Promoting High-Quality Data in OBIS: Insights from the OBIS Data Quality Assessment and Enhancement Project Team

Yi-Ming Gan[‡], Ruben Perez Perez[§], Pieter Provoost[|], Abigail Benson[¶], Ana Carolina Peralta Brichtova[#], Elizabeth Lawrence[|], John Nicholls[¤], Johnny Konjarla[«], Georgia Sarafidou[»], Hanieh Saeedi[ˆ], Dan Lear[ˇ], Anke Penzlin[ˆ], Nina Wambiji[¦], Ward Appeltans[|]

‡ Royal Belgian Institute of Natural Sciences, Brussels, Belgium
§ Flanders Marine Institute, Oostende, Belgium
| Intergovernmental Oceanographic Commission of UNESCO, Ocean Biodiversity Information System, Oostende, Belgium
¶ U.S. Geological Survey, Colorado, United States of America
# Universidad Simón Bolívar, INTECMAR, Caracas, Venezuela
¤ Trinity College Dublin, Dublin, Ireland
« Centre for Marine Living Resources & Ecology, Kochi, India
» HCMR Hellenic Centre for Marine Research, Heraklion, Greece
ˆ Senckenberg Research Institute and Natural History Museum, Frankfurt am Main, Germany
ˇ Marine Biological Association of the United Kingdom, Plymouth, United Kingdom
¦ Kenya Marine and Fisheries Research Institute (KMFRI), Mombasa, Kenya

Corresponding author: Yi-Ming Gan (ymgan@naturalsciences.be), Ruben Perez Perez (ruben.perez@vliz.be)

Abstract

The Ocean Biodiversity Information System (OBIS) (Klein et al. 2019) is a global database of marine biodiversity and associated environmental data, which provides critical information to researchers and policymakers worldwide. Ensuring the accuracy and consistency of the data in OBIS is essential for its usefulness and value, not only to the scientific community but also to the science-policy interface. The OBIS Data Quality Assessment and Enhancement Project Team (QCPT), formed in 2019 by the OBIS steering group, aims to assess and enhance data quality. It has been working on three categories of activities for this purpose:

Data quality enhancement and management

The OBIS QCPT organized data laundry events to identify and address data quality issues of published OBIS datasets. Furthermore, individual OBIS nodes were invited to give their data-processing presentations in the monthly meetings to foster knowledge sharing and collaborative problem-solving focused on data quality. Data quality issues and solutions highlighted in the presentations and data laundry events were documented in a dedicated GitHub repository as GitHub issues. The solutions for data quality issues and marine-specific pre-publication quality control tools, designed to identify the data quality issues, were provided as feedback to the OBIS Capacity Development Task Team.

These inputs were used to create training resources (see OBIS manual, upcoming OBIS training course hosted on OceanTeacher Global Academy) aimed at preventing these issues.

Standardization of OBIS data processing pipeline

As OBIS uses the Darwin Core standard (Wieczorek et al. 2012), the use of standardized tests and assertions in the data processing pipeline is encouraged. To achieve this, the OBIS QCPT aligned OBIS quality checks with a subset of core tests and assertions (Chapman et al. 2020) developed by the Biodiversity Information Standards (TDWG) Biodiversity Data Quality (BDQ) Task Group 2 (TG2) (Chapman et al. 2020) as tracked in this GitHub issue. Not all default parameters of the core tests and assertions are optimal for marine biodiversity data. The OBIS QCPT met monthly to determine suitable parameters for customizing the tests. The pipeline produces a data quality report for each dataset with quality flags that indicate potential data quality issues, enabling node managers and data providers to review the flagged records.

Community engagement

The OBIS QCPT led a survey among data users to gather insights into OBIS data quality issues and bridge the gap between the current implementation and user expectations. The survey findings enabled OBIS to prioritize issues to be addressed, as summarized in Section 2.2.2 of the 11th OBIS Steering Group meeting report. In addition to engaging with data users, the OBIS QCPT also served as a platform to discuss questions related to the use of Darwin Core from the nodes and provided feedback for the term discussions.

In summary, the OBIS QCPT improves marine species data reliability and usability through transparent and participatory approaches, fostering continuous improvement. Collaborative efforts, standardized procedures, and knowledge sharing advance OBIS' mission of providing high quality biodiversity data for research, conservation, and ocean management.

Keywords

quality control, biodiversity standards, biogeography, controlled vocabularies

Presenting author

Yi-Ming Gan

Presented at

TDWG 2023

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Chapman A, Belbin L, Zermoglio P, Wieczorek J, Morris P, Nicholls M, Rees ER, Veiga A, Thompson A, Saraiva A, James S, Gendreau C, Benson A, Schigel D (2020) Developing Standards for Improved Data Quality and for Selecting Fit for Use Biodiversity Data. Biodiversity Information Science and Standards 4 https://doi.org/10.3897/biss.4.50889
- Klein E, Appeltans W, Provoost P, Saeedi H, Benson A, Bajona L, Peralta AC, Bristol RS (2019) OBIS Infrastructure, Lessons Learned, and Vision for the Future. Frontiers in Marine Science 6 https://doi.org/10.3389/fmars.2019.00588
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7 (1). https://doi.org/10.1371/journal.pone.0029715