Expanding the Ocean Biogeographic Information System (OBIS) beyond species occurrences, by including associated environmental data – experiences from the OBIS-ENV-DATA project

Leen Vandepitte, Flanders Marine Institute (Belgium), leen.vandepitte@vliz.be
Daphnis De Pooter, Flanders Marine Institute (Belgium), daphnis.depooter@vliz.be
Klaas Deneudt, Flanders Marine Institute (Belgium), klaas.deneudt@vliz.be
Bart Vanhoorne, Flanders Marine Institute (Belgium), bart.vanhoorne@vliz.be
Pieter Provoost, UNESCO-IOC Project Office for IODE (Belgium), p.provoost@unesco.org
Ward Appeltans, UNESCO-IOC Project Office for IODE (Belgium), w.appeltans@unesco.org
Nicolas Bailly, Hellenic Centre for Marine Research (Greece), nbailly@hcmr.gr
Menashè Eliezer, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (Italy), menashe.eliezer@gmail.com
Ei Fujioka, Duke University (USA), efujioka@duke.edu
Philip Goldstein, Univ. of Colorado Museum of Natural History (USA), philip.goldstein@colorado.edu
Alessandra Giorgetti, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (Italy), agiorgetti@ogs.trieste.it
Mirtha Lewis, Centro Nacional Patagónico (Argentina), mirtha@cenpat-conicet.gob.ar
Marina Lipizer, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (Italy), mlipizer@ogs.trieste.it
Kevin Mackay, National Institute of Water and Atmospheric Research (New Zealand), kevin.mackay@niwa.co.nz
Gwenaelle Moncoiffe, British Oceanographic Data Centre (UK), gmon@bodc.ac.uk.
Stamatina Nikolopoulou, Hellenic Centre for Marine Research (Greece), stamatinan@gmail.com
Shannon Rauch, Woods Hole Oceanographic Institution (USA), smrauch1@gmail.com
Andres Roubicek, CSIRO Oceans and Atmosphere (Australia), Andres.Roubicek@csiro.au
Carlos Torres, Universidad Autónoma de Baja California (Mexico), ctorres@uabc.edu.mx
Anton Van de Putte, Royal Belgian Institute for Natural Sciences (Belgium), antonarctica@gmail.com
Matteo Vinci, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (Italy), matteovincister@gmail.com
Nina Wambiji, Kenya Marine and Fisheries Research Institute (Kenya), nwambiji@gmail.com
Francisco Hernandez, Flanders Marine Institute (Belgium), francisco.hernandez@vliz.be

The Ocean Biogeographic Information System (OBIS) aims to integrate smaller, isolated datasets into a larger, more comprehensive picture of life in our oceans. Therefore, OBIS provides a gateway to many datasets containing information on where and when marine species have been observed. The datasets within OBIS are contributed by a network of hundreds of institutes, projects and individuals, all with the common goal to gain more scientific knowledge and to make these data and knowledge easily available to the public. Until recently, OBIS had solely focused on biogeographic data, in the form of presence of marine species in space and time.

Data collected for biological studies however often include more than just the biological parameters such as presence or abundance. Data collected at the same time can include physical and chemical measurements which can provide insights into the environmental conditions the species live in. Details on the nature of the sampling methods, equipment used and effort can also be of major importance.

In early 2015, the IOC Committee on International Oceanographic Data and Information Exchange (IODE) recommended the establishment of OBIS-ENV-DATA as a pilot project of IODE, under the OBIS umbrella. This project will run for 2 years and involves 11 institutions from 10 countries in North-America, South-America, Europe, Africa and Australia. The project will investigate how 'combined

datasets' – datasets containing biological parameters in combination with additional physical and chemical measurements and observations of the habitat – can be kept together. The project will also demonstrate how institutes holding combined datasets can collaborate on the joint management and exchange of such data, and to show the benefit of this approach for marine sciences, biological analysis and modelling. In addition, the project will seek solutions to maximize data sharing with regional and global data systems (e.g., OceanDataPortal or the World Ocean Database), optimize interoperability and minimize the risks of data duplication.

A first workshop was organized in October 2015, bringing together data providers that manage combined datasets. Several project partners contributed pilot datasets to the workshop. The goal was to run through the different data conversion steps to come to a suitable format, discuss the problems encountered and write a preliminary manual that can be used as a guideline for further data exchange. Prior to the workshop, four options were proposed. All pilot datasets were transformed into one or several of these formats, to illustrate and explore potential issues. During the workshop, two additional options were identified, bringing a total of 6 options to the discussion table.

Workshop participants mostly struggled with the fact that the Darwin Core Archive standard (DwC-A) maintains a star schema, where a core file links to different extension files. The general perception was that a simple link between the Occurrence and MeasurementOrFacts extensions would simplify the format and workload. The chosen option includes a customized or extended MeasurementOrFacts Extension (eMoF), allowing a simple link to the Occurrence Extension (Fig. 1). In addition, it includes a hierarchy in events, thereby dealing with a lot of previously identified problems in the management of such data. As the original DwC star schema is still maintained – it only includes an additional link between the two extensions – there is no conflict or incompatibility with IPT, the preferred data exchange mechanism to transfer data to OBIS.

This new OBIS data standard allows for the management of sampling methodology, animal tracking and telemetry data. It also makes it possible to include environmental measurements such as nutrient concentrations, sediment characteristics, climate variables or other abiotic parameters measured during sampling, to characterize the environment where the biogeographic data was collected. The new standard builds on the DwC-A standard and on practices adopted by the Global Biodiversity Information Facility (GBIF). It consists of a DwC Event Core in combination with a DwC Occurrence Extension and a proposed enhancement modification to the DwC MeasurementOrFacts Extension. This new structure enables the linkage of measurements or facts - quantitative or qualitative parameters - to both sampling events and species occurrences, and includes additional fields for parameter standardization. The effectiveness of this new format is demonstrated by the pilot datasets available at http://ipt.iobis.org/obis-env/.

The workshop report will be published as a scientific paper to promote the project and its potential for future data management and data exchange. OBIS data providers will be encouraged to submit the environmental data associated with the occurrence data in this newly defined format and OBIS will make these available through their portal and other regional and global repositories such as SeaDataNet and EMODnet.

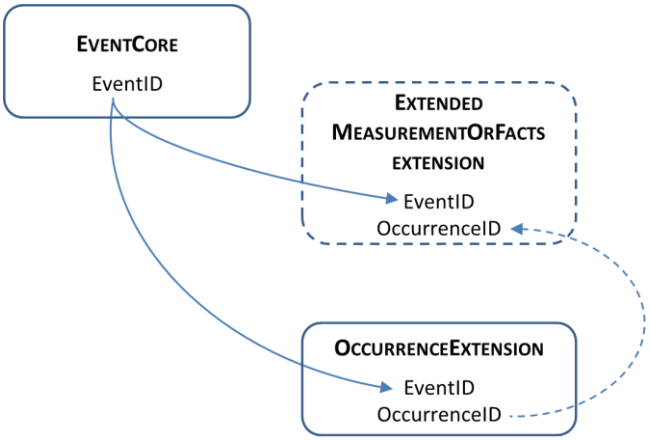

Fig. 1: Schematic representation of the OBIS-ENV schema. The characteristic IPT star schema is maintained. A linkage for the biological data is incorporated between the Occurrence extension and the newly developed "Extended MeasurementOrFacts" extension (eMoF).