

Fit for purpose? Identifying and resolving quality issues with marine biodiversity datasets in R

Theme: Biodiversity tools & data

Presentation format: Oral

Abstract (200 words):

Millions of marine species occurrences and abundances can be accessed through the Ocean Biogeographic Information System (OBIS), which are then often combined with data from additional sources such as GBIF, citizen science projects, scientific literature and personal communications. However, the quality of the available data is variable and it thus needs to be scrutinized in order to get a dataset that is fit for purpose.

To help this process, as well as increase the quality of the data before they are published in OBIS, we developed the *obistools* R package. It allows users to identify and resolve common data errors such as taxonomic, spatial, temporal and measurement issues. The package combines and builds on existing services made available by the World Register of Marine Species as well as some new OBIS home-made services. The interactive interface provides a series of strict and fuzzy quality checks ranging from longitude/latitude checks to environmental outlier detection. These checks, in combination with pre-defined constraints based on for instance the physiological knowledge of the species and the expected spatial extent, can then be used to evaluate if specific records can be used in the analysis to obtain the final dataset for the analysis.

Keywords: quality control, public biodiversity data, OBIS

Short description (120 characters):

We present *obistools*, an R package for identifying and resolving quality issues in marine biodiversity data.

Authors:

Samuel Bosch, Pieter Provoost, Ward Appeltans (w.appeltans@unesco.org)

Ocean Biogeographic Information System (OBIS), Intergovernmental Oceanographic Commission (IOC) of UNESCO, IOC Project Office for IODE, Oostende, Belgium