AQUARES

## Aquatic species Register Exchange and Services

PROJECT DESCRIPTION

The use of organism names is ubiquitous in a wide range of scientific, environmental management and policy domains. Specialist curated taxonomic databases and tools to query these data are therefore essential for ensuring the quality of biological data from collection and generation to data curation and management. Species information systems for monitoring status and trends of biodiversity and those dealing with policy concern – Natura 2000 species, commercial, invasive alien species and pest species – benefit from such high quality tools and databases ensuring the interoperability of the data.
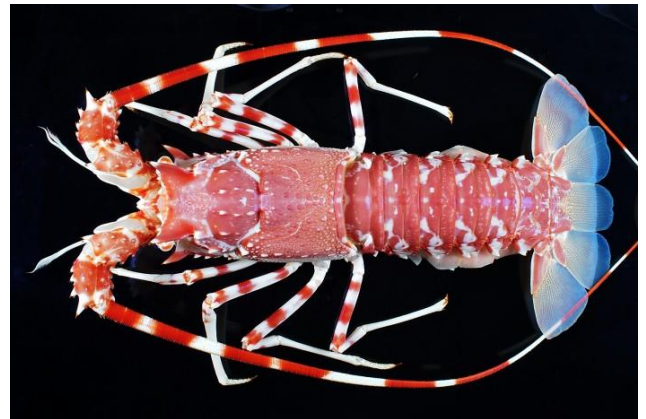
The World Register of Marine Species (WoRMS), the Register of Antarctic Marine Species (RAMS) and the Freshwater Animal Diversity Assessment (FADA) database are three major Global Species Directories (GSD) hosted in Belgium. These data collections consist of authoritative taxonomic data, curated by international experts and contribute to initiatives such as Catalogue of Life (CoL) and LifeWatch, and potentially to the Pan-European Species directories Infrastructure (PESI) and other regional and national species lists. Most of these initiatives rely on a wide array of specialists' contributions to independent checklists and require extensive interactions with a wide expert network. Given the potential overlap in taxonomic specialists and the complex nature of the data, exchanging expertise and data among these initiatives is highly beneficial for all parties involved.

The main objective of this project is therefore to ensure and enhance the interoperability and public availability of these aquatic species databases through the development of a set of web services. Such services can guarantee the automatic and timely exchange of data between WoRMS, RAMS and FADA, but also expose the data for use in other initiatives and applications such as Encyclopedia of Life (EoL), Catalogue of Life (CoL), Global Biodiversity Information Facility (GBIF) and e-Science initiatives such as Biodiversity Virtual e-Laboratory (BioVeL) and LifeWatch.

To ensure the quality of the data exposed through those web services, we aim to improve the data import and exchange procedures into the partner databases and will develop a data entry interface to facilitate the entry of more complete distribution information. These procedures and tools will be tested and used during a hands-on workshop with taxonomic experts. To stimulate their involvement and advertise the free and open publication of their data, we will implement a tool for generating a checklist paper, which can be published in a scientific journal and provides more straightforward solution for properly citing and tracking citations of the data.
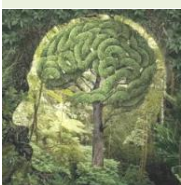
AQUARES

On top of the web services exposing the data from each of the databases, we will implement a joint data cache on which we can build a range of tools to perform data quality control of species related data. A central tool, which is relevant for both internal quality control of taxon data and for the curation of occurrence data or other species related data, is the so-called Taxamatch tool. This tool uses fuzzy matching algorithms for searching phonetically matching or very similar scientific species names and will be modified and improved for running on top of the joint data cache. Other tools, which are more specifically targeting species occurrence data include a service for mapping and validating occurrence data in comparison to expert checked distribution ranges and tools for checking technical errors in data files (e.g. incorrect date format, missing required fields). While these services will run as independent tools, they are highly relevant for a wide range of users and could, in the framework of the European Biodiversity Virtual e-Laboratory (BioVeL) project, be combined into a data workflow together with other web services available in the biodiversity informatics community.

Throughout this project, we will organise regular consultations with a wide range of potential users to document their requirements and get their feedback on the developed tools and services. Data from the FP7 BioFresh project, the European Ocean Biogeographic Information System (EurOBIS) and the Antarctic Biodiversity information Facility (AntaBIF) will be used as specific test cases to validate and improve the tools. Further tests with data from biological collections and ecological monitoring data are envisaged to ensure that these services are of interest to a wide range of institutes and researchers dealing with aquatic species data.

CONTACT INFORMATION

Coordinator

Aaike DE WEVER
Royal Belgian Institute of Natural Sciences (RBINS)
Operational Directorate Natural Environments
aaike.dewever@naturalsciences.be

Partners

Leen VANDEPITTE
Vlaams Instituut voor de Zee (VLIZ)
Datacentrum
leen.vandepitte@vliz.be

Bruno DANIS
Université libre de Bruxelles (ULB)
Marine Biology Lab
bdanis@ulb.ac.be

500 µm