



Biological data management at the British Oceanographic Data Centre

Gwenaëlle Moncoiffé
British Oceanographic Data Centre, Joseph Proudman Building, 6 Brownlow Street, Liverpool, L3 5DA, UK
www.bodc.ac.uk

As the UK's National Oceanographic Data Centre, BODC began banking biological data in the late 1980s. This was in response to the challenges posed by increasingly sophisticated multi-disciplinary programmes such as the UK-NERC North Sea Project (NSP) and the UK-JGOFS Biogeochemical Ocean Flux Study (BOFS). The activity provided a unified approach to data management, delivering an integrated quality controlled data set, including the physics and chemistry of the water column. The system adopted was conceived to fully integrate biological data with concomitant oceanographic measurements. Since then, BODC's biological data collections, including taxonomic data sets, have grown steadily, integrating a large number of highly diverse data from UK- and European-funded multi-disciplinary fieldwork programmes in the Atlantic Ocean, Arabian Sea, western and north western European margins and in shelf seas around the UK.

BODC's oceanographic databases

BODC's data managing system is centred around two main types of database, both implemented using the Oracle Relational Database Management System (RDBMS):

- the National Oceanographic Data Bank (NODB) designed to manage oceanographic data as dataseries: e.g. moored instrument time series, ships continuous underway measurements, 2D profiling instrument series, fully worked up CTDs, etc.; inventories, metadata and documentation are fully managed under Oracle; datacycles are stored as files in BODC's standard format (NetCDF equivalent).

and

- the Community Research Project database (CRP) best designed for discrete data collections; contains data from water or core sampling, net tows, CTD and other profiling instruments, incubations, etc.; metadata and data are stored in Oracle with full text documentation currently stored externally.

Currently, most of our biological data collections are held in the CRP database. However, continuous profiles and time series of biological measurements are now banked in the NODB.

Both database models are supported by a common parameter coding system designed to cover an extensive range of multi-disciplinary parameters and defined in the BODC Parameter Dictionary.

Biological data in the BODC Parameter Dictionary

BODC's Parameter Dictionary is a powerful data markup tool which allows the association of any data value to its parameter name and methodology using a single 8-bytes parameter code. The dictionary has been undergoing major re-structuring and standardisation work recently with, in particular:

- re-structuring of the parameter definitions using a semantic model,
- standardisation of our taxonomic parameters against the standards set up by the Integrated Taxonomic Information System.

Fig.1. Distribution of BODC's parameter codes according to major data types.

To date, BODC parameter dictionary contains 16560 parameter codes, 74% of them biological parameters. As illustrated on this plot, parameter codes related to biodiversity information and taxonomy represent a large fraction of the biological parameter group.

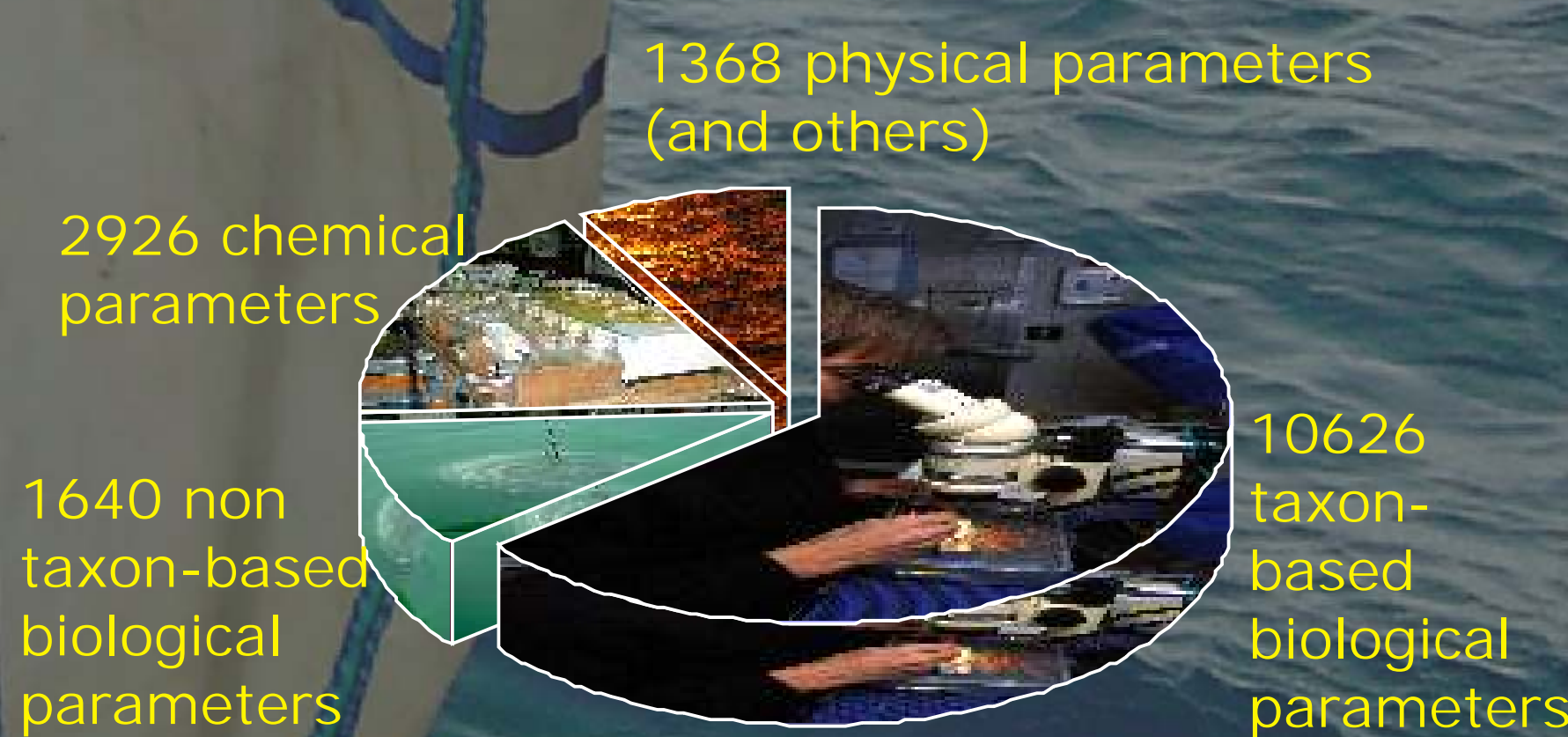


Fig.2. Distribution of BODC's taxonomic parameter codes among major groups of marine organisms.

79% of BODC's taxonomic codes have now been mapped to ITIS;

3%, commonly used in biological oceanography, do not comply with international taxonomic nomenclature and are unmappable to ITIS;

18% have not yet been recorded in the ITIS database.



BODC Biological Data Collections

Current biological data holdings include taxon-based and non taxon-based measurements of abundance, biomass, production rate, loss rate, biochemical composition, molecular information and biogeochemical cycling for planktonic and benthic organisms ranging from viruses and bacteria to micro- and meso-zooplankton collected by over 100 different sampling gears and recording instruments from over 1800 cruises.

The Project Database currently holds data corresponding to some 357 thousand sampling events. Over 10% of these sampling events are related to a biological measurement.

Table 1. Distribution of BODC's biological data holdings. Current data holdings are mainly concentrated in the north-eastern Atlantic Ocean and in areas around the British Isles and the western European margins. Some areas were repeatedly sampled during successive projects.

		zooplankton	microzooplankton	phytoplankton	bacteria	pigment composition	pelagic production and biogeochemical cycling	benthic fauna	benthic fluxes
NE Atlantic and W. European margins									
OWS INDIA	1970								
OWS INDIA	1971-1974	✓	✓	✓					
OWS INDIA	1975	✓	✓	✓					
BOFS	1989-1991	✓	✓	✓	✓	✓	✓		
OMEX I (Norwegian S)	1994	✓	✓	✓					
OMEX I (W European S)	1993-1994	✓	✓	✓	✓	✓	✓		
OMEX I (W European S)	1994-1995	✓	✓	✓	✓	✓	✓		
LOIS (SES)	1995		✓	✓				✓	
LOIS (SES)	1996		✓	✓				✓	
PRIME	Jun-Jul 1996	✓	✓	✓	✓	✓	✓		
PRIME	Dec 1996	✓	✓	✓	✓	✓	✓		
OMEX II	1997-1998	✓	✓	✓	✓	✓	✓		
ACSOE	Jun-Jul 1998	✓	✓	✓	✓	✓	✓		
OMEX II	1999	✓	✓	✓	✓	✓	✓		
MARPROD	Nov-Dec 2001	✓	✓	✓	✓	✓	✓		
MARPROD	Apr-Dec 2002	✓	✓	✓	✓	✓	✓		
Irish Sea, W. English Channel and Celtic Sea									
Bristol Channel	1971-1975	✓	✓						
Bristol Channel	1976-1978	✓	✓						
PML core programmes	1991			✓			✓		
PML core programmes	1992			✓			✓		
PML core programmes	2000	✓	✓	✓	✓	✓	✓		
PML core programmes	2002	✓	✓	✓	✓	✓	✓		
North Sea									
North Sea Project	1988-1989	✓	✓	✓			✓		
North Sea Project	Apr 1989	✓	✓	✓			✓		
North Sea Project	1990	✓	✓	✓			✓		
LOIS (RACS)	1992		✓	✓			✓		
LOIS (RACS)	1993-1995		✓	✓			✓		
LOIS (RACS)	1996		✓	✓			✓		
PROVESS	1998	✓	✓	✓	✓	✓	✓		
PROVESS	1999	✓	✓	✓	✓	✓	✓		
PML core programmes (DISCO)	Jun 1999	✓	✓	✓	✓	✓	✓		
Atlantic Ocean									
Atlantic Meridional Transect	1995-2000	✓	✓	✓	✓	✓	✓		
Atlantic Meridional Transect	2001-2005	✓	✓	✓	✓	✓	✓		
Indian Ocean									
ARABESQUE	Sep+Nov 1994	✓	✓	✓	✓	✓	✓		
M&FMB (AMBITION)	Sep 2001	✓	✓	✓	✓	✓	✓		
Southern Ocean									
BOFS	Nov-Dec 1992	✓	✓	✓	✓	✓	✓		

Spatial and temporal coverage is expected to increase rapidly within the next few years as stricter enforcement of data management and access policies by the Natural Environment Research Council (NERC), the UK main funding body for environmental research, and the UK and European Governments will ensure that copies of most data collected at sea on NERC's research vessels are deposited with the BODC and made available via its databases.

Improving access to our data inventories and data holdings

As part of a major effort in enhancing its web services, BODC is developing web interfaces which will provide online access to its data inventories and data collections via its web database. Of particular interest to oceanographic biological data users will be the development of a web application providing access to data held in the CRP database. Access will require registration and authorised users will be able to search our data holdings and selectively extract data.

This is now 'work-in-progress' scheduled for release in 2005. In the meantime information about our biological data collections may be obtained by contacting BODC directly.