

Appendix 1

Data Management Requirements and Specifications

Subsurface Monitoring Unit
MC252 Deepwater Horizon Oil Spill

Prepared by:

NOAA

EPA

BP

Updated Sunday, September 18, 2010

Version 4.05

ALL CHANGES TO THESE DATA MANAGEMENT REQUIREMENTS AND SPECIFICATIONS MUST BE APPROVED BY UNIFIED COMMAND.

Change requests should be submitted to Data.SMU@NOAA.gov.

Requirement 1: Each cruise must have a designated Data Management Coordinator Onboard

This Shipboard Data Management Coordinator will perform the following duties:

- ✓ Ensure that Daily Deliverables are properly generated and submitted.
- ✓ Ensure that the data in the Daily Sampling Log matches the information found on the relevant Chains of Sample Custody and Sample Labels.
- ✓ Upload raw data to an FTP site for archiving.

Each Shipboard Data Management Coordinator must email the following information to the SMU Data Manager before the start of each cruise:

- Their name
- The name of their vessel
- Their email address
- Their vessel's satellite phone number (can be delivered upon arrival to vessel)
- A cell phone number that can be used to reach them when in port
- The start date of cruise
- The projected end date of cruise

Requirement 2: Each Shipboard Data Management Coordinator must organize the cruise's data products

The Shipboard Data Management Coordinator will maintain all data files generated during the cruise in a series of folders and subfolders. Each vessel will have its own Tier 1 folder for each cruise. Within this Tier 1 folder for the cruise will be three Tier 2 subfolders named Data Collection, Documents, GIS.

Tier 2 Data Collection Subfolder

Within the Data_Collection subfolder there should be additional Tier 3 subfolders for each category of data that is collected (such as CTD, LISST and Location data). Some default Tier 3 subfolders have already been created for new cruises but if your ship collects a category of data for which there is not a subfolder, please add it.

Within each Tier 3 subfolder are additional Tier 4 subfolders for each day that data files are generated. The naming convention of each Tier 4 subfolder is:

<DATE>

where <DATE> is the date that the data was generated in the format YYYY_MMDD (e.g. 2010_0710).

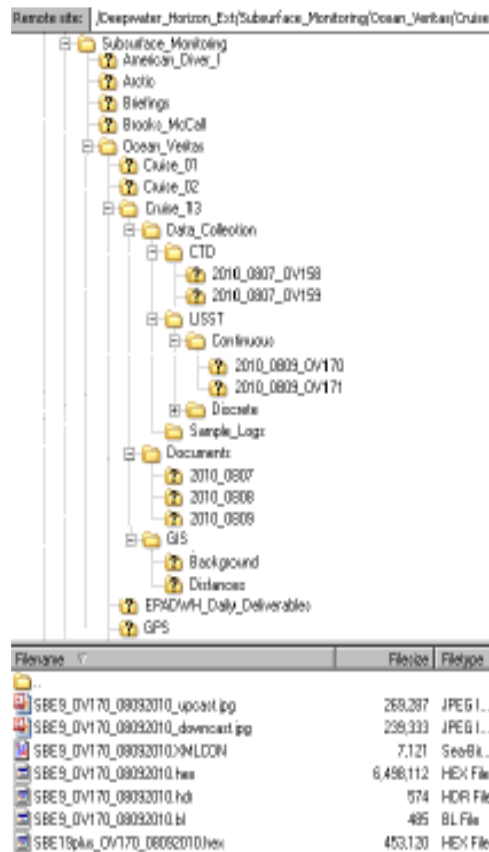
Within each Tier 4 daily subfolder should be the Tier 5 subfolders that actually hold the data files. The naming convention of each Tier 5 subfolder is:

<DATE>_<VESSEL CODE><STATION ID>

where <DATE> is the date that the data was generated in the format, <VESSEL CODE> is the 2-character abbreviation for the vessel (such as BM, OV, or RC), and <STATION_ID> is the Station ID or location where the data was collected (e.g. 2010_0710_OV098).

The overall goal is to provide organized and logical data. Please provide feedback to the SMU Data Manager with suggestions for improvement.

An example of this file structure is:



Inside each Tier 4 CTD subfolder should be Tier 5 subfolders for each type of CTD instrument that is used. : For example, the CTD 9 is a real-time data streaming device and the CTD 19 is the CTD mounted on the rosette that has to be manually downloaded at the end of the cast. If possible, put the raw data files (which you will get from the CTD technician) into the correct folder at the conclusion of each day. Examples of file types the CTD 9 will produce for each station include:

- .xmlcon
- .hex
- .bl
- .hdr
- 2- .jpg
- .zip

For the CTD 19, you should get an .hex file that has SBE (if the equipment is from Seabird) and the site name and Niskin/Go-Flo bottle. There is also a LISST folder for the CTD for each site. In the LISST sub-folder there will be an Excel file for the site as well (ex. LISST_OV01901.xls) and another folder with a date. You will get this file from the CTD technician. Use similar logic for all sensor data.

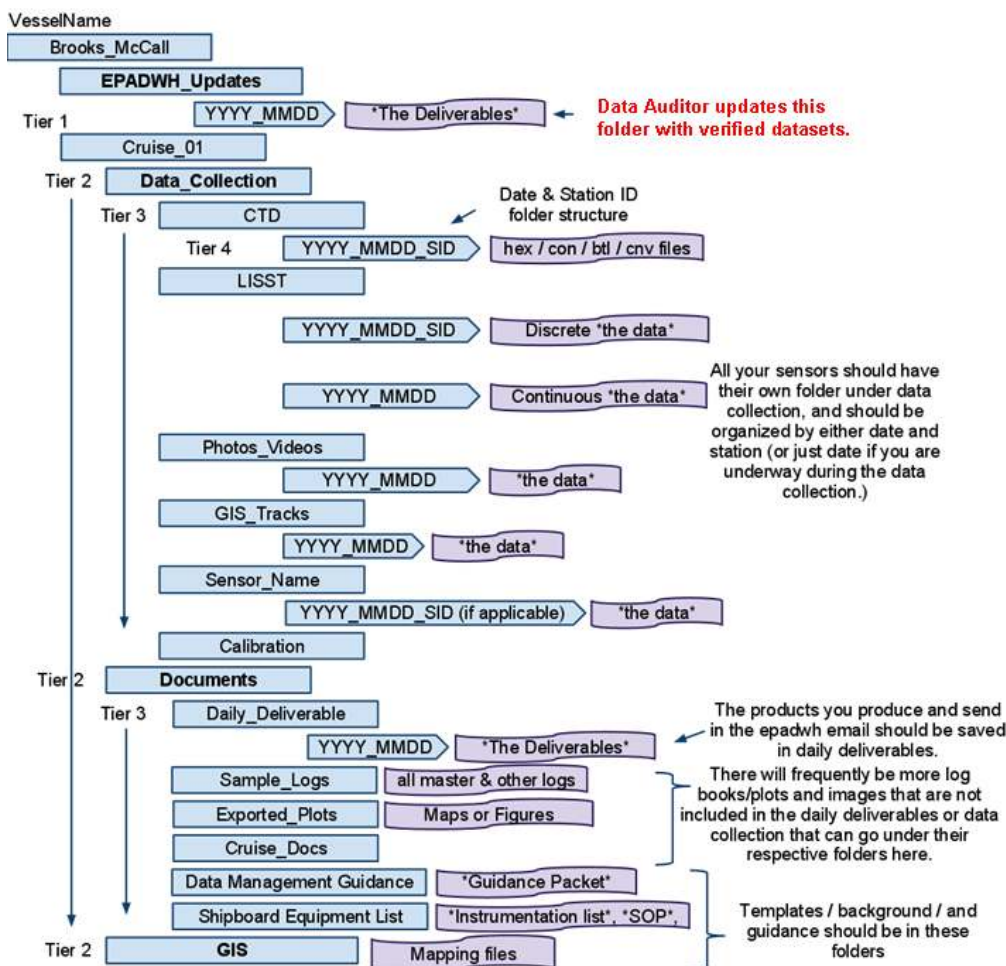
Tier 2 Documents Subfolder

The Daily Deliverables sent to epadwh@gmail.com (see Requirement #3) will be uploaded shoreside to the daily deliverables folder "Daily_Deliverables" under a date folder (YYYY_MMDD) for the deliverables on a daily bases - this is at the root of the ship folder. However, the cruise Data Manager Coordinator will also save the daily deliverables that they generate and submit in the Tier 2 Document subfolder > Tier 3 subfolder named "Daily Deliverables" > Tier 4 subfolder that is created for each day that data is collected. The naming convention of each Tier 4 subfolder is the Date in the format YYYY-MMDD (e.g. 2010_0602). The Data Management Coordinator should place all Daily Deliverables in this folder.

Tier 2 GIS Folder

There is one more sub-folder in the Cruise main folder called GIS. Cruise specific Mapping, GIS or Google earth material should go in here. Please pass this on to the next cruise Data Manager. There is also a GIS folder on the SFTP site that can be used for shipboard mapping.

Figure 1: Organization of Subsurface Monitoring Data on Secure FTP Site



Requirement 3: Each cruise will submit data deliverables on a daily basis.

Daily data packages must be emailed to the SMU Data Auditor (smudwh@gmail.com and epadwh@gmail.com) by midnight central time. These data packages must meet the following specifications. **If any data deliverable that is submitted does not meet these requirements, the entire daily data package will be rejected and a message will be sent to the shipboard Data Management Coordinator indicating what information needs to be corrected.**

Daily Deliverable #1: Status Report

Required

Format: Microsoft Word (.doc) Document

File Naming Convention: <Date (in the format MM-DD-YYYY)>_<Vessel Name>_Status Report

Name of Example File: 05-23-10_Brooks McCall_Status Report

Description: A concise log and report of the vessel's activities during the previous 24-hour operating period. This report must include a status update of ongoing rototoxicity tests (if these tests are conducted onboard) to include a summary of what samples are currently being tested.

Daily Deliverable #2: Data Table of Locations, Samples, Results & Observations and Rototox worksheets

The Data Auditor will then convert these worksheets into separate csv files for uploads to Scribe.

Worksheet 1: Information – Contains a list of all worksheets and corresponding worksheet column descriptions. There are also entries for the following information:

- Data Manager's Name
- Vessel Name
- Data Manager's email address
- Vessel's satellite phone number
- A cell phone number that can be used to reach the vessel when in port
- Start date of cruise
- Projected end date of cruise

Worksheet 2: Locations

Worksheet3: Results and Observations

Worksheet4: RotoTox data

Worksheet5: Hapsite

Worksheet6: Gas Chromatography

Worksheet7: Lab Results

Note: Do not edit the column headers OR move the columns. These are fixed locations that the processing programs look for. They reject the data if information is not found in the assigned locations of the spreadsheet.

Daily Deliverable #3: LISST Particle Analysis Report

Optional

Format: Microsoft Word (.doc) Document

File Naming Convention: <Date (in the format MM-DD-YYYY)>_<Vessel Name>_LISST

Name of Example File: 05-23-2010_Brooks McCall_LISST

Description: A report of LISST analysis activities. This report should be generated by LISST technicians if they are onboard during a cruise.

Daily Deliverable #4: Annotated CTD Plots

Required

Format: .jpg

File Naming Convention: <Date (in the format MM-DD-YYYY)><Vessel Name>_CDT<StationID>

Name of Example File: 05-23-10_Brooks McCall_CTD_BM014

Description: This deliverable is an image (.jpg) file of a CTD plot. This plot is produced by a Seabird technician. This technician must program the following parameters into the SeaBird processing software so that the plots that are produced are consistent:

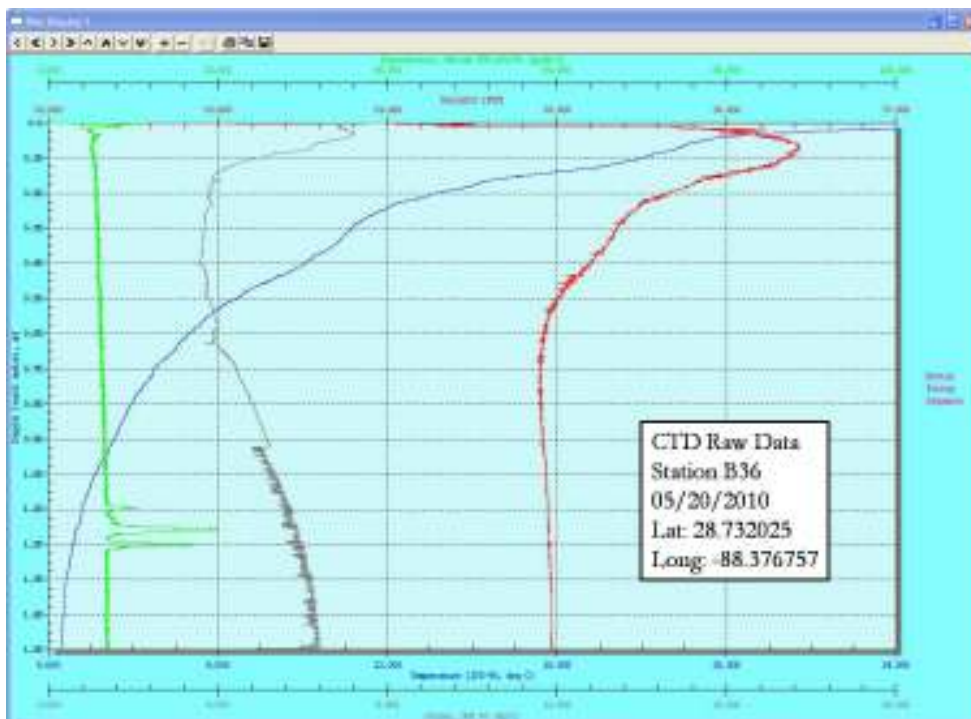
Variable	Units	Axis minimum	Axis maximum
Fluorescence	mg/m ³	0	50
Dissolved Oxygen	mg/L	0	20
Temperature	deg C	0	26
Salinity	PSU	32	37

For ease of viewing and to generate a smooth data scan, only the downscan for each of the above variables should be included in the plot. This can be accomplished by splitting the data and generating the plot or stopping the plot generation at the maximum depth.

Once the plot is exported as an image from the SeaBird software, an annotation tag must be added to each plot using Microsoft Powerpoint (or other suitable software) and the resulting diagram should be saved as a JPEG image. The annotation tag must include only the following information (see example):

- “CTD Raw Data”
- StationID
- Date (in the format MM/DD/YYYY)
- CTD Latitude in decimal degrees
- CTD Longitude in decimal degrees

Figure 2: Example CTD Plot



Additional Specifications for CTD Instrumentation Setup

The table below details additional CTD instrumentation setup and detail data specification for the Sea-Bird software. Configuring the CTD scan using these parameters will ensure consistency in data analysis between cruises and between vessels.

It is important to ensure that the parameters specified here are configured exactly as detailed e.g. the order of the “name” fields is important so that this data can be mapped to the target specification. An example header of the original hex file header (needed for metadata – so make sure the tech updates the file!):

Source File Data	Comment
* Sea-Bird SBE 9 Data File:	
* FileName = D:\03 June_SBE 9 data\SBE 9 Data\OV02101.hex	
* Software Version Seasave V 7.20d	
* Temperature SN = 2234	
* Conductivity SN = 1815	
* Number of Bytes Per Scan = 30	
* Number of Voltage Words = 4	
* Number of Scans Averaged by the Deck Unit = 1	
* System UpLoad Time = Jun 04 2010 12:12:29	
** Vessel: Ocean Veritas	
** Station ID: OV02101	Format: <VESSEL_ID><STATION_ID>
** Date / Time: 4 Jun 2010 / 12:12	
** Nav. Datum-NAD-27	
** Lat: 28.706530	Latitude in decimal degrees
** Lon: -88.348059	Longitude in decimal degrees
** Operator: WJ Bekins	

Order of variables in the file (#name 0 is the first column of data) built after the “datacvn” processing:

# nquan = 15	
# nvalues = 3000	
# units = specified	
# name 0 = t090C: Temperature [ITS-90, deg C]	Ensures that temperature will be presented in column 1 of the detail output
# name 1 = prDM: Pressure, Digiquartz [db]	Ensures that pressure will be presented in column 2 of the detail output
# name 2 = sbeox0V: Oxygen Voltage, SBE 43	Ensures that oxygen voltage will be presented in column 3 of the detail output
# name 3 = flECO-AFL: Fluorescence, Wetlab ECO-AFL/FL [mg/m ³]	Ensures that fluorescence will be presented in column 4 of the detail output
# name 4 = v1: Voltage 1	Ensures that voltage 1 will be presented in column 5 of the detail output
# name 5 = v2: Voltage 2	Ensures that voltage 2 will be presented in column 6 of the detail output
# name 6 = v3: Voltage 3	Ensures that voltage 3 will be presented in column 7 of the detail output
# name 7 = c0S/m: Conductivity [S/m]	Ensures that conductivity will be presented in column 8 of the detail output
# name 8 = depSM: Depth [salt water, m], lat = 29	Ensures that depth will be presented in column 9 of the detail output
# name 9 = sal00: Salinity, Practical [PSU]	Ensures that salinity will be presented in column 10 of the detail output
# name 10 = density00: Density [density, Kg/m ³]	Ensures that density will be presented in column 11 of the detail output
# name 11 = svCM: Sound Velocity [Chen-Millero, m/s]	Ensures that sound velocity will be presented in column 12 of the detail output
# name 12 = sbeox0Mg/L: Oxygen, SBE 43 [mg/l], WS = 2	Ensures that oxygen will be presented in column 13 of the detail output
# name 13 = sbeox0ML/L: Oxygen, SBE 43 [ml/l], WS = 2	Ensures that oxygen will be presented in column 14 of the detail output
# name 14 = flag: flag	
# span 0 = 4.3540, 28.0873	
# span 1 = 3.020, 1518.186	
# span 2 = 1.8009, 3.7157	
# span 3 = 3.1479, 42.4415	
# span 4 = 1.8014, 3.7191	
# span 5 = 0.0781, 0.4918	
# span 6 = 4.9426, 4.9426	
# span 7 = 3.345567, 5.778150	
# span 8 = 3.000, 1502.500	
# span 9 = 33.3517, 36.6291	
# span 10 = 1021.1595, 1034.6632	
# span 11 = 1486.90, 1542.73	
# span 12 = 3.73192, 7.85966	
# span 13 = 2.61137, 5.49972	
# span 14 = 0.0000e+00, 0.0000e+00	

Additional CTD Processing Specifications

The CTD processing steps in the Seabird Software are:

Datcnv > filter > alignctd > celltm > loopedit > wildedit > Derive > binavg > ascii_out > Bottle Summary

Specifics are outlined for each step below (see the seabird website for more detailed instructions on how to use their free software.) The .blt file will be created with the “bottle summary”. Create the bottle summary with the same field order as the ascii file.

```
# interval = meters: 1                               Interval specified at 1 meter
# start_time = Jun 04 2010 12:12:29
# bad_flag = -9.990e-29
# datcnv_date = Jun 07 2010 13:04:28, 7.20c
# datcnv_in = C:\C16408\4-Jun-10\ctd re-run\OV02101.hex
C:\C16408\4-Jun-10\ctd re-run\OV02101.XMLCON
# datcnv_skipover = 0
# datcnv_ox_hysteresis_correction = yes
# filter_date = Jun 07 2010 13:06:20, 7.20c
# filter_in = C:\C16408\4-Jun-10\ctd re-run\OV02101.cnv
# filter_low_pass_tc_A = 0.030
# filter_low_pass_tc_B = 0.150
# filter_low_pass_A_vars =
# filter_low_pass_B_vars = prDM
# alignctd_date = Jun 07 2010 13:07:09, 7.20c
# alignctd_in = C:\C16408\4-Jun-10\ctd re-run\OV02101-f.cnv
# alignctd_adv = sbeox0V 3.500, c0S/m 0.073
# celltm_date = Jun 07 2010 13:07:23, 7.20c
# celltm_in = C:\C16408\4-Jun-10\ctd re-run\OV02101-f-a.cnv
# celltm_alpha = 0.0300, 0.0000
# celltm_tau = 7.0000, 0.0000
# celltm_temp_sensor_use_for_cond = primary,
# loopedit_date = Jun 07 2010 13:07:38, 7.20c
# loopedit_in = C:\C16408\4-Jun-10\ctd re-run\OV02101-f-a-c.cnv
# loopedit_minVelocity = 0.200
# loopedit_surfaceSoak: do not remove
# loopedit_excl_bad_scans = yes
# wildedit_date = Jun 07 2010 13:07:53, 7.20c
# wildedit_in = C:\C16408\4-Jun-10\ctd re-run\OV02101-f-a-c-l.cnv
# wildedit_pass1_nstd = 2.0
# wildedit_pass2_nstd = 20.0
# wildedit_pass2_mindelta = 0.000e+000
# wildedit_npoint = 100
# wildedit_vars = t090C prDM v2 v3 c0S/m
# wildedit_excl_bad_scans = yes
# Derive_date = Jun 07 2010 13:09:51, 7.20c
# Derive_in = C:\C16408\4-Jun-10\ctd re-run\OV02101-f-a-c-l-w.cnv
C:\C16408\4-Jun-10\from ftp\C16408 Data\SBE 9 21-
26\test2601.XMLCON
# derive_time_window_docdt = seconds: 2
# derive_ox_tau_correction = yes
# binavg_date = Jun 07 2010 13:12:34, 7.20c
# binavg_in = C:\C16408\4-Jun-10\ctd re-run\OV02101-f-a-c-l-w-d.cnv
# binavg_bintype = meters
# binavg_binsize = 1                               Bin size must be 1 meter
# binavg_excl_bad_scans = yes
# binavg_skipover = 0
# binavg_surface_bin = no, min = 0.000, max = 0.000, value = 0.000
# file_type = ascii
*END*                                             Required to indicate the end of the header section
```


Additional Specifications related to BP “Target” Format

Header Record

Field Name	Description	Format/Unit of Measure
Orig Datum Used*	Original Datum Used	See note below
Date	Date	MMDDYYYY
Start Time	Start Time – in GMT	HHMMSS
End Time	End Time - in GMT	HHMMSS
Longitude	Longitude	Decimal Degrees (8 decimal places)
Latitude	Latitude	Decimal Degrees (8 decimal places)

* To ensure the integrity of all spatial data associated with the Deepwater Horizon Project, any coordinates supplied in Latitude and Longitude must state which Datum the coordinates are on. In almost all cases this will be WGS84, NAD83 or NAD27.

Specifications related to Detail Record

Field Name	Description	Format/Unit of Measure
Long	Longitude	Decimal Degrees (8 decimal places)
Lat	Latitude	Decimal Degrees (8 decimal places)
Depth_m	Depth	Meters
Temp_C	Temperature	Celsius
Sal_PSU	Salinity	PSU
Den_Kg_m3	Density	Kg/m3
Vp_m_s	Sound Velocity	m/s
Pres_db	Pressure	Db
O2_mg_l	Oxygen	Mg/l
O2_ml_l	Oxygen	ml/l
Fluoro_mg_m3	Fluorescence	Mg/m3
V1 (to VN as req'd)	Voltage Readings	V

Specifications related to Source to Target Field Mapping

The table below provides the source to target field mapping assignment for the header data. Note that any NULL records must be converted to -99999.

Target Field Name	Source Field Name
Orig Datum Used*	** Nav.
Date	** Date / Time:
Start Time	# start_time
End Time	
Longitude	** Lon:
Latitude	** Lat:

The table below provides the source to target field mapping assignment for the detailed data.

Target Field Name	Source Field Name
Long	** Lon:
Lat	** Lat:
Depth_m	Column 9
Temp_C	Column 1
Sal_PSU	Column 10
Den_Kg_m3	Column 11
Vp_m_s	Column 12
Pres_db	Column 2
O2_mg_l	Column 13
O2_ml_l	Column 14
Fluoro_mg_m3	Column 4
V1	Column 5
V2	Column 6
V3	Column 7