



MareGraph
Towards an Interoperable
Marine Knowledge Graph

EU-FUNDED PROJECT

MAREGRAPH

**TOWARDS AN INTEROPERABLE
MARINE KNOWLEDGE GRAPH**

(DATA MODELS CHARTER)

OVERVIEW

Start date	02/05/2023
End date	31/04/2025
Project team	<p>ISTC-CNR:</p> <ul style="list-style-type: none"> • Giorgia Lodi • Alessandro Russo • Francesco Poggi <p>Digitaal Vlaanderen:</p> <p>OSLO</p> <ul style="list-style-type: none"> ○ Lucas Cornette ○ Lorenzo Vyliders <p>VSDS</p> <ul style="list-style-type: none"> ○ Simon Claus <p>VLIZ:</p> <ul style="list-style-type: none"> • Joanna Goley • Marc Portier • Lennert Tyberghein • Leen Vandepitte • Lennert Schepers
Public workshops/webinars	Maximum 5 public workshops, if consensus reached
Decision criteria	Unanimity minus one (U-1)
License of the data models	Creative Commons Attribution 4.0 (CC-BY 4.0) or less restrictive open licence (COMMISSION IMPLEMENTING REGULATION (EU) 2023/138 of 21 December 2022)
Location of documentation	GitHub Maregraph
Issue logging	GitHub issues

1 CONTEXT

WHAT

The objective of the MareGraph project is to increase the semantic, technical, and legal interoperability of the following selected data:

- the taxonomic dataset of the World Register of Marine Species (WoRMS database);
- a geospatial dataset about Marine Regions;
- the data about biogeographic observations (EurOBIS database).

These High Value Datasets (HVDs) fit under the categories ‘geospatial’ and ‘Earth observation and environment’ of the European [Open Data Directive \(1024/2019\) - “Open Data and Reuse of the Public Sector Information”](#) and related [implementation regulation](#).

The scope of the work is to create and publish the linked open data version of the World Register of Marine Species (WoRMS) and Marine Regions (commonly referred to as the “taxonomic backbone”, essential reference datasets used as core term lists within the marine domain) as well as of EurOBIS, the European Node of the international Ocean Biodiversity Information System.

More specifically, the primary objective is to effectively and timely create a large and comprehensive European knowledge base that leverages the benefits of a semantic graph-based data structure and high-quality related data to address issues related to the aquatic biogeographical environment and its ecosystems. The knowledge graph will be easily accessible through a set of APIs to serve a variety of development needs.

WHY

The extension of the historical role of these datasets in the field of semantic interoperability is essential for their continued applicability, dissemination in the context of the Semantic Web, and for effectively enabling data integration to increase and discover knowledge in the specific marine domain. By providing these integrated data, semantically enriched through the definition of formal ontologies that are going to be published and made available as reusable assets, a significant contribution can be made to the creation of a European "Marine Knowledge Graph" that can be further enriched and linked with datasets from the European Commission, Member States and other initiatives where marine biodiversity, and other marine data types, are considered.

Finally, since the datasets subject of this charter are considered High Value Datasets (HVDs) according to the above-mentioned EU Open Data Directive, their integration in a knowledge graph, also according to rules defined in the implementing regulation for HVDs, can be better leveraged for both European and national Data Spaces. At the European level, in fact, the integrated and semantically enriched datasets can be used to possibly feed the “Green Deal” Data space, while at the national level, these HVDs can be successfully used for the integration, for example,

in the so-called Water Data Space, which is part of the Flanders Smart Data Space. This data space is currently being set up by the Digital Flanders Agency and it uses principles, technical building blocks and components of the European Data Space, based on the linked data paradigm.

2 USE CASES

The following main use cases are defined as starting points of the modelling process; some of these use cases are also possible through the combination of the datasets WoRMS, EurOBIS and Marine Regions:

- Being able to query over the datasets mentioned in this charter simultaneously in order to get the benefit of the data integration implemented through the “Marine Knowledge Graph” (e.g., “give me the a list of taxonomic species identifiers occurring in a sea region, the percentage of all known species in World Heritage Marine Sites (or other protected areas), the number of IUCN red list species in a particular areas of interest, the number of invasive species in harbours where ships fill their ballast tanks,...);
- Being able to query the taxonomic data of WoRMS to get the taxonomic information for the [commercial designations information system](#);
- Being able to obtain the most up-to-date information about genetic information for marine species;
- Being able to feed the data of this charter into the Flanders Water Domain, thus making it interoperable with that data space. The Flanders Smart Data Space is building components to standardize, publish and reuse data based on Linked Data Event Stream (LDES). Data space principles of technical components of the data space are being applied in the Water Domain in Flanders, part of the Smart Data Space. The Water Data Space can be, in fact, considered as a thematic (“vertical”) data space that publishes standardized, linked data on water (e.g., water quality, water quantity or water use, etc.);
- Being able to link the data with other linked open data already available from other European projects (e.g., [WHOW – Water Health Open Knowledge](#)).

3 SCOPE

The main objective of the working groups that will be created based on this charter and formed by a variety of domain experts, is to analyse, standardise and map to semantic web standards the data related to WoRMS and EurOBIS information systems, considering, in this work, the already existing Marine regions linked open dataset. Specifically, whereas WoRMS and EurOBIS data is still not available according to semantic web standards, for Marine Regions it has been done an effort in the past to provide the data as Linked Open Data with also different APIs, and in particular LDES – Linked Data Event Stream APIs. Therefore, in context of this charter, Marine Regions’ data model



and data will be used and, if needed, adapted in order to better support the integration with the data models that will be designed and published for WoRMS and EurOBIS. The ultimate goal is to integrate data represented in these datasets into a European "Marine Knowledge Graph" that can be queried through a variety of possibilities (at least using SPARQL and in the form of LDESs).

The working groups will define the reference use cases and collect the requirements that can then guide the definition of data models for WoRMS and EurOBIS data. The data models can encompass OWL/RDFS ontologies and linked data based controlled vocabularies to be defined for specific typologies of data.

To achieve these objectives, we will rely on existing standards, including, for example, already defined semantic models for taxa, their names and their occurrence (e.g., BioSchemas.org, Darwin Core, etc.).

4 STAKEHOLDERS

Possible stakeholders of this process include:

Stakeholder type	Examples
Research bodies (in the Environment Field)	<p>Italian National Institute for Environmental Protection and Research (ISPRA)</p> <p>VLIZ</p> <p>CNR IRBIM</p> <p>LifeWatch ERIC Service Centre</p> <p>FORTH</p>
Biodiversity and Marine Species information provider	<p>Ocean Biodiversity Information System (OBIS)</p> <p>Global Biodiversity Information Facility (GBIF)</p> <p>FishBase</p> <p>SeaLifeBase</p> <p>Encyclopedia of Life (EOL)</p> <p>Catalogue of Life (COL)</p> <p>Food and Agriculture Organization (FAO)</p> <p>Ocean Info Hub (OIH)</p>
European and international fisheries institutes and organisations	<p>The International Council for the Exploration of the Sea (ICES)</p> <p>Low impact fishers of Europe (LIFE)</p> <p>National Oceanic and Atmospheric Administration (NOAA)</p> <p>Marine Stewardship Council (MSC)</p>
Institutional bodies for agriculture and fisheries matter	ILVO
Companies in the fishery domain	
No profit organisations (also for data altruism)	

5 CRITERIA FOR SUCCESS

This trajectory will be considered a success when the deliverables are widely used and applied. Initially within the MareGraph project, but also beyond. In particular, we list the following criteria:

1. There has been maximum coordination with all stakeholders who are represented in at least one of the working group sessions.
2. The working group sessions result in a stable candidate standard representing a consensus of all participants.
3. Tools are available to support and validate implementations.
4. The specification has been implemented in at least one pilot (first implementation) which proves the added value of the specification in practice.

6 DELIVERABLES

The working group will provide the following deliverables:

- Prepare an overview of information needs based on analysis of available documentation and existing standards.
- Organizing a business workshop with stakeholders to validate and further expand the information needs.
- Organizing and facilitating 2 to 3 workshops with the working group composed of domain experts, collecting and processing feedback.
- Preparation of reusable documentation for the data models and publication on data.vlaanderen.be, possibly on schema.gov.it (the Italian national catalogue of semantic assets for semantic interoperability) and MAREGRAPH dedicated web spaces:
 - RDF vocabulary serialized at least in RDF/Turtle
 - HTML documentation for the vocabulary with terms, definitions and other axioms defined in the vocabulary
 - UML and Graffoo diagrams (two types of diagrams, one for UML notation and one for a notation developed for OWL/RDF semantic web standards)
 - HTML documentation for the UML diagram representing the application profile
 - SHACL validation rules
 - JSON-LD format availability
- Integration into the OSLO and possibly schema.gov.it systems of semantic assets

5 MILESTONES

Date	Milestone
19/06/2023	First business workshop with stakeholders about data models for WoRMS and EurOBIS – discussion, preliminary list of requirements, points of focus
30/08/2023	First Thematic Workshop for WoRMS data only
29/09/2023	Draft of the data model (also organized as a network of ontological modules) for WoRMS to be published on GitHub
19/10/2023	Second Thematic Workshop for WoRMS data only
20/11/2023	Revision of the data model for WoRMS for public review to be published on GitHub
22/11/2023	Public Review Starts
22/12/2023	Public Review Ends
01/02/2024	WoRMS data model (also organized as a network of ontological modules) ready and published on GitHub and other relevant web sites
30/04/2024	First Thematic Workshop for EurOBIS data only
24/06/2024	Draft of the data model (also organized as a network of ontological modules) for EurOBIS to be published on GitHub
03/07/2024	Second Thematic Workshop for EurOBIS data only
20/09/2024	Revision of the data model for WoRMS for public review to be published on Github
23/09/2024	Start Public Review
13/01/2025	End Public Review
14/04/2025	EurOBIS data model (also organized as a network of ontological modules) ready and published on Github and other relevant web sites