



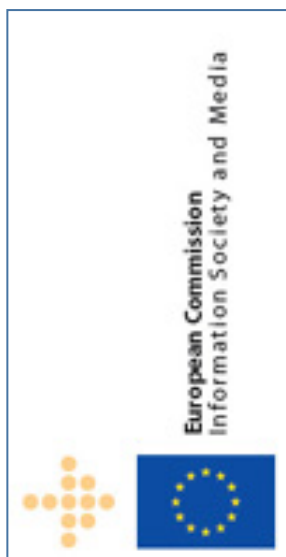
<i>Project Acronym</i>	<i>iMarine</i>
<i>Project Title</i>	<i>Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources</i>
<i>Project Number</i>	<i>283644</i>
<i>Deliverable Title</i>	<i>Virtual Research Environments Activity Report</i>
<i>Deliverable No.</i>	<i>D6.4</i>
<i>Delivery Date</i>	<i>July 2013</i>
<i>Author</i>	<i>Anton Ellenbroek, Leonardo Candela, Pasquale Pagano</i>

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.

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.	D6.4
Deliverable Title	Virtual Research Environments Activity Report
Contractual Delivery Date	June 2013
Actual Delivery Date	July 2013
Author(s)	Anton Ellenbroek, Leonardo Candela, Pasquale Pagano
Editor(s)	Anton Ellenbroek
Reviewer(s)	Nicholas Bailly
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/>).

Copyright © The iMarine Consortium 2011. See <http://www.i-marine.eu/Content/About.aspx?id=6cc695f5-cc75-4597-b9f1-6ebea7259105> for details on the copyright holders.

For more information on the project, its partners and contributors please see <http://www.i-marine.eu/>. You are permitted to copy and distribute verbatim copies of this document containing this copyright notice, but modifying this document is not allowed. You are permitted to copy this document in whole or in part into other documents if you attach the following reference to the copied elements: "Copyright © The iMarine Consortium 2011."

The information contained in this document represents the views of the iMarine Consortium as of the date they are published. The iMarine Consortium does not guarantee that any information contained herein is error-free, or up to date. THE IMARINE CONSORTIUM MAKES NO WARRANTIES, EXPRESS, IMPLIED, OR STATUTORY, BY PUBLISHING THIS DOCUMENT.

DELIVERABLE SUMMARY

Virtual Research Environments are among the key products that the iMarine project is requested to deliver to serve the needs of the Ecosystem Approach Community of Practice. The deployment and operation of a Virtual Research Environment is a task that involves 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 activities performed to provide the Ecosystem Approach Community of Practice with the set of Virtual Research Environments hosted by the iMarine portal in June 2013.

EXECUTIVE SUMMARY

Virtual Research Environments (VRE's) are “systems” that provide their users with a web-based set of facilities (including services, data and computational facilities) to accomplish a set of tasks by dynamically relying on the underlying infrastructure. VRE's are among the products to be delivered by the iMarine project to support the business cases of the iMarine Ecosystem Approach Community of Practice (EA-CoP).

The development of VRE's is based on three main activities: (i) the development of software artefacts that realise a set of functions, (ii) the deployment of these artefacts in an operational infrastructure, and (iii) the deployment and operation of these Virtual Research Environments that exploit the facilities offered by the deployed infrastructure to achieve objectives specified in the Business Cases of the Community of Practice.

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 as per June 2013. The second aspect – i.e. infrastructure deployment – is captured by another document, deliverable D5.3 “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 business cases identified by the Ecosystem Approach Community of Practice (EA-CoP).

As of June 2013, 14 VREs were active, with several new VREs added to support specific collaborative activities since the previous report. In particular, *FishFinderVRE* was deployed to support a well-defined group of biologists, *iMarineBoard* was deployed to serve the homonymous community operated by the project, *TCom* was deployed to serve project members forming the Technical Committee, *ScalableDataMining* was deployed to offer as a gCube App an environment for data mining. All VREs have been substantially enriched with new generic services, and the number of users continues to grow. In average, it has been observed an increase of +34% (147 at Jul '12 vs 197 at Jun '13) in the number of users served by FARM VREs and of +114% (112 at Jul '12 vs 240 at Jun '13) in the number of users served by gCubeApps VREs.

During the reporting period, examples of software artefacts that were newly released, or substantially altered following recommendations resulting from EA-CoP validation include the Workspace, Species Products Discovery portlet, GeoExplorer/GISViewer, the Species View portlet, Occurrence Management, Report Manager, Template Manager, and Social tools.

Besides portlets and VREs, 2 mobile applications interfacing with the services of the iMarine infrastructure have been developed. One is oriented to make available the “search” facility on a mobile phone. The other, AppliFish, combines information from several data-owners in fact-sheets of species important to marine fisheries and aquaculture, or high-profile species. The app combines factsheets with dynamically downloaded information such as species maps.

The report describes the set of activities performed to provide the target Community of Practice with a modified set of Virtual Research Environments. During the reporting period, VRE's that were made

available to the EA-CoP included the iMarine Board VRE, and the FishFinderVRE. These, and all other VREs, benefitted from new and improved software artefacts mentioned above.

The deliverable captures the activities between September 2012 and June 2013. In order to do that, the deliverable builds upon D6.3 and updates it thus to describe the state of the art at June 2013. The activities are dynamic and reach 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

1	INTRODUCTION	10
2	VIRTUAL RESEARCH ENVIRONMENTS RESOURCES AND TOOLS	12
2.1	DATA RESOURCES	12
2.2	TOOLS.....	14
2.2.1	<i>The Workspace</i>	<i>14</i>
2.2.2	<i>The social portal and social data sharing</i>	<i>15</i>
2.2.3	<i>The GeoExplorer Portlet.....</i>	<i>16</i>
2.2.4	<i>The GISViewer Portlet</i>	<i>17</i>
2.2.5	<i>The Species Products Discovery Portlet.....</i>	<i>18</i>
2.2.6	<i>The Species View Portlet</i>	<i>19</i>
3	VIRTUAL RESEARCH ENVIRONMENTS DEPLOYMENT AND OPERATION.....	20
3.1	DESIGN, DEPLOYMENT AND OPERATION.....	20
3.1.1	<i>The AquaMaps VRE.....</i>	<i>21</i>
3.1.2	<i>The BiodiversityResearchEnvironment VRE</i>	<i>22</i>
3.1.3	<i>The DocumentWorkflow VRE.....</i>	<i>23</i>
3.1.4	<i>The EcologicalModelling VRE.....</i>	<i>23</i>
3.1.5	<i>The FCPPS – Fisheries Country Profiles Production System VRE.....</i>	<i>25</i>
3.1.6	<i>The FishFinderVRE.....</i>	<i>25</i>
3.1.7	<i>The Off-line FishFinderVRE.....</i>	<i>27</i>
3.1.8	<i>The ICIS – Integrated Capture Information System VRE</i>	<i>28</i>
3.1.9	<i>iMarine Board VRE.....</i>	<i>29</i>
3.1.10	<i>Scalable Data Mining VRE</i>	<i>31</i>
3.1.11	<i>The TimeSeries VRE.....</i>	<i>31</i>
3.1.12	<i>The VesselActivitiesAnalyser VRE.....</i>	<i>32</i>
3.1.13	<i>The VME-DB VRE.....</i>	<i>33</i>
3.1.14	<i>The VTI – Vessel Transmitted Information VRE.....</i>	<i>34</i>
3.1.15	<i>Mobile Applications</i>	<i>35</i>
4	CONCLUDING REMARKS	37

TABLE OF FIGURES

Figure 1. The Workspace portlet	15
Figure 2. Social tool.....	16
Figure 3. The GeoExplorer Portlet	17
Figure 4. The GISViewer Portlet.....	18
Figure 6. The Species View Portlet.....	19
Figure 7. The AquaMaps VRE Homepage.....	21
Figure 8. The BiodiversityResearchEnvironment VRE Homepage	22
Figure 9. The DocumentsWorkflow VRE Homepage.....	23
Figure 10. The EcologicalModelling VRE Homepage.....	24
Figure 11. The FCPPS VRE Homepage	25
Figure 12. The FishFinderVRE Homepage	26
Figure 13. The ICIS VRE Homepage.....	28
Figure 14. The iMarine Board VRE Homepage.....	30
Figure 15. The Scalable Data Mining VRE Homepage	31
Figure 16. The TimeSeries VRE Homepage	32
Figure 17. The VesselActivitiesAnalyser VRE Homepage	33
Figure 18. The VME-DB VRE Homepage	33
Figure 19. The VTI VRE Homepage.....	34

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. VRE’s are the key products to be delivered by the iMarine project to meet the needs of the iMarine Ecosystem Approach Community of Practice.

This deliverable – D6.4 ‘*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 (EA-CoP). It build upon D6.3 [3] released in the previous period and updates it to describe the state of art at June 2013.

All tools related to EA-CoP activities have been discussed by the iMarine Board. The WP3 Wiki page summarizes 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 described VREs were expected to evolve (and have) 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, GeoNetwork for geospatial data discovery and access, FIMES schema support for publication of species fact-sheets species information, Hadoop and WPS for data processing, and X-Search for semantic search.

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. Most of this integration is performed dynamically, with data only loaded on demand, and respecting the data providers’ access and sharing policies. These cover several domains, including:

- Biodiversity: OBIS: the Ocean Bio-geographic Information System, GBIF: the Global Biodiversity Information Facility¹, CoL: Catalogue of Life², WoRMS: World Register of Marine Species³, IRMNG, WoRDS;
- Statistical: FAO Code lists exposed through the FAO SDMX registry⁴, code lists exposed through any SDMX registry like IRD and Eurostat, and FAO capture data sets;
- Geospatial data exposed via OGC standards like Web Map Service for map images, and WFS for features for any compliant source the EA-CoP wishes to access.

The remainder of this report is organised as follows. Section 2 describes 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 realize Virtual Research Environments it is fundamental to equip these working environments with an effective mix of facilities and data. This section reports how (a) the data sources that have been integrated in the infrastructure can be used in a Virtual Research Environment (cf. Section 2.1), and (b) the set of tools that, in addition to the gCube technology [8][9][10][11], has been developed to provide Virtual Research Environment users with facilities supporting their tasks (cf. Section 2.2).

2.1 DATA RESOURCES

In many applications users have to be provided with data. The effectiveness of a Virtual Research Environment might be affected if the user must acquire the data 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 over time and the user should be provided with an up to date version or dynamically accessed data. Because of this, a number of facilities have been developed (namely in the context of the Data Management area [9]) 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. 2.2.3). Very often – actually whenever possible – such mediators are built by relying on standards and protocols for data access and discovery including OAI-PMH⁵, TAPIR⁶, SDMX⁷, and DarwinCore⁸.

By relying on the gCube technologies, the most important data sources needed to support the applications discussed in the context of the EA-CoP have been linked to the iMarine infrastructure and made available 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 sources 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 than 32 million records on species and 1,000+ datasets;
- *Catalogue of Life*¹¹: this data source offers an integrated checklist and a taxonomic hierarchy of more than 1.3 million species of animals, plants, fungi and micro-organisms;

⁵ www.openarchives.org/pmh/

⁶ wiki.tdwg.org/TAPIR/

⁷ sdmx.org

⁸ rs.tdwg.org/dwc/

⁹ www.gbif.org

¹⁰ www.iobis.org

¹¹ www.catalogueoflife.org

- *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 from 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.
- *FAO GeoNetwork*²²: This data source exposes spatial data maintained by FAO and its partners;
- *SDMX Codelists* either directly accessed from the FAO Registry²³, or manually uploaded through the facility developed in the context of ICIS;

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

²² <http://geonetwork.fao.org>

- World Ocean Atlas²⁴: This data source give access to a number of environmental variables. In particular, iMarine focuses on some indicators including Apparent Oxygen Utilisation, Dissolved Oxygen, Nitrate, Oxygen Saturation, Phosphate, Sea Water Salinity, Sea Water Temperature, and Silicate;
- myOceans²⁵: This data source give access to a number of environmental variables. In particular, iMarine focuses on some indicators including ice concentration, ice thickness, ice velocity, mass concentration of chlorophyll in sea water, meridional velocity, mole concentration of dissolved oxygen in sea water, mole concentration of nitrate in sea water, mole concentration of phosphate in sea water, mole concentration of phytoplankton expressed as carbon in sea water, net primary production of carbon, salinity, sea surface height, temperature, zonal velocity, wind speed, and wind stress;
- PenSoft Journals: This data source give access to a number of open-access journals. In particular, iMarine focuses on BioRisk, Comparative Cytogenetics, International Journal of Myriapodology, Journal of Hymenoptera Research, MycoKeys, Nature Conservation, NeoBiota, PhytoKeys, Subterranean Biology, and ZooKeys.
- Various Knowledge Base including FAO FLOD²⁶, DBPedia²⁷, FactForge²⁸ and YAGO2²⁹.

2.2 TOOLS

In addition to the rest of gCube technology [8][9][10][11] 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. 2.2.1) to specific environments supporting well defined tasks (e.g. species distribution maps discovery and visualisation – cf. Sec. 2.2.6). In the remainder of this section, the major tools that have either be developed or reinforced between September 2012 and June 2013 are reported.

2.2.1 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.

²³ <http://www.fao.org/figis/sdmx/>

²⁴ http://www.nodc.noaa.gov/OC5/WOA09/pr_woa09.html

²⁵ <http://www.myocean.eu/>

²⁶ <http://www.fao.org/figis/flod/>

²⁷ <http://dbpedia.org/About>

²⁸ <http://factforge.net/>

²⁹ <http://www.mpi-inf.mpg.de/yago-naga/yago/>

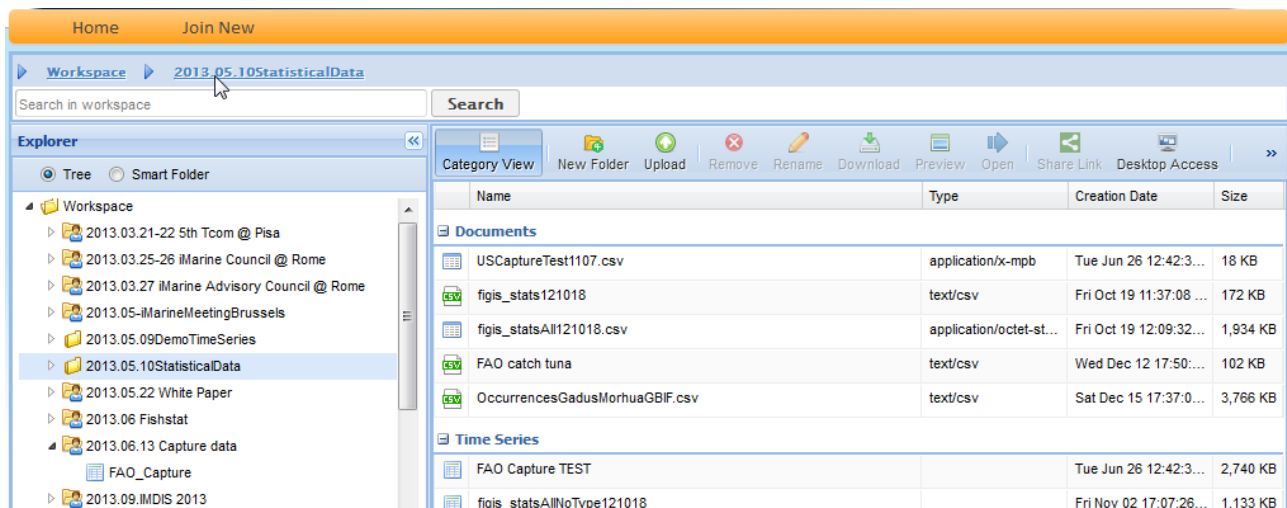


Figure 1. The Workspace portlet

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 WebDAV³⁰ based mechanism making it possible to integrate the entire workspace in the user file system.

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.

2.2.2 THE SOCIAL PORTAL AND SOCIAL DATA SHARING

The *Social Portal* aims to provide users with all facilities required by and expected in a modern multi-user environment to provide an immersive collaborative toolset for sharing, communicating and interacting. In addition, the users now are provided with social data sharing through messages and mail directly from the main bar from any application. Such tools allow interaction with other users according to any specific need. Every user of Virtual Research Environment is provided with these tools that resemble widely used systems.

³⁰ <http://www.webdav.org/>

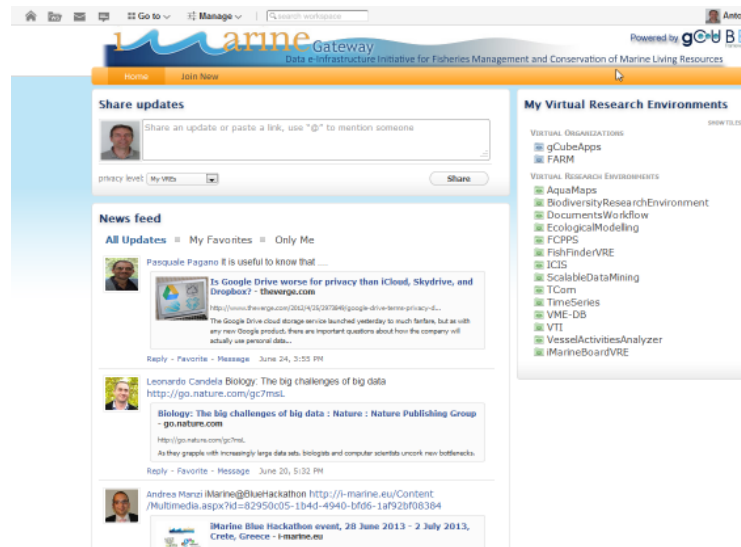


Figure 2. Social tool

The added value of this interactive toolset is that it brings cohesion to both the infrastructure as well as the EA-CoP, who encounter a very similar environment in all produced VREs. This builds trust and confidence, and users quickly familiarize with the item types in a seamless way. The tools range from messaging through email and a posting and commenting tool, to object sharing and publication facilities. Also the upload and download facility, the publication facility and the remote folder options all aim to lower the threshold that users have to engage with the system.

The social tool has a key function to develop an EA-CoP. In the context of the Virtual Research Environments, it is the glue that brings users together, enables their interaction, and allows them to ‘take the message home’.

2.2.3 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).

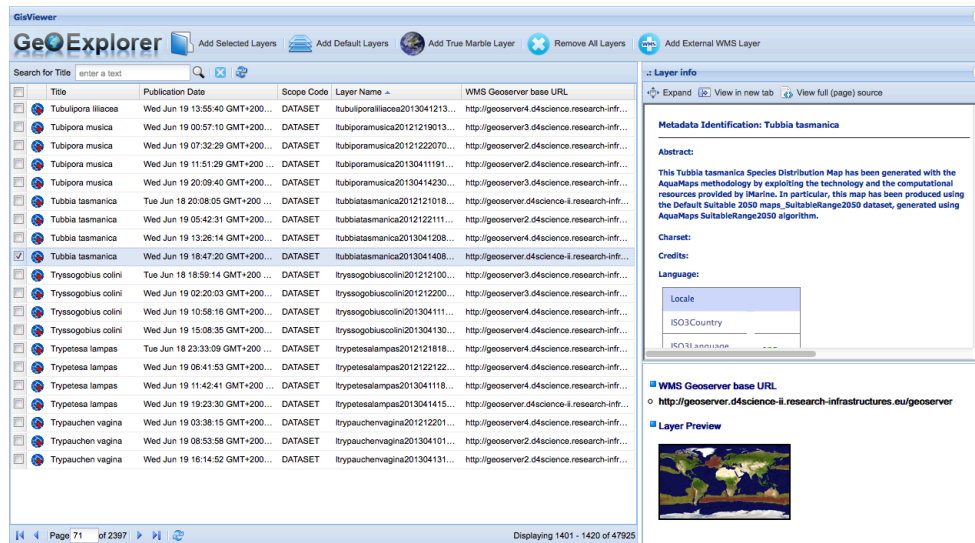


Figure 3. The GeoExplorer Portlet

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. 2.2.4). 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.

2.2.4 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.

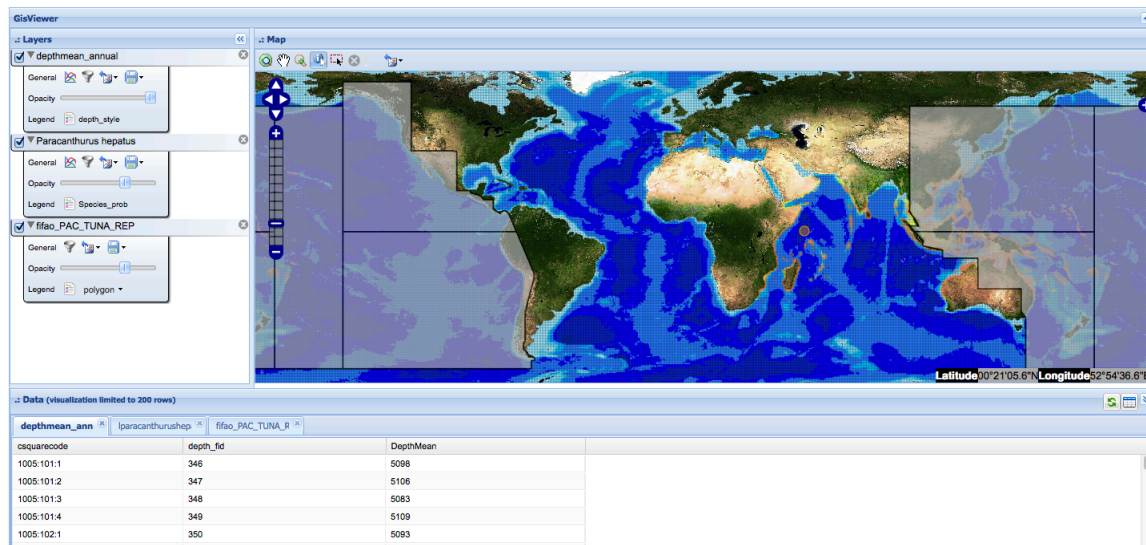


Figure 4. The GISViewer Portlet

2.2.5 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.

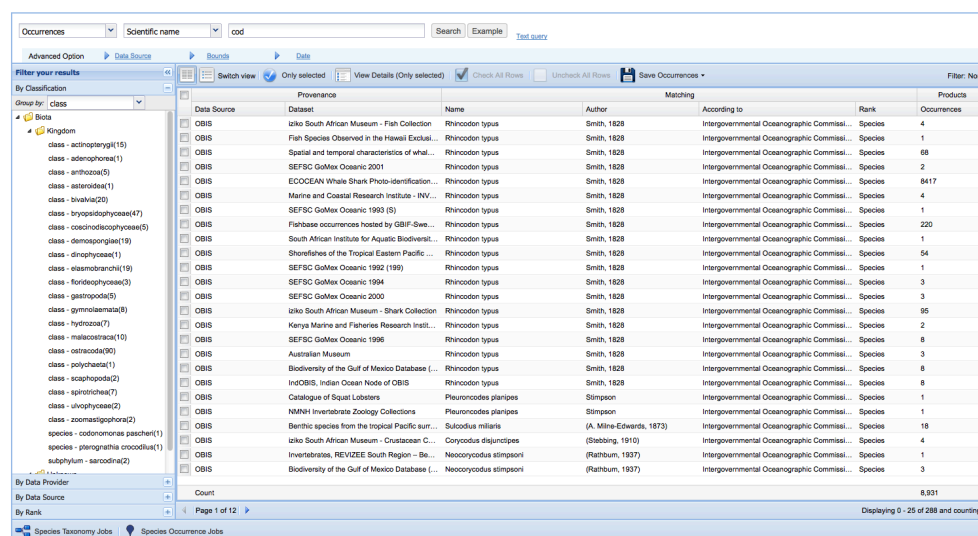


Figure 5. The Species Product Discovery Portlet

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).

2.2.6 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.

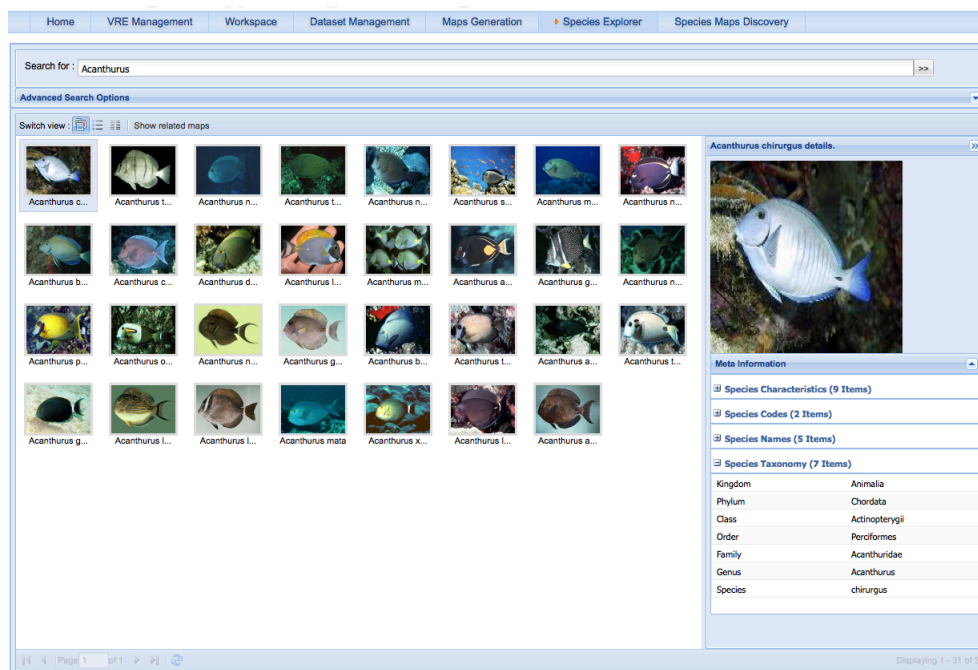


Figure 6. The Species View Portlet

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.

In addition to VRE's, the EA-CoP is served with 2 mobile applications (cf. Sec. 3.1.15). One app demonstrates search capability, the other, AppliFish, serves a mash-up of content related to important marine species.

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 June 2013. 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	54
BiodiversityResearchEnvironment	gCubeApps	Jul 2012	41
DocumentWorkflow	gCubeApps	Nov 2011	32
EcologicalModelling	gCubeApps	Nov 2011	46
FCPPS – Fisheries Country Profiles Production System	FARM	Nov 2011	38
FishFinderVRE	FARM	March 2013	10
ICIS – Integrated Capture Information System	FARM	Nov 2011	48
iMarineBoardVRE	FARM	Feb 2013	16
ScalableDataMining	FARM	Oct 2012	27
TCom	gCubeApps	Apr 2013	36
TimeSeries	gCubeApps	Nov 2011	36
VesselActivitiesAnalyser	gCubeApps	Nov 2011	33
VME-DB	FARM	Jul 2012	12
VTI – Vessel Transmitted Information	FARM	Nov 2011	26

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

All the VREs are provided with:

- A *shared 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. 2.2.1);
- A *VRE Management facility* 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;
- A *social networking facility* to enable users to use the common facilities promoted by social networks – e.g., posting news, commenting on posted news – yet adapted to settings of the iMarine environment (cf. Sec. 2.2.2). News can be posted by users as well as applications;
- A *notification facility* to alert users on an as-it-happens basis. These notifications offer a sense of anticipation and create a productivity boost. Users receive an alert (through a priori selected channels, e.g., email, web portal, twitter) notifying them when something of interest has happened in their VRE(s);
- A *messaging facility* to provide users with a common email environment as-a-Service. The distinguishing feature is represented by its integration with the rest, e.g., it is possible to send as attachment any information object residing in the workspace (although “big” and “complex”) without consuming bandwidth.

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.

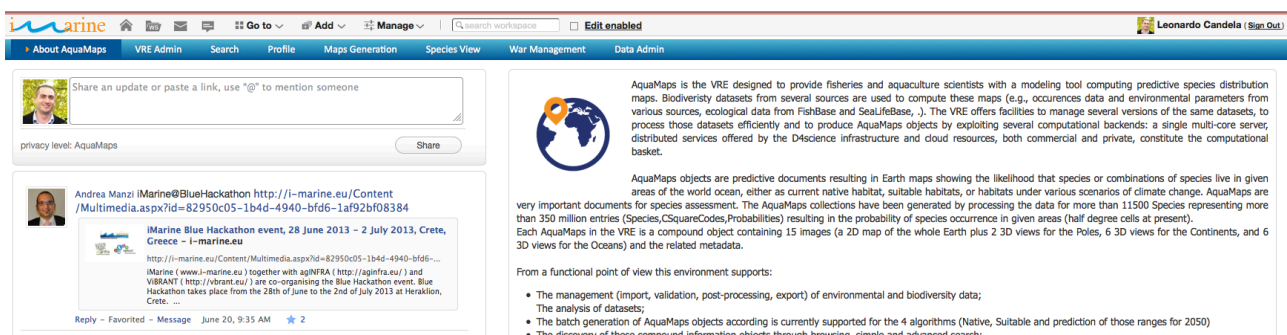


Figure 7. The AquaMaps VRE Homepage

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);

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.

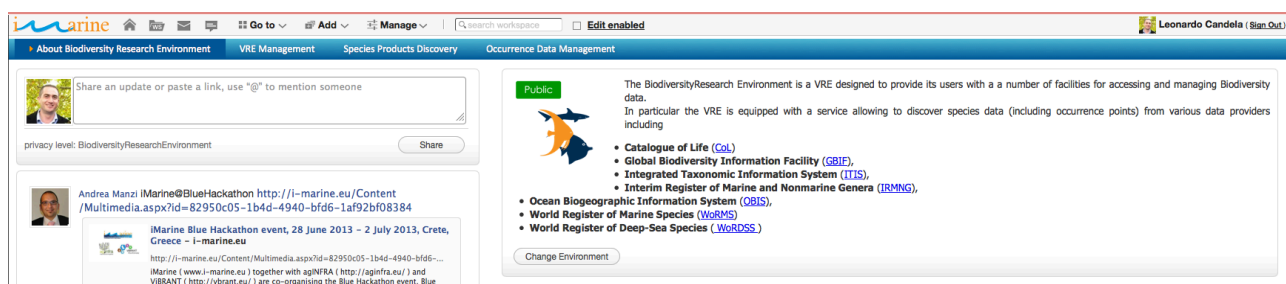


Figure 8. The BiodiversityResearchEnvironment VRE Homepage

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. 2.2.5). Once discovered, objects can be stored in the workspace for future uses;
- Occurrence Data Management: to enable users to perform data processing tasks on Occurrence Data. In particular, tasks like duplicates detection and removal, datasets union, intersection, and subtraction are supported;

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 world's 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 exploits the facilities offered by the gCube Business Documents Workflow Management Suite enabling the production of reports that require a collaborative activity of several actors.

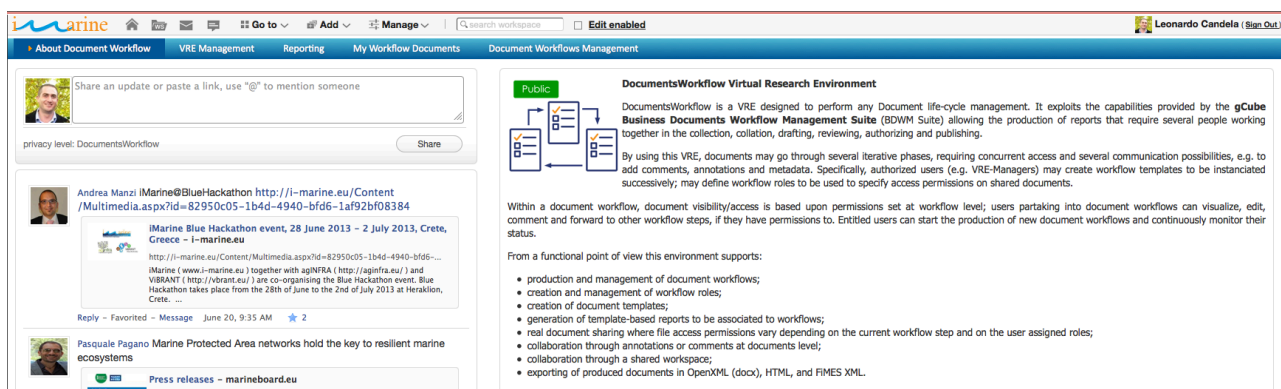


Figure 9. The DocumentsWorkflow VRE Homepage

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;

3.1.4 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.

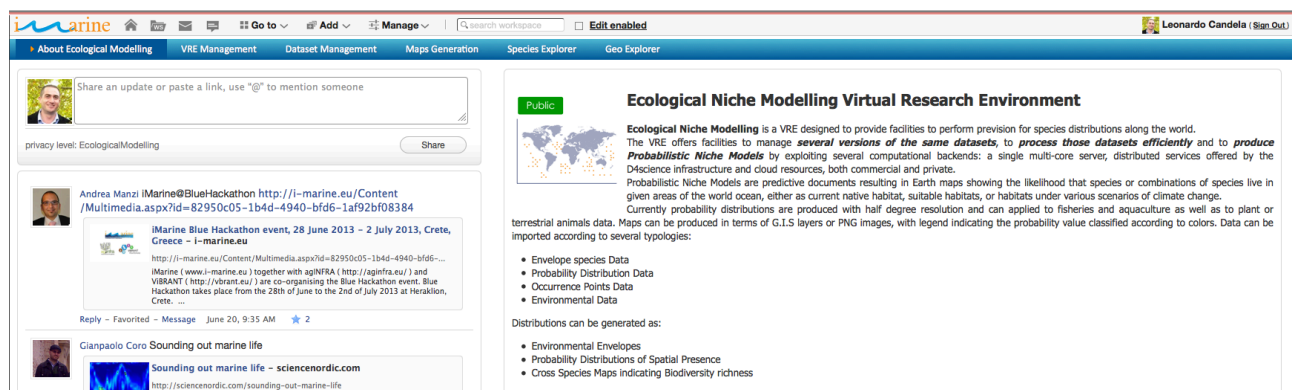


Figure 10. The EcologicalModelling VRE Homepage

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. 2.2.6);
- **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. 2.2.3) 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;

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).

3.1.5 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.

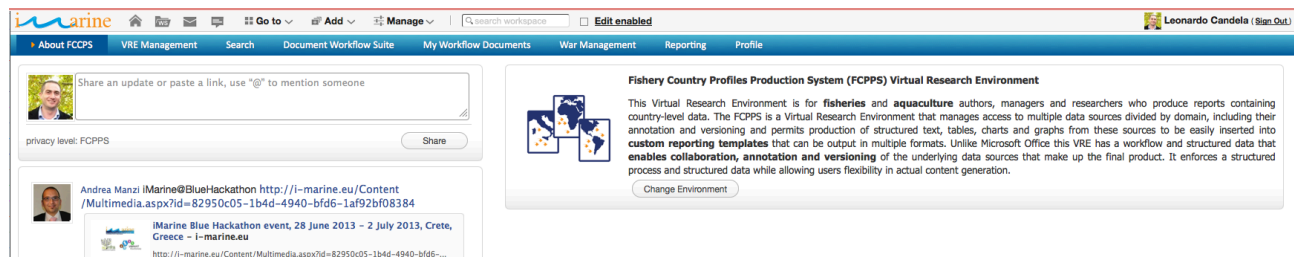


Figure 11. The FCPPS VRE Homepage

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;

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).

3.1.6 THE FISHFINDERVRE

The FishFinderVRE is established to elaborate Species Fact Sheets, fill / view their metadata, and select data for download and-or display in the stand-alone version of this VRE (see 3.1.7). The VRE extends the previously released FCPPS VRE.

The explicit purpose of this VRE is to enable some 50 authors to prepare hundreds of species fact sheets. The tools will be operated by one VRE manager in FAO, with globally distributed authors. The FishFinderVRE allows defining, pre-filling, and elaborating a fact sheet using a template. Fact-sheet

compilation will be marshaled by a work flow, with one VRE manager communicating with dozens of authors globally distributed that can access their work-flow to find the tasks assigned to them.

A fact-sheet will be given a scientific species name from a controlled list. The fact-sheet structure will rely on a template structure that is identical for all individual fact-sheet / reports. The user fills this template, for taxonomic names, biological characteristics, etc. Data can be either visible data or invisible metadata. The users will be assigned their tasks on particular fact-sheets through the work-flow features of the reporting environment.

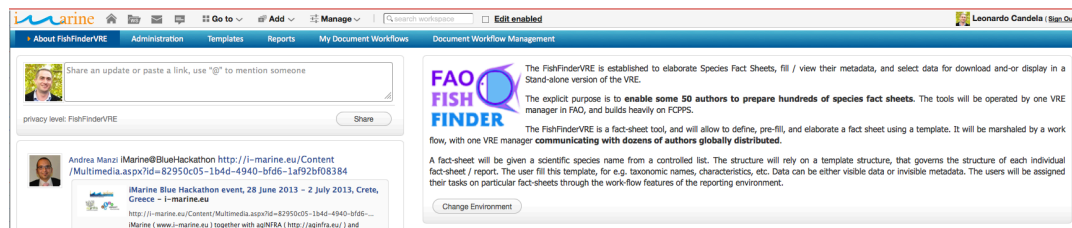


Figure 12. The FishFinderVRE Homepage

The specific facilities this VRE offers are:

- Reporting facilities: to enable users to produce reports as 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;

The VRE is an adaptation of the FCPPS VRE to serve the specific needs of practitioners in the field of species fact sheets production. Biologists are expected to produce fact sheets on hundreds of species, with the VRE offering access to background information on such species, such as through the SPD and AquaMaps. The entire process is governed by a data manager in charge of the documents work-flow.

To enable the participation of scientist devoid of high capacity internet connections, an off-line version was produced with the exact same capacities as the on-line version, except for those services that require such connectivity (cf. Sec. 3.1.7). To maintain the document work-flow, the off-line version is equipped with tailor made communication protocols via either e-mail, or in the absence of any connectivity, by file exchange.

The produced and validated reports will be uploaded to the FAO infrastructure, and the conversion to FiMES compliant XML has already been prepared.

3.1.7 THE OFF-LINE FISHFINDERVRE

The off-line version of FishFinderVRE is established to elaborate Species Fact Sheets. User can fill / view their reports and associated metadata, and exchange reports with the master VRE in a stand-alone version of the VRE. It is intended also as a pilot for the development of data-collection components that are loosely coupled to the infrastructure. In the future these data collection components can evolve as stand-alone VRE's or any App, website, or any service capable of synchronizing data over the established protocols.

The explicit purpose of this VRE is to enable off-line species fact-sheet editors to participate to collaborative VRE processes integrated in the iMarine infrastructure. It does so by capturing locally the reports produced by authors that can be periodically interchanged with the e-infrastructure by manually uploading or sending the reports to this infrastructure. Most facilities are identical to the on-line version, with the exception of those services relying on the infrastructure. The VRE enables only fact-sheet editing, and will communicate with the master VRE FAO only.

The Off-line FishFinderVRE is a pure fact-sheet editing tool, and thus a data collection component. It allows defining, pre-filling, and elaborating a fact sheet using a template. It will be marshaled by a work flow, with one VRE manager communicating with dozens of authors globally distributed.

A fact-sheet will be given a scientific species name from a controlled list. The structure will rely on a template structure that governs the structure of each individual fact-sheet / report. The user fills this template, for e.g. taxonomic names, characteristics, etc. Data can be either visible data or invisible metadata. The users will be assigned their tasks on particular fact-sheets through the work-flow features of the reporting environment.

The specific 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;
- Download and Upload facilities: to enable users to receive/send specific instances of species fact-sheets to and from the VRE Manager.

The VRE is an adaptation of the on-line to serve the needs of practitioners that do not have reliable internet connections. It's functionality is a sub-set of the on-line version; it enables biologists to produce fact sheets on species. The entire process is governed by a data manager in charge of the documents work-flow.

The off-line version misses the services that require connectivity. To maintain the document work-flow, the off-line version is equipped with tailor made communication protocols via either e-mail, or in the absence of any connectivity, by file exchange.

3.1.8 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. During the reporting period, several activities to further the development and use of ICIS were performed;

Discussion of WP3 VRE-validation results with WP6 and other WP representatives. Some of the comments received required the intervention of developers to enrich the VRE with facilities to support the capturing of data-structures, enrich the curation facilities, and improve integration with other gCube leveraged components, such as for data mining and data processing. A major addition is the capacity to capitalize on SDMX code-lists, in fact, a complete code list management solution is being implemented, and the integration of an SDMX registry.

The feed-back from the EA-CoP was also instrumental to propose several completely new VRE's (listed below) that stand evidence for the interest in the EA-CoP of data driven integrated solutions.

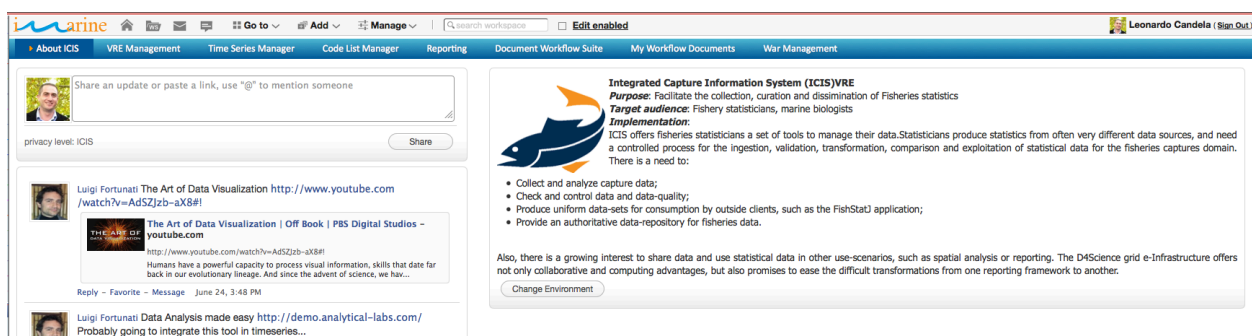


Figure 13. The ICIS VRE Homepage

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 supports 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 R³¹ 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 SDMX³² 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;
- **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;

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).

3.1.9 IMARINE BOARD VRE

The iMarineBoardVRE is designed to provide the members of the iMarine Board with collaboration tools, project information and publications, and as a demonstrator of infrastructure facilities.

The iMarine Board Members are involved in the decision making process for the implementation of a robust digital framework to support the Ecosystem Approach to Fisheries and Natural resources Management. The iMarine e-infrastructure offer a several high-capacity facilities for accessing, sharing, managing and analyzing data, deploying entire applications and services, and this VRE is intended to assist with real-life facilities this decision making process.

It is a responsibility of the Ecosystem Approach Communities of Practice (EA-CoP) to help develop a sustainable governance model, and the guidelines and policies regulating the management and operation of the e-infrastructure. This VRE therefore includes those services that put into effect iMarine governance models and policies, such as (i) the collaboration suite including a shared workspace and messaging system, (ii) services for accessing biodiversity data from several major databases, and (iii) services for managing tabular data (e.g. catch statistics) and code lists.

³¹ <http://www.r-project.org/>

³² sdmx.org

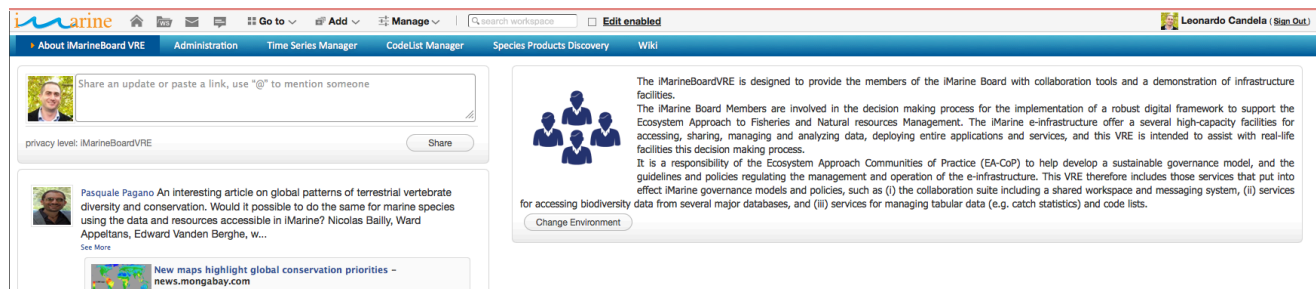


Figure 14. The iMarine Board VRE Homepage

The main facilities this VRE offers are:

- Access through the workspace of all iMarine Board documentation. Previously, these were fragmented in several locations. Through the VRE, communication with and within the Board is facilitated by mail, messaging, sharing, and learning facilities.
- Time Series Management: to enable users to import, curate and manage time series. This is a comprehensive and feature-rich environment that supports data managers during the whole life cycle of data management from capture to publishing and visualisation. It enables 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 R³³ 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 enables users to import CSV files or existing code lists from SDMX³⁴ 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;
- A wiki for elaborating Board specific documentation. This wiki allows the board to develop a shared position transparent to all members, while ensuring confidentiality and security. For instance, reviews of iMarine policy documents that are initially for circulation in the Board, or proposals for collaboration at board level are proposed as wiki pages.

³³ <http://www.r-project.org/>

³⁴ sdmx.org

3.1.10 SCALABLE DATA MINING VRE

Scalable Data Mining is a VRE designed to apply Data Mining techniques to biological data. The algorithms are executed in a distributed fashion on the e-Infrastructure nodes or on local multi-core machines. Scalability is thus meant as distributed data processing but even as services dynamically provided to the users. The system is scalable in the number of users and in the size of the data to process. Statistical data processing can be applied to perform Niche Modelling or Ecological Modelling experiments. Other applications can use general purpose techniques like Bayesian models. Time series of observations can be managed as well, in order to classify trends, catch anomaly patterns and perform simulations. The idea under the distributed computation for data mining techniques is to overcome common limitations that can happen when using statistical algorithms.

The logic behind this VRE is realized in several work-packages, and WP6 deploys these in an organized fashion to the users.

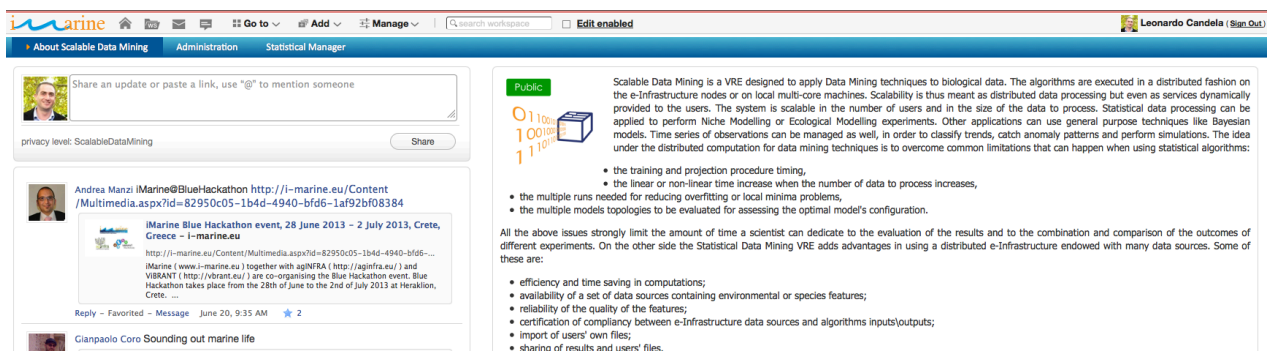


Figure 15. The Scalable Data Mining VRE Homepage

The main facilities this VRE offers are:

- **Statistical manager:** To enable users to select their data using a template for pre-filtering, the required algorithm that can be applied to the selected data, the resources required to perform the computation if foreseen by the algorithm, and a process monitor that allows the user to track the process, and intervene when necessary, and manage the output. This three-phased approach to computational requirements is extremely versatile, and can be extended by users wishing to bring their own data and algorithms.

3.1.11 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;

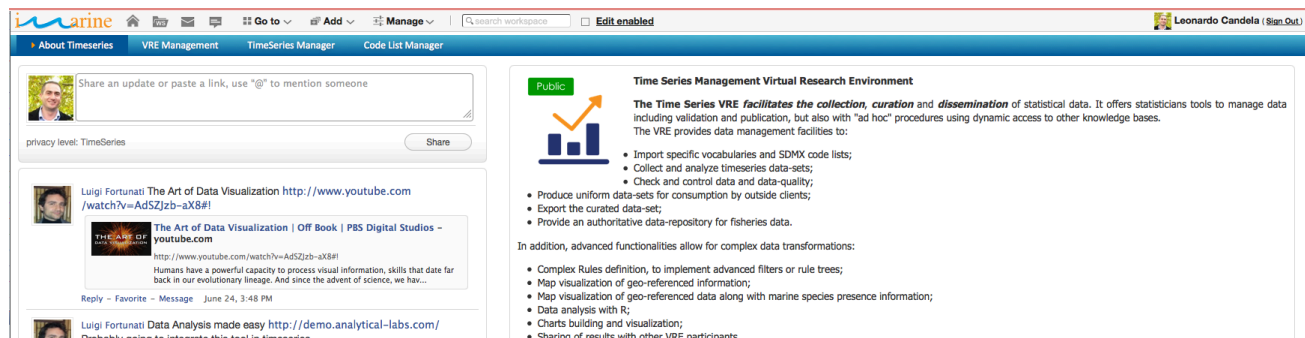


Figure 16. The TimeSeries VRE Homepage

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 R³⁵ 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 SDMX³⁶ 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;

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).

3.1.12 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.

³⁵ <http://www.r-project.org/>

³⁶ sdmx.org

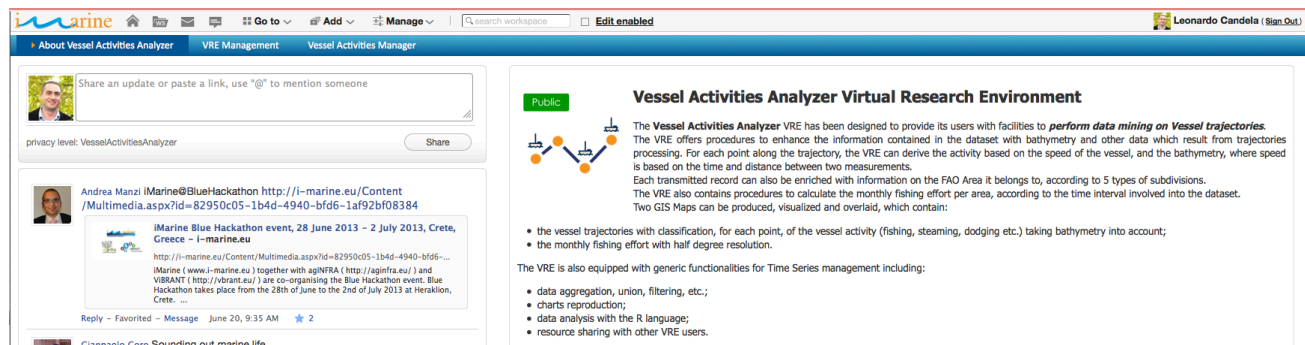


Figure 17. The VesselActivitiesAnalyzer VRE Homepage

The specific 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 R³⁷ environment.

The VRE offers a base for extension for EA-CoP members in need of a geospatial explicit analysis of their often confidential data. At the time of writing this document, several initiatives were aware of the potential, but with limiting data access and sharing policies could not yet contribute data and algorithms for further development, implementation and use of the facility.

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).

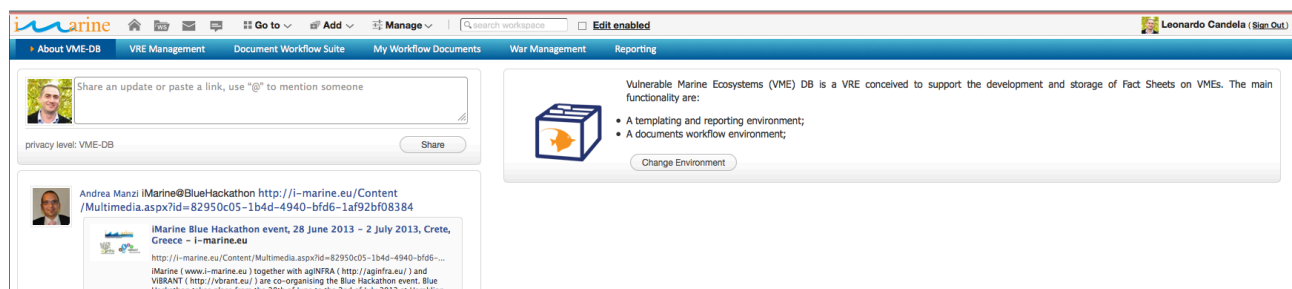


Figure 18. The VME-DB VRE Homepage

³⁷ <http://www.r-project.org/>

The specific 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.

This VRE is much in development in order to capture the evolving requirements from the EA-CoP behind. The VME project in FAO expects a VRE that can produce web-content through an editing work-flow, yet itself is still (June 2013) in the process of defining the template for VME's, the document workflow, and the integration with the FAO web-infrastructure. A working version is expected December 2013.

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.

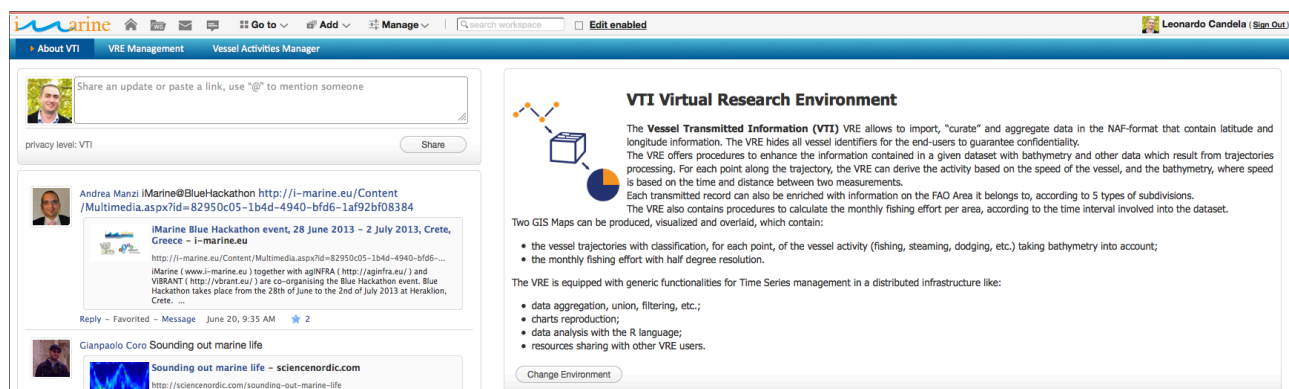


Figure 19. The VTI VRE Homepage

The specific 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 R³⁸ environment.

The deployed version of the VRE has been validated with the EA-CoP that required such functionality. This has resulted in a request for further functionality that is still being developed. These include environmental enrichment of vessel positions in order to better understand the vessel movement in adverse weather (high winds) or environmental conditions (currents), the transformation of VMS data into structured and aggregated data formats to comply with the DG MARE FLUX framework, and more advanced spatial analysis of vessel tracks, that might be offered through the statistical service.

3.1.15 MOBILE APPLICATIONS

Two mobile applications have been developed to make it possible for users to exploit the iMarine services via their mobile phones.

The **AppliFish** Mobile app for Android and iOS combines species fact sheets with information from other resources. Over 4,300 downloads are reported since its release in January. The more than 550 fact-sheets inform on species and their distribution, common names, maps and images. They can be used by consumers to better understand the status of a species and make a conscious choice when choosing fish.

The specific facilities this App offers are:

- Species Fact Sheets: AppliFish is built from five major data components: FAO Species Fact-sheets, OBIS and WoRMS common names, AquaMaps, FishBase / SeaLifeBase, and FAO Capture Statistics. This data is combined in an input process for acquiring data, and brought together in a database which is then incorporated in the mobile application framework;
- Map Download: To reduce the size of the App, maps are dynamically downloaded from the e-Infrastructure.

The **gCube Mobile Search App** is an application that brings the federated search facilities of gCube System, that empowers iMarine Project, into the reach of the mobile user. Through this application, scientists can use their familiar mobile device to easily locate data coming from sources registered in D4Science infrastructure.

For the anonymous user, the mobile application offers services over the publicly available subset of the data sources registered in the D4Science infrastructure:

- Free text search: access to search in a manner similar to internet search services, yet focused on the aforementioned data sources.
- Results formulation: the metadata of elements located via search are presented to the end-user's mobile device.

³⁸ <http://www.r-project.org/>

- Content retrieval: the content (images, documents) that corresponds to the aforementioned results is delivered to end-user's mobile device for presentation.

iMarine registered users can access more features. An advanced user interface is available to offer the ability to sign into the infrastructure and access the advanced search functionality, i.e. filtering data by the content of specific fields.

Currently gCube Mobile Search is available for the Android Platform. An iOS release is planned.

4 CONCLUDING REMARKS

Virtual Research Environments are among the key products to be delivered by the iMarine project to meet the needs of the iMarine Ecosystem Approach Community of Practice. They 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.

This deliverable detailed the deployed Virtual Research Environments in terms of community tools integrated, resources involved, and user exploitation. It described 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. Moreover, it listed the entire offering of iMarine in terms of Virtual Research Environments as of June 2013. Overall 14 VREs have been deployed and operated to serve a total of 455 users. The infrastructure is also serving the mobile apps community, e.g. as of today there are more than 4,300 instances of AppliFish.

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] L. Candela, A. Ellenbroek, P. Pagano. Virtual Research Environments Activity Report. D6.3 iMarine Project Deliverable. July 2012
- [4] A. Manzi, L. Candela. iMarine Data e-Infrastructure Operation Report. iMarine D5.2 Project Deliverable. September 2012
- [5] A. Manzi, L. Candela. iMarine Data e-Infrastructure Operation Report. iMarine D5.3 Project Deliverable. June 2013
- [6] Y. Laurent. ICIS Assessment for Tuna Atlas Use-Case. iMarine internal document. July 2012
- [7] A. Gentile. FCPPS Assessment for VME-DB Use-Case. iMarine internal document. August 2012
- [8] M. Simi. iMarine Data Infrastructure Enabling Software. iMarine D8.2 Project Deliverable. August 2012
- [9] G. Coro, F. De Faveri, L. Lelii. iMarine Data Management Software. iMarine D9.1 Project Deliverable. June 2012
- [10] G. Farantatos. iMarine Data Consumption Software. iMarine D10.3 Project Deliverable. June 2012
- [11] M. Simi, R. Tsantouli. Application Programming Interface Software. iMarine D11.2 Project Deliverable. June 2012