Article:

Webb, T.J. orcid.org/0000-0003-3183-8116, Beja, J., Bejarano, S.J.F. et al. (6 more authors) (2025) Realising the potential of interoperable data products to improve the outlook for marine biodiversity: lessons from the European marine observation and data network. Marine Policy, 173. 106578. ISSN 0308-597X

https://doi.org/10.1016/j.marpol.2024.106578

Marine Policy 173 (2025) 106578

Contents lists available at ScienceDirect

Marine Policy

journal homepage: www.elsevier.com/locate/marpol

Realising the potential of interoperable data products to improve the outlook for marine biodiversity: Lessons from the European marine observation and data network

Thomas J. Webb [a,*], Joana Beja [b], Salvador Jesús Fernández Bejarano [b], Elvira Ramos [c], Samuel Sainz-Villegas [c], Karline Soetaert [d], Willem Stolte [e], Charles Troupin [f], Benjamin Weigel [g]

[a] Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield S10 2TN, UK
[b] Flanders Marine Institute, Jacobsenstraat 1, 8400, Ostend, Belgium
[c] IHCantabria - Instituto de Hidráulica Ambiental de la Universidad de Cantabria, Santander 39011, Spain
[d] Department of Estuarine and Delta Systems, Royal Netherlands Institute for Sea Research (NIOZ), Yerseke, the Netherlands
[e] Marine and Coastal Systems, Deltares, Delft, the Netherlands
[f] GeoHydrodynamics and Environment Research, FOCUS Research Unit, University of Liège, Liège, Belgium
[g] INRAE, EABX, 50 avenue de Verdun, Cestas 33612, France

ARTICLE INFO

Keywords:
FAIR data
Marine biodiversity monitoring
European seas
Ecoinformatics
Computational tools
Ecological indicators

ABSTRACT

Policies responding to increasing pressures on marine biodiversity require adequate data to support their implementation and to monitor their effectiveness. Marine biodiversity science has made significant progress generating and aggregating biodiversity data, however turning this into evidence-based knowledge useful to decision makers remains a significant challenge. 'Data products' provide processed data to address specific user needs, and are widely used in climate science, geosciences, and remote sensing, but the development of biodiversity data products is challenging due to the complexity of biological systems and of the data derived from surveys designed without explicit biodiversity policy or management guidance. A wide range of potential products of interest may include distributional data for thousands of individual taxa, requiring advanced statistical methods to model patterns in biodiversity using heterogeneous and sparse source data with biases in spatial, temporal, and taxonomic coverage. We illustrate these challenges using data products created within the Biology thematic lot of the European Marine Observation and Data Network (EMODnet), and we propose that the EMODnet Biology approach, which involves providing clear and open documentation of the product creation process with a strong emphasis on the computational tools needed to link source data to higher-level data products, can productively support decision making at the European scale. Furthermore, this approach provides part of the essential infrastructure required to maximise the financial benefits of FAIR data, and data products play a key role in empowering users to make maximum use of existing biodiversity data to help to understand and manage our seas.

1. Introduction

Our seas are subject to increasing pressures from demands for food, materials, and space [1]. In the European context, a wide range of policies at national, regional, and international scales have been developed to conserve and maintain marine biodiversity in the face of these accelerating and often competing demands. These include the Birds and Habitats Directives [2], the Water Framework Directive (WFD, 2000/60/EC), the Marine Strategy Framework Directive (MSFD, 2008/56/EC), and the EU Biodiversity Strategy for 2030 (COM/2020/380). Norway and the United Kingdom have similar national regulations, such as the Norwegian Integrated Management of the Marine Environment of the North Sea and Skagerrak [3], and the UK Marine Strategy [4–6]. Several Regional Seas Conventions relate to European seas, including the OSPAR Commission in the northeast Atlantic (https://www.ospar.org/), the Baltic Marine Environment Protection Commission (HELCOM, https://helcom.fi/), the Mediterranean Action Plan (MAP, https://www.unep.org/unepmap/), and The

* Corresponding author.
E-mail address: t.j.webb@sheffield.ac.uk (T.J. Webb).

https://doi.org/10.1016/j.marpol.2024.106578
Received 27 June 2024; Received in revised form 18 December 2024; Accepted 20 December 2024
Available online 8 January 2025
0308-597X/

Commission on the Protection of the Black Sea Against Pollution (http://www.blacksea-commission.org/). Cooperative agreements on taxonomic groups include the North Atlantic Marine Mammal Commission (NAMMCO, https://nammco.no/). EU Member States have also agreed to protect 30 % of their seas under the Kunming-Montreal Global Biodiversity Framework during the Convention on Biodiversity in 2022 (https://www.cbd.int/gbf), with an additional ambition for strict protection stemming from the European Biodiversity Strategy for 2030 (COM/2020/380). On top of this - either by legal commitments or voluntarily - many European coastal states have nature restoration in general terms on their agenda, which includes enhancing, mitigating and compensation measures for species and habitats impacted by human activities.

Although these policies have differing origins and diverse goals, all of them require adequate data to support their implementation and to monitor their effectiveness. This includes data and 'data products' (see below for formal definition) for the systematic monitoring of the status of biodiversity (or indicators thereof) in space and time. Such indicators have been codified in MSFD Descriptors of Good Environmental Status (GES). Descriptor 1, for example, requires that "Biological diversity is maintained" and that "The quality and occurrence of habitats and the distribution and abundance of species are in line with prevailing physiographic, geographic and climatic conditions" (http://mcc.jrc.ec.europa.eu/main/dev.py?N=19&O=118&titre_chap=D1%20Biological%20diversity). OSPAR Common Indicators (http://www.ospar.org/work-areas/cross-cutting-issues/ospar-common-indicators) also include a range of indicators related to biodiversity, including the condition of benthic communities, changes in plankton communities, and abundance of certain key species of marine mammal and seabird. More generally, Essential Biodiversity Variables (EBVs) have been suggested as a unified way of capturing simultaneously the distribution or abundance of multiple species in space and time [7], while the related Essential Ocean Variables (EOVs) have been developed to assess and detect changes in marine biodiversity and ecosystems with relevance to societal needs [8]. Robust implementation of these measures is dependent upon transparent, open, and harmonised cross-boundary monitoring data.

Fortunately, in parallel with the data requirements for effective marine biodiversity policy becoming apparent, marine biodiversity science has made enormous progress in collating existing data and generating new data [9,10], through historical data rescue by scientists and citizen scientists [11,12], sustained monitoring programmes [8,13], aggregation of data from scientific surveys, structured citizen science programmes, and opportunistic observations [9]; and increasingly by harnessing new technologies (e.g. environmental DNA, remote sensing, biologgers, automated vehicles; [14,15]). The supporting infrastructure, including servers, software, and tools, required to store, serve, process and interpret all this data have also improved massively [10]. However, many of the surveys - which are the ultimate sources of marine biodiversity data - have not been designed with explicit biodiversity policy or management requirements in mind, so that turning the raw biodiversity data into information useful to decision makers remains a significant challenge. Ensuring that all evidence-based knowledge about biological systems that is created for policy purposes is 'climate-aware' adds another layer of complexity and urgency [16,17].

The creation of *data products* - "an instance of persistent [meta]data which has been processed to be offered to external users" [18,19] - can help to facilitate this translation of raw data into evidence-based information. Data products are very commonly used in fields such as climate science, geosciences, and remote sensing (e.g. global surface temperature anomaly data products [20], global elevation models [21], and other global environmental and bioclimatic layers [22,23]). Hardisty et al. [24] provide a framework under which the concept of data products can be extended to biodiversity data, ultimately leading to reliable and interoperable EBV products. They state that "EBV data products that are sufficiently large (e.g., in terms of data volume, coverage, granularity) and comprehensive (in terms of temporal and spatial scales) would facilitate forecasting and assessing the impact of management interventions on biodiversity from national to global scales" [24], thus playing a crucial role in the management of marine ecosystems. This idea has been further developed in the specific case of European marine biodiversity data products [19] (see below). However, despite the existence of this framework, biological data products - in particular those derived from biodiversity data aggregators, rather than from individual standardised surveys - have typically proved challenging to develop and have not been widely taken up by science or policy stakeholders. As we discuss below, this is in part due to the complexity of biological systems and the high and heterogeneous dimensionality of biological data, resulting in a wide range of potential products spanning the movement of individual organisms, the distributions of populations and species, through to assemblage-level metrics of biodiversity change in space and time.

Here, we review the philosophy and process behind the creation of biological data products within the Biology thematic lot of the European Marine Observation and Data Network (EMODnet, https://emodnet.ec.europa.eu/en/biology). Using a selection of products varying in spatial, temporal, and taxonomic scope, as well as the methodology used and the 'level' (*sensu* Lear et al. [19]; see Fig. 1) of the product produced, we outline how the products that have been developed have been - or could be - applied to support decision making at the European scale, as well as acting as templates and catalysts to the development of further products. We highlight mismatches between the stakeholder desires as documented by [19], the data available to EMODnet Biology, and the products of most interest to the scientific community, and we discuss how these discrepancies can be addressed. We also document how a common set of web services serving data from across the various EMODnet thematic lots can enable biological data products to be developed together with data from the other thematic pillars (chemistry, physics, bathymetry, geology, seabed habitats, and human activities), leading to more synthetic outputs and insight. Throughout we emphasise the importance of transparent documentation of the products themselves, and of computer code and associated tools required to produce them. Our overall aim is to show how the lessons from EMODnet Biology can be applied more generally to the development, dissemination, and uptake of data products that improve understanding and management of our seas in the face of multiple and increasing pressures on marine biodiversity.

2. Data Products for the Marine Environment: EMODnet Biology as a case study

The European Marine Observation and Data Network (EMODnet, https://emodnet.ec.europa.eu/) is a network of organisations supported

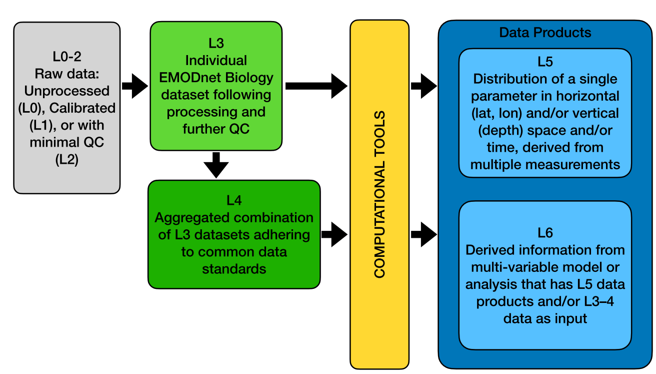

Fig. 1. From data to products via computational tools. Representation of the levels of data products (L0-L6) as set out by Lear et al. [19] and defined by the wider EMODnet community, illustrating the key role of computational tools in developing higher-level biological data products. Data, products, and tools at each level will have associated descriptive metadata to ensure that each level remains FAIR.

by the EU's integrated maritime policy, and is the EU's operational service for marine *in-situ* data. EMODnet is committed to processing data obtained from marine observations according to international standards, and to make the information freely available both as interoperable data layers and as data products, following the FAIR principles of data stewardship – namely that data and research objects should be Findable, Accessible, Interoperable, and Reusable [25]. Information about the data ingestion process developed by EMODnet, including criteria for inclusion of datasets into the network, is provided by [26]. EMODnet is organised around seven themes: Bathymetry, Biology, Chemistry, Geology, Human Activities, Physics and Seabed Habitats. The themes differ in the nature of the data that they have aggregated at L0–4 (Fig. 1), which are then served to the community via the Central Portal product catalogue (https://emodnet.ec.europa.eu/geonetwork/srv/eng/catalog.search); a selection of these are available to subset, merge and download via the Central Portal map viewer (https://emodnet.ec.europa.eu/geoviewer/). EMODnet metadata records are associated with data and products at all levels of the product hierarchy

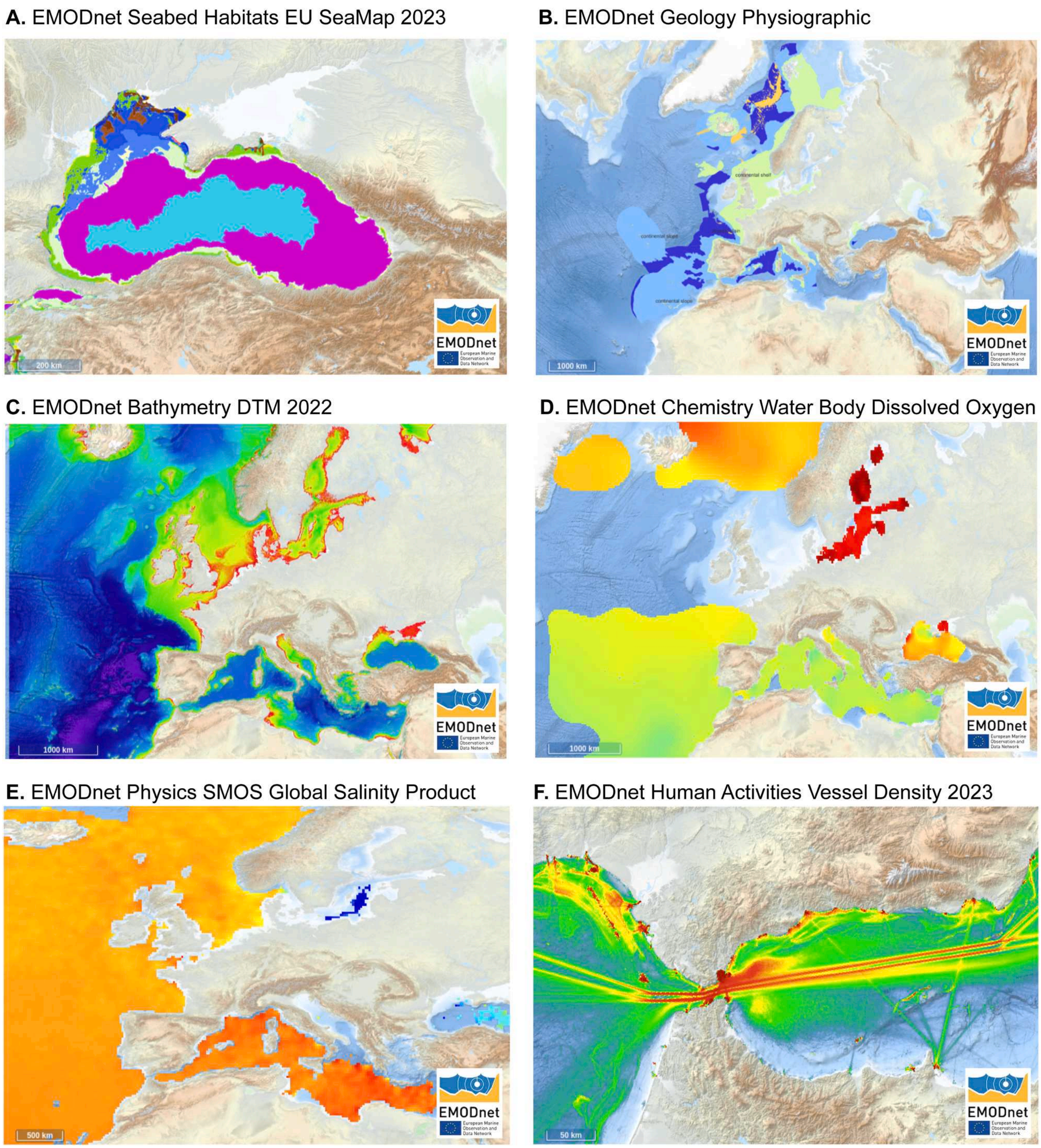

Fig. 2. Examples of non-Biology data products produced by EMODnet Thematic Lots. Polygon data is represented as **A** classifications of seabed habitats (EMODnet Seabed Habitats; [28]) and **B** general physiographic features (EMODnet Geology; [30]), with colours indicating different classifications. Examples of gridded field data are **C** bathymetry (EMODnet Bathymetry; [31]), **D** water body dissolved oxygen (EMODnet Chemistry; [32]), **E** sea surface salinity (EMODnet Physics; [33]), and **F** vessel density (EMODnet Human Activities; [34]). Gridded fields have a continuous colour scale from blue through yellow to red. Full details of the data displayed in each panel can be found in the sources cited.

(Fig. 1), and these comply with the INSPIRE (Infrastructure for Spatial Information in the European Community) Directive 2007/2/EC, which defines how the geospatial data should be formatted and made available (see [27]). While these standards impose some constraints on how data and data products can be classified, they ensure that the product catalogue provides access to all available resources (>5500 at the time of writing) across all EMODnet thematic lots. The extensive documentation provides users with a clear description of the criteria for inclusion of data within any specific data product. Resources can be filtered by selecting the thematic lot providing the resource or by free text search. All thematic lots have produced at least one data product (L5–6 in Fig. 1), with most also hosting external data products that have been created by organisations or individuals from outside of the EMODnet consortia. The number of external products is expected to increase due to the European Research and Innovation projects that are required to make their outputs (data and/or data products) available in EMODnet. For instance, projects funded under the Biodiversity and Ecosystem Services call (HORIZON-CL6–2022-BIODIV-01) include commitments to incorporate relevant marine data and data products into EMODnet.

Products from most EMODnet thematic lots often involve the spatial representation of one or a small number of variables. This is true even of some of the more complex products. For example, EUSeaMap 2021 [28] is a broad scale seabed habitat map of European seas, derived from a wide range of survey data and classified into EUNIS (European Union Nature Information System; [29]) habitat categories. In essence, however, this is an L5 product (Fig. 1) depicting the distribution of a single parameter (seabed habitat) in space, and it is relatively easily viewed (both conceptually and literally) as a standard map, with polygons to represent different seabed habitat types (Fig. 2A). Additional complexity, such as the associated measures of certainty in habitat type assignations, can also be mapped in an equally straightforward way (see [28]). Similar maps, depicting either polygons of specific features (e.g. general physiographic features from EMODnet Geology [30]; Fig. 2B) or gridded fields (Fig. 2C-F) including bathymetry from EMODnet Bathymetry [31], water body dissolved oxygen from EMODnet Chemistry [32], sea surface salinity from EMODnet Physics [33], and vessel density from EMODnet Human Activities [34]. While some of these products can include dimensions other than two-dimensional space - commonly time and / or depth - these remain relatively straightforward to represent in standard data structures, and many of them are very well suited to formats such as Network Common Data Form (NetCDF) and associated Climate and Forecasting (CF) metadata standards widely used for multi-dimensional gridded data [35–37]. Time-varying spatial data can also be visualised; depending on the viewer or software used, this may include mapping only a specific time period of interest, averaging over time, or animating maps to show changes over time.

3. Data Standards and Non-Standard Data: Dimensionality, complexity, and heterogeneity of marine biodiversity data products

Some Biology products fit easily into the same general framework as the non-Biology products shown in Fig. 2, in terms of both the underlying data structure and the ability to display them on maps. For instance, we can use the methods of species distribution modelling (SDM) [38] to map the current and predicted future geographic distributions of species of key ecological or societal interest. A good example is non-native or invasive species, such as marine seaweeds that have spread outside their natural boundaries, sometimes being transported into new oceans and water bodies by human activities, with impacts on biodiversity and the economy [39]. However, the inherent characteristics of these species, such as the lack of niche equilibrium, make modelling their distributions a challenging process where the model configuration, the environmental predictors selected, and the quality of the input data are key aspects to be considered [40,41]. An EMODnet Biology data product has created SDMs to model the historical and projected distributions of four invasive macroalgae (*Asparagopsis armata*, *Caulerpa taxifolia*, *Sargassum muticum* and *Undaria pinnatifida*) along the European coasts (Fig. 3A; [42]). The product consists of modelled distributions of each species on a 5 arcmin resolution grid (c. 0.08° or 9.2 km at the equator) at each of two time periods (historic and 2100 IPCC RCP 8.5 scenario), which fits well into standard data structures (e.g. lat-lon-time NetCDF) and can also be mapped with a manageable number of options. This data product could be of interest to managers to anticipate, prevent and reduce the potential negative impacts of these alien species in non-colonised ecosystems, as it helps to identify the potential areas of expansion with models based on reliable information and optimised configurations.

Modelling species distributions also demonstrates how the process of developing data products can lead to improvements in the methodologies used to create them. For example, another EMODnet Biology data product models the distribution of two ecologically important species of planktonic copepods in the North Atlantic, *Calanus finmarchicus* and *C. helgolandicus* (Fig. 3B; [43]). This product is built using a single large constituent dataset of EMODnet Biology (i.e., L3 data in Fig. 1), the Continuous Plankton Recorder (CPR) survey [44], and creates smoothly interpolated distribution maps using the DIVAnd tool [45]. DIVAnd is an evolution of the DIVA (Data-Interpolating Variational Analysis) tool [46], which enables interpolation in an arbitrary number of dimensions, typically longitude, latitude, time and depth. DIVAnd was initially applied to produce gridded temperature and salinity maps for different European regions and eutrophication-related variables (e.g. nitrate concentration, dissolved oxygen concentration) within EMODnet Chemistry. Extending this approach to modelling zooplankton distributions required the method to be modified to allow for multivariate analysis: instead of a single variable processed at a time, now DIVAnd can incorporate a second variable for which observations are available to improve spatial coverage. In this specific case, because *C. finmarchicus* and *C. helgolandicus* are known to be sensitive to sea water temperature [47], the interpolated maps included temperature to improve spatial interpolations. Because of the high frequency at which the CPR survey is conducted, the *Calanus* distribution maps have been created at monthly time intervals for the years 1959–2018, so there are potentially 720 maps per species. In addition to the gridded density field, DIVAnd outputs also include an error field associated with each density map as an indicator of confidence in the interpolation process, with this confidence generally increasing with the density of observations (Fig. 3B). Although the large number of maps can present challenges for online map viewers, conceptually this is similar to serving environmental variables such as temperature data at high temporal resolution. The underlying data structure remains simple to represent in standard formats such as NetCDF, as for each species there are only regular dimensions of space and time.

Expanding the taxonomic and spatial scale of Biology data products typically requires aggregation and combination of many L3 datasets (*sensu* [19]; Fig. 1) from multiple sources, covering many hundreds or thousands of individual species sampled in different places and at different times, by different teams of surveyors using different methodologies. This process of combining data can present problems, for instance in distinguishing 'true' absences (a species was searched for at a location but not found) from 'pseudo' absences (e.g. when surveys at a location did not search for all species). Extensive quality control - which can be automated to some extent, but which still relies on expert input and interpretation - can overcome many of these issues. For instance, an EMODnet Biology data product [48] has identified constituent datasets which comprise whole-community surveys of the macrozoobenthos, allowing absence to be inferred - if a macrozoobenthic species is not recorded in a dataset in which we know the whole macrozoobenthos community was routinely surveyed, it is a reasonable assumption that the species was genuinely absent, rather than present but not recorded. Thus, multiple - but crucially not all - L3 datasets including relevant taxa were combined into a single product documenting presence and

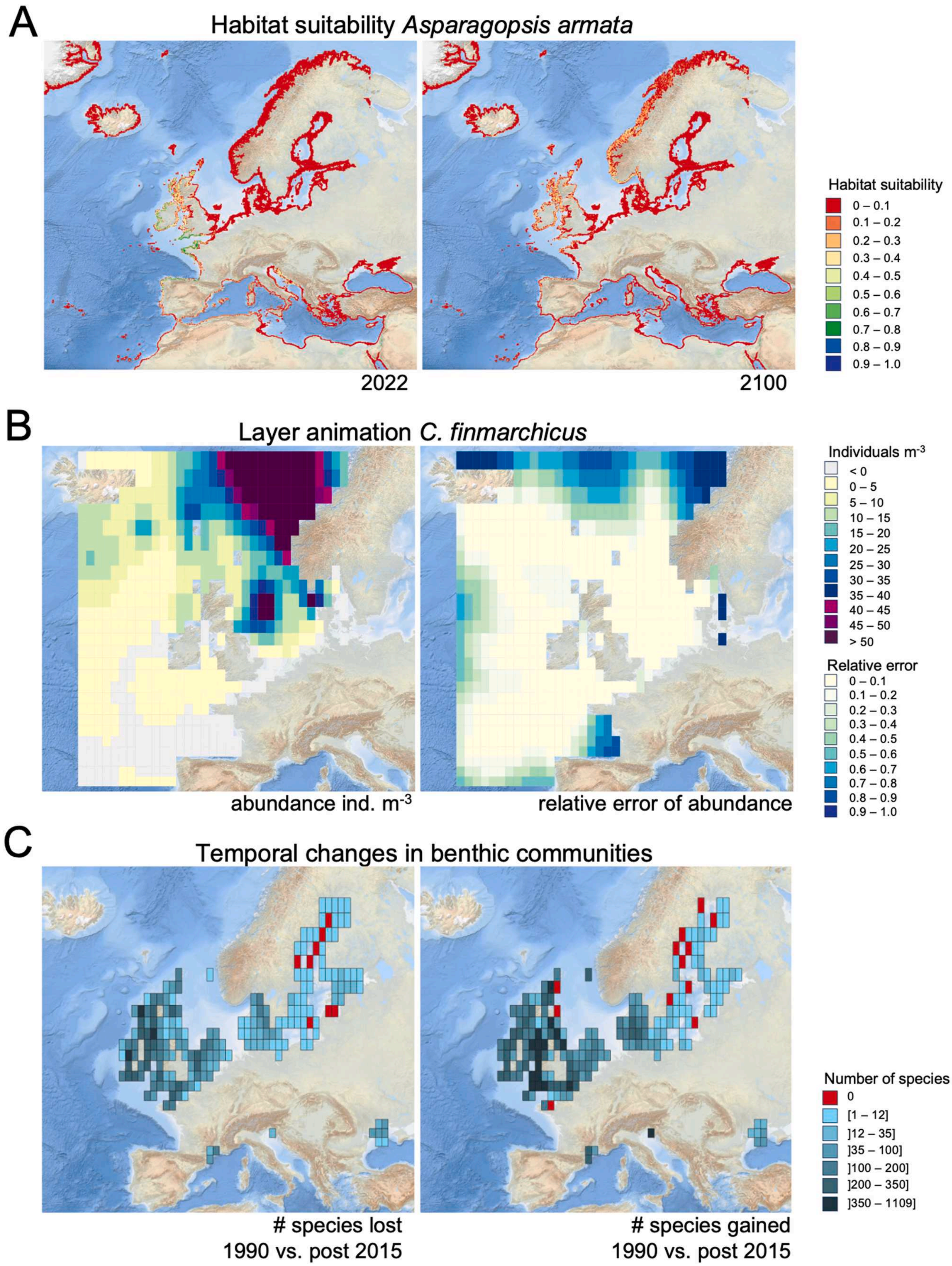

Fig. 3. Examples of Biology data products extracted from the EMODnet MapViewer (https://emodnet.ec.europa.eu/geoviewer/). **A** Habitat suitability predictions for the seaweed *Asparagopsis armata* in 2022 (left panel) and 2100 (right panel) following RCP 8.5. **B** Snapshot of the layer animation of *Calanus finmarchicus* abundance (left panel) and the associated error fields (right panel) in 1970 (animation covers 1959–2018), based on the DIVAnd interpolation tool. **C** Temporal changes in benthic communities in terms of species lost (left panel) and species gained (right panel), comparing the 1990s with years post 2015.

reconstructed absence data for 10,830 benthic taxa at 292,332 sampling stations throughout European seas [48]. The addition of (inferred) absence data greatly expands the scope of the data for modelling species distributions as a function of environmental or human impact variables, for instance in distinguishing between different reasons why a species is not present at a given site [49]. Herman's [48] benthos presence-absence product has enabled the production of detailed distribution models of certain species of high importance, such as reef-forming species in the North Sea [50]. However, a large part of its value derives from its very extensive spatial and taxonomic coverage, enabling subsequent analyses of the entire macrozoobenthic community at continental scales.

The spatial coverage is standard for EMODnet products, but the survey-level spatial resolution (i.e. point samples) means that some level of aggregation is required to coerce the data into gridded latitude-longitude structures, to avoid creating a prohibitively large, very sparsely populated grid with billions of cells representing every possible combination of latitude, longitude and taxon. There are options within the CF convention to represent taxonomic abundance time series, but it remains the case that forcing sparse point data into a gridded structure is neither straightforward nor practical from a user perspective. In addition, it can be challenging to incorporate other relevant dimensions in biology such as season or future climate change scenario, as well as supplementing species-level abundance or occurrence data with additional biological information. This may include life stage (especially important for meroplanktonic species) or more detailed information on species' biological traits, which has important applications in assessing benthic ecosystem functioning [51]. Other complex data structures may arise when combining data of different types, for example, by mapping species occurrences onto seabed habitats to derive sets of species lists by habitat type. More generally, there are no clear standards to store biological occurrences in multidimensional data cubes. The interim solution in the case of the benthos presence-absence product was to accept that a single data standard is not appropriate for the full variety of Biology data products, and to store this data product as linked tables in vector format using Darwin Core standards [52]. In future new initiatives such as B-Cubed (https://b-cubed.eu/) or EDITO-INFRA (https://edito-infra.eu/) may provide more robust and generalisable solutions or standards that are applicable to this problem.

4. Products upon products

One of the advantages of the data product model is that new products can be created either as summaries or analyses of existing products, or through the combination of multiple existing products. Derived products may be designed to overcome the complexities outlined above - for example, by summarising measured or estimated species richness derived from the benthic presence-absence product, which would be much simpler to map than the distributions of each individual species – but they can also introduce additional complexity and/or other technical challenges to the simple display of product outcomes. An example of how derived products can add complexity is the EMODnet Biology product that examines the temporal turnover of European macrobenthic communities by mapping changes in community composition through time [53], using the benthos presence-absence product [48] as its source data (Fig. 3C). Serving this product in a viewer is challenging for two reasons. First, the temporal coverage of the source data is uneven, making it impossible to compare regular time periods (e.g. every year, or every decade). Instead, six time periods are defined based on data availability (before 1990, 1990–1999, 2000–2004, 2005–2009, 2010–2014, and 2015 and after). To be included in the product, individual grid cells required benthic data from at least two of these time periods; the number of comparisons for a given grid cell therefore varies from 1 (where a cell contains observations from only two time periods) to a maximum of 15 (where a cell contains observations in all six time periods). Representing all pairwise comparisons in a viewer requires either a list of each individual comparison (e.g. 'pre-1990 versus 2005–2009'), that can lead to an awkward user interface, or a more sophisticated interface allowing the user to select both start and end points.

Adding to this complexity, there is no single measure of community turnover. The benthic turnover product [53] therefore presents several metrics, including three measures of beta diversity (total beta diversity, turnover due to species replacement, and turnover due to changes in species richness; [54]), and five measures of species turnover (species shared between time periods, species lost, species gained, and proportion of species lost and gained). This results in 120 possible combinations of the 15 temporal comparisons and 8 turnover metrics. Therefore, although conceptually this product could be mapped - e.g. 'show me how many species have been lost and gained from each 1 degree grid cell between the 1990s and the 2000s' (Fig. 3C) - in practice the data structure is challenging to represent. A single NetCDF file can be created (and is currently used to serve this product via the EMODnet Central Portal viewer; see Fig. 3C), but enabling a user to interpret the dimensions and variables requires substantial additional explanation. Alternate formats - e.g. a single GeoTIFF for each turnover metric, with layers for each temporal comparison - may be more easily understood by users, but this would require separate documentation and would not fit the aspiration of a single standard for all products.

Another example of a product built by combining two existing products involved matching the benthic presence-absence data product to the broadscale seabed habitat map EUSeaMap 2021 [28] (Fig. 2A). The aim of this derived product was to summarise the habitat affinities of benthic species, considering the habitats in which they have been recorded as both present and absent [55]. Although both input products are very large, the primary derived product is simply a 3.4MB csv file containing habitat summary data for each of 3287 benthic species, with each species in a separate row, and each habitat variable in a separate column. Habitat summaries include various habitat descriptors, including for instance the proportion of all presence records for a species occurring within each EUNIS habitat type. Because there are a large number of habitat variables - including categorical EUNIS types as well as quantitative estimates of sediment composition derived from [56] - in total, 147 variables are recorded for each species. Code also exists to generate summary infographics for each species (Fig. 4), but neither the derived data file, nor these plots, contain explicit and mappable spatial information. Thus, this product cannot be displayed in a map viewer, which limits its visibility to potential end users, and which in turn presents a challenge to its uptake (see *Use and Uptake of Biodiversity Data Products*).

5. Computational tools as an interface between biodiversity data and biodiversity data products

Although the specific details of EMODnet Biology products reflect the interests and expertise of the partners who have developed them, a significant effort has focused on providing open and reproducible code and related tools to allow users to interact with EMODnet data and expand upon the products that have been developed. This could include, for instance, creating updates of existing products to incorporate new data ingested into EMODnet, either from new sources or as regular updates from ongoing surveys. These computational tools provide a critical link between datasets and L5–6 data products (Fig. 1). In addition to code and documentation used to build, enrich, and apply metadata standards to all EMODnet Biology products (see https://github.com/EMODnet), we identify two primary groups of tools that can be considered as products in their own right: software packages to access EMODnet data by interfacing with web services, and standalone packages that serve up subsets of data with associated statistical and visualisation methods.

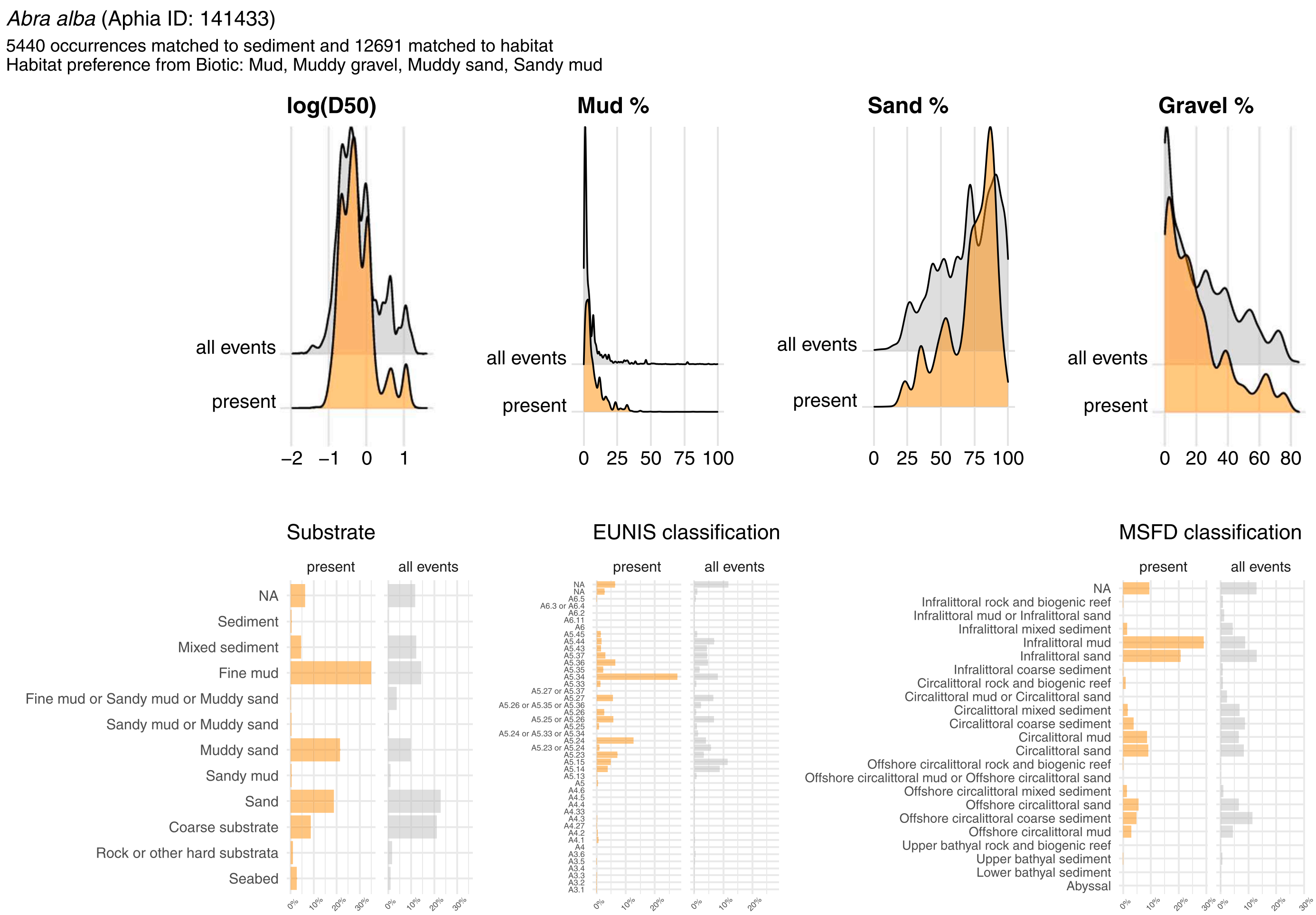

Fig. 4. Example of an infographic generated by the EMODnet Biology data product [55] that matches Herman's [48] benthos presence-absence data product to the broadscale seabed habitat map EUSeaMap 2021 [28] and a sediment composition dataset [56]. The graphic provides the species name (*Abra alba*) obtained from the provided WoRMS Aphia ID, and summarises the number of presence records derived from [48] that were successfully matched to the habitat and sediment datasets (which have different spatial extent). Qualitative habitat preference information from the Biological Traits Information Catalogue (BIOTIC; [57]) are also provided if available. The top row of plots shows the density distribution of median sediment grain size, and percentage mud, sand, and gravel, from [56], for surveys where *A. alba* was present, and for all unique survey locations in [48]. The bottom row similarly plots frequencies of *A. alba* presences and all survey locations by various habitat classifications available within [28].

5.1. Products to access EMODnet data

All higher-level (*sensu* [19]) data products require access to data at lower levels. EMODnet thematic lots provide web services to allow programmatic access to all data and data products. Facilitating access to EMODnet data via these web services, within the R computing environment [58], has been a significant focus within EMODnet Biology. R is very widely used in ecology and biodiversity science [59–61], and has a number of features (including the ability to link with other platforms, e.g. Git, Docker) designed to promote open and reproducible research (e.g. [62–64]), which enable products to be developed and documented in alignment with open science principles. Tools to better expose EMODnet data to R users therefore have significant potential to stimulate innovative use cases.

EMODnet thematic lots serve data in two primary formats. Geographic vector data is served via EMODnet Web Feature Services (WFS), while geographic raster (i.e., gridded) data is served via EMODnet Web Coverage Services (WCS). EMODnet Biology has developed R packages to make both WFS and WCS data available to the vast community of R users. The EMODnetWFS R package [65] provides a direct interface to fine-grained geographic vector information at the feature and feature property level, allowing them to be queried, filtered, manipulated and visualised in R as 'simple features' using the popular sf R package [66,67]. Furthermore, EMODnetWFS enables geographic vector data to be queried and filtered on the server using GeoServer's Enhanced Common Query Language (ECQL, https://docs.geoserver.org/latest/en/user/filter/ecql_reference.html), allowing the user to interact with very large spatial datasets without the need to read the full dataset into R (see https://emodnet.github.io/EMODnetWFS/articles/ecql_filtering.html for details).

Although many EMODnet datasets are served via WFS, there are also many gridded data products that are instead served via WCS, a standard created by the Open Geospatial Consortium (OGC, https://www.ogc.org/) that is analogous to WFS, but for raster data. The EMODnetWCS R package [68] has been created to allow interrogation of and access to EMODnet's geographic raster data in R. WCS is able to return metadata relating to EMODnet's WCS services and available coverage (e.g. geographic, temporal and elevation extents, grid sizes and resolution, coverage coordinate reference) and EMODnetWCS allows the user to download compiled metadata and to extract individual metadata in more usable form (see https://emodnet.github.io/EMODnetWCS/articles/metadata.html). As with EMODnetWFS, EMODnetWCS interacts directly with the most relevant geospatial R packages - in this case returning coverages in the SpatRaster format used by the terra package [69], now the primary R package for working with gridded (raster) data.

5.2. Packaging data into products

The Biology data products that can be built on EMODnet data using the EMODnetWFS and EMODnetWCS packages may still fit neatly into standard spatial product formats that are straightforward to visualise. However, these packages also facilitate linking different data types together, and the resulting products may be better served in different formats. EMODnet Biology includes examples of these highly derived products, such as the product documenting habitat affinities of benthic species [55] (Fig. 4). Other products created and published by the other thematic lots can also be queried and downloaded using these two R packages. One way to improve the useability of such products is to present them as self-contained packages. The Btrait R package [70] is an example of this, packaging up data on the distribution (density and biomass) of macrobenthos in the North Sea together with information on their taxonomy and biological traits. The package provides functions for users to interact with these data, creating summaries at taxon or trait level over time and space. It is supplemented by the package BtraitWeb which provides an interactive interface to the data via a Shiny app. R Shiny apps facilitate collaboration between scientists, and between scientists and stakeholders - including those with minimal specialist technical knowledge - on datasets up to moderate size [71]. As such they may be a useful way of disseminating some of the more complex biological data products described here, and of encouraging use and uptake across a broad range of users.

6. Use and uptake of biodiversity data products

The data products produced by EMODnet Biology are designed to be used, and significant input from end users has been actively sought, for instance in dedicated workshops [72], in an effort to direct the development of these products to make them useful. However, as is commonly the case with open data [73], use and uptake beyond the product creators has been limited, and is challenging to track. In this section we evaluate features of the data products from EMODnet Biology that have been most widely used, how these compare to widely-used data products produced elsewhere, and whether the complexity of biology data products described above necessarily represents a barrier to their uptake. We consider the extent to which existing data products meet the needs identified by stakeholders [19], and whether this is still impacted by a lack of open data in some key areas. Finally, we propose guidelines to improve use and uptake of future biology data products developed by EMODnet Biology or by other networks and consortia.

6.1. Actual and potential use of EMODnet Biology data products

EMODnet Biology has played a key role in mobilising marine biodiversity data and aggregating it into a quality-controlled database of species occurrences which forms a significant part of the global Ocean Biodiversity Information System [74]. OBIS in turn is widely used to inform global assessments of the marine environment, including the UN's World Ocean Assessments [75,76], as well as in more targeted assessments, e.g. of the contribution of UNESCO World Heritage Sites to biodiversity conservation [77]. It is by providing access to biodiversity data as a Level 4 product (Fig. 1), as well as facilitating the flow of a range of new or non-standard data types (e.g. environmental DNA, images, citizen science data) into the aggregated dataset, that EMODnet Biology has gained greatest traction. Data products at higher (L5-L6) levels have so far been used less frequently, although there are documented use cases, such as using the benthos presence-absence product [48] to create improved models of the distribution of reef-building benthos in the North Sea [50]. Here we identify two reasons why higher-level data products have not been more widely used, and suggest potential developments that may improve future uptake.

First, there is no obvious candidate for a single (or small set of) archetypal synthetic biodiversity data product. This contrasts with products from other themes, where 'doing one thing well' is closer to being realised. For instance, it is quite clear that an obvious product related to seabed habitats is a synthetic map of seabed habitats, such as has been achieved with EUSeaMap [28]. Likewise, a map of bathymetry is something that combines both a clear concept and high utility. While other themes may have many potential variables of interest (e.g. measurements of multiple physical or chemical properties of seawater), the form that a product should take to maximise utility - a gridded field of the value of this relatively small set of parameters in space and/or time - is often unambiguous. For biology products, as documented above, there is considerable added complexity, encompassing the distributions of many thousands of individual taxa as well as summary statistics (e.g. species richness, species turnover) in space and time. Because of significant spatial and temporal bias in biodiversity sampling (e.g. [9]), creating smooth interpolated distribution maps is a major statistical challenge even for individual species, and although this can be addressed (as it has been in some EMODnet Biology products, e.g. [43]), there will remain considerable uncertainty in predicted distributions, with the extent of this uncertainty varying between taxonomic and functional groups as well as in different marine habitats. Fully quantifying gaps and biases in existing databases aggregated from heterogeneous sources, and developing robust statistical methods to use them to create synthetic variables to map the current distribution and temporal trends in biodiversity, remains an active area of research [78–83]. A consequence is that analysts have addressed this complexity in different ways, creating the wide range of biology data products described above; and the use of any one of these will be limited given the necessary compromises made with regards to spatial, temporal, and taxonomic scope.

The second challenge to the use and uptake of data products is the extent to which they map onto the needs of key stakeholders. For biodiversity themed data products, the needs of many of these stakeholders will be driven by the reporting requirements of legal frameworks such as the MSFD or the Regional Seas Conventions. The indicators used under such frameworks to track changes in the state of marine biodiversity are now quite well established, and it has previously been suggested that EMODnet Biology has an important role in providing the data and products required to calculate and monitor the value of these indicators [19]. Efforts to document the relevance of EMODnet Biology data products to MSFD descriptors in the Baltic Sea [84] did identify products with potential applicability to biodiversity and to non-indigenous species descriptors, but the extent to which these could be operationalised as functional indicators is unclear. For instance, the benthos presence-absence product [48] is identified as relevant to the biodiversity descriptor, however the lack of an explicit temporal dimension to this product, and the uneven temporal coverage of the underlying data, make the quantification of change from a baseline a significant challenge, and limit its utility as an indicator of the maintenance of biodiversity.

6.2. Identifying a niche for EMODnet Biology data products

The major strengths of EMODnet Biology data - the increased spatial and taxonomic coverage achieved by aggregating many individual datasets - may be less relevant to users needing to track the status and trends of specific components of biodiversity using individual datasets obtained with consistent methodologies, for which they may already have developed bespoke data and analysis pipelines. Some taxonomic groups of high interest are also the most challenging to build robust data products for. For instance, OSPAR's Biodiversity Monitoring and Assessment work (https://www.ospar.org/work-areas/bdc/biodiversity-monitoring-assessment-1) includes six indicators of marine mammal status, but contributing data products to aid these assessments is complicated by a reluctance of some data providers to make their survey data FAIR, as well as the fact that sightings and occurrence data alone are of limited use in assessing populations and distributions of marine mammals - rather, considerable time and analytical expertise is required to control for multiple factors impacting detectability in order to produce credible outputs (e.g. [82]). As a result of these and other data quality and useability issues [80], marine mammal data products built naively on EMODnet Biology data may lack credibility among key stakeholders.

This potential credibility gap is one reason why building new products to meet existing needs or requirements risks low uptake, as individual users may already have preferred data sources (which may or may not be FAIR) and workflows that meet specific data quality and auditing requirements. Rather than attempting to replicate existing indicators and assessments with data not designed for that purpose, an alternative and potentially more productive approach may be for initiatives like EMODnet Biology in future to work with relevant stakeholders to co-design new indicators that play to the strengths of the available data. A similar process has previously proved effective in the development of an indicator of plankton communities to track changes in broad classes of plankton life forms [85]. This approach has a number of similarities with EMODnet Biology data product creation, including the combination of heterogeneous data (in this case, combining time series from numerous fixed point coastal sampling schemes with the spatially extensive Continuous Plankton Recorder survey), and enriching data on species distribution and abundance with information on species traits to construct the suite of lifeforms, or groups of species that carry out similar functional roles in the ecosystem, such as large phytoplankton, gelatinous zooplankton, and fish larvae/eggs [85]. Using the complementarity and synergies between different constituent datasets to build synthetic indices at a broad scale may be the best way of maximising the contribution of EMODnet Biology data to inform management and policy, rather than attempting to replicate more localised or taxonomically specific metrics. The key to uptake here may be for Biology data products to more clearly and explicitly emphasise how they relate to existing metrics such as MSFD descriptors.

7. Conclusions

Data products are commonly created in a range of fields to process complex observational data into a format that is more generally useful to external users. In this paper, we have used the data products created within the Biology thematic lot of the European Marine Observation and Data Network (EMODnet) to illustrate some of the advantages of a product-led approach to sharing marine biodiversity data to support decision making at the European scale. Building on the typology presented by [19] (Fig. 1) we have emphasised in addition the vital importance of computational tools in linking source data to higher-level data products. We have shown how inherent features of biological data, as well as methodological challenges in dealing with it, add significant complexity to the process of data product creation. These include high dimensionality (e.g. products that include distributional data for thousands of individual taxa), the wide range of potential products of interest, and the statistical challenge of modelling patterns in biodiversity using heterogeneous and sparse source data that may include known or unknown biases in spatial, temporal, and taxonomic coverage. A response to these challenges may be to develop products that are narrow in scope, useful for some specific tasks but with limited utility outside of those specific purposes. Alternatively, product designers may lean into the strengths of EMODnet Biology data, such as its highly extensive spatial, temporal, and taxonomic coverage, to develop products that are extremely broad in scope but which may have shortcomings when applied to a specific local use case. In essence this is the macroecological approach, whereby interesting and important features of large-scale data only emerge by 'ignoring the details' [86] - the extent to which those details are of interest to end users may limit the eventual uptake of these broader-scale products. However, by providing clear and open documentation of the product creation process, an audit trail is available, as well as a bank of code for those wishing to adapt or extend products to their own specific purposes. This provides part of the essential infrastructure required to maximise the financial benefits of FAIR data, which run to billions of euros a year [87]. Ultimately, empowering users to make maximum use of existing biodiversity data to help to understand and manage our seas is a key role of EMODnet Biology, and data products are a crucial component of making the data accessible.

CRediT authorship contribution statement

Thomas J. Webb: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Funding acquisition, Conceptualization. **Joana Beja:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Salvador Jesús Fernández Bejarano:** Writing – review & editing, Visualization, Software, Methodology. **Elvira Ramos:** Writing – review & editing, Writing – original draft, Software, Conceptualization. **Samuel Sainz-Villegas:** Writing – review & editing, Software. **Karline Soetaert:** Writing – review & editing, Software, Conceptualization. **Willem Stolte:** Writing – review & editing, Writing – original draft, Software, Conceptualization. **Charles Troupin:** Writing – review & editing, Writing – original draft, Software, Conceptualization. **Benjamin Weigel:** Writing – review & editing, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The European Marine Observation and Data Network (EMODnet) is financed by the European Union under Regulation (EU) 2021/1139 of the European Parliament and of the Council of 7 July 2021 establishing the European Maritime, Fisheries and Aquaculture Fund and its predecessor, Regulation (EU) No. 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund. For the DIVA maps, computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

Data Availability

No data was used for the research described in the article.

References

[1] J.-B. Jouffray, R. Blasiak, A.V. Norström, H. Österblom, M. Nyström, The blue acceleration: the trajectory of human expansion into the ocean, One Earth 2 (2020) 43–54, https://doi.org/10.1016/j.oneear.2019.12.016.

[2] European Commission, Directorate-General for Environment, K. Sundseth, The EU birds and habitats directives – For nature and people in Europe, 2015. https://data.europa.eu/doi/10.2779/49288.

[3] Norwegian Ministry of the Environment, Norwegian Ministry of the Environment (2013) Integrated Management of the Marine Environment of the North Sea and Skagerrak (Management Plan), 2013. https://www.regjeringen.no/en/dokumenter/meld.-st.-37-2012-2013/id724746/?ch= 1.

[4] Department for Environment, Food and Rural Affairs, Marine strategy part one: UK initial assessment and good environmental status, 2012. https://www.gov.uk/government/publications/marine-strategy-part-one-uk-initial-assessment-and-good-environmental-status.

[5] Department for Environment, Food and Rural Affairs, Marine strategy part two: UK marine monitoring programmes, 2014. https://www.gov.uk/government/publications/marine-strategy-part-two-uk-marine-monitoring-programmes.

[6] Department for Environment, Food and Rural Affairs, Marine strategy part three: UK programme of measures, 2020. https://www.gov.uk/government/publications/marine-strategy-part-three-uk-programme-of-measures.

[7] W. Jetz, M.A. McGeoch, R. Guralnick, S. Ferrier, J. Beck, M.J. Costello, M. Fernandez, G.N. Geller, P. Keil, C. Merow, C. Meyer, F.E. Muller-Karger, H.

M. Pereira, E.C. Regan, D.S. Schmeller, E. Turak, Essential biodiversity variables for mapping and monitoring species populations, Nat. Ecol. Evol. 3 (2019) 539–551, https://doi.org/10.1038/s41559-019-0826-1.

[8] P. Miloslavich, N.J. Bax, S.E. Simmons, E. Klein, W. Appeltans, O. Aburto-Oropeza, M. Andersen Garcia, S.D. Batten, L. Benedetti-Cecchi, D.M. Checkley, Jr, S. Chiba, J.E. Duffy, D.C. Dunn, A. Fischer, J. Gunn, R. Kudela, F. Marsac, F.E. Muller-Karger, D. Obura, Y.-J. Shin, Essential ocean variables for global sustained observations of biodiversity and ecosystem changes, Glob. Chang. Biol. 24 (2018) 2416–2433, https://doi.org/10.1111/gcb.14108.

[9] G.J. Edgar, A.E. Bates, T.J. Bird, A.H. Jones, S. Kininmonth, R.D. Stuart-Smith, T. J. Webb, New Approaches to Marine Conservation Through the Scaling Up of Ecological Data, Ann. Rev. Mar. Sci. 8 (2016) 435–461, https://doi.org/10.1146/annurev-marine-122414-033921.

[10] T.J. Webb, B. Vanhoorne, Linking dimensions of data on global marine animal diversity, Philos. Trans. R. Soc. Lond. B Biol. Sci. 375 (2020) 20190445, https://doi.org/10.1098/rstb.2019.0445.

[11] S. Faulwetter, E. Pafilis, L. Fanini, N. Bailly, D. Agosti, C. Arvanitidis, L. Boicenco, T. Capatano, S. Claus, S. Dekeyzer, T. Georgiev, A. Legaki, D. Mavraki, A. Oulas, G. Papastefanou, L. Penev, G. Sautter, D. Schigel, V. Senderov, A. Teaca, M. Tsompanou, EMODnet Workshop on mechanisms and guidelines to mobilise historical data into biogeographic databases, Res. Ideas Outcomes 2 (2016) e9774, https://doi.org/10.3897/rio.2.e9774.

[12] L. Vandepitte, D. Mavraki, R. Perez Perez, G. Sarafidou, S. Paragkamian, V. Gerovasileiou, L. Marquez, D2.3: Report on efforts undertaken in rescuing historical data through citizen science (2022) https://emodnet.ec.europa.eu/sites/emodnet.ec.europa.eu/files/public/Biology/D2.3_0-Report%20on%20efforts%20undertaken%20in%20rescuing%20historical%20dat%20through%20citizen%20science.pdf.

[13] N. Mieszkowska, H. Sugden, L.B. Firth, S.J. Hawkins, The role of sustained observations in tracking impacts of environmental change on marine biodiversity and ecosystems, Philos. Trans. A Math. Phys. Eng. Sci. 372 (2014), https://doi.org/10.1098/rsta.2013.0339.

[14] M. Takahashi, M. Saccò, J.H. Kestel, G. Nester, M.A. Campbell, M. Van Der Heyde, M.J. Heydenrych, D.J. Juszkiewicz, P. Nevill, K.L. Dawkins, C. Bessey, K. Fernandes, H. Miller, M. Power, M. Mousavi-Derazmahalleh, J.P. Newton, N. E. White, Z.T. Richards, M.E. Allentoft, Aquatic environmental DNA: A review of the macro-organismal biomonitoring revolution, Sci. Total Environ. (2023) 162322, https://doi.org/10.1016/j.scitotenv.2023.162322.

[15] M. Dornelas, E.M.P. Madin, M. Bunce, J.D. DiBattista, M. Johnson, J.S. Madin, A. E. Magurran, B.J. McGill, N. Pettorelli, O. Pizarro, S.B. Williams, M. Winter, A. E. Bates, Towards a macroscope: Leveraging technology to transform the breadth, scale and resolution of macroecological data, Glob. Ecol. Biogeogr. 28 (2019) 1937–1948, https://doi.org/10.1111/geb.13025.

[16] A. Baudron, F. Bastardie, A. Belgrano, L. Bergström, B. Berx, S. Birchenough, E. Bresnan, J. Bueno-Pardo, C.J. Byron, A. Cooper, C. Cosgrove, M. Dickey-Collas, E. Glyki, K. Hamon, O. Henriksen, M. Hidalgo, K. Holsman, K. Hunter, C. Johnston, J. Kellner, J. Kempf, E. Klein, L. Li, K. Longo, R. López, I. Martinez, S. Meseck, K.E. Mills, G. Nolan, H. Ojaveer, J.K. Pinnegar, B. Planque, D. Purchase, D. Reid, M.-J. Roux, M. Rub, M. Rust, K. Schleit, S. Theuerkauf, M. Tomczak, B. Townhill, V. Trenkel, K. Van De Wolfshaar, L. Vaughan, Workshop on pathways to climate-aware advice (WKCLIMAD), (2023). https://doi.org/10.17895/ices.pub.22196560.

[17] G. Rilov, S. Fraschetti, E. Gissi, C. Pipitone, F. Badalamenti, L. Tamburello, E. Menini, P. Goriup, A.D. Mazaris, J. Garrabou, L. Benedetti-Cecchi, R. Danovaro, C. Loiseau, J. Claudet, S. Katsanevakis, A fast-moving target: achieving marine conservation goals under shifting climate and policies, Ecol. Appl. 30 (2020) e02009, https://doi.org/10.1002/eap.2009.

[18] A. Nieva de la Hidalga, B. Magagna, M. Stocker, A. Hardisty, P. Martin, Z. Zhao, M. Atkinson, K. Jeffery, The ENVRI Reference Model (ENVRI RM) version 2.2, 30th October 2017, 2017. https://doi.org/10.5281/zenodo.1050349.

[19] D. Lear, P. Herman, G. Van Hoey, L. Schepers, N. Tonné, M. Lipizer, F.E. Muller-Karger, W. Appeltans, W.D. Kissling, N. Holdsworth, M. Edwards, E. Pecceu, H. Nygård, G. Canonico, S. Birchenough, G. Graham, K. Deneudt, S. Claus, P. Oset, Supporting the essential - Recommendations for the development of accessible and interoperable marine biological data products, Mar. Policy 117 (2020) 103958, https://doi.org/10.1016/j.marpol.2020.103958.

[20] C.P. Morice, J.J. Kennedy, N.A. Rayner, J.P. Winn, E. Hogan, R.E. Killick, R.J. H. Dunn, T.J. Osborn, P.D. Jones, I.R. Simpson, An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set, J. Geophys. Res. 126 (2021), https://doi.org/10.1029/2019jd032361.

[21] GEBCO Bathymetric Compilation Group, The GEBCO_2022 Grid - a continuous terrain model of the global oceans and land, (2022). https://doi.org/10.5285/E0F0BB80-AB44-2739-E053-6C86ABC0289C.

[22] J. Assis, L. Tyberghein, S. Bosch, H. Verbruggen, E.A. Serrão, O. De Clerck, D. Tittensor, Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling, Glob. Ecol. Biogeogr. 27 (2018) 277–284, https://doi.org/10.1111/geb.12693.

[23] L. Tyberghein, H. Verbruggen, K. Pauly, C. Troupin, F. Mineur, O. De Clerck, Bio-ORACLE: a global environmental dataset for marine species distribution modelling, Glob. Ecol. Biogeogr. 21 (2012) 272–281, https://doi.org/10.1111/j.1466-8238.2011.00656.x.

[24] A.R. Hardisty, W.K. Michener, D. Agosti, E. Alonso García, L. Bastin, L. Belbin, A. Bowser, P.L. Buttigieg, D.A.L. Canhos, W. Egloff, R. De Giovanni, R. Figueira, Q. Groom, R.P. Guralnick, D. Hobern, W. Hugo, D. Koureas, L. Ji, W. Los, J. Manuel, D. Manset, J. Poelen, H. Saarenmaa, D. Schigel, P.F. Uhlir, W. D. Kissling, The Bari Manifesto: An interoperability framework for essential biodiversity variables, Ecol. Inform. 49 (2019) 22–31, https://doi.org/10.1016/j.ecoinf.2018.11.003.

[25] M.D. Wilkinson, M. Dumontier, I.J.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M. E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data 3 (2016) 160018, https://doi.org/10.1038/sdata.2016.18.

[26] B. Martín Míguez, A. Novellino, M. Vinci, S. Claus, J.-B. Calewaert, H. Vallius, T. Schmitt, A. Pititto, A. Giorgetti, N. Askew, S. Iona, D. Schaap, N. Pinardi, Q. Harpham, B.J. Kater, J. Populus, J. She, A.V. Palazov, O. McMeel, P. Oset, D. Lear, G.M.R. Manzella, P. Gorringe, S. Simoncelli, K. Larkin, N. Holdsworth, C. D. Arvanitidis, M.E. Molina Jack, M. del M. Chaves Montero, P.M.J. Herman, F. Hernandez, The European marine observation and data network (EMODnet): Visions and roles of the gateway to marine data in Europe, Front. Mar. Sci. 6 (2019) 434155, https://doi.org/10.3389/fmars.2019.00313.

[27] G. Bartha, S. Kocsis, Standardization Of Geographic Data: The European Inspire Directive, EJG 2 (2011) https://eurogeojournal.eu/index.php/egj/article/view/36 (accessed April 12, 2024).

[28] M. Vasquez, H. Allen, E. Manca, L. Castle, H. Lillis, S. Agnesi, Z. Al Hamdani, A. Annunziatellis, N. Askew, T. Bekkby, L. Bentes, V. Doncheva, V. Drakopoulou, G. Duncan, J. Gonçalves, R. Inghilesi, L. Laamanen, V. Loukaidi, S. Martin, F. McGrath, G. Mo, P. Monteiro, M. Muresan, C. Nikilova, R. Pesch, J. Pinder, J. Populus, A. Ridgeway, D. Sakellariou, A. Teaca, F. Tempera, V. Todorova, L. Tunesi, E. Virtanen, EUSeaMap 2021. A European broad-scale seabed habitat map, (2021). https://archimer.ifremer.fr/doc/00723/83528/ (accessed September 14, 2023).

[29] I. Galparsoro, D.W. Connor, A. Borja, A. Aish, P. Amorim, T. Bajjouk, C. Chambers, R. Coggan, G. Dirberg, H. Ellwood, D. Evans, K.L. Goodin, A. Grehan, J. Haldin, K. Howell, C. Jenkins, N. Michez, G. Mo, P. Buhl-Mortensen, B. Pearce, J. Populus, M. Salomidi, F. Sánchez, A. Serrano, E. Shumchenia, F. Tempera, M. Vasquez, Using EUNIS habitat classification for benthic mapping in European seas: present concerns and future needs, Mar. Pollut. Bull. 64 (2012) 2630–2638, https://doi.org/10.1016/j.marpolbul.2012.10.010.

[30] H.T.V. Vallius, A.T. Kotilainen, K.C. Asch, A. Florentino, M. Judge, H.A. Steward, B. Pjetursson, Discovering Europe's seabed geology: the EMODnet concept of uniform collection and harmonization of marine data, Geol. Soc., Lond., Spec. Publ. 505 (2020) 7–18, https://doi.org/10.1144/SP505-2019-208.

[31] EMODnet Bathymetry Consortium, EMODnet Digital Bathymetry (DTM 2022), (2022). https://doi.org/10.12770/ff3aff8a-cff1-44a3-a2c8-1910bf109f85.

[32] University of Liège, GeoHydrodynamics and Environment Research, European Seas - DIVAnd 4D monthly analysis of Water body dissolved oxygen concentration 1960/2020, (2023). https://doi.org/10.13120/37218601-e946-42d7-bdef-f94b92be7b5b.

[33] E. Olmedo, C. González-Haro, N. Hoareau, M. Umbert, V. González-Gambau, J. Martínez, C. Gabarró, A. Turiel, Nine years of SMOS sea surface salinity global maps at the Barcelona Expert Center, Earth Syst. Sci. Data 13 (2021) 857–888, https://doi.org/10.5194/essd-13-857-2021.

[34] EMODnet Human Activities & CLS, EMODnet Human Activities Vessel Density Map, (2019). https://emodnet.ec.europa.eu/geonetwork/srv/eng/catalog.search#/metadata/0f2f3ff1-30ef-49e1-96e7-8ca78d58a07c.

[35] Rew, Davis, Data Management: NetCDF: an Interface for Scientific Data Access, 10 (1990) 76–82. https://doi.org/10.1109/38.56302.

[36] S. Hankin, J. Blower, T. Carval, K. Casey, C. Donlon, O. Lauret, T. Loubrieu, A. Srinivasan, J. Triñanes, Ø. Godøy, R. Mendelssohn, R. Signell, J.D. L. Beaujardière, P. Cornillon, F. Blanc, R. Rew, J. Harlan, NetCDF-CF-OPeNDAP: Standards for ocean data interoperability and object lessons for community data standards processes 2 (2010) 450–458, https://doi.org/10.5270/OCEANOBS09.CWP.41.

[37] D. Hassell, J. Gregory, J. Blower, B.N. Lawrence, K.E. Taylor, A data model of the Climate and Forecast metadata conventions (CF-1.6) with a software implementation (cf-python v2.1), Geosci. Model Dev. 10 (2017) 4619–4646, https://doi.org/10.5194/gmd-10-4619-2017.

[38] J. Elith, J.R. Leathwick, Species Distribution Models: Ecological Explanation and Prediction Across Space and Time, Annu. Rev. Ecol. Evol. Syst. 40 (2009) 677–697, https://doi.org/10.1146/annurev.ecolsys.110308.120159.

[39] S. Katsanevakis, I. Wallentinus, A. Zenetos, E. Leppäkoski, M.E. Çinar, B. Oztürk, M. Grabowski, D. Golani, A.C. Cardoso, Impacts of invasive alien marine species on ecosystem services and biodiversity: a pan-European review, Aquat. Invasions 9 (2014) 391–423, https://doi.org/10.3391/ai.2014.9.4.01.

[40] C.F. de la Hoz, E. Ramos, A. Puente, J.A. Juanes, Temporal transferability of marine distribution models: The role of algorithm selection, Ecol. Indic. 106 (2019) 105499, https://doi.org/10.1016/j.ecolind.2019.105499.

[41] S. Sainz-Villegas, C.F. de la Hoz, J.A. Juanes, A. Puente, Predicting non-native seaweeds global distributions: The importance of tuning individual algorithms in ensembles to obtain biologically meaningful results, Front. Mar. Sci. 9 (2022), https://doi.org/10.3389/fmars.2022.1009808.

[42] E. Ramos, S. Sainz-Villegas, C.F. de la Hoz, A. Puente, J.A. Juanes, Species Distribution Models for invasive macroalgae, (2023). https://github.com/EMODnet/EMODnet-Biology-invasive-macroalgae-sdm.

[43] C. Troupin, Spatial interpolation of *Calanus finmarchicus and Calanus helgolandicus* observations in the North Sea, Integrated data products created under the European Marine Observation Data Network (EMODnet) Biology project Phase IV (EMFF/2019/1.3.1.9/Lot 6/SI2.837974), funded by the by the European Union under Regulation (EU) No 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund, 2023. https://github.com/EMODnet/EMODnet-Biology-Interpolation-Calanus.
[44] D. Lear, SAHFOS, Continuous Plankton Recorder Survey (CPR Survey), (2019). https://doi.org/10.17031/1629.
[45] A. Barth, J.-M. Beckers, C. Troupin, A. Alvera-Azcárate, L. Vandenbulcke, Divand-1.0: *n*-dimensional variational data analysis for ocean observations, Geosci. Model Dev. 7 (2014) 225–241, https://doi.org/10.5194/gmd-7-225-2014.
[46] C. Troupin, A. Barth, D. Sirjacobs, M. Ouberdous, J.-M. Brankart, P. Brasseur, M. Rixen, A. Alvera-Azcárate, M. Belounis, A. Capet, F. Lenartz, M.-E. Toussaint, J.-M. Beckers, Generation of analysis and consistent error fields using the Data Interpolating Variational Analysis (DIVA), Ocean Model 52–53 (2012) 90–101, https://doi.org/10.1016/j.ocemod.2012.05.002.
[47] P. Helaouët, G. Beaugrand, Macroecology of Calanus finmarchicus and C. helgolandicus in the North Atlantic Ocean and adjacent seas, Mar. Ecol. Prog. Ser. 345 (2007) 147–165, https://doi.org/10.3354/meps06775.
[48] P.M.J. Herman, Summary presence/absence maps of macro-endobenthos in European Seas, based on the EMODNET Biology database, Integrated data products created under the European Marine Observation Data Network (EMODnet) Biology project Phase IV (EMFF/2019/1.3.1.9/Lot 6/SI2.837974), funded by the by the European Union under Regulation (EU) No 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund, 2022. https://github.com/EMODnet/EMODnet-Biology-Benthos-European-Seas.
[49] J.M. Lobo, A. Jiménez-Valverde, J. Hortal, The uncertain nature of absences and their importance in species distribution modelling, Ecography 33 (2010) 103–114, https://doi.org/10.1111/j.1600-0587.2009.06039.x.
[50] P.M.J. Herman, F.F. van Rees, Mapping Reef forming North Sea Species, Deltares, Delft, The Netherlands, 2022. ⟨https://www.noordzeeloket.nl/publish/pages/199260/id55-mapping-reef-forming-north-sea-species.pdf⟩.
[51] O. Beauchard, M.S.A. Thompson, K.E. Ellingsen, G. Piet, P. Laffargue, K. Soetaert, Assessing sea floor functional biodiversity and vulnerability, Mar. Ecol. Prog. Ser. 708 (2023) 21–43, https://doi.org/10.3354/meps14270.
[52] J. Wieczorek, D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, D. Vieglais, Darwin Core: an evolving community-developed biodiversity data standard, PLoS One 7 (2012) e29715, https://doi.org/10.1371/journal.pone.0029715.
[53] T. Webb, Temporal turnover of macrobenthos in European seas, Integrated data products created under the European Marine Observation Data Network (EMODnet) Biology project Phase IV (EMFF/2019/1.3.1.9/Lot 6/SI2.837974), funded by the by the European Union under Regulation (EU) No 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund, 2023. https://github.com/EMODnet/EMODnet-Biology-benthos-trends.
[54] P. Cardoso, F. Rigal, J.C. Carvalho, BAT–Biodiversity Assessment Tools, an R package for the measurement and estimation of alpha and beta taxon, phylogenetic and functional diversity, Methods Ecol. Evol. (2015), https://doi.org/10.1111/2041-210X.12310.
[55] T. Webb, Benthic occurrences, habitat maps, and species traits, Integrated data products created under the European Marine Observation Data Network (EMODnet) Biology project (EASME/EMFF/2017/1.3.1.2/02/SI2.789013), funded by the by the European Union under Regulation (EU) No 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund, 2021. https://github.com/EMODnet/EMODnet-Biology-Benthic-Habitats-Occurrences-Traits.
[56] R.J. Wilson, D.C. Speirs, A. Sabatino, M.R. Heath, A synthetic map of the north-west European Shelf sedimentary environment for applications in marine science, Earth Syst. Sci. Data 10 (2018) 109–130, https://doi.org/10.5194/essd-10-109-2018.
[57] MarLIN, BIOTIC - Biological Traits Information Catalogue, (2006). www.marlin.ac.uk/biotic.
[58] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2023. ⟨https://www.R-project.org/⟩.
[59] J. Lai, D. Cui, W. Zhu, L. Mao, The Use of R and R Packages in Biodiversity Conservation Research, Diversity 15 (2023) 1202, https://doi.org/10.3390/d15121202.
[60] J. Lai, C.J. Lortie, R.A. Muenchen, J. Yang, K. Ma, Evaluating the popularity of R in ecology, Ecosphere 10 (2019) e02567, https://doi.org/10.1002/ecs2.2567.
[61] M. Grenié, E. Berti, J. Carvajal-Quintero, G.M.L. Dädlow, A. Sagouis, M. Winter, Harmonizing taxon names in biodiversity data: A review of tools, databases and best practices, Methods Ecol. Evol. 14 (2023) 12–25, https://doi.org/10.1111/2041-210x.13802.
[62] A. Peikert, A.M. Brandmaier, A Reproducible Data Analysis Workflow With R Markdown, Git, Make, and Docker, Quant. Comput. Methods Behav. Sci. 1 (2021) 27, https://doi.org/10.5964/qcmb.3763.
[63] C.H.M. Tso, M. Hollaway, R. Killick, P. Henrys, D. Monteith, J. Watkins, G. Blair, Advancing reproducible research by publishing R markdown notebooks as interactive sandboxes using the learnr package, R. J. 14 (2022) 255–263, https://doi.org/10.32614/rj-2022-021.
[64] C. Gandrud, Reproducible Research with R and RStudio (Third Edition), Chapman and Hall, 2020.
[65] A. Krystalli, S. Fernández-Bejarano, M. Salmon, EMODnetWFS: Access EMODnet Web Feature Service data through R. R package version 0.0.2, Integrated data products created under the European Marine Observation Data Network (EMODnet) Biology project Phase IV (EMFF/2019/1.3.1.9/Lot 6/SI2.837974), funded by the by the European Union under Regulation (EU) No 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund, 2020. https://github.com/EMODnet/EMODnetWFS.
[66] E. Pebesma, Simple Features for R: Standardized Support for Spatial Vector Data, R. J. 10 (2018) 439–446, https://doi.org/10.32614/RJ-2018-009.
[67] E. Pebesma, R. Bivand, Spatial Data Science: With Applications in R, Chapman & Hall/CRC, London, England, 2023.
[68] A. Krystalli, EMODnetWCS: Access EMODnet Web Coverage Service data through R. R package version 0.0.0.9012, Integrated data products created under the European Marine Observation Data Network (EMODnet) Biology project Phase IV (EMFF/2019/1.3.1.9/Lot 6/SI2.837974), funded by the by the European Union under Regulation (EU) No 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund, 2022. https://github.com/EMODnet/EMODnetWCS.
[69] R. Hijmans, terra: Spatial Data Analysis. R package version 1.7-39, 2023. https://CRAN.R-project.org/package=terra.
[70] K. Soetaert, O. Beauchard, Btrait: Working with Biological density, taxonomy, and trait composition data, Data product created under the European Marine Observation Data Network (EMODnet) Biology Phase IV, 2023. https://github.com/EMODnet/Btrait.
[71] P. Kasprzak, L. Mitchell, O. Kravchuk, A. Timmins, Six years of shiny in research - collaborative development of web tools in R, R. J. 12 (2020) 155, https://doi.org/10.32614/rj-2021-004.
[72] D. Lear, H. Millburn, A. Pasquale, C. Lynham, W.D. Kissling, D. Vaughan, F. Hernandez, F. MullerKarger, G. Canonico, G. Graham, G. Vanhoey, H. Lillis, I. Shepherd, I. Makarenko, J. Kaitaranta, J. Martin, K. Deneudt, L. Avellan, L. Dodds, M. Lipizer, M. Dickey-Collas, P. Herman, W. Appeltans, J. Hill, S. Claus, EMODnet Biology: Essential Data Products Workshop, The Marine Biological Association, Plymouth, 2017. ⟨https://emodnet.ec.europa.eu/sites/emodnet.ec.europa.eu/files/public/reports/Essential_Data_Products_Workshop_0.pdf⟩.
[73] A. Quarati, J.E. Raffaghelli, Do researchers use open research data? Exploring the relationships between usage trends and metadata quality across scientific disciplines from the Figshare case, J. Inf. Sci. Eng. 48 (2022) 423–448, https://doi.org/10.1177/0165551520961048.
[74] OBIS, Ocean Biodiversity Information System, (2024). Intergovernmental Oceanographic Commission of UNESCO https://obis.org.
[75] United Nations, ed., The First Global Integrated Marine Assessment: World Ocean Assessment I, Cambridge University Press, Cambridge, England, 2017. https://www.un.org/regularprocess/content/first-world-ocean-assessment.
[76] United Nations Office of Legal Affairs, ed., The Second World Ocean Assessment, 2021. https://doi.org/10.18356/9789216040062.
[77] UNESCO, IUCN, World Heritage: a unique contribution to biodiversity conservation, UNESCO, Paris, 2023. https://doi.org/10.58337/LSRE8424.
[78] A.-S. Bonnet-Lebrun, M. Sweetlove, H.J. Griffiths, M. Sumner, P. Provoost, B. Raymond, Y. Ropert-Coudert, A.P. Van de Putte, Opportunities and limitations of large open biodiversity occurrence databases in the context of a Marine Ecosystem Assessment of the Southern Ocean, Front. Mar. Sci. 10 (2023), https://doi.org/10.3389/fmars.2023.1150603.
[79] J.A. Smith, A.L. Benson, Y. Chen, S.A. Yamada, M.C. Mims, The power, potential, and pitfalls of open access biodiversity data in range size assessments: Lessons from the fishes, Ecol. Indic. 110 (2020) 105896, https://doi.org/10.1016/j.ecolind.2019.105896.
[80] V. Moudrý, R. Devillers, Quality and usability challenges of global marine biodiversity databases: An example for marine mammal data, Ecol. Inform. 56 (2020) 101051, https://doi.org/10.1016/j.ecoinf.2020.101051.
[81] A. Zizka, F. Antunes Carvalho, A. Calvente, M. Rocio Baez-Lizarazo, A. Cabral, J.F. R. Coelho, M. Colli-Silva, M.R. Fantinati, M.F. Fernandes, T. Ferreira-Araújo, F. Gondim Lambert Moreira, N.M.C. Santos, T.A.B. Santos, R.C. Dos Santos-Costa, F.C. Serrano, A.P. Alves da Silva, A. de Souza Soares, P.G. Cavalcante de Souza, E. Calisto Tomaz, V.F. Vale, T.L. Vieira, A. Antonelli, No one-size-fits-all solution to clean GBIF, PeerJ 8 (2020) e9916, https://doi.org/10.7717/peerj.9916.
[82] J.J. Waggitt, P.G.H. Evans, J. Andrade, A.N. Banks, O. Boisseau, M. Bolton, G. Bradbury, T. Brereton, C.J. Camphuysen, J. Durinck, T. Felce, R.C. Fijn, I. Garcia-Baron, S. Garthe, S.C.V. Geelhoed, A. Gilles, M. Goodall, J. Haelters, S. Hamilton, L. Hartny-Mills, N. Hodgins, K. James, M. Jessopp, A.S. Kavanagh, M. Leopold, K. Lohrengel, M. Louzao, N. Markones, J. Martínez-Cedeira, O. Ó Cadhla, S.L. Perry, G.J. Pierce, V. Ridoux, K.P. Robinson, M.B. Santos, C. Saavedra, H. Skov, E.W.M. Stienen, S. Sveegaard, P. Thompson, N. Vanermen, D. Wall, A. Webb, J. Wilson, S. Wanless, J.G. Hiddink, Distribution maps of cetacean and seabird populations in the North-East Atlantic, J. Appl. Ecol. 57 (2020) 253–269, https://doi.org/10.1111/1365-2664.13525.
[83] T.F. Johnson, A.P. Beckerman, D.Z. Childs, T.J. Webb, K.L. Evans, C.A. Griffiths, P. Capdevila, C.F. Clements, M. Besson, R.D. Gregory, G.H. Thomas, E. Delmas, R. P. Freckleton, Revealing uncertainty in the status of biodiversity change, Nature (2024) 1–7, https://doi.org/10.1038/s41586-024-07236-z.
[84] D. Obaton, L. Crosnier, L. Bastide, J. Beja, A. Giorgetti, C. Altobelli, A. Pititto, A. Novellino, M. Vasquez, Copernicus Marine & EMODnet Data Catalogue for the Marine Strategy Framework Directive, 2021. https://emodnet.ec.europa.eu/sites/emodnet.ec.europa.eu/files/public/PDF/EMODnet_CMEMS_productPortfolio_MSFD_FINAL.pdf.
[85] A. McQuatters-Gollop, A. Atkinson, A. Aubert, J. Bedford, M. Best, E. Bresnan, K. Cook, M. Devlin, R. Gowen, D.G. Johns, M. Machairopoulou, A. McKinney,

A. Mellor, C. Ostle, C. Scherer, P. Tett, Plankton lifeforms as a biodiversity indicator for regional-scale assessment of pelagic habitats for policy, Ecol. Indic. 101 (2019) 913–925, https://doi.org/10.1016/j.ecolind.2019.02.010.
[86] J.H. Lawton, Are There General Laws in Ecology? Oikos 84 (1999) 177–192, https://doi.org/10.2307/3546712.
[87] European Commission, Directorate-General for Research and Innovation, Cost-benefit analysis for FAIR research data – Cost of not having FAIR research data, 2018. https://data.europa.eu/doi/10.2777/02999.