= Integer(IDYC) - Mn
End;
```