Mobilizing historical marine data within EMODnet Biology

Vasilis Gerovasileiou, Hellenic Centre for Marine Research (HCMR) (Greece), vgerovas@hcmr.gr
Stamatina Nikolopoulou, Hellenic Centre for Marine Research (HCMR) (Greece), snikolo@hcmr.gr
Dimitra Mavraki, Hellenic Centre for Marine Research (HCMR) (Greece), dmavraki@hcmr.gr
Christos Arvanitidis, Hellenic Centre for Marine Research (HCMR) (Greece), arvanitidis@hcmr.gr
Leen Vandepitte, Flanders Marine Institute (VLIZ) (Belgium), leen.vandepitte@vliz.be
Gizem Poffyn, Flanders Marine Institute (VLIZ) (Belgium), gizem.poffyn@vliz.be
Joana Beja, Flanders Marine Institute (VLIZ) (Belgium), joana.beja@vliz.be
Ruben Perez Perez, Flanders Marine Institute (VLIZ) (Belgium), ruben.perez@vliz.be
Laura Boicenco, National Institute for Marine Research and Development "Grigore Antipa" (NIMRD) (Romania), lboicenco@alpha.rmri.ro
Marina Lipizer, National Institute of Oceanography and Applied Geophysics (OGS) (Italy), mlipizer@inogs.it
Menashè Eliezer, National Institute of Oceanography and Applied Geophysics (OGS) (Italy), meliezer@inogs.it

Historical marine data are often scattered in old publications, reports and expedition logbooks, either in the form of hard copies or in simple and unorganized spreadsheets in electronic storage media. Such sources provide an invaluable resource of biological and environmental information that could be used for reconstructing and modelling past environmental conditions or predicting future trends and shifts in distribution range, biological invasions and regional species extinctions. This is extremely important for regional European Seas and adjacent marine regions, which are vulnerable to an ever-increasing number of human activities and pressures.

Recognizing the importance for mobilizing historical marine data, EMODnet Biology has developed a long-term data archaeology and rescue strategy, under a dedicated work package (WP3). The overall objective of this initiative is to fill spatial and temporal gaps in aquatic species occurrences and make the rescued historical data available freely through the EMODnet portal and global biogeographic and biodiversity information systems (e.g. OBIS and its regional nodes, WoRMS and its sub-registers, LifeWatch Species Information Backbone). The standard WoRMS taxonomic mapping makes these data usable for long term time series. Special focus is given to understudied, "data poor" regions which are particularly susceptible to environmental alterations and biological invasions such as the South-Eastern Mediterranean Sea and the Black Sea.

A non-exhaustive pool of approximately 240 historical marine archaeological (1890s to 1950s) and rescue datasets (1960s to 2000s) was assembled by HCMR and is continuously being updated and annotated with metadata. A similar process was carried out by VLIZ, resulting in 95 datasets being identified for rescue; the datasets range from 1930s to mid-1990s, with the bulk of covering the 1970s decade and describing data collected mainly in the Belgian North Sea area. In addition, paper archives, primary protocols and grey literature from the Romanian Black Sea waters were scanned and rescued by NIMRD, containing data on phytoplankton, zooplankton and macrozoobenthos dating back to the 1950s, which remained unavailable to the wider research community.

A set of purpose-build criteria was adopted for the prioritization and selection of datasets to be digitized and mobilized, including thematic and taxonomic cover, temporal and geographic scope, language and readiness in their availability (e.g. online digital files versus hard copies in libraries). So far 90 historical datasets, published during the period 1868-1999, which included 110,230 occurrence records have been rescued and mobilised (Figure 1). The experience gained has revealed several challenges at all stages of the data "life-cycle", from dataset identification and metadata extraction to the digitization, standardization and quality control of historical datasets, such as: lack of

standardization, georeferencing accuracy, taxonomic inconsistencies and updates, misspellings of taxa and locations and poorly documented or missing sampling protocols (Figure 2). Nevertheless, facing these challenges is of paramount importance since loss of such data equals to loss of unique resources required to understand global changes and ultimately to the loss of our natural wealth.
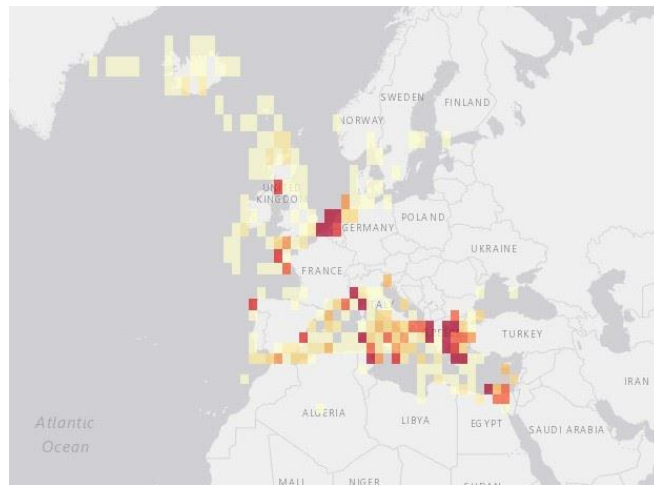
Figure 1: Distribution of historical occurrence data mobilised in OBIS (in total 110,230 occurrences from 90 datasets). Colour shades represent the number of occurrence records (higher values in red)

Figure 2: Original presentation of sampling metadata of the historical "Thor" Expedition to the Mediterranean Sea and adjacent waters (from: Schmidt 1912 Høst & Son, Copenhagen, 49 pp.)

Acknowledgments