iMarine Catalogue of Applications

The iMarine initiative provides a **data infrastructure** aimed at **facilitating open access, the sharing of data, collaborative analysis, processing and mining processing, as well as the dissemination of newly generated knowledge**. The iMarine data infrastructure is developed to support decision making in high-level challenges that require policy decisions typical of the ecosystem approach.

iMarine has developed a **series of applications** which can be clustered in four main thematic domains (the so called **Application Bundles**, set of services and technologies grouped according to a family of related tasks for achieving a common objective).

Virtual Research Environments (VREs) are dynamically created by selecting and combining several iMarine applications to support diverse communities of practice. Virtual research environments offer flexible and secure web-based, community-centric platforms, so researchers can work together on common challenges. Each VRE in the infrastructure is tightly integrated with the underlying gCube enabling software, and can access and re-purpose data from other iMarine applications.

To date 15 VREs are operational and serve hundreds of users (**Register to the iMarine Gateway** https://portal.i-marine.d4science.org to access the iMarine VREs).

BiolCube Bundle

BiolCube applications help to **generate new knowledge from data** (e.g. modeling and analyzing distribution data, comparing checklists, and producing maps).

Who is it for?
Practitioners, including biologists and marine ecologists, working with species occurrence data and taxonomic profiles.

BiolCube Applications

Occurrence and Taxonomic Data Discovery

This application offers facilities for **discovering and accessing species occurrence and taxonomic data** from major repositories and information systems, e.g. OBIS, GBIF, Catalogue of Life, WoRMS.
The discovery mechanism is simple (based on species common names or scientific names) yet powerful since it supports query expansion and additional filters. The identified datasets are enriched with links to other species, can be displayed on a map as well as saved in standard formats (e.g. DarwinCore, DarwinCore-Archive, CSV) for future uses.

Occurrence Data Processing

This application offers facilities for **species occurrence datasets processing**. These facilities include algebraic operations (union, intersection, subtraction, and duplicates deletion) based on spatial and syntactic similarity measurements, clustering (e.g. density based algorithms such as DBScan, distance based algorithms such as K-means), outliers detection (e.g. Local Outlier Factor approach), occurrence points representativeness (e.g. Habitat Representativeness Score technique), and occurrence points enrichment with chemical and physical environmental parameters.

Species Distribution Modeling

This application offers facilities for **building a rich array of species distribution models** by relying on algorithms and approaches including AquaMaps. AquaMaps is actually a family of approaches (e.g. suitable, native) producing species distribution probabilities on half-degree cells by relying on (i) a table containing species envelops (HSPEN), (ii) a table containing environmental parameters (HCAF) and (iii) a table containing species occurrences points (half-degree cells). Moreover, it offers methods for producing new versions of HSPEN and HCAF.

Species Distribution Maps Discovery

This application offers facilities for **retrieving species distribution maps**. It offers: (i) simple search by species name, (ii) advanced search by supporting criteria on species name, code, taxonomy, and/or characteristic, and (iii) support for visualisation of products by image, by detailed record, and by scientific record.

Taxonomic Data Comparison

This application offers facilities for **comparing two taxonomic checklists in DarwinCore-Archive format**. In particular, the application provides the user with a flexible environment for detecting, analysing and reporting relationships among taxa of the compared checklists (e.g. corresponds, includes, overlaps, not found in).

Taxonomic Data Matching

This application offers facilities for **taxonomic data matching**, i.e. the identification of most suitable scientific names given a list of species names. In particular it offers a workflow base approach that enable users to combine a number of matchers (e.g. GSay, FuzzyMatcher, Levensthein, Trigram) and tune their contribution while identifying the proper scientific names recognised by a number of authoritative sources.

ConnectCube Bundle

ConnectCube applications are a comprehensive suite of tools, which support a **collaborative, standards-oriented data publication environment, including semantic technologies.**

Who is it for?
Practitioners who need to discover and access information from heterogeneous data sources and collaboratively produce new data products out of this information.

ConnectCube Applications
Enhanced Documents Management
This application offers facilities for **creating and managing enhanced documents**, i.e. rich information objects resembling documents yet aggregating multiple parts. Parts include images, datasets, maps, and graphs. It offers functionalities for defining templates the documents should adhere to, as well as facilities for defining and monitoring workflows driving the collaborative production of these "documents".
Fact-sheets Management
This application offers a flexible environment for the **management of fact-sheets**, a special kind of document designed to report key information on a given subject, e.g. fisheries and aquaculture-related subjects. This application builds on the facilities offered by Enhanced Documents Management application and complements them with fact-sheets specific ones, e.g. automatic generation out of customizable database schema, semi-automatic business metadata generation.
Information Object Discovery
This application offers facilities for **retrieving information objects from multiple collections and information systems in a seamless way**. It offers both a Google-like approach and an advanced search allowing users to characterise in detail the information objects they are looking for. Moreover, it includes facilities for presenting the results according to semantic-based clustering.
Messaging
This application offers facilities for **exchanging messages among users**. It resembles an email environment with the distinguishing feature of being integrated with the rest, e.g., it is possible to send as attachment any dataset residing in the workspace without consuming bandwidth.
Shared Workspace
This application offers facilities **for organising a workspace**. It resembles a folder-based file system for managing information objects. The added value is represented by the type of information objects it can manage in a seamless way. It supports items ranging from binary files to information objects representing datasets, workflows, species distribution maps, time series, and comprehensive research products. Through it, data sharing is fostered, making results, workflows, annotations and documents immediately available to co-workers, to anyone provided with URIs, and to any other person authorized via WebDAV.
Social Networking Facilities
This application offers facilities for **integrating social networking in research environments**. It is conceptually close to the common facilities promoted by social networks – e.g., posting news, commenting on posted news, re-share news – yet adapted to promote large scale collaboration and cooperation on comprehensive scientific products, datasets, theories, and tools.

www.i-marine.eu/ApplicationBundles

GeosCube Bundle

GeosCube applications help practitioners dealing with **geospatial information to properly access, consume, and produce data.**

Who is it for?
Practitioners dealing with geospatial information yet tailored to the needs of the fisheries and biodiversity communities.

GeosCube Applications
Geospatial Data Discovery
This application offers facilities for **browsing and visualising geospatial data**. In particular, these include facilities to navigate, search and discovery layers from a GeoNetwork instance via the OGC CSW protocol. Moreover, these include facilities to interactively explore, manipulate, visualize, compare, and analyse geospatial data.
Geospatial Data Processing
This application offers facilities for **executing a rich array of data processing tasks on geospatial data**. The current set of supported tasks includes maps comparison algorithms (supported formats include WFS, OPeNDAP and ASC), an intersection algorithm, i.e. an approach that computes the percentages of overlap between different areas on the compared maps, and many more. Moreover, the application enables the invocation of services exposing their capabilities via the OGC WPS protocol.

StatsCube Bundle

StatsCube applications make up a complete **data-life-cyclesupporting framework**, including data validation, data enrichment and efficient analytical tools

Who is it for?
Practitioners working with a rich array of information, ranging from observational data to statistical data.

StatsCube Applications

Code List Discovery

This application offers facilities for **browsing and discovering available Code Lists from a set of repositories**.
It offers both a Google-like approach and an advanced search allowing users to characterise in detail the information they are looking for.

Code List Management

This application offers facilities for **managing code lists**, i.e. recognised controlled vocabularies.
It includes facilities for code lists creation (also via ingestion), collaborative curation, and publishing.

Statistical Engine

This application offers facilities for **efficiently and effectively executing a rich array of statistical data processing algorithms**. The application relies on the distributed and elastic computing capacities offered by the underlying infrastructure. It offers a set of off-the-shelf algorithms including clustering algorithms such as DBScan. Moreover, it enables a simple integration and execution of user-defined algorithms expressed in a number of programming and scripting languages including R. It currently embeds more than 70 different algorithms ranging from Anomalies Detection, Classification, Clustering, Simulation, Training, Bayesian Methods, Trends, and many more. These algorithms are executed on a distributed infrastructure by completely hiding the complexity of such an execution while ensuring robustness, throughput, fault-tolerance, and privacy.

Tabular Data Discovery

This application offers facilities for **browsing and discovery available Tabular Data resources from a set of repositories**. It offers both a Google-like approach and an advanced search allowing users to characterise in detail the information they are looking for.

Tabular Data Enrichment

This application offers facilities for **augmenting a tabular dataset having geospatial and temporal attributes with selected physical and chemical environmental parameters acquired dynamically by authoritative sources**. These parameters might include salinity, temperature, ice concentration, etc. The application properly adapts the parameters to the spatial and temporal resolution of the tabular dataset.

Tabular Data Management

This application offers facilities for **managing tabular data**. In particular, it offers facilities for supporting the entire workflow of tasks on tabular data including tabular data creation, collaborative curation and publishing.
Moreover, it offers a number of facilities for tabular data manipulation including filtering, grouping, unions and intersections. In addition to that, it is equipped with a powerful mechanism for versioning and a rich set of metadata for describing the tabular data resource including provenance.

Tabular Data Processing

This application offers facilities for **performing data mining tasks on tabular data**. In particular, it relies on the Statistical Service to offer effective tabular data manipulation facilities including geocoding, maps projection, clustering, outlier identification, hidden trends, trends comparison, and many more.