Labelling and metadata standards for Shelf Sea Biogeochemistry (SSB) (updated version 2)

Updated: 27 May 2014
Updates include the removal of GEARCODE and addition of links to the SSB website and documentation.

The following labelling and metadata standards have been compiled by the British Oceanographic Data Centre (BODC) for the Shelf Sea Biogeochemistry (SSB) project. It is aimed at facilitating data exchange between cruises and work packages. BODC asks PSOs and participants to use the following metadata and labelling standards for all events and samples collected during SSB. For further information please contact Louise Darroch (lorr@bodc.ac.uk) or Sean Gaffney (sgaf@bodc.ac.uk).

Contents

1   Identifying stations

1.1   Central stations

Each station that occurs on the Shelf Sea Biogeochemistry (SSB) project must be described by a unique combination of alphanumeric codes recorded in a central cruise log (or event log). This metadata (data about data) scheme is adapted from WOCE and is shown in Example 1 and Appendix C. Stations are major scientific events that are recognised by the bridge. They are usually instrument deployments (e.g. CTD, corers), rig deployments (e.g. moorings). They can sometimes be ship surveys (e.g. swath bathymetry, vessel-mounted ADCP). The resulting log should be appended to the end of each cruise report. A template can be downloaded from the SSB website (see http://www.uk-ssb.org/research_cruises/documentation/).

Example 1. Code structure to identify stations

Structure	CRUISECODE | STNNBR | {CASTNO}	(For station events)
Example	*JC090 | 040 | 02*	(example of a delimited format for ascii or excel)

CRUISECODE [Cruise ID]: Each cruise in the project is given a unique identifier. This is defined by BODC in Appendix A.

STNNBR [Station number (3 digits)]: The second identifier is a number used to describe each station occupied during the cruise. Station number should not be confused with fixed stations (or sites) described in section 1.3 below. The station number must be sequential and preferably between 001 to *nnn*. Numbers must be 3 digits long. Where numbers are shorter than this they must be padded with proceeding zeroes (eg. station 1 = 001) to facilitate any electronic sorting of events. Station numbering should be continuous for each cruise (i.e. station numbers should not be reset at the beginning of each cruise leg) and should follow the chronological order of events.

CASTNO [Cast number (2 digits){optional}]: The third identifier is an optional number. Some PSOs may wish to use the cast number to describe instances where there is more than one deployment per station (STNNBR). Some PSOs may wish to use the cast number to denote the *n*th deployment of a particular type of instrument. Others may prefer to only use STNNBR for all deployments for simplicity, even if more than one deployment occurs at a particular site. The cast number is

preferably between 01 to *nn* and preferably 2 digits long, although this can be extended if necessary. Where numbers are shorter than 2 digits they must be padded with proceeding zeroes (e.g. cast 1 = 01) to facilitate any electronic sorting of events.

1.2 Underway discrete sampling stations

Underway discrete sampling stations occupied during SSB must be described by a unique combination of alphanumeric codes as shown in Example 2. Underway sampling (such as discrete samples from the underway non-toxic seawater supply or trace metal tow fish) is not necessarily recognised by the bridge and can generally occur at any time of the day which distinguishes them from central stations. They are not recorded in the event log but samples withdrawn from the non-toxic seawater supply should be recorded in a communal underway station log. The communal underway log should be appended to the end of the cruise report. A template can be downloaded from the SSB website (see http://www.uk-ssb.org/research_cruises/documentation/)

Example 2. Code structure to identify underway stations

Structure	CRUISECODE | DATE | TIME	(For underway samples)
Example	*JC090 | 2013JAN24 | 1324*	(example of a delimited format for ascii or excel)

CRUISECODE [Cruise ID]: As described in section 1.1 above

DATE [Date in UTC (YYYYMONDD)]: This identifier is specific to underway discrete samples. It is used to describe the date of the underway sampling event. Date must be UTC. The date format is reversed (i.e. year first) to facilitate any electronic sorting of samples. Months are expressed as 'MON' (e.g. JAN, JUL) to distinguish between American and British date formats.

TIME [Time in UTC (hh24mm)]: This identifier is specific to underway discrete samples. It is used to describe the hours and minutes of the underway sampling event. Time must be UTC and correspond to the 24 hour clock.

1.3 Fixed stations (or sites)

Fixed stations (or sites) are nominally-fixed, geographic locations which might be areas of interest or are locations visited for repeated measurements over time. It is recommended that fixed stations (and their nominal geographic locations) are identified and named prior to the start of the project. It may also be wise to identify the deviation in the position of point fixed stations (e.g. ± 500 m in latitude and longitude) if you wish to reduce the time required for a ship to occupy these locations using dynamic positioning. Fixed stations do not necessarily need to be included in the alphanumeric code structure used to identify stations described above (shown in sections 1.1 and 1.2) but they can

be reported in supporting information (e.g. the central station log) and supplied in data files submitted to BODC. Fixed mooring stations for SSB are reported in Appendix B .

Example 3. Code to identify fixed stations

Structure	{SITE}	(For fixed stations)

Examples	*JC090 | 2013JAN24 | 1324 |PAP_NW1*	(examples of a delimited
	JC090 | 040 | 02 | PAP_NW1	formats for acsii or excel)

SITE [Fixed station (up to 40 characters){optional}]: Site is a nominal code used to identify fixed stations. It must be up to 40 alphanumeric characters long. Each SITE must be unique across SSB.

1.4 Summary

By using controlled alphanumeric codes, each station carried out during SSB should be uniquely identifiable, not just within a cruise but across the whole programme and should facilitate data exchange between disciplines and work packages. Events are identified by their cruise and station (for central stations), or their cruise, date and time for underway discrete stations. Fixed stations are nominal codenames used to identify sites which are repeatedly visited for sampling. If appropriate, the codes can also be used to identify data filenames. For guidance on the file structure of data submissions to BODC see 'Guidance on providing data to BODC for Shelf Sea Biogeochemistry'.

2 Identifying physical samples

Each physical sample collected during the SSB programme must be accompanied by the combination of alphanumeric codes which identify the sampling event. Ideally, they should also be identified by a sample identifier. This is particularly important when there is more than one physical sample collected in an event, such as 24 rosette bottles on a CTD cast or 12 core tubes on a megacorer.

Example 4. Code structure to identify physical samples

Structure	CRUISECODE | STNNBR | {CASTNO} | SAMPNO	(For general sampling stations)
	CRUISECODE | STNNBR | {CASTNO} | SAMPNO | {SEGNO}	(For core samples)
	CRUISECODE | DATE | TIME | SAMPNO	(For underway samples)

Examples	*JC090 | 040 | 02 | 112*	(examples of delimited formats
	JC090 | 003 | 04 | A	for ascii or excel files)
	JC090 | 2013JAN24 | 1324 | UW1	

CRUISECODE [Cruise ID]: As described in section 1.1 above

STNNBR [Station number (3 digits)]: As described in section 1.1 above

CASTNO [Cast number (2 digits){optional}]: As described in section 1.1 above

DATE [Date in UTC (YYYYMONDD)]: As described in section 1.2 above

TIME [Time in UTC (hh24mm)]: As described in section 1.2 above

SAMPNO [Sample number (alphanumeric)]: The sample number (SAMPNO) is used to identify the collection of a physically separate sample by a sampling device, such as a core or niskin bottle. SAMPNO is optional if there is only one physical sample collected in an event because the event itself is sufficient to identify the sample (e.g. an underway sample or a boxcore). SAMPNO is not optional if there is more than one physical sample collected in an event (e.g. multiple rosette bottles on a CTD cast, multiple SAPs or marine snow catchers in one event). There are different schemes which can be used for SAMPNO. The aim is to ensure that each sample can be identified from each event. For example, many groups use the rosette position of CTD bottles as the sample identifier for CTD casts. Some groups may use the serial number or the deployment depth of a sampling device. Other groups assign unique and consecutive sample numbers to every physical sample collected during a cruise or from a type of platform, such as samples withdrawn from the non-toxic, underway seawater supply (e.g. UW1, UW2). The scheme used will depend on the situation. Sometimes physical samples are combined to make larger sample volumes, such as combining two niskin bottles or combining two cores. In these situations, the SAMPNO must contain the identifiers of all physical samples that were combined. Whatever scheme is used, the sample number must distinguish each physical sample collected in an event.

SEGNO [Segment number (alphanumeric) {optional}]: SEGNO is optional and is used to identify sub-divisions of the physical sample, specifically core samples (i.e. when cores are vertically sub-sectioned). Segment number might be the section depth or might be an alternative unique identifier such as a letter or section number. SEGNO must be unique for each core.

In summary, each physical sample collected during SSB must be linked to the alphanumeric codes used to describe the sampling event. They should also be identified by a sample number if there is more than one physical sample collected in an event. Vertical sub-sections of cores can also be identified by an additional segment identifier. These codes should be included in any data files compiled from the programme. For guidance on the file structure of data submissions to BODC see 'Guidance on providing data to BODC for Shelf Sea Biogeochemistry'.

3 Identifying incubations

Physical samples are sometimes used to obtain further measurements in time-course incubations (experimental or turnover rates). All incubations must be identified by unique alphanumeric codes

to identify the incubation, start date and time because incubations can take place sometime after the collection of a sample. They must also be accompanied by the alphanumeric codes which identify the physical sample(s) used and event (as shown in Example 5). The codes can be used for any type of incubation (e.g. on-deck, laboratory or *in-situ*).

Example 5. Codes to identify incubations

EXPREF:	*EXPLIGHT01*
STARTDATE:	*2013-JAN-24-13:24*
Source(s):	*JC090_040_02/112 (low-light 10m, high-light 10m), JC090_040_02/109 (low-light 30m, high-light 30m)*

EXPREF [Incubation reference (EXP[up to 17 alphanumeric characters])]: This code is obligatory and is used to identify the incubation. A unique reference is assigned to each incubation. The code is a user-defined alphanumeric string (up to 20 characters in total) but must start with 'EXP'.

STARTDATE [start date and time in UTC (YYYY-MON-DD -hh24:mm)]: This code is also obligatory and is also used to identify the incubation. It is the start date and time of the incubation which often occurs sometime after the collection of the physical sample (potentially on land after the cruise). The date and time must be in UTC. The date format is reversed to facilitate any electronic sorting. Months are expressed as 'MON' (e.g. JAN, JUL) to prevent any ambiguity between American and British date formats.

Source(s) [Physical samples (see Example 4)]: All incubations carried out during SSB must be accompanied by the corresponding code(s) that identify the physical sample(s) used. If applicable, these should be clearly linked to their respective treatments. The codes should be structured as shown in Example 4.

In summary, all incubations used to obtain further measurements from physical samples must be identified by a unique EXPREF and the start date and time of the incubation. The activities should also be accompanied by the alphanumeric codes used to identify the physical samples used. These codes should be included in data files arising from the incubation. For guidance on the file structure of data submissions to BODC see 'Guidance on providing data to BODC for Shelf Sea Biogeochemistry'.

4   Operational considerations

The following are recommendations which could help labelling events and samples, and identifying suspect samples. They are methods which have already been put into practice by PSOs, PIs and data managers.

4.1   Using a central whiteboard

A central whiteboard is a useful tool to keep a running order of central station numbers. It will not be as detailed as the event log itself. It will merely summarise the station number (and, if appropriate, the cast number) and activity as and when they occur. It is used as a quick central reference for the current station number (or cast) number.

4.2   Event scribing errors

Graphically plotting stations by position and in order of their time helps to identify scribing errors during the course of a cruise. Scribing errors can occur frequently, especially during busy sampling periods. Applications such as Google Earth can be used.

4.3   Using deployment log sheets

Generic deployment log sheets are a quick tool for recording the timings and positions of deployments on-deck. The log sheets are then stored in a central folder and can be used to update the event log when convenient. An example is shown in Appendix C. A variety of log sheets including this generic sheet, are available on the SSB website (see http://www.uk-ssb.org/research_cruises/documentation/).

4.4   Using sticky labels for CTD sampling

We have found that sticky labels can be an efficient and error-free way to label samples withdrawn from CTD casts. The method suggested below is adapted from WOCE and the sticky label scheme used by Ed Mawji on UK GEOTRACES. It was used on DY008.

For DY008, we obtained A4 sheets of durable sticky labels. Prior to each CTD cast, we printed multiple duplicate labels for each niskin bottle fired (see Example 6). The labels identified the event - the cruise (CRUISECODE), the station number (STNNBR) and the optional cast number (CASTNO) – and finally, the rosette position (SAMPNO). A corresponding sticky label was then removed and tacked to each sub-sample withdrawn from each niskin bottle. Labels were used by all disciplines sampling from the CTD. In this way, each sub-sample was easily identified without scribing errors. The duplicate labels were also used for any sub-samples, such as particulate filters or sub-fractions. To help this, the labels were also tagged with a dotted line to allow users to write on the label (Example 6). We used a Microsoft Word template (provided by the manufacturer) which was setup at the beginning of the cruise. Prior to each CTD, we would use the 'Find and Replace' function to change the station number (004). On other cruises, such as UK GEOTRACES, scientists used labels

with unique and consecutive sample numbers for each niskin bottle fired during the cruise. They linked sample number to the bottle firing information (such as rosette position) by using on-deck log sheets. In this way, all niskin bottles fired on the cruise were identified by a unique SAMPNO.

Example 6. A portion of a label template

DY008/004 CAST: 01 ROS: 1 _____	DY008/004 CAST: 01 ROS: 2 _____	DY008/004 CAST: 01 ROS: 3 _____	DY008/004 CAST: 01 ROS: 4 _____
DY008/004 CAST: 01 ROS: 1 _____	DY008/004 CAST: 01 ROS: 2 _____	DY008/004 CAST: 01 ROS: 3 _____	DY008/004 CAST: 01 ROS: 4 _____
DY008/004 CAST: 01 ROS: 1 _____	DY008/004 CAST: 01 ROS: 2 _____	DY008/004 CAST: 01 ROS: 3 _____	DY008/004 CAST: 01 ROS: 4 _____
DY008/004 CAST: 01 ROS: 1 _____	DY008/004 CAST: 01 ROS: 2 _____	DY008/004 CAST: 01 ROS: 3 _____	DY008/004 CAST: 01 ROS: 4 _____
DY008/004 CAST: 01 ROS: 1 _____	DY008/004 CAST: 01 ROS: 2 _____	DY008/004 CAST: 01 ROS: 3 _____	DY008/004 CAST: 01 ROS: 4 _____

Sticky labels do not only have to apply to CTD events. It could be adapted to other communal events, for individual disciplines or certain platforms, especially where there are likely to be many simultaneous samples collected. For example, we also used sticky labels to identify several sample bottles withdrawn at a station from the non-toxic underway, pumped-seawater supply.

4.5   Labelling physical sampling devices with serial numbers

Serial numbers can be used to index all physical sampling devices of certain types taken on a cruise. It can help to identify malfunctioning sampling devices and in some cases, it can also be used as sample number (SAMPNO). CTD niskin bottles are usually engraved with a serial number on NMF ships and we have found that it is useful to record this number as well as the rosette bottle position. For example, if one niskin bottle consistently leaks or produces suspect results, it can be removed from the available pool. Labelling devices with serial numbers is optional and does not need to be included in the alphanumeric codes to identify events or samples, or in the data files submitted to BODC.

5   Appendix A

5.1   SSB CRUISECODEs

Cruise identifiers are accessible through the BODC Cruise Inventory search form. This includes all NMF and partner cruises.

https://www.bodc.ac.uk/data/information_and_inventories/cruise_inventory/search/

6   Appendix B

6.1   Mooring fixed stations for work package 1

SITE	Nominal position		Deviation in position		Comments
	lat	*lon*	*lat*	*lon*	
Candyfloss	49.40	-8.60	Not set	Not set	Mooring site 1

6.2   Mooring fixed stations for work package 2

SITE	Nominal position		Deviation in position		Comments
	lat	*lon*	*lat*	*lon*	
East of Haig Frais	50.5691	-7.0223			Mooring site 2
Nymph Bank	51.0433	-6.600			Mooring site 3
Celtic Deep	51.1372	-6.5673			Mooring site 4
East of Celtic Deep	51.1258	-6.1770			Mooring site 5

7 Appendix D

7.1 Example of a generic deployment/recovery log sheet

Generic deployment/recovery log

EVENT DESCRIPTION
ADCP mooring

Cruise (CRUISECODE)	DY008
Station ID (STNNBR)	142
Cast ID (CASTNO)	01
Date (UTC)	24/04/2014
Sampling site (SITE)	Candyfloss

Sequence of events

SUB-EVENT (date change)	TIME (UTC)	LAT	LON	Depth (m)
Buoy in	10:28	51·29465	-7·20547	104
Sensor frame in	12:14	51·29878	-7·20589	119
Reposition ship	12:25	51·29477	-7·20544	104
Anchor in	12:55	51·29475	-7·20541	103
Hooked on (25/04/2014)	09:16	51·29476	-7·20543	103
All on deck	11:46	51·29478	-7·20546	104

COMMENTS (e.g. missing sensors, failures etc.)
Oxygen sensor missing on recovery

7.2   Example of an event log

CRUISECODE	STNNBR	CASTNO	SITE	Description	STATION START				At BOTTOM				STATION END				SEA FLOOR DEPTH at bottom		CONTACT for event	COMMENTS
					Date (UTC)	Time (UTC)	Latitude	Longitude	Date (UTC)	Time (UTC)	Latitude	Longitude	Date (UTC)	Time (UTC)	Latitude	Longitude	Uncorr. (m)	Corr. (m)		
DY008	001	01	East of Celtic Deep	Guard buoy	21/03/2014	09:30	51.12890	-6.15210	-	-	-	-	21/03/2014	00:43	51.12630	-6.15920	105		Sivyer	Deployment (end = anchor in)
DY008	002	01	East of Celtic Deep	CEFAS mini lander	21/03/2014	14:05	51.12370	-6.16300	-	-	-	-	21/03/2014	14:26	51.12190	-6.16510	103		Sivyer	Delayed deployment (start = lander in)
DY008	003	01	East of Celtic Deep	Titanium	21/03/2014	15:10	51.12100	-6.16622	21/03/2014	15:18	51.12100	-6.16620	21/03/2014	15:32	51.12100	-6.16627	103		Statham	Test
DY008	004	01	East of Celtic Deep	Titanium	21/03/2014	15:52	51.12115	-6.16623	21/03/2014	15:56	51.12100	-6.16623	21/03/2014	16:17	51.12100	-6.16623	105		Statham	WP1/WP3
DY008	005	01	East of Celtic Deep	Day grab	-	-	-	-	21/03/2014	17:32	51.12087	-6.16600	-	-	-	-	107		Fones	Mud
DY008	006	01	Benthic A1	Day grab	-	-	-	-	21/03/2014	19:44	51.21358	-6.13700	-	-	-	-	-		Fones	Silty mud
DY008	007	01		Day grab	-	-	-	-	21/03/2014	20:47	51.21292	-6.13628	-	-	-	-	110		Fones	Mud
DY008	008	01		Day grab	-	-	-	-	21/03/2014	21:09	51.21219	-6.13543	-	-	-	-	112		Fones	Failure
DY008	009	01		Day grab	-	-	-	-	21/03/2014	21:23	51.21219	-6.13541	-	-	-	-	101		Fones	Silty mud
DY008	010	01		Day grab	-	-	-	-	21/03/2014	21:44	51.21151	-6.13448	-	-	-	-	111		Fones	Failure
DY008	011	01	Benthic A2	Day grab	-	-	-	-	21/03/2014	21:56	51.21149	-6.13448	-	-	-	-	110		Fones	Silty mud
DY008	012	01		Day grab	-	-	-	-	21/03/2014	22:15	51.21084	-6.13354	-	-	-	-	109		Fones	Silty mud
DY008	013	01		Day grab	-	-	-	-	21/03/2014	22:34	51.21026	-6.13247	-	-	-	-	107		Fones	Silty mud
DY008	014	01		Day grab	-	-	-	-	21/03/2014	22:52	51.20970	-6.13131	-	-	-	-	110		Fones	Silty mud
DY008	015	01	Benthic A3	Day grab	-	-	-	-	21/03/2014	23:12	51.20905	-6.13003	-	-	-	-	108		Fones	Silty mud
DY008	016	01	Benthic A4	Day grab	-	-	-	-	21/03/2014	23:41	51.21357	-6.13008	-	-	-	-	109		Fones	Silty mud
DY008	017	01	Benthic A5	Day grab	-	-	-	-	22/03/2014	00:11	51.20903	-6.13728	-	-	-	-	109		Fones	Silty mud
DY008	018	01	Nymph Bank	Guard buoy	22/03/2014	00:31	51.04500	-6.59400	-	-	-	-	22/03/2014	08:11	51.04400	-6.59740	105		Fones	Deployment (end = anchor in)
DY008	019	01	Nymph Bank	CEFAS mini lander	22/03/2014	10:04	51.04330	-6.59990	-	-	-	-	22/03/2014	10:58	51.04250	-6.60580	106		Sivyer	Deployment (start = lander in)
DY008	020	01	Nymph Bank	Titanium	22/03/2014	11:57	51.04233	-6.60730	22/03/2014	12:02	51.04232	-6.60733	22/03/2014	12:16	51.04230	-6.60732	104		Statham	WP1/WP3
DY008	021	01	East of Haig Frais	Guard buoy	22/03/2014	16:56	50.59400	-7.16100	-	-	-	-	22/03/2014	17:20	50.59500	-7.01900	109		Sivyer	Deployment (end = anchor in)
DY008	022	01	East of Haig Frais	CEFAS mini lander	22/03/2014	19:43	50.59610	-7.02210	22/03/2014	19:51	50.59600	-7.02210	22/03/2014	20:00	50.59650	-7.02570	109		Sivyer	Deployment (start = lander in)
DY008	023	01	East of Haig Frais	Titanium	22/03/2014	20:57	50.59707	-7.03000	22/03/2014	21:02	50.59710	-7.02993	22/03/2014	21:10	50.59708	-7.02993	111		Statham/Sivyer	Calibration for mini lander
DY008	024	01	Benthic D1	Day grab	-	-	-	-	22/03/2014	22:31	50.50369	-7.05816	-	-	-	-	112		Fones	Sandy mud
DY008	025	01		Day grab	-	-	-	-	22/03/2014	22:55	50.50369	-7.05120	-	-	-	-	111		Fones	Failure
DY008	026	01		Day grab	-	-	-	-	22/03/2014	23:05	50.50370	-7.05122	-	-	-	-	111		Fones	Failure