Project Acronym	***iMarine***
Project Title	***Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources***
Project Number	***283644***
Deliverable Title	***Virtual Research Environments Activity Report***
Deliverable No.	***D6.3***
Delivery Date	***July 2012***
Author	***Anton Ellenbroek, Leonardo Candela, Pasquale Pagano***

Abstract*: This report documents the activity needed to provide the iMarine Ecosystem Approach Community of Practice with a set of Virtual Research Environments aiming at serving the scenarios and requirements discussed by such community. In particular, the report describes (a) the set of Virtual Research Environments that have been deployed and (b) the development of specific applications and tools that are needed to realize the expected Virtual Research Environments in tandem with the rest of gCube technology.*

iMarine (RI – 283644) is a Research Infrastructures Combination of Collaborative Project and Coordination and Support Action (CP-CSA) co-funded by the European Commission under the Capacities Programme, Framework Programme Seven (FP7).

DOCUMENT INFORMATION

PROJECT	
Project Acronym	iMarine
Project Title	Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources
Project Start	1st November 2011
Project Duration	30 months
Funding	FP7-INFRASTRUCTURES-2011-2
Grant Agreement No.	283644
DOCUMENT	
Deliverable No.	D6.3
Deliverable Title	Virtual Research Environments Activity Report
Contractual Delivery Date	July 2012
Actual Delivery Date	19 October 2012
Author(s)	Anton Ellenbroek, Leonardo Candela, Pasquale Pagano
Editor(s)	Anton Ellenbroek
Reviewer(s)	Pasquale Pagano
Contributor(s)	N/A
Work Package No.	WP6
Work Package Title	Virtual Research Environments Deployment and Operation
Work Package Leader	CNR
Work Package Participants	ERCIM, NKUA, FAO, FIN, UNESCO, CRIA, IRD
Estimated Person Months	1.5
Distribution	Public
Nature	Report
Version / Revision	1.0
Draft / Final	Final
Total No. Pages (including cover)	37
Keywords	Virtual Research Environment, Community Tools, Applications

DISCLAIMER

iMarine (RI – 283644) is a Research Infrastructures Combination of Collaborative Project and Coordination and Support Action (CP-CSA) co-funded by the European Commission under the Capacities Programme, Framework Programme Seven (FP7).

The goal of iMarine*, Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources*, is to establish and operate a data infrastructure supporting the principles of the Ecosystem Approach to Fisheries Management and Conservation of Marine Living Resources and to facilitate the emergence of a unified Ecosystem Approach Community of Practice (EA-CoP).

This document contains information on iMarine core activities, findings and outcomes and it may also contain contributions from distinguished experts who contribute as iMarine Board members. Any reference to content in this document should clearly indicate the authors, source, organisation and publication date.

The document has been produced with the funding of the European Commission. The content of this publication is the sole responsibility of the iMarine Consortium and its experts, and it cannot be considered to reflect the views of the European Commission. The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

The European Union (EU) was established in accordance with the Treaty on the European Union (Maastricht). There are currently 27 member states of the European Union. It is based on the European Communities and the member states' cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice, and the Court of Auditors (http://europa.eu.int/).

DELIVERABLE SUMMARY

Virtual Research Environments are among the main products the iMarine project is requested to deliver in order to serve the needs of the Ecosystem Approach Community of Practice. The deployment and operation of a Virtual Research Environment is a task involving the exploitation of technologies that have been developed as well as the development of new technologies aiming at offering new facilities. This report describes the set of activities performed to provide the Ecosystem Approach Community of Practice with the set of Virtual Research Environments hosted by the iMarine portal in September 2012.

EXECUTIVE SUMMARY

Virtual Research Environments are “systems” aiming at providing their users with web-based working environments offering the entire spectrum of facilities (including services, data and computational facilities) needed to accomplish a given task by dynamically relying on the underlying infrastructure. Such a kind of systems are among the products to be delivered by the iMarine project in order to meet the needs of the iMarine Ecosystem Approach Community of Practice.

The development of such systems is based on three main activities: *(i)* the development of software artefacts aimed at realising a set of well defined functions, *(ii)* the deployment of such software artefacts to build and operate an infrastructure, and *(iii)* the deployment and operation of a set of Virtual Research Environments that exploit the facilities offered by the deployed infrastructure to serve scenarios raised by the target Community of Practice. The expected result is an offering of services and resources in an economy-of-scale oriented-approach.

This report documents the first and the last of the above three activities – i.e. dedicated software development and VREs deployment – as implemented in the context of the iMarine project. The second aspect – i.e. infrastructure deployment – is captured by another document, deliverable D5.2 “iMarine Data e-Infrastructure Operation Report”. Thus, this report describes the set of software artefacts that, in addition to the core technology, has been developed to serve the specific needs identified by Ecosystem Approach Community of Practice scenarios. Examples of such software artefacts include the Species Products Discovery portlet, the GIS Viewer and the Species View portlet. Moreover, the report describes the set of activities performed to provide the target Community of Practice with an initial set of Virtual Research Environments aiming at realising the tasks defined by the scenarios.

The deliverable captures the activities of the first months of the project, up to September 2012. However, the whole activity is very dynamic and it actually spans across the boundaries of WP6. In fact, the development and operation of Virtual Research Environments is performed in the context of WP6 in close cooperation with: *(a)* WP3 for what concerns the interaction with the Ecosystem Approach Community of Practice, *(b)* WP5 for what concerns the deployment and operation of the underlying infrastructure, and *(c)* WP8, WP9, WP10 and WP11 for what concerns the development of the core technology needed to enable the deployment of the infrastructure and the Virtual Research Environments. In addition to the WP6 activities, the deliverable documents such a network of cooperation and provides the reader with a list of related documents, Wiki pages and TRAC tickets allowing to build a comprehensive understanding of the overall process and related activities.

TABLE OF CONTENTS

TABLE OF FIGURES

GLOSSARY

Code list: A predefined list from which some (statistical) coded concepts take their values.

EA-CoP: Ecosystem Approach Community of Practice.

Community of Practice: A term coined to capture an "activity system" that includes individuals who are united in action and in the meaning that "action" has for them and for the larger collective. The communities of practice are "virtual", *i.e.*, they are not formal structures, such as departments or project teams. Instead, these communities exist in the minds of their members, are glued together by the connections they have with each other, as well as by their specific shared problems or areas of interest. The generation of knowledge in communities of practice occurs when people participate in problem solving and share the knowledge necessary to solve the problems.

OAI-PMH: An HTTP based protocol consisting of a set of six verbs or services that make it possible for (i) *Data Providers* to expose structured metadata on their resources and (ii) *Service Providers* to harvest that metadata and offer enhanced services on that.

Occurrence data: A set of observations of the presence (sometimes also absence) of a species or other taxonomic entity, usually in a specified location. These data are often contained in a repository that can be queried using de-facto standards for species observational data.

SDMX: An international Standard for the exchange of statistical (mainly aggregated) datasets.

Virtual Research Environment: A "*system"* with the following distinguishing features: *(i)* it is a Web-based working environment; *(ii)* it is tailored to serve the needs of a Community of Practice; *(iii)* it is expected to provide a community of practice with the whole array of commodities needed to accomplish the community's goal(s); *(iv)* it is open and flexible with respect to the overall service offering and lifetime; and *(v)* it promotes fine-grained controlled sharing of both intermediate and final research results by guaranteeing ownership, provenance, and attribution.

VRE: see Virtual Research Environment.

1 INTRODUCTION

Virtual Research Environments are "systems" aiming at providing their users with web-based working environments offering the entire spectrum of facilities (including services, data and computational facilities) needed to accomplish a given task by dynamically relying on the underlying infrastructure. Such a kind of systems are among the products to be delivered by the iMarine project in order to meet the needs of the iMarine Ecosystem Approach Community of Practice.

This deliverable – D6.3 *'Virtual Research Environments Activity Report*' – details the deployed Virtual Research Environments in terms of community tools integrated, resources involved, and user exploitation. It describes the set of software artefacts that, in addition to the core technology, has been developed to serve the specific needs identified by Ecosystem Approach Community of Practice scenarios.

The iMarine Board plays a key role in this process being the project body in charge of building and constantly interacting with the Ecosystem Approach Community of Practice. All tools relate to EA-CoP activities have been discussed by the iMarine Board. The WP3 Wiki page summarizes the iMarine Board Work plan by describing the expected VREs in terms of functional requirements and on the EA-CoP exploitation planning, requirements and desiderata:

http://wiki.i-marine.eu/index.php/Ecosystem_Approach_Community_of_Practice:_VRE_planning

The Board plan for new and adjusted VRE's is based on a template that captures requirements in the functional and technical domains, and collects information useful to cost-benefit analysis.

The described VREs are expected to evolve for the entire duration of the project. New VREs, community tools and applications and data sources will be described and discussed, and existing community tools will be subject to change and review. At a later stage, delivered community tools will also be validated and commented.

Once the EA-CoP desiderata have reached a stable state, they are thoroughly assessed from the technology perspective. The results are captured in wiki-pages representing the Virtual Research Environment development plan [1] and the Applications and Tools development plan [2]. Here, planning and implementation goes through 3 well identified steps; Analysis, Development, and Deployment. The requirements can be implemented as any of the following functional components: VREs, (Community) Tools, or integration of Community Data and Data Tools.

The e-infrastructure is equipped with tools that facilitate Community tools integration. Here the opportunity is to operate in the e-infrastructure existing frameworks, applications, or components that leverage services or support use-cases that the community wants to port to the e-infrastructure. These expand on already integrated e-infrastructure resources such as PostgreSQL/PostGIS and Geoserver. Examples of integrated tools are: OpenSDMX for statistical data access and transport, THREDDS for geospatial data access, R for tabular data analysis, AquaMaps for projection of species distribution algorithms, FiMES schema support for publication of fact-sheets and aggregated species information, Hadoop and WPS for data processing.

Finally, the planning has to cope with the integration or interoperation of Community data and data tools that will be critical for the exploitation by the EA-CoP of the e-infrastructure. These include GBIF: the Global Biodiversity Information Facility[1], CoL: Catalogue of Life[2], WoRMS: World Register of Marine Species[3], FAO Code lists exposed through the FAO SDMX registry[4], code lists expose through any SDMX registry like IRD and Eurostat, geospatial data exposed via OGC standards like Web Map Service for map images, Fisheries Linked Open Data (FLOD) knowledge base.

The remainder of this report is organised as follows. Section 2 describe the technologies and resources that have been developed and integrated to support the deployment of Virtual Research Environments. Section 3 describes the Virtual Research Environments that have been deployed and operated during the reporting period. For each Virtual Research Environment the deliverable describes the goal, the data available through such an environment and the set of facilities supported.

[1] http://www.gbif.org/
[2] http://www.catalogueoflife.org/
[3] http://www.marinespecies.org/
[4] http://www.fao.org/figis/sdmx/

2 VIRTUAL RESEARCH ENVIRONMENTS RESOURCES AND TOOLS

In order to make it possible to realize Virtual Research Environments suitable for serving diverse application scenarios it is fundamental to be able to equip such innovative working environments with an effective mix of facilities and data that are needed to accomplish the expected tasks. In this section, it is reported *(a)* the data sources that have been integrated in the infrastructure and can be used while defining a Virtual Research Environment (cf. Section 2.1), and *(b)* the set of tools that, in addition to the rest of gCube technology [6][7][8][9], have been developed to provide Virtual Research Environment users with facilities supporting their tasks (cf. Section 2.2).

2.1 DATA RESOURCES

In many application scenarios it is of paramount importance for the users to be provided with the data that are needed to accomplish certain tasks. The effectiveness of a Virtual Research Environment might be severely affected if the user is requested to acquire the data he/she needs by its own and with no support from the Virtual Research Environment.

In the majority of cases, the data are of diverse types and are scattered among a number of heterogeneous data sources. Moreover, data evolve along the time and the user should be provided with an up to date version of such material. Because of this characteristic, a number of facilities have been developed (namely in the context of the Data Management area [7]) with the aim to act as mediators between the data sources and services/clients aiming at consuming and making available such data in a seamless way (e.g. Species Products Discovery – cf. Sec. 2.2.1). Very often – actually whenever possible – such mediators are built by relying on standards and protocols for data access and discovery including OAI-PMH[5], TAPIR[6], SDMX[7], and DarwinCore[8].

By relying on the gCube technologies, the most important data sources needed to support the application scenarios discussed in the context of the EA-CoP have been linked to the iMarine infrastructure and made available for use in various Virtual Research Environments (see Section 3 for a list of data made available in the context of every VRE).

In particular, the focus was on the following data source that have been identified by the EA-CoP:

- *Global Biodiversity Information Facility (GBIF)*[9]: this data source offers more than 377 million of records on species and more than 10,000 datasets aggregated from 400+ publishers;
- *Ocean Biogeographic Information System (OBIS)*[10]: this data source offers more that 32 million records on species and 1,000+ datasets;

[5] www.openarchives.org/pmh/
[6] wiki.tdwg.org/TAPIR/
[7] sdmx.org
[8] rs.tdwg.org/dwc/
[9] www.gbif.org

- *Catalogue of Life*[11]: this data source offers an integrated checklist and a taxonomic hierarchy of more that 1.3 million species of animals, plants, fungi and micro-organisms;
- *Integrated Taxonomic Information System (ITIS)*[12]**:** this data source offers authoritative taxonomic information on plants, animals, fungi, and microbes of North America and the world;
- *World Register of Marine Species (WoRMS)*[13]: this data source offers species `names' for more than 200,000 species including 300,000+ species names and synonyms and 400,000+ taxa;
- *National Center of Biotechnology Information (NCBI) Taxonomy*[14]: this data source offers a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet;
- *Interim Register of Marine and Nonmarine Genera (IRMNG)*[15]**:** this data source offers access to over 465,000 genus names covering all types of biota and 1.6 million species names;
- *Aquatic Commons*[16]: this data source offers access to thematic material covering natural marine, estuarine/brackish and fresh water environments via OAI-PMH;
- *DRS at National Institute of Oceanography*[17]: this data source offers institutional publications including journal articles and technical reports via OAI-PMH;
- *WHOAS*[18]: this data source offers the production of Woods Hole scientific community including articles and data sets via OAI-PMH;
- *Central and Eastern European Marine Repository (CEEMar)*[19]: this data source offers material covering marine, brackish and fresh water environments via OAI-PMH;
- *OceanDocs*[20]: this data source offers research and publication materials in Marine Science by aggregating content form 256 repositories via OAI-PMH;
- *PANGAEA*[21]: this data source offers georeferenced data from earth system research via OAI-PMH. The system guarantees long-term availability of its content through a commitment of the operating institutions. The aggregated repositories are 475.
- *IRD UMR EME/Observatoire Thonier SDMX Registry and Repository*: This data source exposes (a) the Sardara database that contains tuna captures data from several countries, aggregated according to CWP statistical squares (1'x1' or 5'x5') and (b) the ObServe database that contains tuna and bycatches captures observed by scientific observers on-board French industrial purse seiners.

---

[10] www.iobis.org
[11] www.catalogueoflife.org
[12] www.itis.gov
[13] www.marinespecies.org
[14] www.ncbi.nlm.nih.gov/taxonomy
[15] www.obis.org.au/irmng
[16] aquaticcommons.org
[17] drs.nio.org/drs
[18] www.mblwhoilibrary.org/services/whoas-repository-services
[19] www.ceemar.org/dspace
[20] www.oceandocs.net
[21] www.pangaea.de

2.2 TOOLS

In addition to the rest of gCube technology [6][7][8][9] that is mainly conceived to offer core facilities, a number of common tools and user interfaces have been developed by relying on such core facilities to provide Virtual Research Environment users with instruments supporting their tasks. These facilities range from generic environments supporting basic tasks (e.g. a shared Workspace providing users with a file and folder-oriented view over the information objects they are willing to work with – cf. Sec. 2.2.5) to specific environments supporting well defined tasks (e.g. species distribution maps discovery and visualisation – cf. Sec. 2.2.4). In the remainder of this section, a description of the major tools that have either be developed or reinforced during up to September 2012 is reported.

2.2.1 THE GEOEXPLORER PORTLET

The *GeoExplorer* portlet is a user interface conceived to support the search and browse of GIS layers available in the infrastructure independently of the repository that physically stores them (by relying on a GeoNetwork instance).

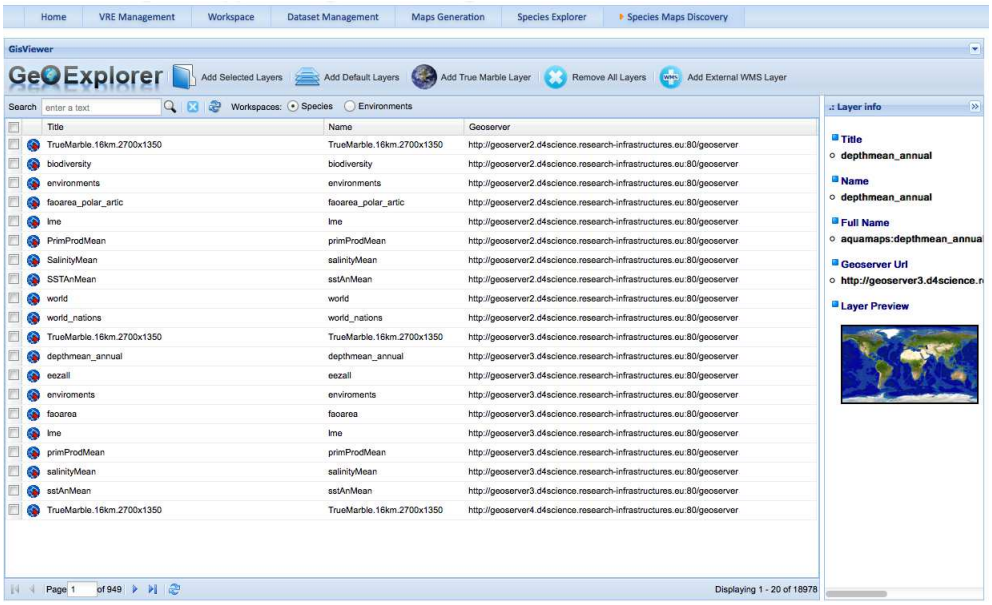

Figure 1. The GeoExplorer Portlet

Besides the discovery, this portlet supports some layers management facilities. Among them, there is the possibility to register a layer physically residing in a repository that is not linked to the infrastructure yet.

As regards layers visualization, this portlet is conceived to work in tandem with the GISViewer portlet (cf. Sec. 2.2.2). After the discovery, it is possible to visualise the selected layers in the GISViewer portlet and exploit all the facilities supported by such a portlet.

2.2.2 THE GISVIEWER PORTLET

The GISViewer is a user interface conceived to support the visualisation of a series of layers on a map. Besides the visualisation, the portlet supports a number of management operations like for each layer including the setting of the layers opacity, the selection of alternative layer style, the execution of a CQL filter, the storage of a layer snapshot into the workspace, the dynamic generation of a transect graph on the selected layer to analyse the plotted values.

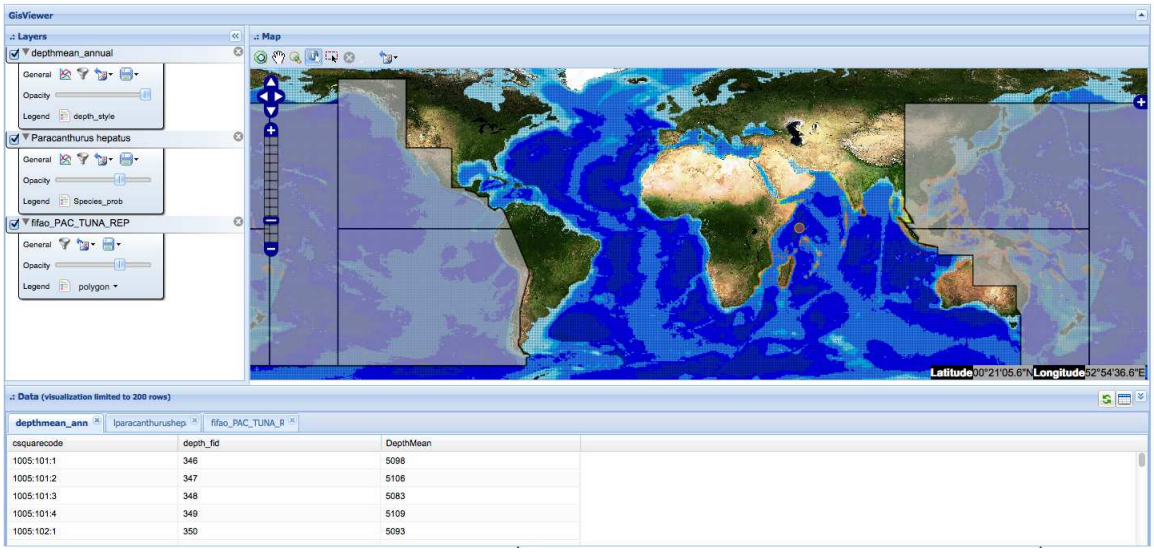

Figure 2. The GISViewer Portlet

2.2.3 THE SPECIES PRODUCTS DISCOVERY PORTLET

The *Species Products Discovery* portlet is a user interface conceived to support the discovery and management of species data including taxa names and occurrence points from a number of providers in a seamless way.

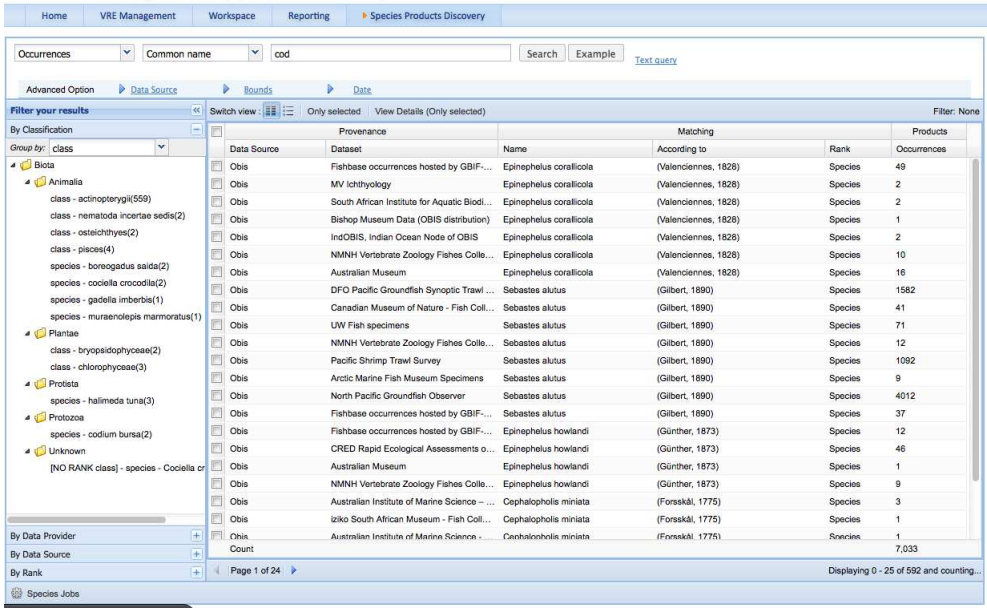

Figure 3. The Species Product Discovery Portlet

As regards the discovery facility, the portlet supports the specification of search criteria based on species scientific name or common name as well as on the type of product the user is interested in, i.e. occurrence points or taxa names. In addition to that, the user can specify *(i)* the data sources he/she is willing to use among the available ones, *(ii)* the geographical area he/she is interested in (via a bounding box) and *(iii)* the time interval he/she is interested in. Diverse clustering of the results are supported including those by classification, by data provider, by data source, and by rank.

With regards to species data management, the portlet supports diverse facilities depending on the type of product to be managed. As regards taxa names, the portlet make it possible to have a detailed description of each selected name including the classification, to save the discovered objects in the workspace as to use them in other contexts (e.g. taxa names comparison), to produce entire checklists of part of a classification by starting from a given taxa name. As regards occurrence points, the portlet make it possible to have a detailed description of the selected occurrence point datasets, to dynamically visualize the selected occurrence points on a map, to store the selected data in the workspace as to use them in future activities (e.g. niche modeling).

2.2.4 THE SPECIES VIEW PORTLET

The *Species View* portlet is a user interface conceived to support an advanced search and browse of species maps produced via the AquaMaps application. 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.

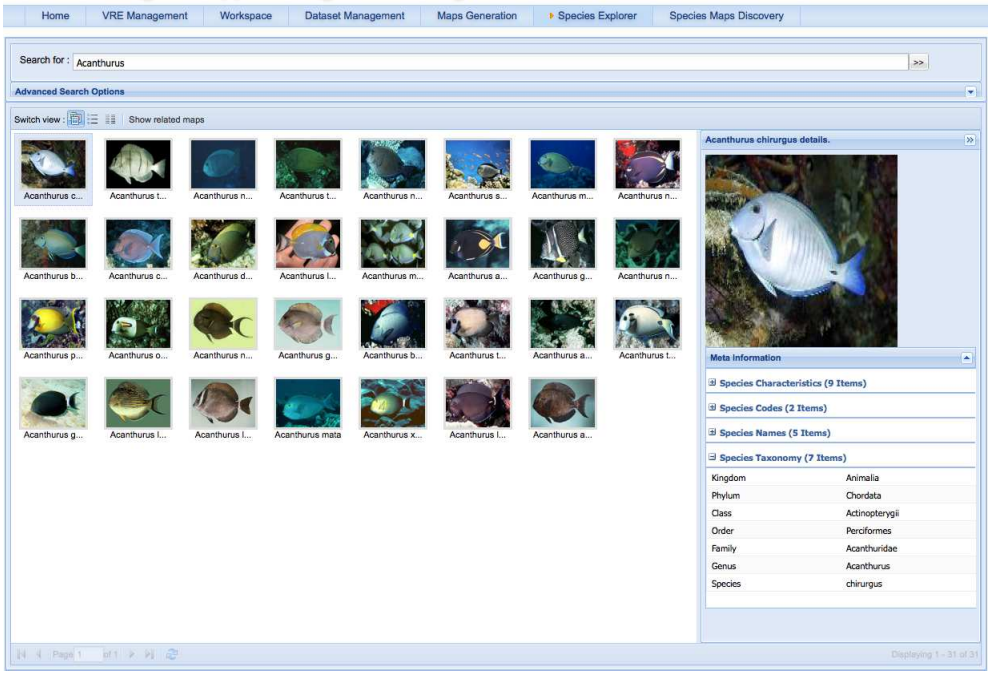

Figure 4. The Species View Portlet

2.2.5 THE WORKSPACE

The *Workspace* portlet is a user interface conceived to provide its user with a collaborative area for storing, exchanging and organizing information objects according to any specific need. Every user of any Virtual Research Environment is provided with this area that resembles a classic folder-based file system.

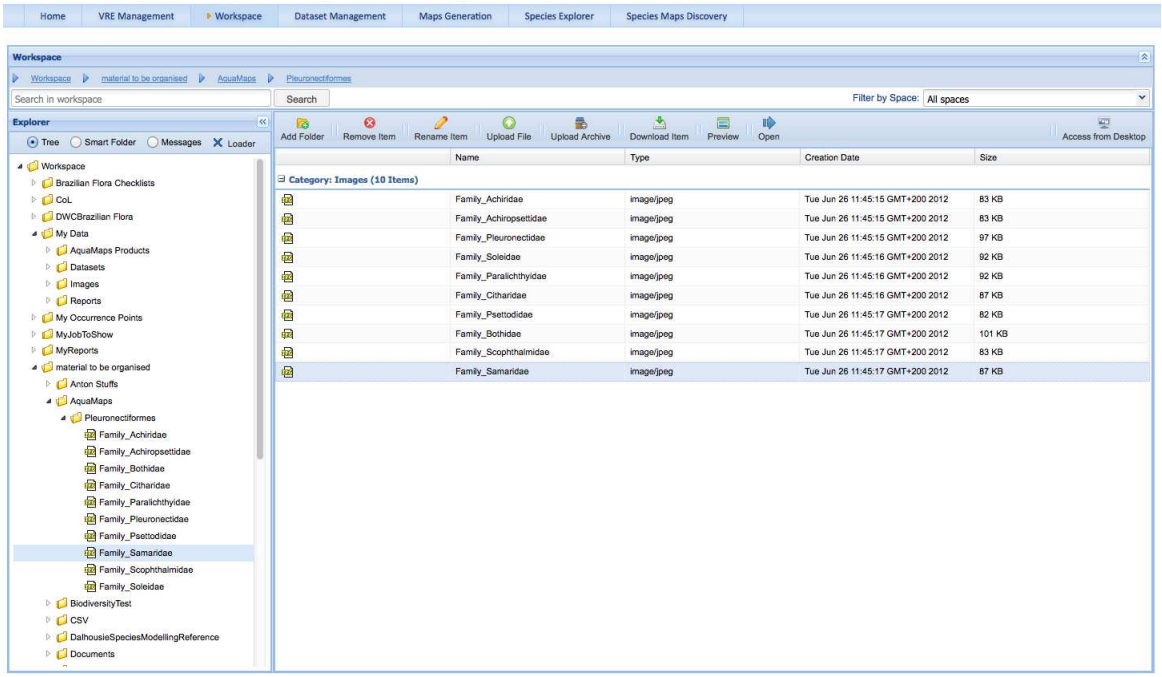

Figure 5. The Workspace Portlet

The added value of this collaborative area is represented by the item types it can manage in a seamless way. They range from binary files to compound information objects representing tabular data, species distribution maps, time series. Every item in the workspace is equipped with a rich metadata including bibliographic information like title and creator as well as lineage data.

In addition to information objects storage and organisation, the portlet allows to easily exchange objects among users as well as to import/export object from/to the user file system so as to make it possible to process such objects by relying on facilities offered by the infrastructure as well as on the facilities a user might have on its own computer. The Workspace is also equipped with a WebDAV[22] based mechanism making it possible to integrate the entire workspace in the user file system.

Another feature supporting cooperation and collaboration is the that enabling messages exchange. Via the Workspace it is possible to create rich messages including attachments to be delivered to co-workers.

Overall, the workspace plays a very central role in the context of the Virtual Research Environments. It is conceived to be the working area users and applications can rely on to acquire objects to be processed or to store objects resulting from any processing activity.

[22] http://www.webdav.org/

3 VIRTUAL RESEARCH ENVIRONMENTS DEPLOYMENT AND OPERATION

The iMarine Board Work Plan contains requests for facilities in support of managing EA-CoP data workflows spanning the statistical, geospatial, and biodiversity domains. Some of these facilities can only be defined after others have been released, as not only the technology requirements will evolve, but also the exploitation scenario may change, e.g. when the Board identifies new exploitation opportunities in the Business Cases. This implies that the description will only be finalized after a facility is released.

In order to provide the EA-CoP members with concrete realisation of the expected facilities, the iMarine project deployed and operated a number of Virtual Research Environments as described in the remainder of this section.

3.1 DESIGN, DEPLOYMENT AND OPERATION

The activity leading to the deployment and management of a Virtual Research Environment is driven by a dedicated development plan [1]. According to such a plan which is dynamic and evolving, the WP6 team is requested to analyse the requirements posed by the EA-CoP on various data management workflows and to put in place three types of activities: *(i)* the development of a new or enhanced technology (Services, software libraries, portlets) needed to support the specific need; *(ii)* the modification of an existing Virtual Research Environment to make available the new facility and/or the data that are needed in a given facility; and *(iii)* design and deploy a new Virtual Research Environment. Table 1 reports the list of the Virtual Research Environments operational in September 2012. Some of these Virtual Research Environments have been inherited by previous projects (namely D4Science-II) and because of this are operational since the beginning of the project.

Table 1. Virtual Research Environments

VRE Name	VO	Start Date	# Users
AquaMaps	FARM	Nov 2011	47
BiodiversityResearchEnvironment	gCubeApps	Jul 2012	21
DocumentWorkflow	gCubeApps	Nov 2011	16
DRIVER	Ecosystem	Nov 2011	19
EcologicalModeling	gCubeApps	Nov 2011	30
EM – Environmental Monitoring	Ecosystem	Nov 2011	15
FCPPS – Fisheries Country Profiles Production System	FARM	Nov 2011	34
HEPGateway	Ecosystem	Nov 2011	17
ICIS – Integrated Capture Information System	FARM	Nov 2011	41
TimeSeries	gCubeApps	Nov 2011	24
TryIt	Ecosystem	Nov 2011	39
VesselActivitiesAnalyser	gCubeApps	Nov 2011	24
VME-DB	FARM	Jul 2012	8
VTI – Vessel Transmitted Information	FARM	Nov 2011	21

A brief description of each available VRE is reported in the following sections.

3.1.1 THE AQUAMAPS VRE

The AquaMaps Virtual Research Environment is for providing fisheries and aquaculture scientists with facilities for producing and accessing species predictive distribution maps showing the likelihood that a certain species or a combination of species will live in specific regions or areas.

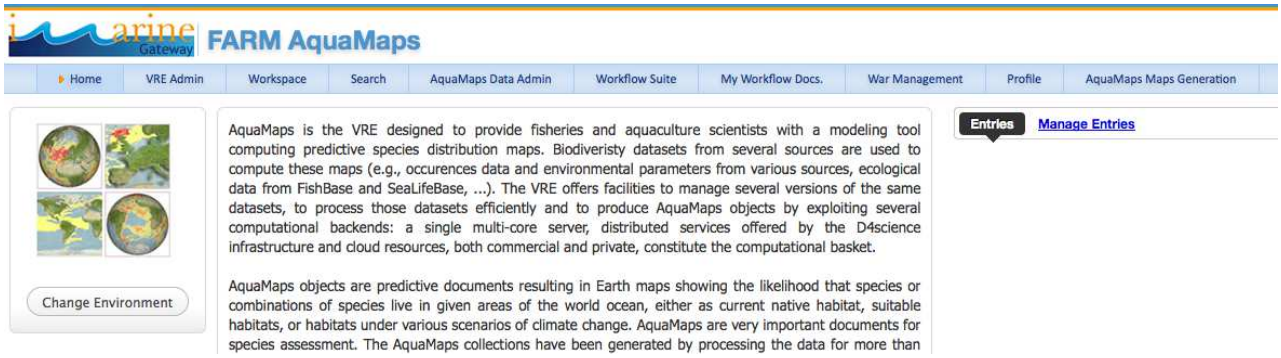

Figure 6. The AquaMaps VRE Homepage

The main facilities this VRE offers are:

- AquaMaps Data Administration: to enable VRE Data Managers to produce new versions of the AquaMaps datasets (the datasets exploited by the AquaMaps service to produce species distribution maps). Such datasets include HSPEC (an estimation of species occurrence by species and cell), HSPEN (an envelope representing the preference of species for environmental ranges), and HCAF (environmental parameters by cell) while the algorithms that can be used to produce new versions of them include Linear or Parabolic Interpolations, Native and Suitable Range, Native 2050 and Suitable 2050 Range;
- AquaMaps Maps Generation: to enable users to produce species and biodiversity predictive distribution maps. The portlet make it possible to select the set of species to analysed, to define the data to use and to submit massive generation tasks leading to the production of AquaMaps objects representing the maps eventually including their GIS version;
- Search: to enable users to discover information objects over a number of collections via a keyword-based Google-like search, an advanced query consisting of diverse criteria per field, or via browsing. Collections are either materialised set of information objects or virtual collections resulting from the interaction with existing data sources (e.g. via harvesting or via query);
- Documents Workflow facilities: to enable users (i) to define complex workflows (including steps and roles users should have to perform certain steps) governing the production of gCube documents, (ii) to instantiate such workflows to actual documents to be collaboratively created, and (iii) to monitor workflow execution;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);
- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers can *(i)* authorise users in accessing the VRE, *(ii)* assign or withdraw roles to users, *(iii)* remove users, and *(iv)* send a communication to the current users.

The main datasets that are available via the services hosted by this VRE include datasets needed to the AquaMaps algorithm (i.e. HSPEC, an estimation of species occurrence by species and cell; HSPEN, an envelop representing the tolerance of species wrt environmental parameters; and HCAF, environmental parameters by cell), time series graphs produced by FAO (aquaculture, capture, production, and trade), fact sheets produced by FAO on introduced species and cultured aquatic species, maps produced by FAO on country, Current National Legislation Overview (NALO) and National Aquaculture Sector Overview (NASO).

3.1.2 THE BIODIVERSITYRESEARCHENVIRONMENT VRE

The BiodiversityResearchEnvironment Virtual Research Environment is conceived to provide biodiversity scientist with facilities for seamless access to a rich array of biodiversity data including occurrence points and taxa records from established providers including GBIF, Catalogue of Life, and OBIS.

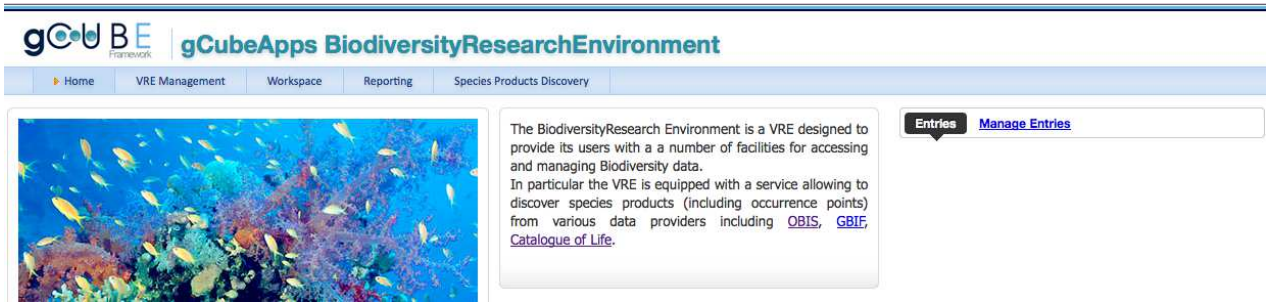

Figure 7. The BiodiversityResearchEnvironment VRE Homepage

The main facilities this VRE offers are:

- Species Products Discovery: to enable users to discover and manage species products (occurrence data and taxa names) from a number of heterogeneous providers in a seamless way (cf. Sec. 2.2.3). Once discovered, objects can be stored in the workspace for future uses;
- Reporting facilities: to enable users to collaboratively produce reports consisting in complex documents characterised by well defined structures (templates). Via this facility, users can define new templates as well as collaboratively create new reports compliant with defined templates. Reports might contain diverse elements ranging from texts to images and tables, and such constituents can result from objects stored in the user workspace. Reports can be materialised in multiple formats including PDF, HTML and OpenXML;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);
- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers can (i) authorise users in accessing the VRE, (ii) assign or withdraw roles to users, (iii) remove users, and (iv) send a communication to the current users.

The main datasets that are available via the services hosted by this VRE include:

- Catalogue of Life – offering information on known species of animals, plants, fungi and micro-organisms, namely taxonomic information (cf. Sec. 2.1);

- GBIF – offering information on species, namely occurrence data (cf. Sec. 2.1);
- OBIS – offering information on marine species from all of the worlds oceans, both occurrence and taxonomic information (cf. Sec. 2.1);
- SpeciesLink – offering information on species, namely occurrence data;
- WoRMS – offering information on marine species, namely taxonomic information (cf. Sec. 2.1);

3.1.3 THE DOCUMENTWORKFLOW VRE

The DocumentsWorkflow Virtual Research Environment is conceived to provide its users with a working environment focused on the gCube facilities for managing Document life-cycles. It exploit the facilities offered by the gCube Business Documents Workflow Management Suite enabling the production of reports that require a collaborative activity of several actors.

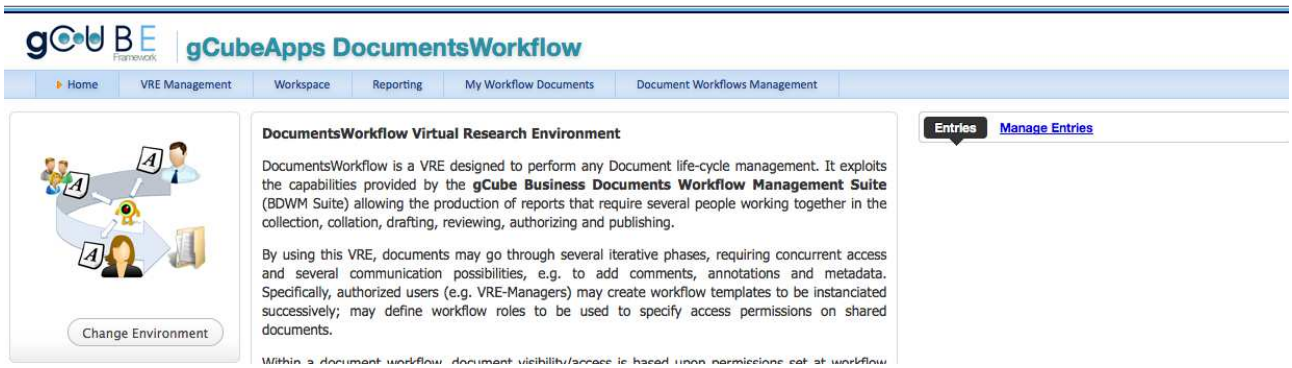

Figure 8. The DocumentsWorkflow VRE Homepage

The main facilities this VRE offers are:

- Reporting facilities: to enable users to collaboratively produce reports consisting in complex documents characterised by well defined structures (templates). Via this facility, users can define new templates as well as collaboratively create new reports compliant with defined templates. Reports might contain diverse elements ranging from texts to images and tables, and such constituents can result from objects stored in the user workspace. Reports can be materialised in multiple formats including PDF, HTML and OpenXML;
- Documents Workflow facilities: to enable users (i) to define complex workflows (including steps and roles users should have to perform certain steps) governing the production of gCube documents, (ii) to instantiate such workflows to actual documents to be collaboratively created, and (iii) to monitor workflow execution;
- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers can (i) authorise users in accessing the VRE, (ii) assign or withdraw roles to users, (iii) remove users, and (iv) send a communication to the current users;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5).

3.1.4 THE DRIVER VRE

The **DRIVER** Virtual Research Environment is for users willing to access the content offered by the DRIVER infrastructure (an infrastructure built by aggregating scientific publications in journal articles, dissertations, books, lectures, report, etc. from 250+ repositories scattered over 30+ countries) in the context of a VRE as to benefit from additional facilities.

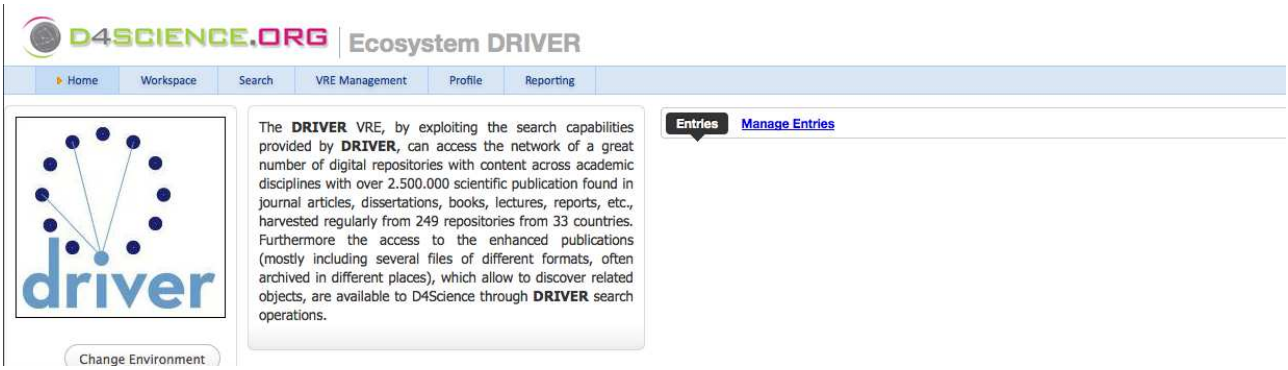

Figure 9. The DRIVER VRE Homepage

The main facilities this VRE offers are:

- Search: to enable users to discover information objects over a number of collections via a keyword-based Google-like search, an advanced query consisting of diverse criteria per field, or via browsing. Collections are either materialised set of information objects or virtual collections resulting from the interaction with existing data sources (e.g. via harvesting or via query);
- Reporting facilities: to enable users to collaboratively produce reports consisting in complex documents characterised by well defined structures (templates). Via this facility, users can define new templates as well as collaboratively create new reports compliant with defined templates. Reports might contain diverse elements ranging from texts to images and tables, and such constituents can result from objects stored in the user workspace. Reports can be materialised in multiple formats including PDF, HTML and OpenXML;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);
- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers can (i) authorise users in accessing the VRE, (ii) assign or withdraw roles to users, (iii) remove users, and (iv) send a communication to the current users.

The main datasets that are available via the services hosted by this VRE include the entire DRIVER collection, i.e. the set of information objects aggregated by the homonymous project building an infrastructure of open access repositories.

3.1.5 THE ECOLOGICALMODELLING VRE

The EcologicalModelling Virtual Research Environment is conceived to provide its users with a working environment focused on the gCube facilities for producing species distribution maps resulting from the processing of data on species characteristics and environmental observations. The resulting maps are actually rich information objects containing PNG images, GIS layers as well as metadata.

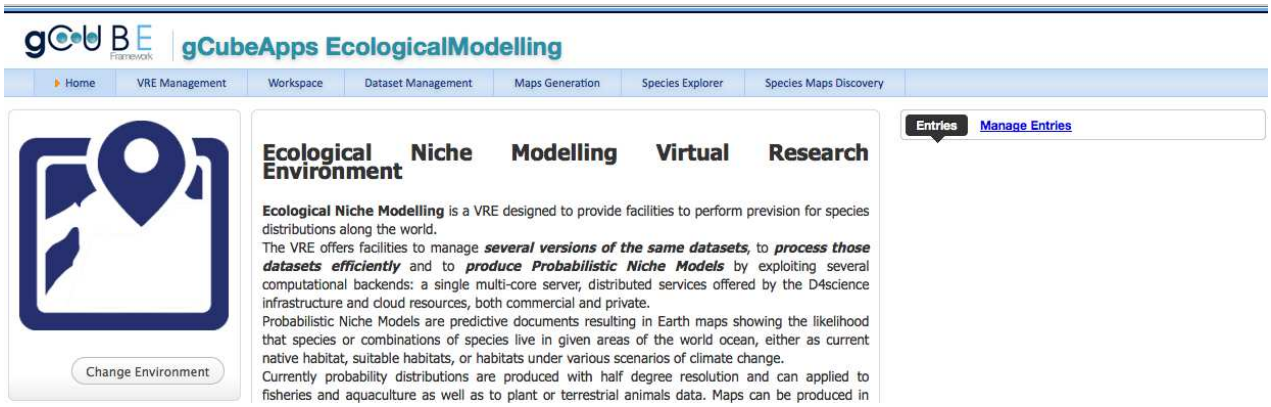

Figure 10. The EcologicalModelling VRE Homepage

The main facilities this VRE offers are:

- AquaMaps Data Administration: to enable VRE Data Managers to produce new versions of the AquaMaps datasets (the datasets exploited by the AquaMaps service to produce species distribution maps). Such datasets include HSPEC (an estimation of species occurrence by species and cell), HSPEN (an envelop representing the tolerance of species wrt environmental parameters), and HCAF (environmental parameters by cell) while the algorithms that can be used to produce new versions of them include Linear or Parabolic Interpolations, Native and Suitable Range, Native 2050 and Suitable 2050 Range;
- AquaMaps Maps Generation: to enable users to produce species and biodiversity predictive distribution maps. The portlet make it possible to select the set of species to analysed, to define the data to use and to submit massive generation tasks leading to the production of AquaMaps objects representing the maps eventually including their GIS version;
- Species View: to enable users to discover and browse species products (namely distribution maps) produced via the AquaMaps Maps Generation facility in an innovative way. This facility supports discovery mechanisms ranging from simple search based on species names to very specific search criterion and it offers a comprehensive set of products visualisation approaches (cf. Sec. 2.2.4);
- Species Maps Discovery: to enable users to discover and visualize GIS layers corresponding to species distribution maps that have been generated and published in the iMarine infrastructure. This facilities relies on the GeoExplorer portlet (cf. Sec. 2.2.1) and make it possible to effectively exploit the generated maps and perform comparisons and analysis of the diverse distributions by enabling maps overlay, transects production and values inspection;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);

- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers are enabled to *(i)* authorise users in accessing the VRE, *(ii)* assign or withdraw roles to users, *(iii)* remove users, and *(iv)* send a communication to the current users.

The main datasets that are available via the services hosted by this VRE include datasets needed to the AquaMaps algorithm (i.e. HSPEC, an estimation of species occurrence by species and cell; HSPEN, an envelop representing the tolerance of species wrt environmental parameters; and HCAF, environmental parameters by cell).

3.1.6 THE EM – ENVIRONMENTAL MONITORING VRE

The Environmental Monitoring (EM) Virtual Research Environment is for making available various satellite data and services for consuming such data.

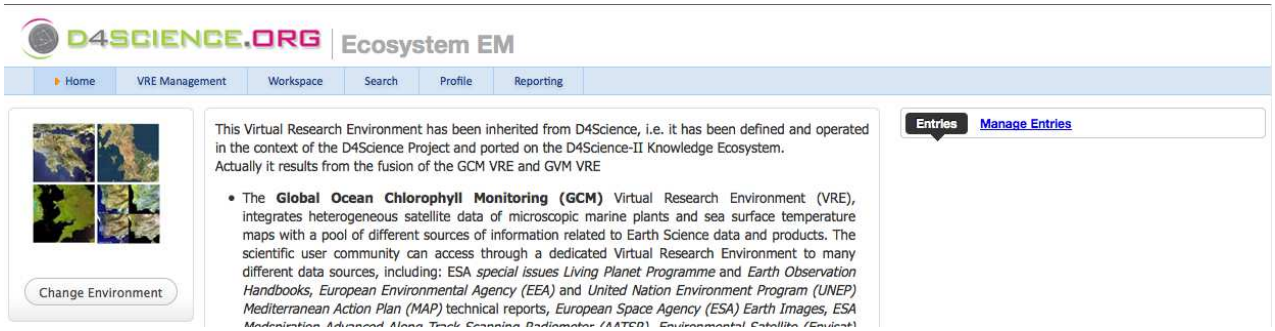

Figure 11. The EM VRE Homepage

The main facilities this VRE offers are:

- Search: to enable users to discover information objects over a number of collections via a keyword-based Google-like search, an advanced query consisting of diverse criteria per field, or via browsing. Collections are either materialised set of information objects or virtual collections resulting from the interaction with existing data sources (e.g. via harvesting or via query);
- Reporting facilities: to enable users to collaboratively produce reports consisting in complex documents characterised by well defined structures (templates). Via this facility, users can define new templates as well as collaboratively create new reports compliant with defined templates. Reports might contain diverse elements ranging from texts to images and tables, and such constituents can result from objects stored in the user workspace. Reports can be materialised in multiple formats including PDF, HTML and OpenXML;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);
- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers are enabled to *(i)* authorise users in accessing the VRE, *(ii)* assign or withdraw roles to users, *(iii)* remove users, and *(iv)* send a communication to the current users.

The main datasets that are available via the services hosted by this VRE include a number of specific collections on Environmental Monitoring material including the European Environment Agency reports,

Earth images collected and distributed by the European Space Agency, level 3 products produced by the MERIS instrument like global vegetation index and chlorophyll in the water.

3.1.7 THE FCPPS – FISHERIES COUNTRY PROFILES PRODUCTION SYSTEM VRE

The **Fisheries Country Profiles Production System (FCPPS)** Virtual Research Environment is for fisheries and aquaculture authors, managers and researchers who produce reports containing country-level data. It provides seamless access to multiple data sources, including their annotation and versioning and permits production of structured text, tables, charts and graphs from these sources to be easily inserted into custom reporting templates that can support multiple output formats.

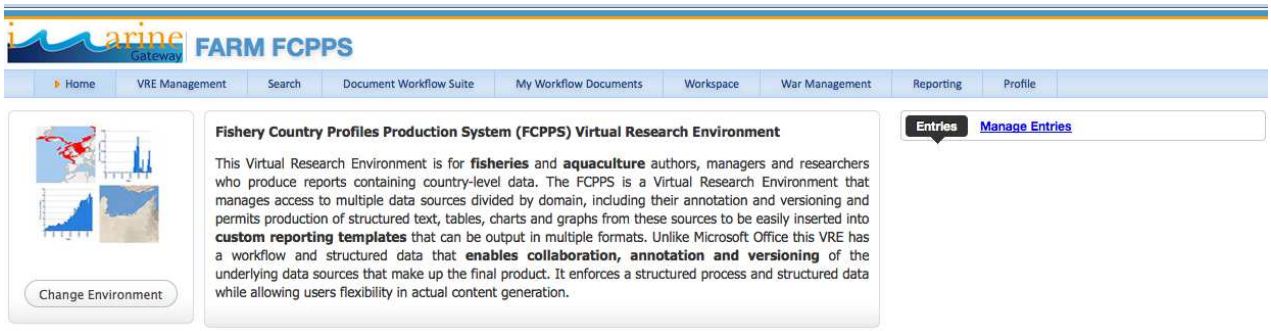

Figure 12. The FCPPS VRE Homepage

The main facilities this VRE offers are:

- Search: to enable users to discover information objects over a number of collections via a keyword-based Google-like search, an advanced query consisting of diverse criteria per field, or via browsing. Collections are either materialised set of information objects or virtual collections resulting from the interaction with existing data sources (e.g. via harvesting or via query);
- Reporting facilities: to enable users to collaboratively produce reports consisting in complex documents characterised by well defined structures (templates). Via this facility, users can define new templates as well as collaboratively create new reports compliant with defined templates. Reports might contain diverse elements ranging from texts to images and tables, and such constituents can result from objects stored in the user workspace. Reports can be materialised in multiple formats including PDF, HTML and OpenXML;
- Documents Workflow facilities: to enable users (i) to define complex workflows (including steps and roles users should have to perform certain steps) governing the production of gCube documents, (ii) to instantiate such workflows to actual documents to be collaboratively created, and (iii) to monitor workflow execution;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);
- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers are enabled to *(i)* authorise users in accessing the VRE, *(ii)* assign or withdraw roles to users, *(iii)* remove users, and *(iv)* send a communication to the current users.

The main datasets that are available via the services hosted by this VRE include material acquired via external repositories like AquaticCommons and WHOAS (cf. Sec. 2.1), time series graphs produced by FAO (aquaculture, capture, production, and trade), fact sheets produced by FAO on introduced species and cultured aquatic species, maps produced by FAO on country, Current National Legislation Overview (NALO) and National Aquaculture Sector Overview (NASO).

3.1.8 THE HEPGATEWAY VRE

The HEPGateway Virtual Research Environment is conceived to support users willing to access the INSPIRE High Energy Physics Information System, i.e. the one-stop-shop for the HEP community research publications. Through this VRE users might benefit from the gCube facilities for managing information objects retrieved through such a service, e.g. store them in their workspace, share with co-workers, annotate.

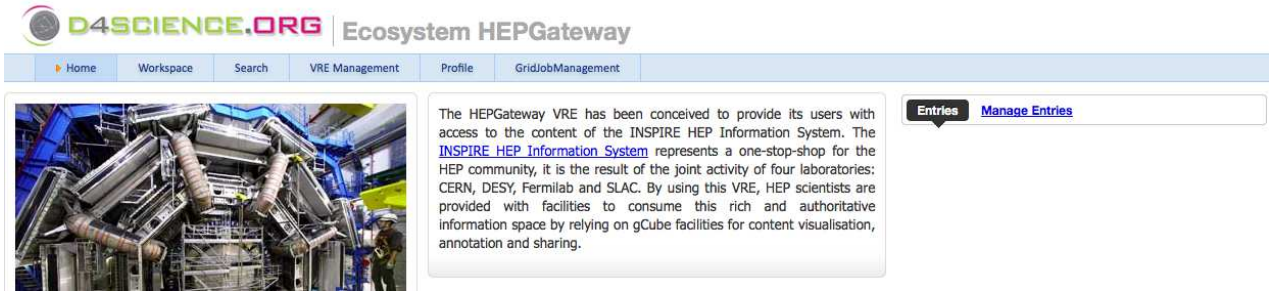

Figure 13. The HEPGateway VRE Homepage

The main facilities this VRE offers are:

- Search: to enable users to discover information objects over a number of collections via a keyword-based Google-like search, an advanced query consisting of diverse criteria per field, or via browsing. Collections are either materialised set of information objects or virtual collections resulting from the interaction with existing data sources (e.g. via harvesting or via query);
- Jobs Management: to enable users to execute and monitor the execution of complex tasks by relying on grid computational resources linked to the iMarine infrastructure;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);
- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers are enabled to *(i)* authorise users in accessing the VRE, *(ii)* assign or withdraw roles to users, *(iii)* remove users, and *(iv)* send a communication to the current users.

The main datasets that are available via the services hosted by this VRE include the INSPIRE collection, i.e. the content of the INSPIRE HEP Information System[23] which is an authoritative database of High-Energy Physics Literature.

3.1.9 THE ICIS – INTEGRATED CAPTURE INFORMATION SYSTEM VRE

The **Integrated Capture Information System (ICIS)** Virtual Research Environment offers fisheries statisticians a set of tools to manage their data. Statisticians produce statistics from often very different data sources, and need a controlled process for the ingestion, validation, transformation, comparison and exploitation of statistical data for the fisheries captures domain;

Figure 14. The ICIS VRE Homepage

The main facilities this VRE offers are:

- Time Series Management: to enable users to import, curate and manage time series. This is a comprehensive and feature-rich environment that support data managers during the whole life cycle of data management from capture to publishing and visualisation. In enable data managers to import and transform CSV files in time series, i.e. tabular data having proper types associated with columns eventually referring to code lists – reference datasets representing recognized value instances of the elements the dataset is about, e.g., species, zones, countries. The environment guarantees that the time series are compliant with the defined types and code lists. Besides the curation, the environment supports the analysis of the data by enabling a user to (i) perform operations like grouping and filtering, (ii) producing charts and GIS maps (if the data have geographic features) and (iii) analysing the data via an R[24] environment. Finally, the environment supports the publishing of time series in the infrastructure by equipping them with rich metadata so that such resources can be used in other application contexts;
- Code Lists Management: to enable the users to import and manage code lists, i.e. reference datasets representing recognised value instances of the elements the dataset is about. Such environment enable users to import CSV files or existing code lists from SDMX[25] repositories, curate them when needed,

[23] http://inspirehep.net/
[24] http://www.r-project.org/
[25] sdmx.org

inspect the current values, and produce and publish new versions that can be used during the curation phase of a time series. Code lists are annotated with rich metadata capturing attribution and lineage;

- Reporting facilities: to enable users to collaboratively produce reports consisting in complex documents characterised by well defined structures (templates). Via this facility, users can define new templates as well as collaboratively create new reports compliant with defined templates. Reports might contain diverse elements ranging from texts to images and tables, and such constituents can result from objects stored in the user workspace. Reports can be materialised in multiple formats including PDF, HTML and OpenXML;
- Documents Workflow facilities: to enable users (i) to define complex workflows (including steps and roles users should have to perform certain steps) governing the production of gCube documents, (ii) to instantiate such workflows to actual documents to be collaboratively created, and (iii) to monitor workflow execution;
- WAR Management: to enable authorised users to upload, deploy and monitor web-based applications packaged and distributed via the Web application ARchive (WAR) format. The upload phase enable a user willing to upload its web application to provide the system with the metadata needed for its management. The deployment and monitoring phase enable a user to select a machine (among the available ones), to deploy on it a web-application and to monitor its operational status;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);
- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers are enabled to *(i)* authorise users in accessing the VRE, *(ii)* assign or withdraw roles to users, *(iii)* remove users, and *(iv)* send a communication to the current users.

The main datasets that are available via the services hosted by this VRE include a series of code lists including FAO and IRD SDMX repositories (cf. Sec. 2.1).

3.1.10 THE TIMESERIES VRE

The TimeSeries Virtual Research Environment is conceived to provide its users with a working environment focused on gCube facilities for managing time series. This environment supports the load of time series objects, the curation and validation by relying on authoritative code lists, the sharing of such objects with co-workers, the production of graphs, the visualization through a GIS service;

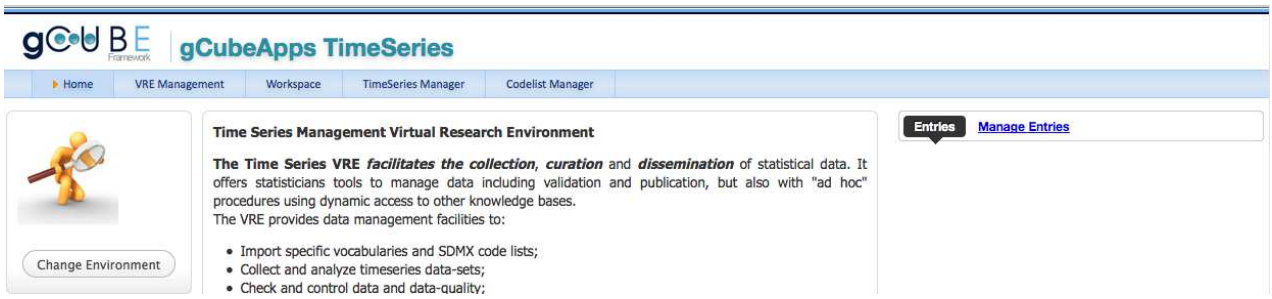

Figure 15. The TimeSeries VRE Homepage

The main facilities this VRE offers are:

- Time Series Management: to enable users to import, curate and manage time series. This is a comprehensive and feature-rich environment that support data managers during the whole life cycle of data management from capture to publishing and visualisation. In enable data managers to import and transform CSV files in time series, i.e. tabular data having proper types associated with columns eventually referring to code lists – reference datasets representing recognized value instances of the elements the dataset is about, e.g., species, zones, countries. The environment guarantees that the time series are compliant with the defined types and code lists. Besides the curation, the environment supports the analysis of the data by enabling a user to (i) perform operations like grouping and filtering, (ii) producing charts and GIS maps (if the data have geographic features) and (iii) analysing the data via an R[26] environment. Finally, the environment supports the publishing of time series in the infrastructure by equipping them with rich metadata so that such resources can be used in other application contexts;
- Code Lists Management: to enable the users to import and manage code lists, i.e. reference datasets representing recognised value instances of the elements the dataset is about. Such environment enable users to import CSV files or existing code lists from SDMX[27] repositories, curate them when needed, inspect the current values, and produce and publish new versions that can be used during the curation phase of a time series. Code lists are annotated with rich metadata capturing attribution and lineage;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);
- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers are enabled to *(i)* authorise users in accessing the VRE, *(ii)* assign or withdraw roles to users, *(iii)* remove users, and *(iv)* send a communication to the current users.

The main datasets that are available via the services hosted by this VRE include a series of code lists including FAO and IRD SDMX repositories (cf. Sec. 2.1).

3.1.11 THE TRYIT VRE

The TryIt Virtual Research Environment is conceived to serve demonstration and training activities. It gives access to a set of sample content and supports the most common functionality a gCube-based VRE might be able to support, e.g. search, browse and personal workspace management.

[26] http://www.r-project.org/
[27] sdmx.org

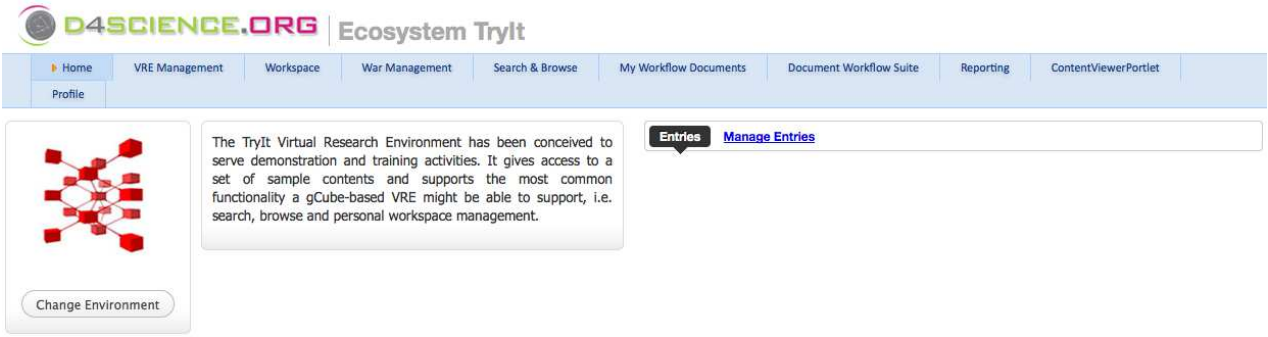

Figure 16. The TryIt VRE Homepage

The main facilities this VRE offers are:

- Search: to enable users to discover information objects over a number of collections via a keyword-based Google-like search, an advanced query consisting of diverse criteria per field, or via browsing. Collections are either materialised set of information objects or virtual collections resulting from the interaction with existing data sources (e.g. via harvesting or via query);
- Reporting facilities: to enable users to collaboratively produce reports consisting in complex documents characterised by well defined structures (templates). Via this facility, users can define new templates as well as collaboratively create new reports compliant with defined templates. Reports might contain diverse elements ranging from texts to images and tables, and such constituents can result from objects stored in the user workspace. Reports can be materialised in multiple formats including PDF, HTML and OpenXML;
- Documents Workflow facilities: to enable users (i) to define complex workflows (including steps and roles users should have to perform certain steps) governing the production of gCube documents, (ii) to instantiate such workflows to actual documents to be collaboratively created, and (iii) to monitor workflow execution;
- WAR Management: to enable authorised users to upload, deploy and monitor web-based applications packaged and distributed via the Web application ARchive (WAR) format. The upload phase enable a user willing to upload its web application to provide the system with the metadata needed for its management. The deployment and monitoring phase enable a user to select a machine (among the available ones), to deploy on it a web-application and to monitor its operational status;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);
- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers are enabled to *(i)* authorise users in accessing the VRE, *(ii)* assign or withdraw roles to users, *(iii)* remove users, and *(iv)* send a communication to the current users.

The main datasets that are available via the services hosted by this VRE include a number of test collections material acquired via external repositories like AquaticCommons (cf. Sec. 2.1), the INSPIRE collection, i.e. the content of the INSPIRE HEP Information System[28] which is an authoritative database of High-Energy Physics Literature, and collections on Environmental Monitoring material including the European Environment Agency reports, Earth images collected and distributed by the European Space Agency, level 3 products produced by the MERIS instrument like global vegetation index and chlorophyll in the water.

3.1.12 THE VESSELACTIVITIESANALYSER VRE

The VesselActivitiesAnalyzer Virtual Research Environment is conceived to provide its users with a working environment focused on gCube facilities for managing vessel trajectories. This environment support users in loading and curating vessel trajectories, enriching such data with bathymetry and FAO Area, sharing with co-workers, analysing such objects by producing maps on vessel activities and fishing monthly effort.

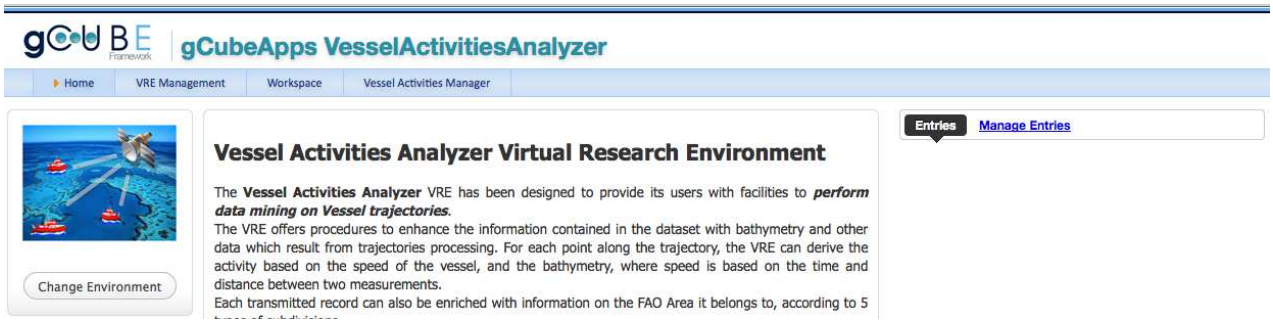

Figure 17. The VesselActivitiesAnalyser VRE Homepage

The main facilities this VRE offers are:

- Vessel Activities Management: to enable users to perform data mining tasks on Vessel trajectories. It enables data managers to import and transform CSV files representing trajectories into well-defined tabular data, to enrich such tabular data with information on FAO areas and bathymetry, to perform mining tasks aiming at deriving the vessel activity by relying on vessel speed and bathymetry. Besides these specific operations, the environment supports the analysis of the data by enabling a user to (i) perform operations like grouping and filtering, (ii) producing charts and GIS maps (if the data have geographic features) and (iii) analysing the data via an R[29] environment;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);
- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers are enabled to *(i)* authorise users in accessing the VRE, *(ii)* assign or withdraw roles to users, *(iii)* remove users, and *(iv)* send a communication to the current users.

[28] http://inspirehep.net/
[29] http://www.r-project.org/

3.1.13 THE VME-DB VRE

The **Vulnerable Marine Ecosystem Database (VME-DB)** Virtual Research Environment is for fisheries and aquaculture authors willing to collaboratively produce Fact Sheets on Vulnerable Marine Ecosystems (VME).

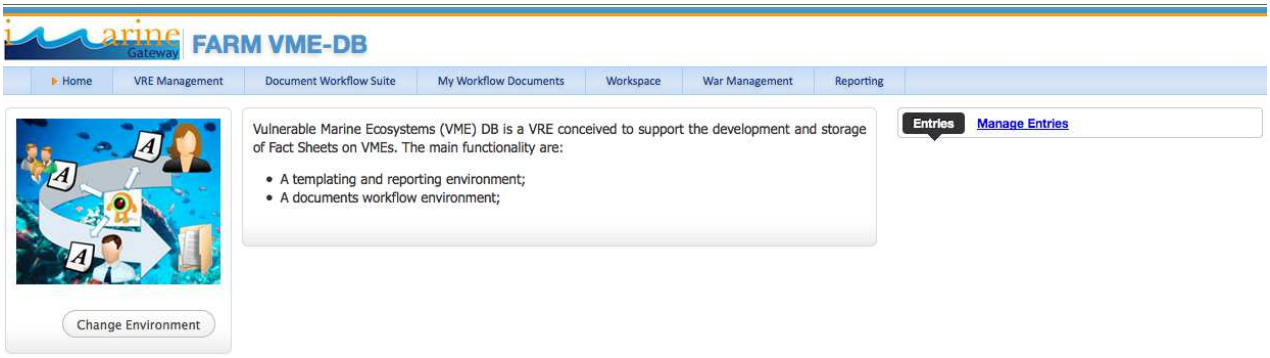

Figure 18. The VME-DB VRE Homepage

The main facilities this VRE offers are:

- Reporting facilities: to enable users to collaboratively produce reports consisting in complex documents characterised by well defined structures (templates). Via this facility, users can define new templates as well as collaboratively create new reports compliant with defined templates. Reports might contain diverse elements ranging from texts to images and tables, and such constituents can result from objects stored in the user workspace. Reports can be materialised in multiple formats including PDF, HTML and OpenXML;
- Documents Workflow facilities: to enable users (i) to define complex workflows (including steps and roles users should have to perform certain steps) governing the production of gCube documents, (ii) to instantiate such workflows to actual documents to be collaboratively created, and (iii) to monitor workflow execution;
- WAR Management: to enable authorised users to upload, deploy and monitor web-based applications packaged and distributed via the Web application ARchive (WAR) format. The upload phase enables a user willing to upload a web application to provide the system with the metadata needed for its management. The deployment and monitoring phase enable a user to select a machine (among the available ones), to deploy on it a web-application and to monitor its operational status;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);
- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers are enabled to *(i)* authorise users in accessing the VRE, *(ii)* assign or withdraw roles to users, *(iii)* remove users, and *(iv)* send a communication to the current users.

3.1.14 THE VTI – VESSEL TRANSMITTED INFORMATION VRE

The **Vessel Transmitted Information (VTI)** Virtual Research Environment enables the analysis of vessel activities over space and time by taking into account environmental data.

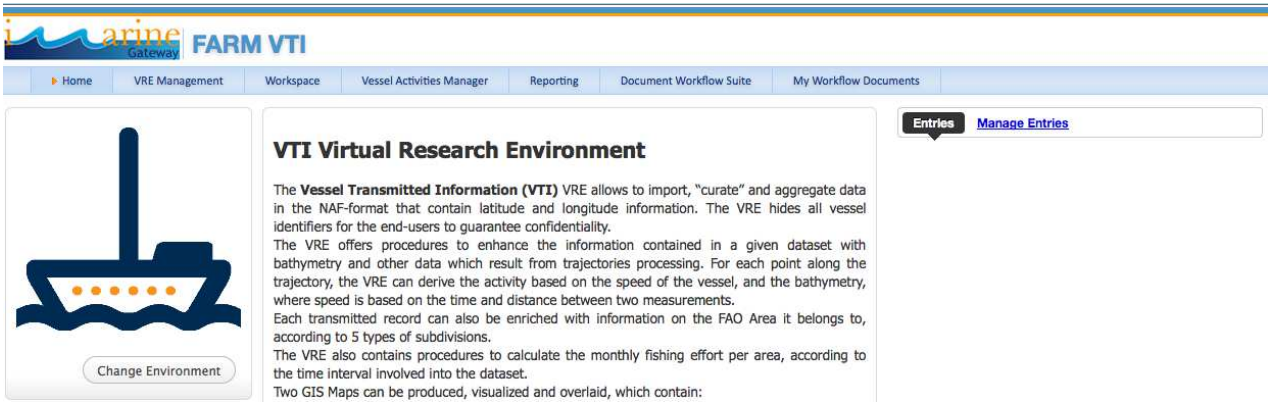

Figure 19. The VTI VRE Homepage

The main facilities this VRE offers are:

- Vessel Activities Management: to enable users to perform data mining tasks on Vessel trajectories. In enable data managers to import and transform CSV files representing trajectories into well-defined tabular data, to enrich such tabular data with information on FAO areas and bathymetry, to perform mining tasks aiming at deriving the vessel activity by relying on vessel speed and bathymetry. Besides these specific operations, the environment supports the analysis of the data by enabling a user to (i) perform operations like grouping and filtering, (ii) producing charts and GIS maps (if the data have geographic features) and (iii) analysing the data via an R[30] environment;
- Reporting facilities: to enable users to collaboratively produce reports consisting in complex documents characterised by well defined structures (templates). Via this facility, users can define new templates as well as collaboratively create new reports compliant with defined templates. Reports might contain diverse elements ranging from texts to images and tables, and such constituents can result from objects stored in the user workspace. Reports can be materialised in multiple formats including PDF, HTML and OpenXML;
- Documents Workflow facilities: to enable users (i) to define complex workflows (including steps and roles users should have to perform certain steps) governing the production of gCube documents, (ii) to instantiate such workflows to actual documents to be collaboratively created, and (iii) to monitor workflow execution;
- Workspace: to enable every user to store and organise the information objects he/she is interested to work with. In addition to that, the user is allowed to collaborate with other users by sharing objects and messages (cf. Sec. 2.2.5);

[30] http://www.r-project.org/

- VRE Management: to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers are enabled to *(i)* authorise users in accessing the VRE, *(ii)* assign or withdraw roles to users, *(iii)* remove users, and *(iv)* send a communication to the current users.

3.2 EXPLOITATION

The actual exploitation of a Virtual Research Environment to perform the institutional tasks a Community of Practice is called to perform starts once the VRE has been validated, the use of the e-infrastructure has been requested and approved by the management and community members have been trained and informed on their responsibilities and support options. Exploitation not necessarily means that large groups of users will use the released functionality. Rather, it means that project partners will have to identify and open a niche for the iMarine products in their data-production work-flow and e-infrastructure.

As an example, the FCPPS VRE is released and was validated by the community in 2011. However, finding resources to use the tool for the production of Fisheries Country Profiles requires that FAO identifies and commits human resources to the production of these reports. This entails more than training a user for the tool, as also country report editors, authors and reviewers must be mobilized.

As of July 2012, the following VREs were analysed in detail and considered ready for use by EA-CoP members:

1. AquaMaps (cf. Sec. 3.1.1);
2. ICIS (cf. Sec. 3.1.9);
3. FCPPS (cf. Sec. 3.1.7);

Several other VRE's have been delivered, but these are not fully used in exploitation activities yet, and the communities are still in the preparatory or validation phases.

In the remainder of this section, details on the exploitation of these VRES is briefly discussed.

3.2.1 AQUAMAPS VRE EXPLOITATION

The current version of the AquaMaps VRE (cf. Sec. 3.1.1) include additional features and enhancements resulting from previous exploitations of this environment and include:

- *Species maps discovery and visualisation* improvements include the enlarged number of layers for end-users to use in visualization results, management of default layers, a vastly improved user interface, and better integration with the workspace;
- *Dataset management* - rich features where added for map search, DarwinCore data support, the table generation, enhanced insight in the analysis performed, and the data table inspector.
- *Map storage and discovery;* allows users to load their own maps, and map-discovery (WMS) and import facility, produce overlays, and download a variety of visual and geospatial formats.

FIN team (the main clients of this VRE) is planning additional validation, also focusing on integration with the new services being developed (e.g. Species Products Discovery). The aim is to have more flexibility in

the AquaMaps generation with respect to the occurrence cell selection, with later addition of more distribution modelling algorithms.

Other than FIN, the intended user communities extend into the ICIS VRE related use-cases, and one particular use-case worth mentioning is the re-allocation of capture statistics. It will be more extensively dealt with in the ICIS components part (cf. Sec. 3.2.2), but the improvement to the map acquisition, management, storage and visualization realized in the context of AquaMaps were aligned with the requirements arising in the ICIS use-case. This exemplifies how larger communities expect to benefit from the e-infrastructure resources.

3.2.2 ICIS AND RELATED COMPONENTS EXPLOITATION

The ICIS VRE (cf. Sec. 3.1.9) is released, but awaits the completion in the communities of data providing services and exploitation work-flows. In addition, the validation revealed several improvements to functional parts of ICIS that are being implemented to facilitate the exploitation of ICIS.

The exploitation of ICIS is scheduled to start in Q4 for the Tuna Atlas use case [4].

Several community components geared towards supplying data resources or other services to the iMarine e-infrastructure continued to be used, were further developed, or installed in community environments.

These components all relate to the statistical cluster, grouped together here as ICIS-Components, since they primarily serve to support the released ICIS VRE. For the exploitation to serve the use-cases of the EA-CoP, several components are expected or have to be contributed by the communities.

As regards Community tools

- *OpenSDMX:* FAO and IRD maintained the facility to publish their code lists as SDMX. These can be imported to e.g. the ICIS code list manager, or any other SDMX capable tool, and provide the base for the code list manager;
- *WPS / SPREAD:* For the re-allocation of data, the Time-Series form the base information set of data. The community has to develop the logic to apply another spatial distribution to this data. The approach has been tested and validated in the community infrastructure, and is now being ported for exploitation to the iMarine e-infrastructure;
- *THREDDS:* In addition to the geospatial features already available, the Thematic Realtime Environmental Distributed Data simplifies the discovery and use of scientific data and allows scientific publications and educational materials to reference scientific data.

As regards new services

- *Statistical service supporting data mining:* for validation, harmonization, large (de)aggregations similarity calculations, and grouping / clustering. The EA-CoP has formulated the requirements for fully exploiting the e-infrastructure. The first results were tested with EA-CoP representatives, and for the Tuna Atlas use case, the planning is to start exploitation in Q4.
- *WAR-Deployer:* this facility to host and operate entire web-applications in the e-infrastructure was initially developed in the context of OpenSDMX. It is also available to other scenarios to exploit, but

since the validation was done for the ICIS use-case it is reported here. The community has already contributed a complete web-application, OpenSDMX, that is managed through this facility.

3.2.3 FCPPS VRE EXPLOITATION

FCPPS (cf. Sec. 3.1.7) is already available for exploitation, and FAO staff has been trained for the use and management of the VRE to produce Fisheries Country Profiles. Country Profiles are one of FAO FI Departments most visited web-product. However, FAO needs to find resources to free staff to start generating updates of existing FCP reports.

The trainees familiarized with the tools for workflow management, template design, and report generation, and tested the facilities. They also learned how to produce reports that comply with the FAO FiMES XML schema.

As a spin-off of the training, they provided feedback on the design, the organization of the workflow processes, and the on-line help. These suggestions have been communicated to the project, and their implementation was being assessed at the time of release of this deliverable.

These comments provide valuable input, not only to improve FCPPS, but also for the training of potential users of the related VME-DB (cf. Sec. 3.1.13). The technical release of the VME-DB VRE materialized in a matter of days, based on a requirements collection effort that was backed by a modified FCPPS VRE. For example, relevant data sources and a reporting template were based on the FCPPS VRE.

REFERENCES

[1] L. Candela. *Virtual Research Environments Plan.* iMarine D6.1 Project Deliverable. January 2012

[2] A. Ellenbroek, L. Candela. *Applications and Tools Development Plan*. iMarine D6.2 Project Deliverable. May 2012

[3] A. Manzi, L. Candela. iMarine Data e-Infrastructure Operation Report. iMarine D5.2 Project Deliverable. September 2012

[4] Y. Laurent. ICIS Assessment for Tuna Atlas Use-Case. iMarine internal document. July 2012

[5] A. Gentile. FCPPS Assessment for VME-DB Use-Case. iMarine internal document. August 2012

[6] M. Simi. iMarine Data Infrastructure Enabling Software. iMarine D8.2 Project Deliverable. August 2012

[7] G. Coro, F. De Faveri, L. Lelii. iMarine Data Management Software. iMarine D9.1 Project Deliverable. June 2012

[8] G. Farantatos. iMarine Data Consumption Software. iMarine D10.3 Project Deliverable. June 2012

[9] M. Simi, R. Tsantouli. Application Programming Interface Software. iMarine D11.2 Project Deliverable. June 2012