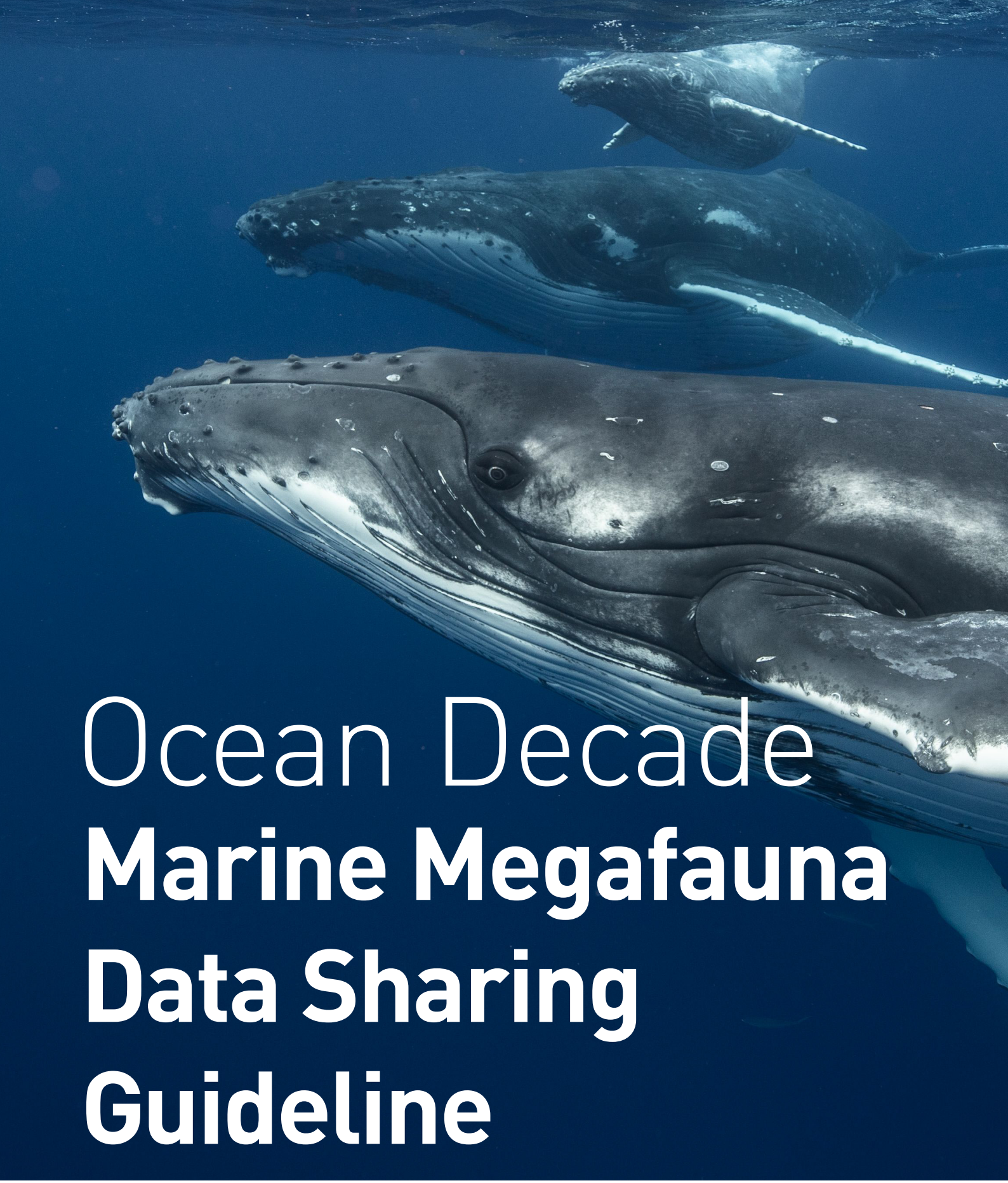

Ocean Decade
Marine Megafauna Data Sharing Guideline

The United Nations
Decade of Ocean Science
For Sustainable Development
2021-2030

2021 2030 United Nations Decade of Ocean Science for Sustainable Development

Published in 2025 by the Intergovernmental Oceanographic Commission of the United Nations Educational, Scientific and Cultural Organization.

7, Place de Fontenoy, 75352 Paris 07 SP, France.

For bibliographic purposes, this publication should be cited as follows:

IOC of UNESCO (2025). *Ocean Decade Marine Megafauna Data Sharing Guideline*. UNESCO, Paris. (The Ocean Decade Series, 68).

Cover: © Cedric Dageville / Kogia*

Graphic design: UNESCO

(IOC/2025/ODS/68)

Acknowledgments

The Intergovernmental Oceanographic Commission (IOC) of UNESCO wishes to thank the following members of the Ocean Decade Corporate Data Group for their time and their invaluable contributions to the development of this guideline:

Alcatel Submarine Networks	Jean Aude, Hayley Adlam
Equinor	Arne Myhrvold, Ana Sofia Aniceto
Fugro	David Millar, Louis Demargne, Dan Smith
HUB Ocean	Anna Silyakova
Mainstream Renewable Power	Karen Brinchmann
Ørsted	Emma Hospes
TotalEnergies	Laurent Cazes, Mathieu Reverberi, Hannah Escoubet
IOC	Peter Burger
Van Oord	Lucie Evaux

The IOC also wishes to thank the following organisations for their time and invaluable contributions during consultations on the topic:

Nicholas School of the Environment, Duke University (SEAMAP-OBIS)

EnerGeo Alliance

Marine Mammal Observers Association

ICES and JCDP

BOEM

SMRU Consulting

OBIS Secretariat

GBIF-OBIS node for Norway

Finally, the IOC wishes to thank the Ocean Decade Coordination Unit and colleagues from the International Oceanographic Data and Information Exchange (IODE) Secretariat, including the Ocean Biodiversity Information System (OBIS) as well as the co-chairs of the Corporate Data Group, Mr Vidar Helgesen, Executive Secretary of IOC, and Mr Mark Heine, CEO of Fugro, for their support and guidance.

Purpose:

This document highlights the opportunities for industrial companies working in the marine environment to make marine megafauna data, collected through their activities, publicly available.

It presents the mutual benefits of data sharing to both science and industry and provides guidelines and best practices on how to achieve this.

This document forms part of a set of data sharing guidelines from the Ocean Decade and should be used in conjunction with more specific manuals and guides from the IOC.

Table of contents

1 Introduction 4

2 Why is sharing marine megafauna data important? 5

3 Sharing marine megafauna data – what, where and how to share? 7

3.1 What to share – Sources of marine megafauna data 7

3.2 What to share – Formats of marine megafauna data 10

3.3 Where to share – Global open repositories 11

3.4 How to share the data 13

4 Important considerations and required action 16

4.1 Authorisation for data sharing 16

4.2 Clause for future contracts to share data directly to GBIF/OBIS 16

5 More information 17

6 References 18

1 Introduction

Marine megafauna is defined as marine animals over 45 kg in body mass (Estes et al., 2016), and usually includes whales, dolphins, sea turtles, manatees, sharks, rays, and some other predatory fish (Moleón et al., 2020). Sharing and making marine megafauna data available for science and policy can contribute to a better understanding and insights on species' presence and migration routes. This allows for more sustainable ocean practices through better policy making, ocean management, and planning of ocean operations by either private or public sector organisations.

Marine megafauna data are also invaluable for understanding the distribution, population trends, and ecological roles of large marine animals, as well as for informing conservation and management efforts to protect species and their habitats.

Industrial data on marine megafauna are routinely collected by many companies as part of their standard operations, and some companies have already actively contributed their data to open-source data repositories. The United Nations Decade of Ocean Science for Sustainable Development 2021-2030 ('the Ocean Decade') aims to foster collaboration among stakeholders for ocean data sharing, promoting the FAIR Data Principles: Findable, Accessible, Interoperable, Reusable (Wilkinson et al. 2016).

This data sharing guideline urges all companies and subcontractors to contribute to the collective understanding of marine ecosystems by sharing both legacy and future marine megafauna data. This collaboration is essential for advancing conservation efforts and fostering sustainable practices in the ocean.

The guideline is designed to support marine operators and companies by addressing fundamental questions about sharing marine megafauna data, such as the types of data, where to share data, data quality and ownership considerations.

2 Why is sharing marine megafauna data important?

Marine megafauna data contribute to the creation of a comprehensive map of animal density, distribution, and migration, and are a key element of all scientific studies of the ocean and the ecosystems it supports. These data also provide fundamental knowledge for operational mitigations regarding marine megafauna safeguarding, including measures to avoid collisions, temporary or permanent hearing damage, reduce behavioural disruptions risk, comply with regulations, and reduce operational downtime and related costs.

Numerous academic studies continue to refine marine megafauna distribution maps. As the climate and ocean ecosystems change - and with marine megafauna representing a highly mobile group of species - there is a growing need for continuous, systematic observational and monitoring data to feed predictive models (e.g. https://shiny.obis.org/distmaps/ and https://mico.eco/system/). To bridge the knowledge gap and enhance our understanding of the ocean's marine megafauna, it is essential to unlock future and existing, yet-to-be-shared data repositories.

Since regulations, license conditions, and practices vary from country to country and project to project, geographical gaps in data collection, existence, and availability can be observed. Aligning with best practices and working towards public marine megafauna data literacy among companies operating on the ocean would allow the filling of such gaps in the required data, information, and knowledge, regardless of whether regulations are in place or not.

Collaboration across multiple stakeholders and disciplines is therefore crucial for advancing marine mammal research and sustainable ocean planning and management.

The main positive impacts of availability and using detailed marine megafauna data include:

1. Scientific research: Detailed seasonal maps of animal distribution and migration paths provide essential insights into trophic interactions, phenology, and ecological processes within marine ecosystems. They also support research on human-wildlife interactions, climate change impacts, ecological modelling, and forecasting - which provides a return benefit to industry (see below).

2. Industry: Detailed information on animal density, distribution, and migration supports more effective offshore operations. Prediction models derived from data help mitigate risks such as collision, hearing damage, and behavioural disruptions caused by human activities. These models also reduce costs by optimising licensing and permitting conditions, predicting, and reducing potential downtime and minimising consenting risk.

3. Environmental conservation: Species distribution patterns guide conservation strategies and habitat protection, contributing to understanding global biodiversity health and ensuring timely action for threatened ecosystems. It would also contribute to Global Biodiversity Framework (GBF) target 3 (30% of marine waters protected).

4. Supporting sustainable development: Accurate data on marine megafauna occurrences enable more effective Marine Spatial Planning (MSP) for Marine Protected Areas (MPA), and offshore industrial projects like wind farms, oil and gas, fisheries, aquaculture, shipping, and tourism. This ensures that industry activities are strategically positioned to minimise ecological impact while meeting regulatory requirements.

5. Enhancing stakeholder and investor confidence: Investors and stakeholders gain confidence in industry practices when decisions are based on robust, data-driven insights that prioritise

both sustainability and operational efficiency in a growing blue economy. Such enhanced confidence ultimately helps lower reputational risks of offshore developments.

Private sector companies are an essential partner in expanding the pool of ocean data generators and users to fill existing data gaps and support the science we need for the ocean we want. The private sector holds a large amount of ocean data that are not yet publicly available. Helping to unlock and facilitate access to privately owned ocean data, including marine megafauna data and information, will improve our collective understanding of the world's ocean. By contributing their ocean datasets, private industry companies can reinforce their position as a key player in the blue economy supporting sustainable development and demonstrating their commitment to Corporate Social Responsibility - and the UN's Sustainable Development Goal 14 (Life Below Water) and GBF target 3.

Regulations and licensing conditions for operations requiring Marine Megafauna **data collection** can vary significantly across countries or jurisdictions. When regulations and licensing conditions for collecting marine megafauna data are in place, they often lack specific details regarding which data should be collected, and which methods should be used for data collection.

In absence of globally approved and used standards in marine megafauna data collection, the design of which data responding to the requirements should be collected is left to the local environmental agency or consultancy and/or country-based practices, however UNESCO has expert panels to provide further guidance on observation system design.[1] Similarly, decisions about which data to collect are typically made by local environmental agencies, consultants, or driven by country-specific practices.

Furthermore, in most cases the environmental agencies or consultancies doing the in-situ megafauna observations onboard a vessel are often not required to provide in-situ observation data back to the private company that contracted them but instead deliver a written report. If the licensing agreement for the private company requires observations data reporting, the data flow from the licensed private company to the regulator and its database. These country databases, however, are often not open source, not standardised, and prone to insufficient quality control or easily accessible.

There are several European data sharing acts supporting the data sharing initiatives outlined within this document, including but not limited to: Open Data Directive (2019/1024), INSPIRE Directive (2007/2/EC), and the Data Governance Act, as well as the IOC Data Policy and Terms of Use (2023).

[1] https://goosocean.org/who-we-are/expert-panels/biology-and-ecosystems-bioeco/

3 Sharing marine megafauna data – what, where and how to share?

3.1 What to share – Sources of marine megafauna data

Private companies engaged in marine activities are often required to collect marine mammal data to meet regulatory requirements established by governmental bodies, such as the Joint Nature Conservation Committee (JNCC) in the United Kingdom or the Bureau of Ocean Energy Management (BOEM) in the United States. These regulations are tied to licenses that define the conditions operators must comply with to operate within a country's jurisdiction. At the 33rd Assembly of the IOC in June 2025, a policy recommendation was adopted. This recommendation urges Member States to collaborate with industry, research, and other data infrastructure stakeholders to standardise ocean data sharing practices through the establishment of national data-sharing policies, regulations, and permissions for all ocean-related activities conducted within their territorial waters and Exclusive Economic Zones. One example is to include the provisions of the IOC Data Policy and Terms of Use in licensing and permitting operations within their jurisdictions.

Marine megafauna information may be collected at various phases of the project (design, executional, operational, and/or decommissioning phase) (Figure 1). Data generating activities would generally be undertaken in support of:

- Environmental (Baseline) surveys - conducted before project activities to understand the existing conditions of the marine environment, including the presence and behaviour of megafauna. These surveys help establish reference points for future comparisons.
- Compliance monitoring - ensures that a project adheres to environmental regulations and permit conditions by tracking the impact on marine megafauna during different phases of the project.
- Operational and maintenance monitoring - measures ongoing impacts on megafauna during the active life of the project, and helps guide adjustments in operations or maintenance to minimise and mitigate for harm.
- Planning compliance monitoring – involves collecting data to support future planning decisions and ensure that environmental commitments made during the planning phase are being met.

Moreover, vessels from cruise lines and various contractors, such as cable laying and logistics companies, may actively survey the waters for megafauna when they are in transit.

The below figure provides insights into the types of marine megafauna-related data types that are generated at each phase of an industrial project at sea.

FIGURE 1

Marine megafauna-related data types per phase of an industrial project at sea.

The data are generated using various methods of observation and registration, leading to different formats and types of knowledge about marine megafauna.

In general, marine megafauna data can be categorised into two main types:

- direct data, which provide direct information on marine megafauna, and
- auxiliary data or indirect data, which complement and enhance the interpretation and usability of actual megafauna-related information.

Direct data include visual observation methods, acoustic monitoring methods, digital observation methods, and genetic monitoring methods:

- **Visual observation methods** (by a person):
    - **Visual marine megafauna sightings** are conducted by trained marine mammal observers onboard vessels using binoculars to spot animals and record key information such as location, behaviour, group size, presence of young and species identification.
    - **Aerial surveys** are a method for monitoring marine mammal populations over large spatial scales using human observers' onboard aircraft to assess population distribution, habitat use, and seasonal movements of marine mammals. Aerial surveys are usually individually designed to monitor target species and operate with no disturbance.
- **Acoustic monitoring methods:**
    - **Passive acoustic monitoring** (PAM) – typically relies on using hydrophones or underwater microphones to record and analyse the sounds generated by marine animals and the environment. PAM systems can be mounted on vessels and Autonomous Underwater and Surface Vehicles (AUVs and ASVs) for mobile monitoring, deployed on buoys, gliders, landers, and moorings for continuous, long-term data collection in fixed locations.
    - **Active acoustic monitoring** (AAM) is a technique used to detect, localise, track, and classify marine life, such as fish and marine mammals, by emitting sound waves and analysing the echoes that bounce back. The data are generally acquired using vessels and ASVs.
    - **Acoustic tagging (acoustic telemetry)** involves an animal carrying a device that transmits its own sound for detection by separate receivers (satellite, buoys) with the aim to track the presence, movement, and behaviour of individual animals over extended periods. Often these animals are also equipped with other sensors to measure oceanographic variables.
- **Digital observation methods**
    - **Digital aerial surveys** utilise high-resolution cameras mounted on aircraft or Unmanned Aerial Vehicles (UAVs) to capture images of the ocean surface, allowing for the detection and identification of marine mammals.

- **Infrared camera detection** helps to monitor marine megafauna particularly in low-visibility conditions such as at nighttime and in rough weather. Infrared cameras detect infrared radiation signatures emitted by marine mammals as they surface to breathe. It is an emerging technology but already widely used in offshore industries.

- **Very high-resolution (VHR) satellite imagery.**

- **Genetic Monitoring methods**

  - **Environmental DNA** (eDNA), e.g. collecting water samples to detect traces of genetic material left behind by animals such as skin cells, mucus, urine, and faeces to identify which species have recently been in the area. It is an emerging method that is already used a lot and will become even more important in the future.

Auxiliary or indirect data include:

- **Effort data** is the recorded information about the conditions, duration, and methods of observation used during a survey, which helps assess how likely it was to detect animals. These include weather data (e.g. wind, tide), but also operational vessel data, oceanographic data from probes or sensors (e.g. for other relevant biological or geoscientific data), deployed while a single marine megafauna survey took place.

- **Ambient and operational noise monitoring** involves listening to and recording existing sounds in the marine environment, capturing both natural background noise and human-made noise. This type of monitoring helps assess noise pollution and its impacts on marine mammals and other species, while also contextualising the acoustic environment of the animal detections reported.

3.2 What to share – Formats of marine megafauna data

Marine megafauna-related data are accepted by open data-repositories in the following general categories:

- **Metadata** contextualises the overall conditions and sampling effort in which the direct data has been collected, such as the external conditions (temporal and spatial context, weather), authorship and protocol information (format, author of a file or observation, sampling methods, provenance, terms of use). This greatly facilitates cataloguing, interoperability, and exchange of data. XML or text are widely used metadata standards to use. Its secondary role is to help future users of the data to identify potential bias in the applied sampling methodology and context.

- **Raw data:** Image, audio, and video data that are publicly hosted on a long-term persistent platform or repository (including free ones, such as Zenodo.org) can be published using the Simple Multimedia extension.

- **Processed data** consists of raw data that has been processed and files from native sensor format converted into common readable format that can be opened by a variety of open-source analytical tools. Example formats of the processed data include .txt, .csv, .netcdf, .ascii, etc. Processed data should include taxonomic identification of marine megafauna species.

The common framework for sharing biodiversity data, including marine megafauna, is the Darwin Core Standard (DwC). Developed and maintained by the Biodiversity Information Standards (TDWG) community, DwC defines a set of standardised terms and structures that support the integration and reuse of biodiversity information across platforms and institutions. Data are submitted in the Darwin Core Archive (DwC-A) format - a compressed package that includes metadata (eml.xml), processed data (CSV files), and a descriptor file (meta.xml) describing the dataset's structure and content. The conversion process to DwC-A and DwC-DP typically involves quality control to ensure data consistency and accuracy. This Darwin Core format is widely adopted by open repositories such as Ocean Biodiversity Information System (OBIS) and Global Biodiversity Information Facility (GBIF) and is preferred for marine megafauna data submission, although alternative approaches are described later.

Quality control is a crucial step before sharing data, ensuring their accuracy and reliability for broader usability. The process involves several straightforward steps:

1. Verify that each data entry includes essential information such as **correct coordinates** (latitude and longitude), the **date and time** of collection, and **species identification**.
2. Ideally, the marine megafauna species observed should be identified by experts, and the name of the species should be matched against an authoritative database, such as the World Register of Marine Species (WoRMS) to confirm correct identification and current taxonomy.

To help and make sure you follow internationally adopted data standards and quality control procedures, you may consult the OBIS manual, including the chapter on how to structure your dataset or contact the OBIS Secretariat helpdesk (helpdesk@obis.org).

3.3 Where to share – Global open repositories

Several globally established open-access repositories are available for submitting marine megafauna and related biodiversity data. Among these, OBIS stands out as the expert network for marine biodiversity data, officially recognised as the Global Data Assembly Centre (GDAC) for biological and ecosystem essential ocean variables under the IOC digital ecosystem. OBIS specialises in marine species and ecosystems, offering rigorous quality control and a comprehensive gateway to data on the diversity, distribution, and abundance of marine life.

GBIF is also a globally respected repository with a broader scope encompassing terrestrial, freshwater, and marine biodiversity.

OBIS is a global open repository for biodiversity data, focusing solely on marine data and runs under the auspices of the IOC and its 152 Member States. It was established in 2000 by the decade long Census of Marine Life program and in 2009 it was adopted by the IOC as part of the International Oceanographic Data and Information Exchange (IODE) program. There are currently 37 OBIS nodes around the world, connecting over a thousand ocean data publishing institutions, providing over 160 million observations of nearly 20,000 marine species as of 2025. OBIS accepts **direct observational data, processed animal detections**, and related **auxiliary effort data as well as environmental data**.

GBIF is an intergovernmental network and data infrastructure, like OBIS, but initially established following a recommendation from the OECD and aimed at providing anyone, anywhere, with open access to data about all types of life on Earth. There are over 100 GBIF nodes worldwide, connecting over 3,000 publishers, encompassing national, regional, and thematic institutions, collectively providing access to more than 3.4 billion species occurrence records across all life forms (as of 2025), both on land and in the ocean (although a fraction of those data are from the ocean).

OBIS and GBIF collaborate closely, using the same data standards and exchange protocols such as the **Integrated Publishing Toolkit (IPT)** for data publication. Datasets are uploaded to an IPT service provided by a regional OBIS or GBIF node (or self-hosted by any institute), and datasets are updated and maintained locally there. The central OBIS and GBIF repositories remotely retrieve data from there, each time a dataset including metadata (including new versions) are uploaded and published publicly.

The protocols for data submission and conversion are described in the OBIS manual. When publishing data via the IPT, you can publish the dataset to both GBIF and OBIS. When the IPT is registered in OBIS and in GBIF, the dataset will simultaneously flow to OBIS and GBIF. If the IPT is registered in GBIF but not in OBIS, there is an option to select OBIS and OBIS will check the dataset before ingestion into OBIS.

While additional national initiatives also exist to address country-specific concerns and needs for marine conservation, the Ocean Decade Corporate Data Group, and following the IOC Data Policy and Terms of Use, recommends submission to (national) data systems that are connected to internationally recognised open data repositories, such as OBIS and GBIF, to enable global collaborative data access and collaboration for the benefit of science and conservation.

Several private companies that are members of the Corporate Data Group (e.g. TotalEnergies and Equinor) have already become data publishers on GBIF and submit their data or require their subcontractors to do so. These companies have established or are establishing a standardised internal practice for submitting data from future operations and have also published data from past projects (Figueira et al., 2020).

FIGURE 2

A snapshot of TotalEnergies metrics on GBIF indicating openly published datasets for non-commercial use, as of March 2025. Source: gbif.org

GBIF accepts four types of datasets, the latter two of which are also accepted by OBIS:

- Resources metadata, which is the simplest level, allows companies to create datasets describing undigitised resources like logbooks, written data sheets, or historical PDF reports. This 'metadata-only' class helps users discover and assess valuable information that

is not yet available online. Each metadata dataset receives a unique Digital Object Identifier (DOI) to ensure proper citation and tracking. The data itself is not made public, only the description of what the data is and any other relevant metadata. 'Metadata-only' datasets can be upgraded to any of the dataset types below by adding data files once the data is ready and digitised.

- Checklist data are lists of species or taxa, organised along taxonomic, geographic, or similar thematic lines. Such checklists are typically created during baseline surveys and environmental assessment studies. Example checklist dataset here.

- Occurrence data provide information about the location of individual organisms in time and space. This includes data points where a marine animal has been observed, identified, and recorded with its precise location and time of observation. Such data are gathered using visual, digital, and acoustic observation methods (see above). Example checklist dataset here.

- Sampling-event data provide additional detail beyond occurrence data, capturing information about the methods, events, and relative abundance of species recorded during a sampling effort. This type of data corresponds to the effort data described in the auxiliary section above. Best practice for sampling-event datasets is to follow the Darwin Core Humboldt Extension reporting format (see above). Example sampling-event dataset here.

This document recommends submitting data to OBIS or GBIF. However, companies may use another publicly accessible repository if preferred; in such cases, the dataset will still be considered in support of the Ocean Decade Marine Megafauna Data Sharing Guideline. While additional national initiatives also exist to address country-specific concerns and needs for marine conservation, the Ocean Decade Corporate Data Group recommends submission to internationally recognised open data repositories, such as OBIS and GBIF, to enable global collaborative data access and collaboration for the benefit of science and conservation.

3.4 How to share the data

Organisations that want to share data through GBIF and OBIS must first register with GBIF to request endorsement as a data publisher. They can select OBIS as the endorsing network.

Before publication, data must be standardised into the Darwin Core format. This means that ideally the company collecting the data should arrange it in a spreadsheet in a flat format with the initial row being columns using standardised names (Darwin Core Terms), and the following rows having one row per data point in standardised formats.

Some conversion may be necessary to align data to the Darwin Core standard, e.g. frequently the collection date-time needs reformatting. Quality checks must be made to ensure consistency in consistent units of measurement and date-time representations. Submission of data can be accommodated by a series of interconnected global open repositories as defined in paragraph 3.3. It is preferred to submit data in DwC format; however, data can still be submitted even if it is not currently within DwC to an OBIS or GBIF node, which can help in the standardisation and formatting. In practice the process of standardisation to Darwin Core can be technically challenging depending on the dataset complexity, in which case it is recommended to seek guidance from the OBIS or GBIF helpdesks. OBIS and GBIF node staff can provide help with standardisation and publishing.

The Integrated Publishing Toolkit will be used as part of the submission process. The OBIS or GBIF helpdesks can help provide the respective OBIS or GBIF node for data submission, which can help in

data processing and facilitating the publishing of the specific data type and data set, and setup an account on their IPT for data upload. Further details regarding the OBIS nodes can be found here.

Non-DwC formatted data can be submitted to OBIS by emailing the OBIS Secretariat at the following address: helpdesk@obis.org. OBIS will then assign the appropriate node to engage with for submission of the respective dataset. For lighter datasets, the data may be submitted directly by email to the OBIS Secretariat.

If your data are not currently in DwC format, please format it to ensure that the minimum recommended format is given in the following table:

OCCURRENCE-ID	EVENTDATE	DECIMAL-LONGITUDE	DECIMAL-LATITUDE	GEODETIC-DATUM	SCIENTIFIC-NAME	SCIENTIFIC-NAMEID	OCCURRENCE-STATUS	BASISOF RECORD
1	2024-03-01T13:00:00Z/-2025-05-11T15:30:00Z	130.45055	8.81694	WGS84	*Tursiops truncatus*	urn:lsid:marinespecies.-org :taxname:137111	Present	Human Observation

Note:

- **Species occurrence ID (occurenceID)** each occurrence record should have a globally unique identifier. These can be easily generated using a Universally Unique Identifier (UUID) generator, such as https://www.uuidgenerator.net/. OccurrenceIDs can also be constructed by combining different fields. It is important that these IDs do not change so it is best to avoid including prefixes like country codes as they may not be stable.

- **Event date (eventDate)** is in UTC (indicated by the 'Z') and describes the start date and time of the sighting and the end date and time of the sighting *(during the example interval between 1 March 2024 1pm UTC and 11 May 2025 3:30pm UTC)*. Ensure that all records have eventDates in the same ISO format.

- **Decimal Longitude and Latitude** are recorded in decimal degrees, preferably in the WGS84 coordinate reference system. If a different reference system is used it must be indicated. The precision and uncertainty of coordinates can also be recorded as needed.

Sensitive species can be protected by generalisation (e.g. of location) as per the Current Best Practices for Generalising Sensitive Species Occurrence Data guidelines. GBIF and OBIS staff can advise on this.

- **Scientific Name:** If the exact species is unknown, the genus, family, or order can be recorded instead, *i.e. Delphinidae.* The scientific name should be matched against an authoritative database, such as the World Register of Marine Species (WoRMS) to confirm correct identification and current taxonomy.

> Some data points are even possible to identify to the individual organism based on identifiable traits (such as the tail fin pattern for whales) using organismName and organismID.

- **Occurrence status:** the default vocabulary is recommended to consist of present and absent, *e.g. If a 60-minute observation is undertaken and no sightings of Delphinidae were present, then 'absent' may be used*. Absence of species is also interesting from a scientific perspective, for example, if a marine megafauna survey was undertaken in the same location at the same time daily for a year and during that time no species of marine megafauna were observed, then this may demonstrate an absence of marine megafauna in the area.

> Additional optional columns related to occurrenceStatus: individualCount or organismQuantity with organismQuantityType.

- Basis of record: This refers to the format by which the data has been collected, i.e. human observation, machine observation *(e.g. passive acoustic monitoring, remote sensing, photograph)*.

Please refer to section 5 for more information on data sharing and standardisation procedures.

4 Important considerations and required action

4.1 Authorisation for data sharing

Marine megafauna data, such as visual sightings, images from aerial surveys, and infrared camera detection, can generally be shared without posing significant legal, financial, or national security risks to the company that collects and shares it. The data is typically not sensitive in terms of company operations or public safety concerns. Examples from private companies that already share marine megafauna data on GBIF demonstrate that the data sharing authorisation process involves affiliates collecting and managing the data, while granting the private company the rights to the data. This process ensures legal ownership and a clear understanding of the risks involved in publicly sharing such data.

Sharing data on marine megafauna presence is essential for advancing scientific understanding, enhancing conservation efforts, and fostering collaboration across a multitude of disciplines. However, it is important to do so responsibly to protect species from potential threats such as poaching or illegal trade. Data can be sensitive for either commercial/reputational risk or revealing the location of CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) species or commercialised wild species which would indirectly feed the illegal wild species trade. To balance the benefits of open data with the need for security, particularly for endangered species, contributors can implement safeguards like anonymising sensitive information by generalising location data to broader regions.

If requested the OBIS Secretariat can help assess the need for security and/or confidentiality and then facilitate the anonymising of sensitive information in a specific data set. Companies can also utilise the following guideline: Current Best Practices for Generalising Sensitive Species Occurrence Data guidelines.

By applying these precautions, we can safely contribute valuable information that supports research and protects vulnerable marine life.

4.2 Clause for future contracts to share data directly to GBIF/OBIS

This guideline recommends that any company in custody of the data takes the advised measures of this document to ensure that the data is made publicly available. It is further advised that when entering into a new contract with a subcontractor who will gather marine megafauna data, the contract can include a provision requiring the subcontractor to prepare for submission of all collected data to an open repository, such as GBIF and OBIS, in accordance with the specified data quality and sensitivity guidelines.

Example text: *"All results from [SAMPLING METHOD] shall be made available in an Excel file provided in the Darwin Core standard as required by GBIF / OBIS format, possibly complemented with other relevant information."*

The subcontractor must then ensure that the data is accurately documented, adheres to established quality standards, and is appropriately anonymised or aggregated to protect any sensitive information. Agreements must clarify that either the company or subcontractor will be responsible for complying with all relevant data authorisation requirements, including securing necessary permissions, and ensuring that the data does not expose vulnerable species or proprietary company information. This clause ensures that valuable data is made accessible for scientific research and conservation efforts while respecting legal, ethical, and security considerations.

5 More information

This document is a collaborative effort across sectors addressing data needs and opportunities in support of ocean science and knowledge generation. Further resources from UNESCO-IOC and the Ocean Decade partners are available in the links below in support of the effective submission and standardisation of marine megafauna and associated datasets.

- Darwin Core templates for different types of data are available on the GBIF website: https://www.gbif.org/dataset-classes and see https://manual.obis.org/formatting.html for more details on how to structure your data in DarwinCore files. General information about Darwin Core Standard can be found at https://dwc.tdwg.org/

- An example of a marine mammal dataset is explained in the OBIS manual: https://manual.obis.org/examples.html#marine-mammals-abundance-and-distribution

- Generate a more detailed template for marine megafauna recording, including a detailed description of each recording variable, e.g. animal behaviour (e.g. foraging, milling, transit), life stage (e.g. adult, juvenile, etc.): https://gbif-norway.github.io/dwc-excel-template-generator-js

- Validate any simple spreadsheets (i.e. single-sheet spreadsheets) using the GBIF validator: https://www.gbif.org/tools/data-validator

- The protocols for data submission and conversion are described in the OBIS manual: https://manual.obis.org

- How to videos on OBIS' YouTube channel: https://www.youtube.com/@oceanbiodiversityinformati6931/playlists

- DarwinCore Archive QC tool: https://rshiny.lifewatch.be/BioCheck/

- The GBIF data mobilisation course: https://docs.gbif.org/course-data-mobilization/en

- The OBIS training course: https://classroom.oceanteacher.org/course/view.php?id=907

- Current Best Practices for Generalising Sensitive Species Occurrence Data guideline

- For testing automatic standardising of simpler observation data GBIF Norway provides a chatbot: https://chatipt.svc.gbif.no

For general information about the Ocean Decade initiative and associated data repositories, please use the following links:

- Ocean Decade: https://oceandecade.org/

- OBIS: https://obis.org/

- GBIF: https://www.gbif.org/

- UN-Oceans and Law of the Sea: https://www.un.org/depts/los/

You may can submit your query or feedback though the contact sheet at: https://oceandecade.org/contact/

6 References

Estes J.A., Heithaus M., McCauley D.J., Rasher D.B., Worm B. (2016). *Megafaunal Impacts on Structure and Function of Ocean Ecosystems (2016)*. Annual Review of Environment and Resources 41:83-116. https://doi.org/10.1146/annurev-environ-110615-085622

Figueira R., Beja P., Villaverde C., Vega M., Cezón K., Messina T., Archambeau A., Johaadien R., Endresen D. & Escobar D. (2020). *Guidance for private companies to become data publishers through GBIF: Template document to support the internal authorization process to become a GBIF publisher.* Copenhagen: GBIF Secretariat. https://doi.org/10.35035/doc-b8hq-me03

GBIF (2024). *OBIS and GBIF endorse joint strategy and action plan for marine biodiversity data - Technical collaboration and capacity building activities to support shared objectives and ensure free and open access to comprehensive information about life in and around the world's oceans.* https://www.gbif.org/news/7w0xoodpNZRwnt1SjWfML0/obis-and-gbif-endorse-joint-strategy-and-action-plan-for-marine-biodiversity-data. Accessed: 2025/08/27

GBIF (2025). *Introducing the Darwin Core Data Package - Diversifying the GBIF Data Model webinar*. 2 June 2025. https://www.gbif.org/composition/3Be8w9RzbjHtK2brXxTtun/introducing-the-darwin-core-data-package. Accessed: 2025/08/27

Igenloff K. (2025). *Survey and Monitoring Data Quick-SIOtart Guide: A how-to for updating a Darwin Core dataset using the Humboldt Extension*. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/doc-7t3p-ve38

Moleón, Marcos & Sánchez-Zapata, José & Donázar, José & Revilla, Eloy & Martín-López, Berta & Gutiérrez-Cánovas, Tano & Getz, Wayne & Morales-Reyes, Zebensui & Campos-Arceiz, Ahimsa & Crowder, Larry & Galetti, Mauro & González-Suárez, Manuela & He, Fengzhi & Jordano, Pedro & Lewison, Rebecca & Naidoo, Robin & Owen-Smith, Norman & Selva, Nuria & Svenning, Jens-Christian & Tockner, Klement. (2020). *Rethinking megafauna. Proceedings of the Royal Society B: Biological Sciences*. 287. 20192643. 10.1098/rspb.2019.2643.

Wilkinson M., Dumontier M., Aalbersberg I. et al. (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

For further information
visit our website:
oceandecade.org

Or follow us on social media:

UNOceanDecade
unoceandecade
OceanDecade
UN-Ocean-Decade

#OceanDecade

unesco
Intergovernmental
Oceanographic
Commission