

A Comprehensive Approach to Marine Biodiversity Data Validation: The Biocheck Tool

Ruben Perez Perez (ruben.perez@vliz.be), Bart Vanhoorne (bart.vanhoorne@vliz.be), Lynn Delgat (lynn.delgat@vliz.be), Joana Beja (joana.beja@vliz.be), Leen Vandepitte (leen.vandepitte@vliz.be)

VLIZ, Flanders Marine Institute (Belgium)

Rationale

Precise and detailed marine biodiversity occurrence data plays a vital role in comprehending and forecasting species population trends. This information serves as the foundation for ecological and conservation research. EMODnet Biology (European Marine Observation and Data Network - Biology lot), in collaboration with other international initiatives such as the Ocean Biodiversity Information System (OBIS) and the Global Biodiversity Information Facility (GBIF), is dedicated to gathering and publishing high-quality biodiversity and sample related data. Maintaining data quality requires careful curation and validation, which can be challenging due to scattered and non-standardized quality check tools tailored to specific user requirements.

The Biocheck tool

The Biocheck tool is a web application that brings the capabilities of the EMODnetBiocheck R package to the wider public, regardless of whether they have knowledge of the R programming language or not. Both the web application and the R package expand on and are built upon the obistools R package developed by OBIS and are a result of the collaborative work of the LifeWatch Belgium and EMODnet Biology projects. The primary objective of this tool is to conduct a thorough exploration and comprehensive quality control of marine biodiversity occurrence (meta)data, aligning with the widely recognized Darwin Core biodiversity data standard. Consequently, it serves as the default quality control tool for all types of marine biodiversity occurrence data.

The Biocheck tool performs a series of essential quality checks, including:

- assessment of (meta)data format and integrity;
- taxonomic verification, using the World Register of Marine Species (WoRMS) as reference;
- evaluation of marine and coastal biogeography;
- scrutiny of parameters linked to occurrences, encompassing organism quantifications and facts, biometrics data and environmental and sampling-related data.

Conclusion

The tool generates exploratory maps and graphs tailored to each dataset, providing valuable insights into the data. Additionally, it compiles and highlights a comprehensive list of records that fail the quality checks, not meeting the required quality standards. The Biocheck tool thus allows its users to identify and fix potentially problematic records and to increase the overall quality of their datasets before or after publication. Therefore, holding a crucial position in enhancing the accuracy and trustworthiness of marine biodiversity occurrence data. This, in turn, aids to our comprehension of species populations and supports well-informed decisions in conservation and ecological matters.



Figure 1 Illustration on the Biocheck tool procedures from Perez Perez et al. [2023].

References

OBIS. <https://obis.org>
 GBIF. <https://www.gbif.org>
 Biocheck tool. <https://rshiny.lifewatch.be/BioCheck>
 EMODnetBiocheck R package. <https://github.com/EMODnet/EMODnetBiocheck>
 Darwin Core. <https://dwc.tdwg.org>
 LifeWatch Belgium. <https://www.lifewatch.be>
 EMODnet Biology. <https://emodnet.ec.europa.eu/en/biology>
 Perez Perez et al., (2023). *The Biocheck tool: Enhancing data quality in marine biodiversity research.* Poster in EMODnet Open Conference 2023.