

Making data management practices compliant with Essential Variables frameworks: A practical approach in the marine biological domain

Martina Zilioli¹, Alessandro Oggioni¹, Paolo Tagliolato¹, Cristiano Fugazza¹, Caterina Bergami³,
Alessandra Pugnetti² and Paola Carrara¹

¹National Research Council of Italy, Institute for Electromagnetic Sensing of Environment (CNR-IREA), Milan

²National Research Council of Italy, Institute of Marine Sciences (CNR-ISMAR), Venice

³National Research Council of Italy, Institute of Marine Sciences (CNR-ISMAR), Bologna

Abstract – Measuring marine biodiversity is essential to proper management of biological resources. Monitoring networks and observatory systems are adopting conceptual frameworks to agree on the fundamental variables describing the biodiversity change and to meaningfully inform policy makers. Biological and ecosystem Essential Ocean Variables (bio-eco EOVs) and Essential Biodiversity Variables (EBVs) both help to harmonize observations of different providers to essential measurements. Conceptual relationships between the two frameworks have been recently outlined to enrich data with complementary information. However, even if bio-eco EOVs and EBVs are adopted among research communities, they represent theoretical approaches whose implementations are often limited and still uncoupled. To evaluate operative reuse of biodiversity measurements, in this paper we identify differences and common aspects between the management recommendations established by the two frameworks in a synoptic comparison. Also, we provide evidences for applying the proposed approach to the marine component of the Long-Term Ecological Research (LTER) networks, by individuating in the Italian network (LTER-Italy) the Northern Adriatic Sea as a case study where compare, in further work, the data handling practices with the identified recommendations.

I. INTRODUCTION

Life in the ocean, besides being a value per se, provides to humankind a wide range of vital benefits and resources, fundamentally depending on biodiversity. In order to catch the significant drivers of marine biodiversity changes, which operate at multiple spatial scales, biological observing programmes have been rising in numbers since the mid-70s [1].

Despite the limited availability of high-tech approaches for automated biological measurements in ocean and coastal ecosystems [2], governments started regulating the reporting of the experts' advices under international agreements to take informed decisions [3].

Nevertheless, the increasing number of programmes and initiatives has not been always accompanied by integration of scientific knowledge in decision making [4]. First, because the exchange of knowledge requires brokering to avoid misalignment between the information asked by policy makers and that provided by scientists [5]. Second, because an agreement on which variables have to be measured or observed to assess the biodiversity status is called into question. Over many decades, one of the principal aims of ecological theory was identification of appropriate measurements of biodiversity, made difficult also by the non-comparability of the indices used to assess it [6]. Moreover, the analysis of organisms with molecular technologies [7] and remote sensing techniques [8] added new description levels in the estimation of biodiversity.

Following the example given by the Essential Climate Variables evolved in the late 1990s [9], two conceptual frameworks have been proposed to pinpoint the minimal set of measurements necessary to understand changes in the biological component of the oceans, and meaningfully inform resource managers.

In 2013 [10], the Group on Earth Observations Biodiversity Observation Network (GEO BON) developed the framework of Essential Biodiversity Variables (EBVs) which groups 22 variables in six main classes¹, each representing a level of biodiversity organization. The framework was endorsed by the Convention on Biological Diversity (CBD) as an abstraction layer of information by which primary observations can be related to the Aichi Biodiversity Targets [11]. Within the GEO BON, Marine BON (MBON) was established in 2016 to frame the EBVs concept in marine realm and to steer evidence-based achievement of the Aichi Biodiversity Targets 6, 10, 11,

¹EBVs classes are: Genetic Composition, Species Populations, Species Traits, Community Composition, Ecosystem Structure, Ecosystem

Function;

which attain to marine resources.

In 2015, the Biology and Ecosystem Panel of Global Ocean Observing System (GOOS) identified a specific set of Essential Ocean Variables (bio-eco EOVs)² to focus observing systems and fundings on technologically feasible measurements of the living components of marine ecosystems [1]. The bio-eco EOVs require complementary data with respect to EBVs. Bio-eco EOVs sort organisms around trophic levels, thus being more informative about top-down pressures and possible causes of changes in the different trophic groups [1]. Differently, EBVs indicate a set of variables through which to describe the taxonomic, spatial and temporal dimensions of the organisms under investigation.

The relationships between the EBVs and the bio-eco EOVs were conceptually described by Muller-Karger [9], suggesting that the two frameworks allow for an alternative use of a common set of scientific measurements. Even if this encourages the reuse of data records, methodologies and instruments are not yet harmonized among research communities in any of these frameworks [12]. Also, standardization of (meta)data schemas and services is yet far from allowing easy integration and processing of multi-source data for both the EBVs [13] and the EOVs.

In this paper, we present a synoptic, up-to-date comparison of the management recommendations for the EBVs and the bio-eco EOVs primary data through a literature review. We highlight differences and common aspects between the two frameworks along with five key-tenets proposed to build the integrated marine biological observing and informatics system [5]. In this way, we identify management actions to be enacted so as to facilitate observing systems in measurement reuse within both frameworks. Finally, we propose to select a case study where to test their feasibility and reliability, in a future work, within the Italian Long-Term Ecological Research Network (LTER-Italy, <http://www.lteritalia.it>). We propose to develop the analysis in the Northern Adriatic Sea (NAS), to match its current practices with the identified recommendations, thus suggesting a roadmap for the management and preservation of biological measurements for both the EBVs and the bio-eco EOVs frameworks.

II. MANAGEMENT RECOMMENDATIONS FOR ESSENTIAL VARIABLES PRIMARY DATA

The goal of the EBVs and the bio-eco EOVs is to align highly heterogeneous researchers to mobilize biological

data or to focus researches on observational gaps in marine ecosystems [11]. Hence, the management and sharing of measurements through web-based means is essential to globally integrate data from any monitoring initiative. The more recent effort to bring together the EBVs and the bio-eco EOVs data management practices has been proposed in the context of the collaboration agreement between the GOOS, the Ocean Biogeographic Information System (OBIS), and MBON. Five key-tenets have been identified by Benson et al. [5] to build a marine biological observing and informatics system, which will facilitate the integration of data and analytical methods shared among professional and lay experts who sample marine biological components. The tenets are the following:

1. *Standardized data collection*: assuring that data for a minimum set of essential variables (EBVs/bio-eco EOVs) are collected using comparable methods across time and multiple sites;
2. *Addressing global reporting needs*: considering national and global policy reporting requirements and regulations;
3. *Making data FAIR*: fitting data to *FAIR Principles* (Findable, Accessible, Interoperable, and Reusable);
4. *Analytical algorithms, tools, and workflow are accessible*: sharing analytical algorithms, workflows and provenance of the data;
5. *Adhering to the Action Ecology principles*: applying the *Action Ecology*³ principles to all the pieces of the system to provide policy-ready suggestions.

As a matter of fact, the historical development of the two frameworks has been independent until few years ago and observations and data management reached distinct coordination outcomes as well as they were approached differently, in accordance to the habits of the research communities involved (i.e. ecologists and oceanographers, respectively). Thus, we argue the importance to analyse the two frameworks separately with respect to the aforementioned tenets: In this way, we highlight differences, commonalities and more mature achievements the data provider should be aware of to enforce the vision illustrated by Benson et al.

In Table 1, we classify the recommendations (i.e. technical solution or technological approach) provided by the EBVs and the bio-eco EOVs for the management of marine biological primary data according to the tenets. As

²The agreed bio-eco EOVs are: microbe biomass and diversity, phytoplankton biomass and diversity, zooplankton biomass and diversity, fish abundance and distribution, benthic invertebrate abundance and distribution, marine turtles, birds, mammals abundance and distribution, hard coral cover and composition, seagrass cover and composition, macroalgae canopy cover and composition, mangrove cover and

composition;

³*Action Ecology* is defined as the research explicitly targeted at providing relatively fast but effective analyses of diversely scaled and multifaceted datasets to support and inform policy and management decision-making about on-going ecological problems [14].

we aim at promptly testing the reuse of marine biological measures, we take into consideration tenets 1-4, and we consider compliance with tenets 1 and 2 to achieve the policy-relevance of findings discussed by tenet 5. Moreover, we describe the open issues which make methodologies documentation inadequate to achieve the data collection standardization addressed by tenet 1. We believe they partially explain the fragmented scenario of data because of which it is difficult to provide normalizing recommendations.

Table 1. Recommendations for marine biological data management (e.g. raw data, experimental/in-situ/remote observations).

5 key-tenets (Benson et al. 2018)	EBVs (Pereira et al. 2013)	Bio-eco EOVs (Miloslavich et al. 2018)
<i>Standardized Data Collection</i> (1)	Missing. Open issue: completeness of methodologies documentation	Missing. Open issues: hidden information about methodologies to collect sensitive data
<i>Addressing Global Reporting Needs</i> (2)	CBD, Ramsar, CMS ⁴ , Habitat Directive, Birds Directive, MSFD ⁵ , WFD ⁶ (Geijzendorffer et al. 2016)	Different conventions and agreements (Miloslavich et al. 2018)
<i>Making Data FAIR</i> (3)	Data standard (e.g. Darwin Core); metadata standard (e.g. EML, ISO19115, ISO 19157); information system (e.g. Global Biodiversity Information Facility); Bari Manifesto (Hardisty et al. 2019)	Data standard (e.g. Darwin Core) metadata standard (e.g. EML); information system (e.g. Ocean Biogeographic Information System); Framework for Ocean Observing (FOO, 2011)
<i>Analytical Algorithms, Tools, and Workflows Are Accessible</i> (4)	Data management and processing workflow (e.g. EBVs data product) (Kissling et al. 2018)	Ocean Data Product (Buck et al. 2019)
<i>Adhering to Action Ecology Principles</i> (5)	Not addressed	Not addressed

In particular, as regards tenet 1, the main issue for harmonizing methodologies is related to the EBVs data sources provided by citizen scientists (e.g., local monitoring communities, volunteers) or long-tail science

(e.g., historical time-series of ecological data, data collected in paper archives), which often lack complete and reusable documentation. Thus, sampling standardization can't be fully achieved. Differently, the public availability of data and methods is one of the issues that hamper this process in the bio-eco EOVs framework. For example, commercial fisheries information, personal identification, and survey coordinates are sensitive data to which restricted access or licences need to be applied. To avoid privacy issues, these data are often aggregated [5] thus hiding methodology information.

Nevertheless, measurements of variables allow researchers to relate their data to different policy reporting conventions. Specifically considering tenet 2, while primary data collected for measuring the EBVs are used to address seven global and national agreements concerning biodiversity [11], a list of twenty-four international conventions attaining ecosystem services, food security, marine spatial planning, and flora and fauna conservation in extreme environments, will benefit from the bio-eco EOVs measurements [1].

As to tenet 3, both frameworks address FAIRness for the independent exploitation of data by users. They recommend adoption of standards for metadata and data as well as to follow a set of formalized interoperability principles. However, providing measurements for the two frameworks can imply different choices about standard selection. Given the expected processing of EBVs datasets, the framework requires measurements being sampled along three dimensions (taxonomical, spatial, and temporal), each detailed in terms of resolution, extent, measurement units, and uncertainty. For such reason, while consolidated standards like Darwin Core (<https://dwc.tdwg.org/>) is suggested for sharing both EBVs and bio-eco EOVs data, EBVs measurements requires high-definition of temporal, geographical, and taxonomical information also at metadata level. Therefore, the Ecological Metadata Language (EML) is recommended by both frameworks, even if in the literature on EBVs metadata schemas such as ISO19115, ISO19157 [15] and INSPIRE Environmental Monitoring Facility [16] are also advised.

Concerning the information system where to deposit data, OBIS is suggested as the repository for registering and finding the bio-eco EOVs due to its historical community of users of marine biological research; instead, the Global Biodiversity Information Facility (GBIF) is recommended for sharing EBVs measurements as it offers standard services to distribute data captured by different initiatives along different taxa and ecosystems (marine, terrestrial and freshwaters). Beside the federation function facilitated by these information systems, different approaches are proposed for the EBVs and the bio-eco EOVs to achieve data interoperability. The Bari Manifesto

⁴ Convention on Migratory Species of Wild Animals;

⁵ Marine Strategy Framework Directive;

⁶ Water Framework Directive;

[17] encapsulates ten principles to outline the current best practices in EBV-focused informatics with the aim to enable interoperability of data and their supplying systems. These principles are related to data management plans, metadata, workflows, ontologies/vocabularies and to the provenance, structure, quality, services, preservation, and accessibility of data, making the operational setting different from that provided by the Framework for Ocean Observing (FOO) [18]. The Bari Manifesto addresses interoperability of research infrastructures so as to assure autonomy of providers in contributing to the EBVs, while in the FOO framework strong attention is given to strengthen collaboration between EOVs providers [5]. In fact, ocean variables other than the bio-eco EOVs are required by a multi-domains community involving not only biology and ecology, but primarily physics and biogeochemistry. Thus, the FOO is intended to conceptually and operationally streamline dialogue between the experts in order to integrate measurements from different domains. For such reason, expanding participation beyond different disciplines as well as from well-resourced countries to a far broader representation of the global community is the primary recommendation [19] given to optimize the exchange and integration of bio-eco EOVs data at global coverage.

As regards tenet 4, the EBVs framework is advancing accessibility of analytical tools by identifying a detailed data management and processing workflow for a small set of EBVs aimed at openly integrating any record in global data products [15]. To demonstrate its feasibility, the workflow was recently tested with data supplied by two biodiversity research infrastructures [13]. This can stimulate the adoption of the approach for analogous products proposed for oceanographic data, which are currently lacking a specific focus on the bio-eco EOVs [4].

III. DISCUSSION

Though EBVs and bio-eco EOVs share no prescriptive approach for the management of related data, from the above comparison along with the five tenets, a set of specific recommendations can be deployed, which must be taken in consideration to enable the reuse of measurements in the scientific practice.

In fact, the five tenets proposed by Benson et al. represent a sound synthetic vision for a coordinated handling of measures and data, but they can be freely interpreted by the data providers and their theoretical contents hide the peculiarities of each framework, which needs to be considered in a conjoint and practical adoption. These recommendations will raise providers' awareness to discern the different technological approaches and to evaluate fit-to-purpose solutions. For example, according to the references analysed which concern tenet 2, a set of policy reporting instruments for biodiversity management are addressed by the EBVs but are not currently satisfied by measuring bio-eco EOVs (e.g. Ramsar Convention on

Wetlands, Habitat Directive, Birds Directive, WFD, MSFD). This affects the reuse of measurements in submitting to the policy obligations of the research program. Also, as highlighted by comparison along with tenet 1, harmonizing methodologies to capture data implies different management issues to face with in programming the reuse and integration of measurements. Instead, concerning tenet 3, a joint management of data and metadata relies on the same recommended solutions (i.e., Darwin Core, EML), facilitating the providers in choosing implementing options. However, additional information can be required according to the variables observed (e.g. additional metadata standards according to the targeted research community in the case of EBVs). Also, according to tenet 4, the maturity of integration processes within the EBVs framework suggests to consider the compliance to formalized workflow steps while planning data collection also for the bio-eco EOVs framework.

IV. RATIONALE OF CASE STUDY SELECTION FOR FURTHER PROJECT DEVELOPMENT

The compliance with data management recommendations needs to be tested in order to bridge the gap between theories and practices. The above section has highlighted that differences exist in the practices of the two frameworks. In order to plan milestones to enact a harmonized data handling for both the EBVs and the bio-eco EOVs frameworks, we identify a suitable case study of a marine observatory system to document its current habits and compare them to the recommendations disclosed in Table 1.

The selection of the case study relies on three requirements: i) presence of a measurements program of marine biological components; ii) adoption of one of the two essential variables frameworks; iii) involvement in an Open-science initiative. We consider four Long-Term Ecological Research (LTER) sites organized in the parent site Northern Adriatic Sea (NAS) (Fig.1). They are 1. the Gulf of Trieste, 2. the Gulf of Venice, 3. the Po Delta and the Romagna Coast, and 4. the Senigallia-Susak transect. The NAS is an ideal analysis context as i) it belongs to LTER-Italy, a research network of collaborative sites (i.e. *in-situ* observation and experimentation facilities) which provide ecosystem measurements with long temporal coverage; ii) LTER-Italy is embracing the EBVs framework to coordinate the parameters selection of individual scientists among multiple monitoring facilities [12] and iii) data collected from both fixed-point observatories (e.g. buoys, pylons) and oceanographic cruises of NAS were recently included in the EcoNAOS project (Ecological Northern Adriatic Open-science Observatory System), which aims at applying Open-Science (OS) principles to marine population time series [20].

- [12] Haase, P, et al. The next generation of site-based long-term ecological monitoring: Linking essential biodiversity variables and ecosystem integrity. *The Science of the Total Environment*, 2018, 613–614: 1376–1384.
- [13] Hardisty, AR, et al. Towards Essential Biodiversity Variables data products for monitoring alien invasive species, 2019, *Environmental Research Letters* 14: 2.
- [14] White, RL, et al. The next generation of action ecology: novel approaches towards global ecological research, 2015, *Ecosphere* 6(8):134.
- [15] Kissling, WD et al. Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale: Building global EBVs, 2018, *Biological Reviews of the Cambridge Philosophical Society* 93(1): 600–625.
- [16] Zilioli, M, et al. Feeding Essential Biodiversity Variables (EBVs): actual and potential contributions from LTER-Italy. *Nature Conservation* 34: 477–503.
- [17] Hardisty, A et al. The Bari Manifesto: An interoperability framework for essential biodiversity variables, 2019, *Ecological Informatics* 49: 22–31.
- [18] Lindstrom, E., Gunn, J., Fischer, A., McCurdy, A., and Glover, L. K. 2012. A Framework for Ocean Observing. By the Task Team for an Integrated Framework
- [19] Miloslavich, P, et al. Challenges for global ocean observation: the need for increased human capacity, 2018, *Journal of Operational Oceanography*
- [20] Minelli et al. The project EcoNAOS: vision and practice towards an open approach in the Northern Adriatic Sea ecological observatory, 2018, *Research Ideas and Outcomes* 4: e24224.
- [21] Mollenhauer et al. Long-term environmental monitoring infrastructures in Europe: Observations, measurements, scales, and socio-ecological representativeness, 2018, *The Science of the Total Environment* 624: 968–978