

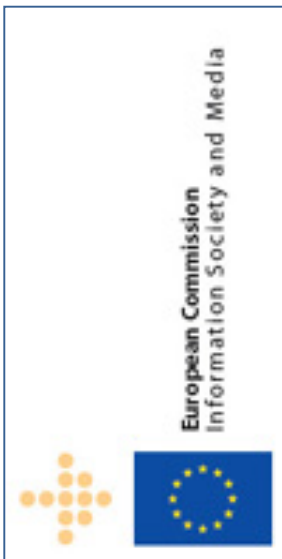


<i>Project Acronym</i>	BlueBRIDGE
<i>Project Title</i>	<i>Building Research environments for fostering Innovation, Decision making, Governance and Education to support Blue growth</i>
<i>Project Number</i>	675680
<i>Deliverable Title</i>	<i>BlueBRIDGE Data Management Plan: Preliminary version</i>
<i>Deliverable No.</i>	<i>D2.1</i>
<i>Delivery Date</i>	<i>February 2016</i>
<i>Authors</i>	<i>Leonardo Candela, Donatella Castelli, Pasquale Pagano, Franco Zoppi</i>

DOCUMENT INFORMATION

PROJECT	
Project Acronym	BlueBRIDGE
Project Title	Building Research environments for fostering Innovation, Decision making, Governance and Education to support Blue growth
Project Start	1st September 2015
Project Duration	30 months
Funding	FP7-INFRASTRUCTURES-2011-2
Grant Agreement No.	675680
DOCUMENT	
Deliverable No.	D2.1
Deliverable Title	BlueBRIDGE Data Management Plan: Preliminary version
Contractual Delivery Date	February 2016
Actual Delivery Date	May 2016
Author(s)	Leonardo Candela (CNR), Donatella Castelli (CNR), Pasquale Pagano (CNR), Franco Zoppi (CNR)
Editor(s)	Leonardo Candela (CNR)
Reviewer(s)	Sara Garavelli (TRUST-IT)
Contributor(s)	Gerasimos Antzoulatos (I2S), Gianpaolo Coro (CNR), Anton Ellenbroek (FAO), Gabriele Giammatteo (ENG), George Kakalettris (UOA)
Work Package No.	WP2
Work Package Title	Project Governance, Exploitation and Sustainability
Work Package Leader	CNR
Work Package Participants	ERCIM, ENG, UOA, FAO, ICES, IRD, FORTH, TRUST-IT, I2S, CITE, CLS, GRID Arendal, PMBret
Estimated Person Months	3
Distribution	Public
Nature	Report
Version / Revision	1.0
Draft / Final	Final
Total No. Pages (including cover)	54
Keywords	Datasets; Management; Data Storage; Data Repository; Data dissemination; Data preservation;

DISCLAIMER



BlueBRIDGE (675680) is a Research and Innovation Action (RIA) co-funded by the European Commission under the Horizon 2020 research and innovation programme

The goal of BlueBRIDGE, *Building Research environments for fostering Innovation, Decision making, Governance and Education to support Blue growth*, is to support capacity building in interdisciplinary research communities actively involved in increasing the scientific knowledge of the marine environment, its living resources, and its economy with the aim of providing a better ground for informed advice to competent authorities and to enlarge the spectrum of growth opportunities as addressed by the Blue Growth societal challenge.

This document contains information on BlueBRIDGE core activities, findings and outcomes and it may also contain contributions from distinguished experts who contribute as BlueBRIDGE 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 BlueBRIDGE 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 BlueBRIDGE Consortium 2015. See <http://www.bluebridge-vres.eu> 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 BlueBRIDGE Consortium 2015."

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

GLOSSARY

ABBREVIATION	DEFINITION
CC	Creative Commons
CC-BY	Creative Commons Attribution Licence
CC-BY-NC	Creative Commons Attribution-Non Commercial Licence
Compound information object	Is an information object formed by either a dataset or of a set of related compound objects semantically forming a single entity
CSV	Comma Separated Values
CSW	OGC Catalogue Service
DMP	Data Management Plan
FCR	Feed Conversion Ratio
ISO 19115	Geographic information -- Metadata
ISO 19139	Geographic information -- Metadata -- XML schema implementation
NetCDF	Network Common Data Form
QAO	Quality Assurance Office
SFR	Specific Feeding Ratio
SGR	Specific Growth Ratio
VRE	Virtual Research Environment
WCS	OGC Web Coverage Service
WFS	OGC Web Feature Service
WMS	OGC Web Map Service
XDR	External Data Representation

TABLE OF CONTENT

DOCUMENT INFORMATION	2
DISCLAIMER	3
GLOSSARY	4
TABLE OF CONTENT	5
DELIVERABLE SUMMARY	7
EXECUTIVE SUMMARY	8
1 Introduction	9
2 Concepts	11
2.1 Per Dataset Information	11
2.2 BlueBRIDGE Data Management Practices and Approaches	12
2.2.1 Data Storage (and Repositories)	12
2.2.2 Data Dissemination (and Catalogues)	13
2.2.3 Data Preservation	13
3 Datasets	14
3.1 Basic Datasets	14
3.1.1 AquaMaps data (GIS maps)	14
3.1.2 ASFIS (Copy)	15
3.1.3 Average annual population to calculate regional GDP data (thousand persons) by NUTS 3 regions	16
3.1.4 Business demography and high growth enterprise by NACE Rev. 2 and NUTS 3 regions	17
3.1.5 Copernicus Marine Environment Monitoring Service Data	17
3.1.6 Derived geospatial datasets by combining Eurostat datasets and the NUTS classification	18
3.1.7 Employment (thousand persons) by NUTS 3 regions	19
3.1.8 Employment (thousand hours worked) by NUTS 2 regions	20
3.1.9 Eurostat Datasets	21
3.1.10 FAO Global Fisheries Production Statistics	22
3.1.11 FCR_Table	23
3.1.12 Financial dataset	24
3.1.13 FIRMS Factsheets on stocks, fisheries and marine resources	26
3.1.14 Fisheries commodities and trade statistics	27
3.1.15 GBIF and OBIS taxa and Observations	27
3.1.16 Global number of Fisheries	28
3.1.17 Global_Seawater_Oxygen18	29
3.1.18 Global Tuna Nominal Catches	30
3.1.19 Gross domestic product (GDP) at current market prices by NUTS 2 regions	31
3.1.20 Gross domestic product (GDP) at current market prices by NUTS 3 regions	32
3.1.21 Investment analysis geospatial search / optimization datasets (maps)	33

3.1.22	Land cover in Greece	34
3.1.23	Mortality_Table	34
3.1.24	(MT) - Turnover and Employment by District	35
3.1.25	NASO Farm locations; National Aquaculture Sector Overviews	36
3.1.26	NUTS 2013 classification	37
3.1.27	Planning Service (NI) - Development Plan - Landuse Zones & Points in UK	38
3.1.28	POSEIDON_DATASETS	39
3.1.29	SamplingToSampling	40
3.1.30	SFR_Table	42
3.1.31	SGR_Table.....	43
3.2	Datasets Resulting from Activities performed within a VRE	44
3.2.1	VRE Content.....	44
3.2.2	Blue Skills VRE Content.....	45
3.3	Project Outputs	46
3.3.1	Deliverables	46
3.3.2	Dissemination and Training Material	47
3.3.3	Scientific Papers	48
3.4	Infrastructure Operation Datasets	48
3.4.1	BlueBRIDGE Data Catalogue	49
3.4.2	BlueBRIDGE Geospatial Data Catalogue	49
3.5	Software	50
3.5.1	gCube Software Artefacts.....	50
3.5.2	gCube Source Code.....	51
4	Conclusion and Next Steps	53
	REFERENCES	54

DELIVERABLE SUMMARY

This deliverable documents the BlueBRIDGE data management strategy. This strategy will be developed by producing three successive versions of this deliverable: a preliminary version at M6, an intermediate version at M18, a final version at M27. This preliminary version contains an initial description of the data collected, processed or generated by the project, the standards and metadata used for them, and an initial plan on how sharing, archiving and preservation of this data will be guaranteed. Roughly 40 datasets have been identified and described. Some of these datasets are actually abstract / meta datasets, i.e. they represent a class of dataset that will manifest in a number of concrete datasets sharing certain needs and characteristics.

EXECUTIVE SUMMARY

Data is a key asset for the economy and our society similar to the classic categories of human and financial resources. In particular, making data publicly available will contribute to offer a series of opportunities, e.g., creating jobs, spurring growth, boosting research productivity and creativity, helping people, engaging citizens [3].

The European Commission is promoting data availability through a series of actions including a specific policy (Art. 29.3 “Open access to research data”) requesting funded projects to “(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate – free of charge for any user – the following: (i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible; (ii) other data, including associated metadata, as specified and within the deadlines laid down in the ‘data management plan’; (b) provide information – via the repository – about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and – where possible – provide the tools and instruments themselves)”. Such a policy goes in tandem with a new element in Horizon 2020 which is requesting projects to develop and use Data Management Plans.

This Data Management Plan outlines how the data “produced” (either generated or collected) during the BlueBRIDGE project are planned to be managed (during the project and after the project completion). In particular, it describes the practices characterising research data handling during and after the project, what data will be collected, processed or generated, what methodology and standards will be applied, whether and how data will be shared / made open access, and how data will be curated and preserved.

In the context of the BlueBRIDGE project a rich array of datasets is expected to be managed including geospatial data, species data, tabular data, software. Because of this heterogeneity BlueBRIDGE is requested to deal with many standards and formats. Methodologies for data management are diverse yet built by relying on a shared data infrastructure (D4Science.org).

1 INTRODUCTION

A Data Management Plan (DMP) is a document outlining how the data “produced” (generated / collected) during a research project are planned to be handled (during the project and after the project completion).

In the Horizon 2020 guidelines it was clarified that DMPs produced by each project are expected to detail “*what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved*”. In addition to that, guidelines recommend to address a number of points (e.g. dissemination, access and preservation policies) on a **per dataset** basis reflecting the current status of maturity within the consortium about the data that will be “produced”. No guidelines on how to “identify” a dataset is given.

This deliverable is intended to be the first version of the BlueBRIDGE Data Management Plan (the second one, D2.2, is expected at Month 18 and the third one, D2.3, at Month 27). It contains an initial description of the datasets collected, processed or generated by the project and an initial plan on how sharing, archiving and preservation of these datasets will be guaranteed.

For the sake of this deliverable the following definition of dataset is used:

A dataset is any set of data (no matter how many files it materialises) that is worth to be considered as a unit for data management activities [1][6]

Examples of possible datasets are the following:

- The set of files (and references to files) stored in a VRE workspace. These may include the “experiments” executed by VRE members, the (reference to) data analysed and the results obtained;
- The set of posts and comments produced by the VRE members;
- Any dataset produced by aggregating data from data providers for the sake of analysing it;
- Any dataset produced by aggregating data from data providers for the sake of building an integrated dataset out of the aggregated data (e.g. this is the case of Knowledge Bases);
- The material of a training course;
- A dataset documenting and providing evidence for either a report or a publication produced in the context of project (and VREs supported) activities.

In the rest an initial set of datasets to be managed by BlueBRIDGE are identified and described. These are clustered in five major classes:

- *Basic Datasets*, i.e., datasets provided by third-parties that are worth to have in the BlueBRIDGE data space because of their potential use when building Virtual Research Environments. The majority of these datasets pre-exist BlueBRIDGE and they are made available by relying on Blue Commons facilities;
- *Datasets Resulting from Activities performed within a VRE*, i.e., products generated by actors while operating in a VREs. These datasets are of very different nature, vary over the time and represent the concrete VRE state;
- *Project Output*, i.e., data products resulting from the different BlueBRIDGE project activities like deliverables, dissemination material, training material, and scientific publications;
- *Infrastructure Operation Datasets*, i.e., datasets resulting from the operation and monitoring of the BlueBRIDGE supporting infrastructure. These datasets represent a valuable information on per-

service-state that has to be managed for multiple reasons, including pure operation activity (e.g., deploy a new instance of a service) and research activity (e.g., data mining);

- Software, i.e., software artefacts and source code produced in the context of the BlueBRIDGE project. Its management is required not only for the specific needs of the project but also for enabling third-parties to discover, access and reuse it. Moreover, once deposited, it may also have a scientific value as “corpora” for various purposes including research tasks, e.g., data mining tasks on source code to discover common patterns.

The rest of the deliverable is organised as follows. Section 2 describes the basic concepts used in this this data management plan, i.e., the set of properties for describing each identified dataset and the BlueBRIDGE practices and facilities exploited for datasets storage (including Repositories), dissemination (including Catalogues) and preservation. Section 3 presents the management plan for the identified datasets by providing a characterisation of each of them according to the concepts identified in Section 2. Finally, Section 4 concludes the report and briefly reports on next steps.

2 CONCEPTS

In addition to the list of datasets described in Sec. 3, the two basic elements that characterize a Data Management Plan are: (a) the set of information to be captured per dataset (cf. Sec. 2.1) and (b) the set of dataset management practices and tools that the project plan to use (cf. Sec. 2.2). This section describes how BlueBRIDGE addresses these two basic elements.

2.1 PER DATASET INFORMATION

For each dataset, the following properties will be collected: (a) dataset reference and name, (b) dataset description, (c) dataset sharing policies, and (d) dataset preservation policies. The last three are compound properties whose elements are described below.

Dataset description

The provided description characterises the dataset, its origin (in case it is collected), nature and scale, to whom it could be useful, and whether it underpins a scientific publication. Requested information includes:

- **Description:** a summary of the dataset content;
- **Generated/Collected:** a specification of whether the dataset is genuinely generated within the project or it is produced by aggregating content out of existing datasets / data sources. In case the dataset is collected the origin(s) has to be indicated;
- **Nature:** the typology of content made available within the dataset. Datasets can either comprise items of a single typology (e.g. images) or they can comprise heterogeneous items having different typologies;
- **Scale:** the size of the dataset (and the expected growth rate). Size can be expressed by reporting the number of entries, the disk space, and any other relevant metrics worth for the specific case;
- **Potential use:** any potential use of the dataset, any community or use case that might benefit from the dataset;
- **Scientific publication:** a specification of whether the dataset is underpinning any scientific publication. It also includes a reference to the associated publication, if any.

Data and Metadata Standards

The most suitable standards for representing the dataset content and the associated metadata should be indicated. Such standards are expected to promote the (re-) use of the dataset. If these do not exist, an outline on how and what data and metadata will be created.

Dataset sharing

Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related). Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.). Requested information include:

- **Access Procedure:** how a user will have access to the dataset (e.g. API, web site);
- **Embargo Periods:** whether there is an embargo period or not and in case, there is, its duration;

- **Dissemination Mechanisms:** how the availability of the dataset will be announced, e.g. publishing the dataset in the BlueBRIDGE Data Catalogue;
- **Software and tools for re-use:** if there is any specific software to be used for consuming the dataset;
- **Access rights:** what are the policies governing access to the dataset, e.g. the dataset is “open”, the dataset is made available to authorised users only;
- **Licence:** the type of permit assigned from the dataset owner to exploit it, e.g. CC Licences;
- **Repository:** the name of the “repository” that support the publishing of the dataset (if any). It may also include BlueBRIDGE ones (e.g. BlueBRIDGE Spatial Computing Infrastructure) ;

Dataset archive and preservation

Description of the procedures and tools that will be put in place for long-term preservation of the data.

Requested information include:

- **Preservation strategy:** the strategy to be implemented and the expected preservation period ;
- **Preservation tool:** the instruments put in place to implement the strategy.

2.2 BLUEBRIDGE DATA MANAGEMENT PRACTICES AND APPROACHES

BlueBRIDGE supports a series of capabilities and approaches for data storage, data access and data preservation. These provide the concrete basis on which dataset management strategies can be implemented. The following sections briefly describe them.

2.2.1 DATA STORAGE (AND REPOSITORIES)

BlueBRIDGE is called to deal with a variety of data types ranging from files to geospatial data and (curated) tabular data. As a consequence, an array of diverse storage solutions and approaches are in place to serve the management of these data types. These solutions include:

- **File-oriented datasets:** are managed by relying on the gCube Storage Manager, i.e. a gCube Java based software library implementing a unique set of methods for services and applications supporting an efficient management of files. It is based on a network of distributed storage nodes managed by specialized open-source software for document-oriented databases. In its current implementation, three possible document store systems can be seamlessly used (i.e. MongoDB, Terrastore and U.STORE) while new ones can be added by implementing a specific mediator.
- **Information objects:** are maintained in the Workspace, that in turn builds on the rest of storage solutions for maintaining the constituents of each information object.
- **Geospatial datasets:** are managed by relying on a federation of repositories based on GeoServer and THREDDS Data Server technologies orchestrated by a GIS publisher library, i.e. a gCube software library enabling a seamless publication of geospatial data on diverse repositories while guaranteeing load balancing, failure management and automatic metadata generation.
- **(curated) Tabular datasets:** are handled by relying on a cluster of RDBMs (PostgreSQL) orchestrated by the gCube Tabular Data Manager, i.e. a service supporting the entire lifecycle of tabular data. It enables a user to ingest data – from a file or a web location – that are represented in formats including CSV, JSON and SDMX. It supports the curation of such data and their processing / analytics.
- **Biodiversity datasets:** are collected on-demand from well-known data providers via a specific mediator service, the gCube Species Product Discovery Service. Once collected, data are stored in the workspace and shared according to their associated policies and licences.

- Social networking and logging datasets: are managed by exploiting NoSQL solutions, namely a Cassandra cluster and a CouchDB cluster.

2.2.2 DATA DISSEMINATION (AND CATALOGUES)

The data dissemination strategy implemented by BlueBRIDGE is supported by a number of catalogues including:

- A geospatial data catalogue – based on GeoNetwork technology – enabling users to search (by keywords) for geospatial datasets stored in the BlueBRIDGE repositories;
- An overall data catalogue – based on CKAN technology – enabling users to perform faceted search on the entire set of datasets managed by the BlueBRIDGE VREs.

The development and enhancement of these catalogues is part of the Blue Commons activities carried out in WP9. Clearly there is a strong dependency between the data management plan and the technology put in place to support it: any data management plan imposes requirements on the management technology and, at the same time, it cannot leave the available technology out of consideration.

Moreover, every information object residing in the Workspace is accessible via a URL. Two diverse URIs can be associated with any of these objects: (a) a restricted one allowing only authorized users to actually access linked the object and (b) a public one enabling any users to access the linked object. This enables a URI-based dissemination mechanism that can be used in several contexts, including social networking ones where the norm is to disseminate content by posting a URL.

2.2.3 DATA PRESERVATION

BlueBRIDGE data preservation strategy ensure that every data is archived in a secure manner. This is done in two different, yet complementary, ways. Every data is constantly copied in a backup area (the periodicity of the backup procedures varies from case by case). Certain storage solutions automatically store the content in multiple copies, e.g. this is the case of the technologies behind the file-oriented storage.

Currently, no format migration strategy or approach is in place, the data are managed with their native format.

Our plan is to address more in-depth data preservation strategies in the coming months.

3 DATASETS

For the sake of clarity of presentation, we have decided to cluster the datasets BlueBRIDGE is called to manage into the following classes:

- Basic Datasets (cf. Sec. 3.1), i.e., the datasets worth to have in the BlueBRIDGE data space because of their potential use in one or more Virtual Research Environments;
- Datasets Resulting from Activities performed within a VRE (cf. Sec. 3.2), i.e., datasets produced as part of the collaboration activities by actors the operation of a Virtual Research Environment;
- Project Output (cf. Sec. 3.3), i.e. datasets resulting from the operation of the BlueBRIDGE project and produced by the BlueBRIDGE consortium;
- Infrastructure Operation Datasets (cf. Sec. 3.4) i.e. datasets resulting from the operation of the BlueBRIDGE supporting infrastructure, datasets underlying infrastructure services;
- Software (cf. Sec. 3.5), i.e. datasets resulting from the software enabling BlueBRIDGE.

The remain of this section describe the management plan for the datasets in the different classes. The first class contain very heterogeneous datasets, with very different properties. In order to make their peculiarities explicit, we list for each of them their characteristics. Fo the remain four classes, instead, we describe the common characteristics and management plan identified so far.

3.1 BASIC DATASETS

At the time of the preparation of this first version of the deliverable the following datasets have already be identified by the VREs designers as necessary for supporting the planned project laboratories. It is expectd that others will be added as the design and the refinement of the requirements will progress.

3.1.1 AQUAMAPS DATA (GIS MAPS)

Dataset Description
<p>This dataset consists of 47,825 niche modelling GIS maps referring to the actual and potential distribution of marine species for today and 2050. Such maps are produced by relying on the homonymous AquaMaps predictive model.</p> <p>Generated/Collected: Generated.</p> <p>Origin(s): N/A</p> <p>Nature: Global scale GIS maps published under WMS and WFS standards.</p> <p>Scale: 100 GB</p> <p>Potential use: Maps comparison, reference in other models and in environmental monitoring algorithms.</p> <p>Scientific publication:</p> <ul style="list-style-type: none"> • Ready, J., Kaschner, K., South, A. B., Eastwood, P. D., Rees, T., Rius, J., et al. (2010). Predicting the distributions of marine organisms at the global scale. <i>Ecological Modelling</i>, 221(3), 467-478. • Coro, G., Magliozzi, C., Ellenbroek, A., Kaschner, K., & Pagano, P. (2015). Automatic classification of climate change effects on marine species distributions in 2050 using the AquaMaps model. <i>Environmental and Ecological Statistics</i>, 1-26.
Data and Metadata Standards
<p>Data and metadata representation standards: Data are made available via GIS standards including Web Feature Service and Web Map Service.</p>

Dataset Sharing

Access Procedure: Through a set of GeoNetwork and GeoServer services with access restricted to controlled VREs members.

Embargo Periods: No embargo period is foreseen.

Dissemination Mechanisms: The constituents of the dataset (i.e., each AquaMaps objects which are datasets per se) are published in both the catalogues offered by BlueBRIDGE (cf. Sec. 2.2.2), namely the geospatial data catalogue and the overall data catalogue.

Software and tools for re-use: GeoNetwork CSW service and GeoServer software mechanisms. Java API are associated to both these services to consume data.

Access rights: The dataset is made available to authorised users only.

Licence: CC-BY-NC.

Repository: The dataset is stored in the BlueBRIDGE Spatial Data Infrastructure.

Dataset archive and preservation

Preservation strategy: Dataset content is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies. Moreover, since the maps are produced / generated according to a defined procedure, a copy of the source data and algorithm leading to a specific version of the dataset is maintained.

Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).

3.1.2 ASFIS (COPY)

Dataset Description

ASFIS List of Species for Fishery Statistics Purposes. The FAO Fisheries and Aquaculture Statistics and Information Service (FIPS) collates world capture and aquaculture production statistics at either the species, genus, family or higher taxonomic levels in 2 189 statistical categories (2013 data) referred to as species items.

ASFIS list of species includes 12,600 species items selected according to their interest or relation to fisheries and aquaculture. For each species item stored in a record, codes (ISSCAAP group, taxonomic and 3-alpha) and taxonomic information (scientific name, author(s), family, and higher taxonomic classification) are provided. An English name is available for most of the records, and about one third of them have also a French and Spanish name. Information is also provided about the availability of fishery production statistics on the species item in the FAO databases.

Generated/Collected: Collected

Origin(s): FAO of the UN

Nature: CSV

Scale: 15,000 records; some 15 fields

Potential use: Reference dataset for data harmonization; master data management for species of interest to fisheries statistical systems.

Scientific publication: N/A

Data and Metadata Standards

Data and metadata representation standards: The dataset is just a csv file.

Dataset Sharing
<p>Access Procedure: The reference version of the dataset is made available by FAO via ftp://ftp.fao.org/FI/STAT/DATA/ASFIS_sp.zip. A copy of this dataset is maintained in the BlueBRIDGE data space and made available via several services as reference data.</p> <p>Embargo Periods: N/A</p> <p>Dissemination Mechanisms: The dataset is published in the BlueBRIDGE overall data catalogue.</p> <p>Software and tools for re-use: N/A</p> <p>Access rights: Open</p> <p>Licence: Free</p> <p>Repository: N/A</p>
Dataset archive and preservation
<p>Preservation strategy: Dataset content is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies. The master (and authoritative) copy is maintained by FAO.</p> <p>Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).</p>

3.1.3 AVERAGE ANNUAL POPULATION TO CALCULATE REGIONAL GDP DATA (THOUSAND PERSONS) BY NUTS 3 REGIONS

Dataset Description
<p>Average annual population to calculate regional GDP data (thousand persons) by NUTS 3 regions (nama_10r_3popgdp).</p> <p>Generated/Collected: Collected</p> <p>Origin(s): Eurostat http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing%3Ffile%3Ddata/nama_10r_3popgdp.tsv.gz</p> <p>Nature: Tabular data, with various data types (integers, doubles, dates, etc.);</p> <p>Scale: NUTS level 3 classification</p> <p>Potential use: Feeding of geospatial search and optimization engine for investment planning analysis</p> <p>Scientific publication: The data sets may support scientific publications in a variety of fields, such as computer scientists, business analysis, environmental monitoring, etc.</p>
Data and Metadata Standards
<p>Data and metadata representation standards: CSV with tab delimiter (TSV)</p>
Dataset Sharing
<p>Access Procedure: API</p> <p>Embargo Periods: none</p> <p>Dissemination Mechanisms: BlueBRIDGE Data Catalogue</p>

<p>Software and tools for re-use: WMS Server. Geospatial search and optimization engine service</p> <p>Access rights: Open</p> <p>Licence: Eurostat License (http://ec.europa.eu/eurostat/about/our-partners/copyright)</p>
Dataset archive and preservation
<p>Preservation strategy: The master (authoritative) version of the dataset is maintained by Eurostat.</p> <p>Preservation tool: N/A</p>

3.1.4 BUSINESS DEMOGRAPHY AND HIGH GROWTH ENTERPRISE BY NACE REV. 2 AND NUTS 3 REGIONS

Dataset Description
<p>Business demography and high growth enterprise by NACE Rev. 2 and NUTS 3 regions (bd_hgnace2_r3).</p> <p>Generated/Collected: Collected</p> <p>Origin(s): Eurostat http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing%3Ffile%3Ddata/bd_hgnace2_r3.tsv.gz</p> <p>Nature: Tabular data, with various data types (integers, doubles, dates, etc.);</p> <p>Scale: NUTS level 3 classification</p> <p>Potential use: Feeding of geospatial search and optimization engine for investment planning analysis</p> <p>Scientific publication: The data sets may support scientific publications in a variety of fields, such as computer scientists, business analysis, environmental monitoring, etc.</p>
Data and Metadata Standards
<p>Data and metadata representation standards: CSV with tab delimiter (TSV)</p>
Dataset Sharing
<p>Access Procedure: API</p> <p>Embargo Periods: none</p> <p>Dissemination Mechanisms: BlueBRIDGE Data Catalogue</p> <p>Software and tools for re-use: WMS Server. Geospatial search and optimization engine service</p> <p>Access rights: Open</p> <p>Licence: Eurostat License (http://ec.europa.eu/eurostat/about/our-partners/copyright)</p>
Dataset archive and preservation
<p>Preservation strategy: The master (authoritative) version of the dataset is maintained by Eurostat.</p> <p>Preservation tool: N/A</p>

3.1.5 COPERNICUS MARINE ENVIRONMENT MONITORING SERVICE DATA

Dataset Description
<p>In situ and satellite data extracted from the Copernicus Marine Environment Monitoring Service.</p> <p>Generated/Collected: Generated/Collected</p> <p>Origin(s): Copernicus Marine Environment Monitoring Service</p> <p>Nature: Raster data and point data in NetCDF format.</p> <p>Scale: About 100GB of disk space is foreseen to host these data.</p> <p>Potential use: Niche modelling, ecological modelling, spatial interpolation, environmental modelling.</p> <p>Scientific publication:</p> <ul style="list-style-type: none"> • Blanc, F. (2008). MyOcean information system. EuroGOOS 2008. • de Dianous, R., Jolibois, T., & Besnard, S. (2015, April). MyOcean Central Information System-Achievements and Perspectives. In EGU General Assembly Conference Abstracts (Vol. 17, p. 5767).
Data and Metadata Standards
<p>Data and metadata representation standards: CSW and ISO 19139 for metadata. NetCDF and WCS for data content representation.</p>
Dataset Sharing
<p>Access Procedure: Dataset is made available via the BlueBRIDGE Spatial Data Infrastructure using OGC standards and Java APIs.</p> <p>Embargo Periods: No embargo period is foreseen.</p> <p>Dissemination Mechanisms: The constituents of the dataset (datasets on its own) are published in both the catalogues offered by BlueBRIDGE (cf. Sec. 2.2.2), namely the geospatial data catalogue and the overall data catalogue.</p> <p>Software and tools for re-use: Datasets are made available via OGC standards, no specific tool is needed apart from the standard ones for dealing with this typology of data.</p> <p>Access rights: The dataset is made available to authorised users only through the Virtual Research Environments access control mechanisms.</p> <p>Licence: CC-BY-NC</p> <p>Repository: BlueBRIDGE Spatial Data Infrastructure.</p>
Dataset archive and preservation
<p>Preservation strategy: Dataset content is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies.</p> <p>Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).</p>

3.1.6 DERIVED GEOSPATIAL DATASETS BY COMBINING EUROSTAT DATASETS AND THE NUTS CLASSIFICATION

Dataset Description
Eurostat collected datasets will be combined with the NUTS 2013 classification dataset and will result to new

<p>geospatial datasets</p> <p>Generated/Collected: Generated</p> <p>Origin(s): Eurostat</p> <p>Nature: Vector Data</p> <p>Scale: Different scale depending on the origin datasets</p> <p>Potential use: Feeding of geospatial search and optimization engine for investment planning analysis</p> <p>Scientific publication: The data sets may support scientific publications in a variety of fields, such as computer scientists, business analysis, environmental monitoring, etc.</p>
Data and Metadata Standards
Data and metadata representation standards: WMS, WFS
Dataset Sharing
<p>Access Procedure: API, Geospatial search and optimization engine user interface (ui)</p> <p>Embargo Periods: none</p> <p>Dissemination Mechanisms: BlueBRIDGE Data Catalogue</p> <p>Software and tools for re-use: WMS and WFS Servers. Geospatial search and optimization engine service</p> <p>Access rights: Open</p> <p>Licence: <i>Dependent on originating datasets</i></p>
Dataset archive and preservation
<p>Preservation strategy: The master (authoritative) version of the dataset is maintained by Eurostat.</p> <p>Preservation tool: N/A</p>

3.1.7 EMPLOYMENT (THOUSAND PERSONS) BY NUTS 3 REGIONS

Dataset Description
<p>Employment (thousand persons) by NUTS 3 regions (nama_10r_3empers).</p> <p>Generated/Collected: Collected</p> <p>Origin(s): Eurostat http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing%3Ffile%3Ddata/nama_10r_3empers.tsv.gz</p> <p>Nature: Tabular data, with various data types (integers, doubles, dates, etc.);</p> <p>Scale: NUTS level 3 classification</p> <p>Potential use: Feeding of geospatial search and optimization engine for investment planning analysis</p> <p>Scientific publication: The data sets may support scientific publications in a variety of fields, such as computer scientists, business analysis, environmental monitoring, etc.</p>
Data and Metadata Standards
Data and metadata representation standards: CSV with tab delimiter (TSV)

Dataset Sharing
<p>Access Procedure: API</p> <p>Embargo Periods: none</p> <p>Dissemination Mechanisms: BlueBRIDGE Data Catalogue</p> <p>Software and tools for re-use: WMS Server. Geospatial search and optimization engine service</p> <p>Access rights: Open</p> <p>Licence: Eurostat License (http://ec.europa.eu/eurostat/about/our-partners/copyright)</p>
Dataset archive and preservation
<p>Preservation strategy: The master (authoritative) version of the dataset is maintained by Eurostat.</p> <p>Preservation tool: N/A</p>

3.1.8 EMPLOYMENT (THOUSAND HOURS WORKED) BY NUTS 2 REGIONS

Dataset Description
<p>Employment (thousand hours worked) by NUTS 2 regions (nama_10r_2emhrw).</p> <p>Generated/Collected: Collected</p> <p>Origin(s): Eurostat http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing%3Ffile%3Ddata/nama_10r_2emhrw.tsv.gz</p> <p>Nature: Tabular data, with various data types (integers, doubles, dates, etc.);</p> <p>Scale: NUTS level 2 classification</p> <p>Potential use: Feeding of geospatial search and optimization engine for investment planning analysis</p> <p>Scientific publication: The data sets may support scientific publications in a variety of fields, such as computer scientists, business analysis, environmental monitoring, etc.</p>
Data and Metadata Standards
<p>Data and metadata representation standards: CSV with tab delimiter (TSV)</p>
Dataset Sharing
<p>Access Procedure: API</p> <p>Embargo Periods: none</p> <p>Dissemination Mechanisms: BlueBRIDGE Data Catalogue</p> <p>Software and tools for re-use: WMS Server. Geospatial search and optimization engine service</p> <p>Access rights: Open</p> <p>Licence: Eurostat License (http://ec.europa.eu/eurostat/about/our-partners/copyright)</p>
Dataset archive and preservation
<p>Preservation strategy: The master (authoritative) version of the dataset is maintained by Eurostat.</p>

Preservation tool: N/A

3.1.9 EUROSTAT DATASETS

Dataset Description

Aquaculture production data are collected and disseminated annually in 5 tables since reference year 2008:

- Production from aquaculture excluding hatcheries and nurseries [fish_aq2a] by species, by FAO major area, by production method, by aquatic environment in Tonnes Live Weight (TLW) and in Euro.
- Production of fish eggs for human consumption from aquaculture [fish_aq2b] by species, by FAO major area, by aquatic environment in TLW, Euro and Euro/Tonne.
- Input to capture-based aquaculture [fish_aq3] by species in Number, TLW, Euro and Euro/Tonne.
- Production of hatcheries and nurseries at eggs stage in life cycle [fish_aq4a] by species and intended uses in Millions.
- Production of hatcheries and nurseries at juveniles stage in life cycle [fish_aq4b] by species and intended uses in Millions.
- Every three years, these data are complemented by Data on the structure of the aquaculture sector [fish_aq5] by species, by FAO major area, by production method, by aquatic environment in Meters, 1000 of M3 and Hectares.

Generated/Collected: Collected

Origin(s): Eurostat

- http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/AppLinkServices?lang=en&appId=bulkdownload&appUrl=http%3A%2F%2Fec.europa.eu%2Feurostat%2Festat-navtree-portlet-prod%2FBulkDownloadListing%3Ffile%3Ddata%2Ffish_aq2a.tsv.gz
- http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/AppLinkServices?lang=en&appId=bulkdownload&appUrl=http%3A%2F%2Fec.europa.eu%2Feurostat%2Festat-navtree-portlet-prod%2FBulkDownloadListing%3Ffile%3Ddata%2Ffish_aq2b.tsv.gz
- http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/AppLinkServices?lang=en&appId=bulkdownload&appUrl=http%3A%2F%2Fec.europa.eu%2Feurostat%2Festat-navtree-portlet-prod%2FBulkDownloadListing%3Ffile%3Ddata%2Ffish_aq3.tsv.gz
- http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/AppLinkServices?lang=en&appId=bulkdownload&appUrl=http%3A%2F%2Fec.europa.eu%2Feurostat%2Festat-navtree-portlet-prod%2FBulkDownloadListing%3Ffile%3Ddata%2Ffish_aq4a.tsv.gz
- http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/AppLinkServices?lang=en&appId=bulkdownload&appUrl=http%3A%2F%2Fec.europa.eu%2Feurostat%2Festat-navtree-portlet-prod%2FBulkDownloadListing%3Ffile%3Ddata%2Ffish_aq4b.tsv.gz

Nature: Three types of codes are assigned to each species item:

- ISSCAAP code
- taxonomic code
- 3-alpha code

Scale: N/A

Potential use: To feed Economy models & algorithms (for experimentation & operation purposes)

Scientific publication: N/A
Data and Metadata Standards
Data and metadata representation standards: xls, csv, Euro SDMX Metadata Structure (ESMS)
Dataset Sharing
<p>Access Procedure: API</p> <p>Embargo Periods: none</p> <p>Dissemination Mechanisms: BlueBRIDGE Data Catalogue</p> <p>Software and tools for re-use: N/A</p> <p>Access rights: Open</p> <p>Licence: Eurostat License</p> <p>Repository: BlueBRIDGE Spatial Data Management Infrastructure, Eurostat</p>
Dataset archive and preservation
<p>Preservation strategy: The master (authoritative) version of the dataset is maintained by Eurostat.</p> <p>Preservation tool:</p>

3.1.10 FAO GLOBAL FISHERIES PRODUCTION STATISTICS

Dataset Description
<p>Global capture statistics as reported by FAO (1950 - 20145). This database contains the volume of aquatic species caught by country or area, by species items, by FAO major fishing areas, and year, for all commercial, industrial, recreational and subsistence purposes. The harvest from aquaculture is also included.</p> <p>Generated/Collected: Collected</p> <p>Origin(s): http://www.fao.org/fishery/statistics/global-production/en</p> <p>Nature: Tabular data described by reference data, or having data types.</p> <p>Scale: 180 * 60 * 50 = approximately 0.5 M observations</p> <p>Potential use: Stock assessment, map production, socio economic analysis</p> <p>Scientific publication: N/A</p>
Data and Metadata Standards
Data and metadata representation standards: SDMX
Dataset Sharing
<p>Access Procedure: Download, API, Website search (filtered), Tabular Data Manager, Statistical manager</p> <p>Embargo Periods: N/A</p> <p>Dissemination Mechanisms: Data Catalogue</p> <p>Software and tools for re-use: FishStat-J and other</p> <p>Access rights: FAO copyright may apply</p>

Licence: <http://www.fao.org/contact-us/terms/en/>

Repository: N/A

Dataset archive and preservation

Preservation strategy: The copy maintained by BlueBRIDGE is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies.

Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).

3.1.11 FCR_TABLE

Dataset Description

The dataset contains the values of the Feed Conversion Ratio (FCR) KPI, which is produced after the statistical modelling that will be done in the VRE at the sampling to sampling dataset. The FCR KPI is a major indicator of growth performance in Aquaculture sector. is an well-known measure which is used in all types of fish farming and indicates how efficient a feed or a feeding strategy can be. Specifically, the F.C.R. is calculated as follows: $F.C.R. = F / (W_f - W_o)$, where F is the weight of food supplied to fish during the study period W_o is the weight of fish at the beginning of the study period W_f is the weight of fish at the end of the study period. In other words, the F.C.R. is the mathematical relationship between the input of the feed that has been fed and the weight gain of a population. There are no measurement units used in writing the F.C.R. The lower the F.C.R, the higher the weight gain obtained from the feed.

The dataset could be in cross-tabular table or tabular data format. In cross-tabular format the rows present the Average Weight in bins and columns present the Temperature degrees in Celsius. In tabular data format each line is a data record of three (3) columns, namely Average Weight, Temperature and the value of FCR KPI.

Generated/Collected: Generated

Origin(s): N/A

Nature: a) Cross-tabular table, b) Tabular data

Scale: a) 30 columns by 20-50 rows (bins of Average Weight), b) up to 1500 rows

Potential use: The specific dataset can be used by VRE owner, so as to produce what-if scenarios about the expected growth. Also, a sensitivity analysis can be made using the FCR table.

Scientific publication: At the moment the dataset will not be actually underpinning any scientific publication.

Data and Metadata Standards

Data and metadata representation standards: Comma-Separated Values (CSV) files

Dataset Sharing

Access Procedure: A user can have access to the dataset via VRE environment.

Embargo Periods: No Embargo Periods

Dissemination Mechanisms:

- BlueBRIDGE Data Catalogue
- Other dissemination actions that will take place related with the estimation of performance in an aqua-farming operation and estimating crucial KPIs.

Software and tools for re-use: It is cross-tabular table (or tabular data) useful at the production process of any Aquaculture company. Also, it can be used for simulation purposes almost by any Production Management System.

Access rights: The dataset is made available only to the VRE owner (Aquaculture company which uses the infrastructure).

Licence: Non-commercial Creative Commons (NC) license

Repository: Repositories and services underlying the Performance Evaluation in Aquaculture VRE and others VRE hosting services for aquafarm performance evaluation.

Dataset archive and preservation

Preservation strategy: The procedures and tools that will be put in place for long-term preservation of the dataset are the same as preservation strategy of its VRE Performance Evaluation in Aquaculture.

Preservation tool: The instruments put in place to implement the strategy, are the same as those which are implement the VRE preservation strategy.

3.1.12 FINANCIAL DATASET

Dataset Description

The dataset contains data and information so as a farmer be able to compose a viable financial analysis of aquaculture businesses and assess the efficiency of economical investment. During the production process incurs expenses, which are categorized as “fixed” or “variable” costs. Fixed costs tend to remain reasonably constant, do not generally vary with the level of production and include such items as depreciation, insurance, licenses, and salaries. Variable costs are those that change with the level of production and species; they include costs for shellfish seed or fish fingerlings, feed, daily or hourly wages, electricity, and chemicals, medications etc. Variable costs are depended on the scale of the infrastructure and production as well as the production strategies and policies. Indicatively, in the above list some financial information to be contained in the dataset are presented:

Fixed costs:

- Labor costs (feeders, other employees, etc.)
- Storage Shed
- Electrical aerator (or air lifts)
- Emergency Generator
- Transport tanks
- Cages
- Dock/Boats
- Other equipment
- Repair and Maintenance cost (components and spare parts and service providers)
- Insurance
- Marketing
- Water acres rental

Variable costs:

- Other labor costs

- Feed cost per kg
- Fry cost per juvenile
- Medicine cost
- Chemicals
- Electrical energy - Electricity cost for aeration
- Utilities costs (electric, phone, etc)
- Diesel/Fuel
- Hired Labor – Hours
- Harvesting
- Transportation Costs of Fingerlings
- Transportation to packaging centers
- Transportation Costs to Market
- Miscellaneous
- Consumables

Other necessary data that should be collected are:

- Fish sale (volumes, prices)
- Gross profit
- Interest payable
- Notes payable
- Capital stock
- Owner equity
- Other financial elements of the balance sheet

Generated/Collected: Collected

Origin(s): VRE Owners

Nature: Tabular data

Scale: 1MB

Potential use: Assess the efficiency of economical investment

Scientific publication: N/A

Data and Metadata Standards

Data and metadata representation standards: Comma-Separated Values (CSV) files

Dataset Sharing

Access Procedure: A user can have access to the dataset via VRE environment.

Embargo Periods: No Embargo Periods

Dissemination Mechanisms:

- BlueBRIDGE Data Catalogue

Software and tools for re-use

Access rights: The dataset is made available only to the VRE owner (Aquaculture company which uses the infrastructure).

Licence: Non-commercial Creative Commons (NC) license

Repository: Repositories and services underlying the Performance Evaluation in Aquaculture VRE and others VRE hosting services for aquafarm performance evaluation.

Dataset archive and preservation

Preservation strategy: The procedures and tools that will be put in place for long-term preservation of the dataset are the same as preservation strategy of its VRE Performance Evaluation in Aquaculture.

Preservation tool: The instruments put in place to implement the strategy, are the same as those which are implement the VRE preservation strategy.

3.1.13 FIRMS FACTSHEETS ON STOCKS, FISHERIES AND MARINE RESOURCES

Dataset Description

The primary aim of the Fisheries and Resources Monitoring System (FIRMS) is to provide access to a wide range of high-quality information on the global monitoring and management of fishery marine resources.

FIRMS is reporting on **Marine Resources** and **Fisheries** according to the data contributed by FIRMS Partners. The coverage depends on FIRMS partners' data contributions and other data collected outside the FIRMS partnership under the Strategy-STF framework.

Generated/Collected: Collected

Origin(s): <http://www.fao.org/fishery/statistics/tuna-catches/en>

Nature: XML based Fact-sheets

Scale: <1K Factsheets

Potential use: Important to Global Record of Stocks and Fisheries.

Scientific publication: N/A

Data and Metadata Standards

Data and metadata representation standards: FiGIS proprietary MXL format

Dataset Sharing

Access Procedure: API, website X-Path search https://support.d4science.org/projects/stocksandfisherieskb/wiki/15-10-23_FIRMS_data_source.

Embargo Periods: N/A

Dissemination Mechanisms: Global record of stocks and fisheries web services.

Software and tools for re-use: Internally, harmonizing the data requires TabMan harmonization templates and MatWare managed semantic knowledge base.

Access rights: Open, but data are under FAO copyright.

Licence: None specific, usage governed by terms of use: <http://www.fao.org/contact-us/terms/en/>

Repository: N/A
Dataset archive and preservation
Preservation strategy: Dataset content is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies. Master (and authoritative) copy is maintained by FAO.
Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3). The tools for maintaining the master copy are operated by FAO.

3.1.14 FISHERIES COMMODITIES AND TRADE STATISTICS

Dataset Description
This database contains statistics on the annual production of fishery commodities and imports and exports (including re-exports) of fishery commodities by country and commodities in terms of volume and value from 1976.
Generated/Collected: Collected
Origin(s): http://www.fao.org/fishery/statistics/global-commodities-production/en
Nature: Tabular data; most column contain reference data, whilst other contain observational data types.
Scale: <1M observations
Potential use: Socio-economic indicator analysis
Scientific publication: N/A
Data and Metadata Standards
Data and metadata representation standards: SDMX
Dataset Sharing
Access Procedure: Download and through FishStatJ, a specialized stand-alone application.
Embargo Periods: N/A
Dissemination Mechanisms: The dataset will be published in the BlueBRIDGE overall data catalogue.
Software and tools for re-use: FishStat-J
Access rights: FAO Policy, http://www.fao.org/contact-us/terms/en/
Licence: None specific, http://www.fao.org/contact-us/terms/en/
Repository: BlueBRIDGE facilities for curated tabula data (cf. Sec. 2.2.1).
Dataset archive and preservation
Preservation strategy: The dataset is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies.
Preservation tool: The dataset is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).

3.1.15 GBIF AND OBIS TAXA AND OBSERVATIONS

Dataset Description
<p>Collection of species observations provided by the GBIF and OBIS data collectors.</p> <p>Generated/Collected: Collected</p> <p>Origin(s): Global Biodiversity Information Facility (GBIF), Ocean Biogeographic Information System (OBIS).</p> <p>Nature: Observation data of marine species, supplied under proprietary formats.</p> <p>Scale: Unknown, the scale is defined by the user download requests.</p> <p>Potential use: Niche modelling, species distribution maps, species presence indication, species commonness estimation.</p> <p>Scientific publication:</p> <ul style="list-style-type: none"> • Candela, L., Castelli, D., Coro, G., Lelii, L., Mangiacrapa, F., Marioli, V., & Pagano, P. (2015). An infrastructure-oriented approach for supporting biodiversity research. <i>Ecological Informatics</i>, 26, 162-172. • Coro, G., Webb, T. J., Appeltans, W., Bailly, N., Cattrijsse, A., & Pagano, P. (2015). Classifying degrees of species commonness: North Sea fish as a case study. <i>Ecological Modelling</i>, 312, 272-280.
Data and Metadata Standards
<p>Data and metadata representation standards: Darwin Core, csv.</p>
Dataset Sharing
<p>Access Procedure: Data will be shared using the facilities provided by BlueBRIDGE, after they will have undergone harmonisation processing through the BlueBRIDGE services. In particular, files are stored in the workspace and made available through it.</p> <p>Embargo Periods: No embargo period foreseen.</p> <p>Dissemination Mechanisms: Each dataset will be published on users' choice on the BlueBRIDGE data catalogue.</p> <p>Software and tools for re-use: Dataset content is produced by gCube Species Product Discovery, it can be reused by any tool capable to use csv files or Darwin Core files.</p> <p>Access rights: The dataset is made available to authorised users only through Virtual Research Environment mechanisms.</p> <p>Licence: CC Licences.</p> <p>Repository: The datasets will be published on users' choice on the BlueBRIDGE data catalogue.</p>
Dataset archive and preservation
<p>Preservation strategy: Dataset content is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies.</p> <p>Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).</p>

3.1.16 GLOBAL NUMBER OF FISHERIES

Dataset Description
<p>This database contains statistics on the total number of people employed annually in commercial and subsistence fishing, by country, by occupational category, by gender and according to the time spent in fishing.</p>

<p>Generated/Collected: Collected</p> <p>Origin(s): FAO Fisheries Circular Numbers of Fishers Circular No.929 (and subsequent revisions). (Trilingual, E-F-S). http://www.fao.org/fishery/statistics/global-fishers/en</p> <p>Nature: Tabular data; most columns are referenced by code lists, others contain observations and have data types.</p> <p>Scale: <1M observations.</p> <p>Potential use: Socio economic analysis.</p> <p>Scientific publication: N/A</p>
Data and Metadata Standards
Data and metadata representation standards: None
Dataset Sharing
<p>Access Procedure: Download</p> <p>Embargo Periods: N/A</p> <p>Dissemination Mechanisms: Data catalogue.</p> <p>Software and tools for re-use: N/A</p> <p>Access rights: FAO policy, http://www.fao.org/contact-us/terms/en/</p> <p>Licence: N/A</p> <p>Repository: N/A</p>
Dataset archive and preservation
<p>Preservation strategy: Dataset content is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies. Master (and authoritative) copy is maintained by FAO.</p> <p>Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).</p>

3.1.17 GLOBAL_SEAWATER_OXYGEN18

Dataset Description
<p>This database is a collection of over 26,000 seawater O-18 values made since about 1950. The fields include longitude, latitude, month, year, depth (m), temperature (C), salinity (psu), d18O, dD, notes and reference. For example, you might be interested in all samples that have an associated salinity, between the depths of 1000 and 2000 meters in the Atlantic Ocean. This is possible by choosing the limits for depth (1000-2000m), salinity (0-40 say) and longitude.</p> <p>Generated/Collected: Collected</p> <p>Origin(s): http://data.giss.nasa.gov/o18data/</p> <p>Nature: The typology of content is fixed width, consisted of real numbers double-precision, integers and strings.</p> <p>Scale: The size of ascii text file is up to 3MB.</p> <p>Potential use: Since this database is an amalgamation of all different laboratories and sources, it can offer to the researchers with not-inconsiderable pain, to take, collect and verify information. It provides useful information about global seawater temperature and oxygen-18 values over the years.</p> <p>Scientific publication: Partial versions of this databases appeared in:</p>

- Schmidt, G. A., Forward modeling of carbonate proxy data from planktonic foraminifera using oxygen isotope tracers in a global ocean model, *Paleoceanography*, 14, 482-497, 1999.
- Bigg, G.R. and E.J. Rohling. An oxygen isotope data set for marine water. *J. Geophys. Res.*, 105, 8527-8535, 2000.

References to the whole database (currently version 1.21) should include the web address as follows:

Schmidt, G.A., G. R. Bigg and E. J. Rohling. 1999. "Global Seawater Oxygen-18 Database - v1.21"
<http://data.giss.nasa.gov/o18data/>

Data and Metadata Standards

Data and metadata representation standards: The External Data Representation (XDR) is the standard that data and metadata are obeyed. XDR is a standard data serialization format, which allows data to be transferred between different kinds of computer systems. It uses a base unit of 4 bytes, floating-point number are represented in IEEE 754 format and also variable-length types such as string and opaque are padded to a total divisible by four bytes.

Dataset Sharing

Access Procedure: A user can have access to the dataset either via the VRE environment or through the NASA GISS website.

Embargo Periods: No Embargo Periods.

Dissemination Mechanisms:

- Publishing the datasets in the BlueBRIDGE Data Catalogue.
- Other dissemination actions that will take place related with the estimation of performance in an aquaculture operation and estimating crucial KPIs.

Software and tools for re-use: VREs software and tools are a consumer of dataset.

Access rights: Open.

Licence: NASA may share information with private organizations as part of a service that provides NASA users with increased capabilities or functionality on the site. On www.nasa.gov, the "Share" function at the top of each page is provided by a third party, AddThis. This firm collects information on visitors who use this feature to share NASA content on their Facebook pages, Twitter feeds or other social media or social networking sites. As noted in their privacy policy AddThis uses this information for its own business purposes, including marketing the information to other parties. AddThis does not collect information from users who do not use the feature.

Repository: BlueBRIDGE Performance Evaluation In Aquaculture.

Dataset archive and preservation

Preservation strategy: The procedures and tools that will be put in place for long-term preservation of the dataset are the same as preservation strategy of its VRE Performance Evaluation in Aquaculture (cf. Sec. 0).

Preservation tool: The instruments put in place to implement the strategy, are the same as those which are implement the VRE preservation strategy (cf. Sec. 0). Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).

3.1.18 GLOBAL TUNA NOMINAL CATCHES

Dataset Description

This database contains, for the principal market tuna species, nominal catches by fishing gear, species, stock, fishing country and year from 1950 to 2010.

Tuna and tuna-like species are very important economically and a significant source of food. They include approximately forty species occurring in the Atlantic, Indian and Pacific Oceans and in the Mediterranean Sea. Their

global production has tended to increase continuously from less than 0.6 million tonnes in 1950 to above 6 million tonnes today.

The so-called principal market tuna species are the most important among the tuna and tuna-like species from the catch weight and economical viewpoints. They are landed in numerous locations around the world, traded on the nearly global scale and also processed and consumed in many locations worldwide. In 2010, their catch was approximately four million tonnes, which represents about 66% of the total catch of all tuna and tuna-like species. Most catches of the principal market tuna species are taken from the Pacific (70.5% of the total catch of principal market tuna species in 2008), with the Indian contributing much more (19.5% in 2010) than the Atlantic and the Mediterranean Sea (10.0% in 2010).

Generated/Collected: Collected

Origin(s): <http://www.fao.org/fishery/statistics/tuna-catches/en>

Nature: Tabular data; reference data as code lists, and data types.

Scale: <1M Observations

Potential use: Important to Global Record of Stocks and Fisheries. Important to Stock Assessment VRE, especially when merged with IRD Sardara (requires advanced data management to identify and remediate differences, overlaps and gaps). Stock assessment. Map productions. Input to e.g. Ecopath and other modeling software.

Scientific publication: N/A

Data and Metadata Standards

Data and metadata representation standards: None

Dataset Sharing

Access Procedure: API, website search (filtered), Map search

Embargo Periods: N/A

Dissemination Mechanisms: Data Catalogue

Software and tools for re-use: None

Access rights: FAO copyright applies.

Licence: None, but there is a terms of use <http://www.fao.org/contact-us/terms/en/>.

Repository: N/A

Dataset archive and preservation

Preservation strategy: Dataset content is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies. Master (and authoritative) copy is maintained by FAO.

Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3). The tools for maintaining the master copy are operated by FAO.

3.1.19 GROSS DOMESTIC PRODUCT (GDP) AT CURRENT MARKET PRICES BY NUTS 2 REGIONS

Dataset Description

Gross domestic product (GDP) at current market prices by NUTS 2 regions (nama_10r_2gdp).

Generated/Collected: Collected

Origin(s): Eurostat <http://ec.europa.eu/eurostat/estat-navtree-portlet->

prod/BulkDownloadListing%3Ffile%3Ddata/nama_10r_2gdp.tsv.gz
Nature: Tabular data, with various data types (integers, doubles, dates, etc.);
Scale: NUTS level 2 classification
Potential use: Feeding of geospatial search and optimization engine for investment planning analysis
Scientific publication: The data sets may support scientific publications in a variety of fields, such as computer scientists, business analysis, environmental monitoring, etc.
Data and Metadata Standards
Data and metadata representation standards: CSV with tab delimiter (TSV)
Dataset Sharing
Access Procedure: API
Embargo Periods: none
Dissemination Mechanisms: BlueBRIDGE Data Catalogue
Software and tools for re-use: WMS Server. Geospatial search and optimization engine service
Access rights: Open
Licence: Eurostat License (http://ec.europa.eu/eurostat/about/our-partners/copyright)
Dataset archive and preservation
Preservation strategy: The master (authoritative) version of the dataset is maintained by Eurostat.
Preservation tool: N/A

3.1.20 GROSS DOMESTIC PRODUCT (GDP) AT CURRENT MARKET PRICES BY NUTS 3 REGIONS

Dataset Description
Gross domestic product (GDP) at current market prices by NUTS 3 regions (nama_10r_3gdp).
Generated/Collected: Collected
Origin(s): Eurostat http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing%3Ffile%3Ddata/nama_10r_3gdp.tsv.gz
Nature: Tabular data, with various data types (integers, doubles, dates, etc.);
Scale: NUTS level 3 classification
Potential use: Feeding of geospatial search and optimization engine for investment planning analysis
Scientific publication: The data sets may support scientific publications in a variety of fields, such as computer scientists, business analysis, environmental monitoring, etc.
Data and Metadata Standards
Data and metadata representation standards: CSV with tab delimiter (TSV)
Dataset Sharing
Access Procedure: API

<p>Embargo Periods: none</p> <p>Dissemination Mechanisms: BlueBRIDGE Data Catalogue</p> <p>Software and tools for re-use: WMS Server. Geospatial search and optimization engine service</p> <p>Access rights: Open</p> <p>Licence: Eurostat License (http://ec.europa.eu/eurostat/about/our-partners/copyright)</p>
Dataset archive and preservation
<p>Preservation strategy: The master (authoritative) version of the dataset is maintained by Eurostat.</p> <p>Preservation tool: N/A</p>

3.1.21 INVESTMENT ANALYSIS GEOSPATIAL SEARCH / OPTIMIZATION DATASETS (MAPS)

Dataset Description
<p>Maps presenting the value of a particular investment indicator over a geographical area.</p> <p>Generated/Collected: Generated</p> <p>Origin(s): N/A</p> <p>Nature: Images & Vector data (alternative representation, to be considered).</p> <p>Scale: Expected size will be in excess of 10GB in the course of the project (either in compressed images or small sized vector data).</p> <p>Potential use: Business planning, Investment planning, Investment model adjustment, Policy adjustment / making.</p> <p>Scientific publication: Scientific publications may be supported by the dataset.</p>
Data and Metadata Standards
<p>Data and metadata representation standards: ISO19115, DublinCore;</p>
Dataset Sharing
<p>Access Procedure: Datasets are published in the BlueBRIDGE overall catalogue. Moreover, they are made available to the members of the VREs equipped with this dataset. In addition to that, maps can be shared by relying on BlueBRIDGE Project Workspace.</p> <p>Embargo Periods: 1 to 6 months, required for dataset validation.</p> <p>Dissemination Mechanisms: Publishing the dataset in the BlueBRIDGE catalogues (both the overall one and the geospatial one). Moreover, the dataset will be published in other catalogues of collaborating initiatives (if not harvesting the BlueBRIDGE catalog).</p> <p>Software and tools for re-use: none.</p> <p>Access rights: Investment analysis data will be generally available to authorized users. Exceptions will apply.</p> <p>Licence: Investment analysis data will be generally available to authorized users. License applied is CC Attribution-ShareAlike 4.0 International.</p> <p>Repository: Dataset will be published in various BlueBRIDGE repositories (depending on the data type) including the Spatial Data Infrastructure (cf. Sec. 2.2.1).</p>
Dataset archive and preservation

Preservation strategy: Dataset content is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies.

Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).

3.1.22 LAND COVER IN GREECE

Dataset Description
<p>A dataset with information regarding land cover in Greece</p> <p>Generated/Collected: Collected</p> <p>Origin(s): http://www.okxe.gr/ http://www1.okxe.gr/geoserver/wms?service=WMS&request=GetCapabilities</p> <p>Nature: Vector Data</p> <p>Scale: N/A</p> <p>Potential use: Feeding of geospatial search and optimization engine for investment planning analysis</p> <p>Scientific publication: The data sets may support scientific publications in a variety of fields, such as computer scientists, business analysis, environmental monitoring, etc.</p>
Data and Metadata Standards
<p>Data and metadata representation standards: WMS, CSW</p>
Dataset Sharing
<p>Access Procedure: API</p> <p>Embargo Periods: none</p> <p>Dissemination Mechanisms: BlueBRIDGE Data Catalogue</p> <p>Software and tools for re-use: WMS Server.</p> <p>Access rights: Open</p> <p>Licence: Creative Commons Attribution (cc-by) 3.0</p>
Dataset archive and preservation
<p>Preservation strategy: The master (authoritative) version of the dataset is maintained by the responsible Hellenic authority.</p> <p>Preservation tool: N/A</p>

3.1.23 MORTALITY_TABLE

Dataset Description
<p>The dataset contains the values of the Mortality Rate KPI, which is produced after the statistical modelling that will be done in the VRE at the sampling to sampling dataset. The Mortality Rate KPI is a major indicator of growth</p>

performance in Aquaculture sector. Fish mortality is a parameter used in fisheries population dynamics to account for the loss of fish in a fish stock through death. **Mortality Rate** is the ratio of fish deaths in an area to the fish population of that area. The dataset could be in cross-tabular table or tabular data format. In cross-tabular format the rows present the Average Weight in bins and columns present the Temperature degrees in Celsius. In tabular data format each line is a data record of three (3) columns, namely Average Weight, Temperature and the value of Mortality Rate KPI.

Generated/Collected: Generated

Origin(s): N/A

Nature: a) Cross-tabular table, b) Tabular data

Scale: a) 30 columns by 20-50 rows (bins of Average Weight), b) up to 1500 rows

Potential use: The specific dataset can be used by VRE owner, so as to produce what-if scenarios about the expected growth. Also, a sensitivity analysis can be made using the Mortality Rate table.

Scientific publication: At the moment the dataset will not be actually underpinning any scientific publication.

Data and Metadata Standards

Data and metadata representation standards: Comma-Separated Values (CSV) files

Dataset Sharing

Access Procedure: A user can have access to the dataset via VRE environment.

Embargo Periods: No Embargo Periods

Dissemination Mechanisms:

- BlueBRIDGE Data Catalogue
- Other dissemination actions that will take place related with the estimation of performance in an aqua-farming operation and estimating crucial KPIs.

Software and tools for re-use: It is cross-tabular table (or tabular data) useful at the production process of any Aquaculture company. Also, it can be used for simulation purposes almost by any Production Management System.

Access rights: The dataset is made available only to the VRE owner (Aquaculture company which uses the infrastructure).

Licence: Non-commercial Creative Commons (NC) license

Repository: Repositories and services underlying the Performance Evaluation in Aquaculture VRE and others VRE hosting services for aquafarm performance evaluation.

Dataset archive and preservation

Preservation strategy: The procedures and tools that will be put in place for long-term preservation of the dataset are the same as preservation strategy of its VRE Performance Evaluation in Aquaculture.

Preservation tool: The instruments put in place to implement the strategy, are the same as those which are implement the VRE preservation strategy.

3.1.24 (MT) - TURNOVER AND EMPLOYMENT BY DISTRICT

Dataset Description

<p>Turnover and Employment by District in Malta</p> <p>Generated/Collected: Collected</p> <p>Origin(s): Malta Information Technology Agency (MITA)</p> <p>http://msdi.data.gov.mt/geoserver/MITA_WMS/wms?service=WMS&version=1.3.0&request=GetCapabilities</p> <p>https://msdi.data.gov.mt/geoserver/MITA_WFS/ows?service=WFS&version=2.0.0&request=GetCapabilities</p> <p>Nature: Vector Data</p> <p>Scale: N/A</p> <p>Potential use: Feeding of geospatial search and optimization engine for investment planning analysis</p> <p>Scientific publication: The data sets may support scientific publications in a variety of fields, such as computer scientists, business analysis, environmental monitoring, etc.</p>
Data and Metadata Standards
Data and metadata representation standards: WMS, WFS, KML
Dataset Sharing
<p>Access Procedure: API</p> <p>Embargo Periods: none</p> <p>Dissemination Mechanisms: BlueBRIDGE Data Catalogue</p> <p>Software and tools for re-use: WMS Server</p> <p>Access rights: Open</p> <p>Licence: No Conditions Apply</p>
Dataset archive and preservation
<p>Preservation strategy: The master (authoritative) version of the dataset is maintained by the responsible Malta Information Technology Agency authority.</p> <p>Preservation tool: N/A</p>

3.1.25 NASO FARM LOCATIONS; NATIONAL AQUACULTURE SECTOR OVERVIEWS

Dataset Description
<p>The main purpose of this map collection is to illustrate, in general, where aquaculture is taking place. Characteristics that accompany the administrative units or individual farms include; cultured species, technology used, culture systems, environments, farm characteristics, production, quantities and values, seed input quantity and characteristics, and main issues (credit, diseases, environmental impact, etc.).</p> <p>The overall objective of this activity is to assist member countries in inventorying and monitoring their aquaculture industry.</p> <p>Generated/Collected: Collected</p> <p>Origin(s): http://www.fao.org/fishery/collection/naso-maps/en</p> <p>Nature: XML based Fact-sheets, for Farm Locations a KML can be requested.</p>

<p>Scale: <10K geolocations</p> <p>Potential use: Important to Aquaculture Atlas Productions system VRE.</p> <p>Scientific publication: N/A</p>
Data and Metadata Standards
<p>Data and metadata representation standards: OGC Compliant dataset</p>
Dataset Sharing
<p>Access Procedure: Datasets generated on request</p> <p>Embargo Periods: N/A</p> <p>Dissemination Mechanisms: Download</p> <p>Software and tools for re-use: Internally, harmonizing the data requires TabMan tabular and geospatial harmonization templates and for exchange with spatial analytical software web services will be developed.</p> <p>Access rights: Open, but data are under FAO copyright</p> <p>Licence: None specific, usage governed by terms of use: http://www.fao.org/contact-us/terms/en/</p> <p>Repository: N/A</p>
Dataset archive and preservation
<p>Preservation strategy: Dataset content is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies. Master (and authoritative) copy is maintained by FAO.</p> <p>Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3). The tools for maintaining the master copy are operated by FAO.</p>

3.1.26 NUTS 2013 CLASSIFICATION

Dataset Description
<p>NUTS classification is the classification of territorial units for statistics. This is a regional classification for the EU Member States providing a harmonised hierarchy of regions: the NUTS classification subdivides each Member State into regions at three different levels, covering NUTS 1, 2 and 3 from larger to smaller areas. Regions have also been defined and agreed with the EFTA and candidate countries on a bilateral basis; these are called statistical regions and follow exactly the same rules as the NUTS regions in the EU, although they have no legal basis.</p> <p>Generated/Collected: Collected</p> <p>Origin(s): Eurostat http://ec.europa.eu/eurostat/cache/GISCO/geodatafiles/NUTS_2013_01M_SH.zip</p> <p>Nature: Vector Data</p> <p>Scale: As described in the NUTS classification</p> <p>Potential use: In combination with other Eurostat Datasets, which are based on NUTS classification</p> <p>Scientific publication: The data sets may support scientific publications in a variety of fields, such as computer scientists, business analysis, environmental monitoring, etc.</p>
Data and Metadata Standards
<p>Data and metadata representation standards: Shape File</p>

Dataset Sharing
<p>Access Procedure: API</p> <p>Embargo Periods: none</p> <p>Dissemination Mechanisms: BlueBRIDGE Data Catalogue</p> <p>Software and tools for re-use: WMS Server</p> <p>Access rights: Open</p> <p>Licence: Eurostat License¹</p> <p>In addition to the general copyright and licence policy applicable to the whole Eurostat website, the following specific provisions apply. The download and usage of these data is subject to the acceptance of the following clauses:</p> <ul style="list-style-type: none"> • The Commission agrees to grant the non-exclusive and not transferable right to use and process the Eurostat/GISCO geographical data downloaded from this page (the "data"). • The permission to use the data is granted on condition that: • the data will not be used for commercial purposes; • the source will be acknowledged. A copyright notice, as specified below, will have to be visible on any printed or electronic publication using the data downloaded from this page. <p>Repository: BlueBRIDGE Spatial Data Management Infrastructure, Eurostat</p>
Dataset archive and preservation
<p>Preservation strategy: The master (authoritative) version of the dataset is maintained by Eurostat.</p> <p>Preservation tool: N/A</p>

3.1.27 PLANNING SERVICE (NI) - DEVELOPMENT PLAN - LANDUSE ZONES & POINTS IN UK

Dataset Description
<p>Planning Service (NI) - Development Plan - Landuse Zones & Points in UK</p> <p>Generated/Collected: Collected</p> <p>Origin(s): UK Land & Property Services https://www.spatialni.gov.uk/wss/service/Planning_Service_NI_Landuse-DLS-I-LIC/WSS</p> <p>Nature: Vector Data</p> <p>Scale: N/A</p> <p>Potential use: Feeding of geospatial search and optimization engine for investment planning analysis</p> <p>Scientific publication: The data sets may support scientific publications in a variety of fields, such as computer scientists, business analysis, environmental monitoring, etc.</p>
Data and Metadata Standards

¹ <http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units>

Data and metadata representation standards: WMS, CSW
Dataset Sharing
<p>Access Procedure: API</p> <p>Embargo Periods: none</p> <p>Dissemination Mechanisms: BlueBRIDGE Data Catalogue</p> <p>Software and tools for re-use: To be discussed.</p> <p>Access rights: Open</p> <p>Licence: UK Open Government License (http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/)</p>
Dataset archive and preservation
<p>Preservation strategy: The master (authoritative) version of the dataset is maintained by the responsible UK Land & Property Services authority.</p> <p>Preservation tool: N/A</p>

3.1.28 POSEIDON_DATASETS

Dataset Description
<p>POSEIDON system (www.poseidon.hcmr.gr) is a unique planning tool in the endeavour for the protection of the marine environment. POSEIDON system operation covers the need of providing timely and reliable information and forecast on the condition of the seas, by establishing a network of observation buoys. The network of observation buoys records continuously the physical, biological and chemical parameters of the Greek seas. Those data are then transmitted to the operational center where they are sorted and fed into forecasting models.</p> <p>The gathered data are easily accessible through a web interface to the system's database. The database is updating in daily basis offering full access both to archived and recent data. The authorized user has accessed to multiple metadata information regarding the buoys sensors, location, operational period and other combined information. More specifically, the authorized user can specify the time period, the desired location(s) and the desired parameters such as:</p> <ul style="list-style-type: none"> • geographical (Latitude, Longitude), • atmospheric (e.g. Air Pressure, Air Temperature, Humidity, Wind Gust, Wind Speed), • hydrological (e.g. Platform average depth, Conductivity, Current Direction, Current Speed, Pressure, Salinity and Water Temperature in various depths, Wind Direction), • waves (e.g. Relative position east, Significant Wave Height, Maximum Wave Height, Average wave direction, Wave period), • biochemical (e.g. Chlorophyll-A in various depths, CO2 Pressure and Temperature, Water Temperature at 3 meters depth recorded by current meter). <p>After submitting user request, the system prompts users to download the results in compressed (zip) file. Data are provided in plain ASCII and excel-compatible formats.</p> <p>Generated/Collected: Collected</p> <p>Origin(s): http://www.poseidon.hcmr.gr/listview.php?id=136</p> <p>Nature: The typology of content is fixed width, consisted of real numbers double-precision and integers.</p>

<p>Scale: The size of the files will not be so big, up to some kilobytes, depending on the users choices.</p> <p>Potential use: Aquafarming community (production people, farm and production managers), researchers, data analysts will have benefit of the using this dataset. The observation of historical and recent data would help them to produce forecasting models which are useful to decision making process.</p> <p>Scientific publication: At the moment the dataset will not be actually underpinning any scientific publication.</p>
Data and Metadata Standards
<p>Data and metadata representation standards: The External Data Representation (XDR) is the standard that data and metadata are obeyed. XDR is a standard data serialization format, which allows data to be transferred between different kinds of computer systems. It uses a base unit of 4 bytes, floating-point number are represented in IEEE 754 format and also variable-length types such as string and opaque are padded to a total divisible by four bytes.</p>
Dataset Sharing
<p>Access Procedure: The user has to be authorised by the system to have accessed to the data, thus access rights have to be requested.</p> <p>Embargo Periods: No Embargo Periods</p> <p>Dissemination Mechanisms: Dissemination actions that will take place related with the estimation of performance in an aqua-farming operation and estimating crucial KPIs.</p> <p>Software and tools for re-use: VREs software and tools will be a consumer of dataset.</p> <p>Access rights: The dataset is made available only to authorised users</p> <p>Licence: CC License</p> <p>Repository: The downloaded files can be published in BlueBRIDGE Performance Evaluation in Aquaculture repository.</p>
Dataset archive and preservation
<p>Preservation strategy: The procedures and tools that will be put in place for long-term preservation of the dataset are the same as preservation strategy of its VRE Performance Evaluation in Aquaculture.</p> <p>Preservation tool: The instruments put in place to implement the strategy, are the same as those which are implement the VRE preservation strategy.</p>

3.1.29 SAMPLINGTOSAMPLING

Dataset Description
<p>The dataset contains sampling information gather from time to time by Aquaculture companies. It is a common procedure in the sector, in order to estimate the number and the average weight of fishes in cages. This kind of datasets can be generated by monitoring systems used by Aquaculture companies. It is consisted of the following attributes (fields):</p> <ul style="list-style-type: none"> • Start Date: date of first sampling; • Start Fish No: number of fish at first sampling; • Start Av. Wt: fish average weight at first sampling; • End Date: date of last sampling • End Fish No: number of fish at last sampling

- End Av. Wt: fish average weight at last sampling
- Av. Temp.: average temperature during the period between particular samplings
- Feed Qty: quantity of food given the period between the particular samplings
- Mortalities: fish mortalities during the particular period
- Harvest #: number of fish harvest the period from sampling to sampling
- Harvest Kg: weight of harvest in kilograms the period from sampling to sampling

Using these attributes will be able to calculate useful KPIs for production, such as Feed Conversion Ratio (FCR) (cf. Sec. 3.1.11), Specific Feeding Ratio (SFR) (cf. Sec. 3.1.30), Specific Growth Ratio (SGR) (cf. Sec. 3.1.31) and Mortality (cf. Sec. **Error! Reference source not found.**).

Generated/Collected: Generated (by each Aquaculture company)

Origin(s): N/A

Nature: Tabular data, with various data types (integers, doubles, dates, etc.);

Scale: 150 entries up to 5000 entries;

Potential use: The specific dataset can be used by any authenticate user or researcher of VRE, so as to produce KPI's tables, such as FCR, SGR, SFR, Mortality tables;

Scientific publication: At the moment the dataset will not be actually underpinning any scientific publication.

Data and Metadata Standards

Data and metadata representation standards: Comma-Separated Values (CSV) files

Dataset Sharing

Access Procedure: A user can have access to the dataset via VRE environment.

Embargo Periods: No Embargo Periods

Dissemination Mechanisms: 1) Publishing the dataset in the BlueBRIDGE Data Catalogue; 2) Other dissemination actions that will take place related with the estimation of performance in an aqua-farming operation and estimating crucial KPIs.

Software and tools for re-use: VREs software is a consumer of dataset, but the outcome tables of KPIs can be utilized by any Production Management System of Aquaculture Company.

Access rights: The dataset is made available only to the VRE owner (Aquaculture company which uses the infrastructure)

Licence: Non-commercial Creative Commons (NC) license;

Repository: Repositories and services underlying the Performance Evaluation in Aquaculture VRE and others VRE hosting services for aquafarm performance evaluation.

Dataset archive and preservation

Preservation strategy: The procedures and tools that will be put in place for long-term preservation of the dataset are the same as preservation strategy of its VRE Performance Evaluation in Aquaculture.

Preservation tool: The instruments put in place to implement the strategy, are the same as those which are implement the VRE preservation strategy.

3.1.30 SFR_TABLE

Dataset Description
<p>The dataset contains the values of the Specific Feeding Ratio (SFR) KPI, which is produced after the statistical modelling that will be done in the VRE at the sampling to sampling dataset (cf. Sec. Error! Reference source not found.). The SFR KPI is a major indicator of growth performance in Aquaculture sector. Specific Feeding Ratio (SFR) indicates the percentage of the weight of food supplied to fish divided by the weight of fish. The dataset could be in cross-tabular table or tabular data format. In cross-tabular format the rows present the Average Weight in bins and columns present the Temperature degrees in Celsius. In tabular data format each line is a data record of three (3) columns, namely Average Weight, Temperature and the value of SFR KPI.</p> <p>Generated/Collected: Generated</p> <p>Origin(s): N/A</p> <p>Nature: a) Cross-tabular table, b) Tabular data</p> <p>Scale: a) 30 columns by 20-50 rows (bins of Average Weight), b) up to 1500 rows</p> <p>Potential use: The specific dataset can be used by VRE owner, so as to produce what-if scenarios about the expected growth. Also, a sensitivity analysis can be made using the SFR table.</p> <p>Scientific publication: At the moment the dataset will not be actually underpinning any scientific publication.</p>
Data and Metadata Standards
<p>Data and metadata representation standards: Comma-Separated Values (CSV) files</p>
Dataset Sharing
<p>Access Procedure: A user can have access to the dataset via VRE environment.</p> <p>Embargo Periods: No Embargo Periods</p> <p>Dissemination Mechanisms:</p> <ul style="list-style-type: none"> • BlueBRIDGE Data Catalogue • Other dissemination actions that will take place related with the estimation of performance in an aqua-farming operation and estimating crucial KPIs. <p>Software and tools for re-use: It is cross-tabular table (or tabular data) useful at the production process of any Aquaculture company. Also, it can be used for simulation purposes almost by any Production Management System.</p> <p>Access rights: The dataset is made available only to the VRE owner (Aquaculture company which uses the infrastructure).</p> <p>Licence: Non-commercial Creative Commons (NC) license.</p> <p>Repository: Repositories and services underlying the Performance Evaluation in Aquaculture VRE and others VRE hosting services for aquafarm performance evaluation.</p>
Dataset archive and preservation
<p>Preservation strategy: The procedures and tools that will be put in place for long-term preservation of the dataset are the same as preservation strategy of its VRE Performance Evaluation in Aquaculture.</p> <p>Preservation tool: The instruments put in place to implement the strategy, are the same as those which are implement the VRE preservation strategy.</p>

3.1.31 SGR_TABLE

Dataset Description
<p>The dataset contains the values of the Specific Growth Ratio (SGR) KPI, which is produced after the statistical modelling that will be done in the VRE at the sampling to sampling dataset. The SGR KPI is a major indicator of growth performance in Aquaculture sector. Specific Growth Ratio (SGR) is used in aquaculture to estimate the production of fish after a certain period. It is a fraction of difference of natural logarithm of final weight and natural logarithm of initial weight by feeding time in days. Formally, $SGR = (\ln W_f - \ln W_o) * 100 / t$. The dataset could be in cross-tabular table or tabular data format. In cross-tabular format the rows present the Average Weight in bins and columns present the Temperature degrees in Celsius. In tabular data format each line is a data record of three (3) columns, namely Average Weight, Temperature and the value of SGR KPI.</p> <p>Generated/Collected: Generated</p> <p>Origin(s): N/A</p> <p>Nature: a) Cross-tabular table, b) Tabular data</p> <p>Scale: a) 30 columns by 20-50 rows (bins of Average Weight), b) up to 1500 rows</p> <p>Potential use: The specific dataset can be used by VRE owner, so as to produce what-if scenarios about the expected growth. Also, a sensitivity analysis can be made using the SGR table.</p> <p>Scientific publication: At the moment the dataset will not be actually underpinning any scientific publication.</p>
Data and Metadata Standards
<p>Data and metadata representation standards: Comma-Separated Values (CSV) files</p>
Dataset Sharing
<p>Access Procedure: A user can have access to the dataset via VRE environment.</p> <p>Embargo Periods: No Embargo Periods</p> <p>Dissemination Mechanisms:</p> <ul style="list-style-type: none"> • BlueBRIDGE Data Catalogue • Other dissemination actions that will take place related with the estimation of performance in an aqua-farming operation and estimating crucial KPIs. <p>Software and tools for re-use: It is cross-tabular table (or tabular data) useful at the production process of any Aquaculture company. Also, it can be used for simulation purposes almost by any Production Management System.</p> <p>Access rights: The dataset is made available only to the VRE owner (Aquaculture company which uses the infrastructure).</p> <p>Licence: Non-commercial Creative Commons (NC) license</p> <p>Repository: Repositories and services underlying the Performance Evaluation in Aquaculture VRE and others VRE hosting services for aquafarm performance evaluation.</p>
Dataset archive and preservation
<p>Preservation strategy: The procedures and tools that will be put in place for long-term preservation of the dataset are the same as preservation strategy of its VRE Performance Evaluation in Aquaculture.</p> <p>Preservation tool: The instruments put in place to implement the strategy, are the same as those which are implement the VRE preservation strategy.</p>

3.2 DATASETS RESULTING FROM ACTIVITIES PERFORMED WITHIN A VRE

VREs are tightly coupled with a rich, heterogeneous and multifaceted data space that that composed by multiple information objects of different nature. Such objects mainly result from the activity performed by VRE members. It consists of (a) the content maintained in the user workspace VRE dedicated folder that is shared among all the VRE members, (b) the set of posts and comments shared through the VRE social networking facilities. The former contains a large variety of objects including files, algorithms and datasets organised in folders. Such material can either pre-exist the VRE and be referred in the Workspace since it is part of the material used by the VRE users to performed their work or be created during the VRE lifetime. The following tables describes the properties of the VRE data space and the plans for its management

3.2.1 VRE CONTENT

Dataset Description
<p>The data space underlying a given Virtual Research Environment includes any data object that is available to VRE members by any of the services equipping the VRE. Such a data space is very heterogeneous and dynamic, it constantly changes during the VRE lifetime. An essential part of this dataset is represented by the data objects that are genuinely produced in the VRE as result of data manipulation and analytics tasks performed by its members</p> <p>Generated/Collected: Generated and Collected</p> <p>Origin(s): It is created by the VRE users within the VRE itself. Some of the generated products are generated by exploiting other information systems and sources. This provenance information is captured by associating appropriate metadata to the specific new object.</p> <p>Nature: The content is heterogeneous. It includes workspace items – single files, sets of files organised in folders, links to objects, compound information objects, e.g. objects representing an entire computation with input data, output data, algorithms, provenance records – as well as posts and comments.</p> <p>Scale: VRE specific. The number of files in the workspace, how big in term of storage they are and the number of posts largely varies according to the VRE application context and usage.</p> <p>Potential use: The primary use of the VRE data space as a whole is by the VRE community members itself. Other potential uses may be possible, e.g., study the patterns and practices governing the functioning of a given community by data mining approaches.</p> <p>Scientific publication: Some of the objects residing in a VRE can support one or more scientific publications.</p>
Data and Metadata Standards
<p>Data and metadata representation standards: Because of the heterogeneity of the data space there is no single standard suitable neither for data nor for metadata. The entire data space is associated with a Dublin Core metadata record. The objects in the data space have their own peculiarities for data and metadata representation. These peculiarities are appropriately exposed via the object themselves.</p>
Dataset Sharing
<p>Access Procedure: The primary access point to the data space is through the VRE services. Clearly, objects that are referenced within the VRE but pre-exist to it may be also be accessed through other object specific alternative channels. VRE users may decide to publish objects through services and with policies that do not require VRE membership to be consumed (e.g., objects are openly available on the BlueBRIDGE global catalogue). Some exploitation scenarios might require that a dump of the data source (or part of it) is created and published thus to enable to access it independently from the VRE services themselves.</p> <p>Embargo Periods: the data space as a whole is not associated with any embargo period while an embargo might apply</p>

<p>to its constituent component.</p> <p>Dissemination Mechanisms: Dissemination is oriented to let potential users to be aware of the VRE content. In case snapshots/dumps of a given VRE are created, they will be published via the BlueBRIDGE overall catalogue by carefully specifying the rationale driving their creation and the time in which the snapshot has been taken.</p> <p>Software and tools for re-use: The data space as a whole can only be consumed by a similar VRE.</p> <p>Access rights: Only VRE authorised members can access the VRE data space content. In the context of the same VRE, diverse users might have different roles and thus have access to certain objects only.</p> <p>Licence: If not specified differently the data space (and its constituents are) is associated with a CC-BY-NC license.</p> <p>Repository: The data space is actually stored by many repositories whose nature depending on the one of its constituents, e.g., geospatial data is stored by the BlueBRIDGE Spatial Data Infrastructure.</p>
Dataset archive and preservation
<p>Preservation strategy: The VRE data space is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies. It is maintained during the entire lifetime of the VRE. Once the VRE is dismissed, the it is made read-only and maintained for a period of time to be agreed with the VRE manager.</p> <p>Preservation tool: The VRE data space is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).</p>

The table above has presented the current management plan for the data space of any VRE. The next table add few more details for the specific VREs creted to support Blue Skill activities.

3.2.2 BLUE SKILLS VRE CONTENT

Dataset Description
<p>This Blue Skill VREs contains course material produced by the instructors (e.g., slides, programmes, notes, tests and exercises) as well as material produced by course attendants (e.g., assignments, notebooks, reports).</p> <p>Generated/Collected: Generated and Collected</p> <p>Origin(s): See 3.2.1</p> <p>Nature: See 3.2.1</p> <p>Scale: See 3.2.1, it largely depends on the duration and complexity of the course as well as on the number of participants.</p> <p>Potential use: The primary use is the course, however other uses can be envisaged, e.g., to analyse the behaviour of actors involved in a course, to systematically assess students' performance.</p> <p>Scientific publication: N/A for the course, while dumps of the dataset can be exploited in certain studies.</p>
Data and Metadata Standards
<p>Data and metadata representation standards: See 3.2.1</p>
Dataset Sharing
<p>Access Procedure: See 3.2.1</p> <p>Embargo Periods: See 3.2.1</p> <p>Dissemination Mechanisms: See 3.2.1. In this case, the primary object of dissemination is expected to be the course (either before it occurs or after).</p>

<p>Software and tools for re-use: See 3.2.1</p> <p>Access rights: See 3.2.1. Different VRE members can have different access rights on the data space, e.g., the instructor has access to the attendants' folders (or part of it) only when granted by the owner(s), attendants have access only to the course material folder other their own.</p> <p>Licence: See 3.2.1. Licences may be different.</p> <p>Repository: See 3.2.1</p>
Dataset archive and preservation
<p>Preservation strategy: See 3.2.1.</p> <p>Preservation tool: See 3.2.1.</p>

3.3 PROJECT OUTPUTS

The activities of the project produce content by themselves that will have to be manage as any other data products. Below we list different type of content and the current plan to manage it.

3.3.1 DELIVERABLES

Dataset Description
<p>This dataset is the collection of the BlueBRIDGE project deliverables.</p> <p>Generated/Collected: Generated</p> <p>Origin(s): N/A</p> <p>Nature: PDF and Word files.</p> <p>Scale: less than 100 objects. Each object is few MBs.</p> <p>Potential use: to be informed on project activities and results.</p> <p>Scientific publication: N/A</p>
Data and Metadata Standards
<p>Data and metadata representation standards: A Dublin Core record will be associated with each item.</p>
Dataset Sharing
<p>Access Procedure: A VRE (BlueBRIDGE Project VRE) has been created for supporting the project operation. The workspace of this VRE contains a folder where deliverables are stored as soon as they are produced (actually submitted to EC). Access to this folder is granted to project members only. However, "public links" can be generated for each deliverable and used to enable "guest users" to access them.</p> <p>Embargo Periods: N/A</p> <p>Dissemination Mechanisms: Deliverables are published via the project website as well as via OpenAIRE and beneficiaries' Institutional repositories.</p> <p>Software and tools for re-use: N/A</p> <p>Access rights: Depends on the per deliverable dissemination level, certain deliverables are public others are restricted.</p> <p>Licence: If not specified differently, deliverables have a CC-BY-NC license.</p> <p>Repository: Deliverables are actually stored in several repositories including the workspace of the VRE created to</p>

support the project and the Institutional repositories of project beneficiaries.
Dataset archive and preservation
Preservation strategy: Deliverables are preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely they are stored in multiple copies.
Preservation tool: Deliverables are preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).

3.3.2 DISSEMINATION AND TRAINING MATERIAL

Dataset Description
<p>This dataset is the collection of training and dissemination material produced by the BlueBRIDGE project.</p> <p>Generated/Collected: Generated</p> <p>Origin(s): N/A</p> <p>Nature: Slides, videos, posters, fliers, news, etc. developed for the dissemination and outreach. Slides, videos, wiki pages, and documentation (both in French and in English) developed for the Training workshops and webinars organised as part of Task 3.4.</p> <p>Scale: The dataset will grow during the project lifetime however it will not belong to the “big data” domain.</p> <p>Potential use: To disseminate the project outcomes and impact. To inform VRE users on: (i) the access and usage of the domain specific services and on the business benefits of the e-infrastructure solutions and market perspectives; (ii) how VREs can be created and maintained.</p> <p>Scientific publication: N/A</p>
Data and Metadata Standards
Data and metadata representation standards: Dublin Core
Dataset Sharing
<p>Access Procedure: This material will be made available through the BlueBRIDGE Project Website.</p> <p>Embargo Periods: The material will be made available as soon as it is produced.</p> <p>Dissemination Mechanisms: The material will be disseminated through the BlueBRIDGE Project Website and other public or partners’ specific channels.</p> <p>Software and tools for re-use: none;</p> <p>Access rights: Open Access</p> <p>Licence: CC Attribution alone (CC-BY);</p> <p>Repository: This material is actually stored in several repositories including the workspace of the VRE created to support the project, the Institutional repositories of project beneficiaries and third party services (e.g. SlideShare).</p>
Dataset archive and preservation
<p>Preservation strategy: Dataset content is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely content is stored in multiple copies. Moreover, part of this material will be published on third-party services (e.g. slides on SlideShare).</p> <p>Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec.</p>

2.2.3).

3.3.3 SCIENTIFIC PAPERS

Dataset Description
<p>This dataset is the collection of the BlueBRIDGE scientific publications.</p> <p>Generated/Collected: Generated</p> <p>Origin(s): N/A</p> <p>Nature: Mainly PDF files yet they can be accompanied by supplemental material including datasets.</p> <p>Scale: Tens of scientific publications are expected to be produced.</p> <p>Potential use: To disseminate the project outcomes and results.</p> <p>Scientific publication: N/A</p>
Data and Metadata Standards
<p>Data and metadata representation standards: Dublin Core, BibTex</p>
Dataset Sharing
<p>Access Procedure: Scientific papers are mainly available / accessible via the Publishers' web site according to the associated access method. BlueBRIDGE promotes "Open Access" thus a machine-readable electronic copy of every publication is expected to be deposited in suitable Open Access repositories.</p> <p>Embargo Periods: The material will be made available as soon as it is produced.</p> <p>Dissemination Mechanisms: Scientific publications will be disseminated through the BlueBRIDGE Project Website as well as through scholarly communication channels, e.g., Publishers/Journals web sites, Institutional Repositories, scholarly communication networks (ResearchGate, Google Scholar).</p> <p>Software and tools for re-use: N/A;</p> <p>Access rights: Open Access</p> <p>Licence: CC Attribution alone (CC-BY);</p> <p>Repository: Beneficiaries must deposit the published version or the final peer-reviewed manuscript accepted for publication in at least an "OpenAIRE compliant" repository. Authors may rely on their Institutional Repositories (if any) as well as on Zenodo. Moreover, a copy of each paper must be uploaded in the project workspace.</p>
Dataset archive and preservation
<p>Preservation strategy: Publications produced in the context of the project is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely they are stored in multiple copies. Moreover, papers will be managed by Publishers according to their preservation strategies.</p> <p>Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).</p>

3.4 INFRASTRUCTURE OPERATION DATASETS

In order to operate the infrastructure a number of data catalogues must be maintained. Below we list the current management plans for the more relevant ones.

3.4.1 BLUEBRIDGE DATA CATALOGUE

Dataset Description
<p>This dataset represents the content of the BlueBRIDGE overall data catalogue. It mainly consists of the collection of the metadata of the datasets published by the catalogue.</p> <p>Generated/Collected: Generated and Collected</p> <p>Origin(s): some of the metadata are produced by collecting them from other catalogues, e.g., the BlueBRIDGE Geospatial Data Catalogue (cf. Sec. 0).</p> <p>Nature: mainly metadata are in the form of (key, value) pairs.</p> <p>Scale: It depends on the granularity of the datasets to be published. However, the catalogue will likely contain hundreds thousands entries / datasets.</p> <p>Potential use: To be informed on the data products available in the BlueBRIDGE domain.</p> <p>Scientific publication: N/A</p>
Data and Metadata Standards
<p>Data and metadata representation standards: Dublin Core / proprietary records.</p>
Dataset Sharing
<p>Access Procedure: The BlueBRIDGE Catalogue Service is exposed through a web GUI and a web API. Certain parts of the catalogue are accessible to authorised groups / members only (e.g., datasets having dissemination level restricted to VRE members are accessible only to them).</p> <p>Embargo Periods: According to the specification provided by the dataset owner.</p> <p>Dissemination Mechanisms: The BlueBRIDGE Catalogue Service URL will be disseminated by the project website via a dedicated page on “data products” as well as via any other suitable channel.</p> <p>Software and tools for re-use: N/A;</p> <p>Access rights: part of the dataset are accessible only to authorised users.</p> <p>Licence: metadata of the datasets are associated with a CC-BY-NC licence;</p> <p>Repository: N/A The dataset is tightly coupled with the service state.</p>
Dataset archive and preservation
<p>Preservation strategy: Dataset content (i.e. datasets metadata) is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely back-up copies are created.</p> <p>Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).</p>

3.4.2 BLUEBRIDGE GEOSPATIAL DATA CATALOGUE

Dataset Description
<p>This dataset represents the content of the BlueBRIDGE geospatial data catalogue. It mainly consists of the metadata of the datasets published by the catalogue.</p> <p>Generated/Collected: Generated and Collected</p>

<p>Origin(s): some of the metadata are produced by collecting them from other catalogues, e.g., the FAO Geonetwork.</p> <p>Nature: mainly metadata in ISO 19139 format</p> <p>Scale: Depending on the granularity of the datasets to be published. However, the catalogue will likely contain hundreds thousands entries / datasets.</p> <p>Potential use: To be informed on the geospatial data products available in the BlueBRIDGE domain.</p> <p>Scientific publication: N/A</p>
Data and Metadata Standards
Data and metadata representation standards: Dublin Core / proprietary records.
Dataset Sharing
<p>Access Procedure: The BlueBRIDGE Catalogue Service is exposed through a web GUI and a web API. Certain parts of the catalogue are accessible to authorised groups / members only (e.g., datasets having dissemination level restricted to VRE members are accessible to them).</p> <p>Embargo Periods: According to the specification provided by the dataset owner.</p> <p>Dissemination Mechanisms: The BlueBRIDGE Geospatial Data Catalogue Service URL will be disseminated by the project website via a dedicated page on “data products” as well as via any other suitable channel.</p> <p>Software and tools for re-use: N/A;</p> <p>Access rights: part of the dataset are accessible only to authorised users.</p> <p>Licence: metadata of the datasets are characterised by CC-BY-NC licence unless explicitly stated.</p> <p>Repository: N/A The dataset is tightly coupled with the service state.</p>
Dataset archive and preservation
<p>Preservation strategy: Dataset content (i.e. datasets metadata) is preserved according to the BlueBRIDGE standard preservation procedure (cf. Sec. 2.2.3), namely back-up copies are created.</p> <p>Preservation tool: Dataset content is preserved by relying on the BlueBRIDGE standard preservation tools (cf. Sec. 2.2.3).</p>

3.5 SOFTWARE

Software is another project artefact that needs to be managed in order to make it discoverable, accessible and reusable by third-parties. Below the solutions identified so far for both software artefacts and source code produced in the context of the BlueBRIDGE project.

3.5.1 GCUBE SOFTWARE ARTEFACTS

Dataset Description
<p>This dataset refers to the software releases of the gCube system. The software is organized in components. The dataset includes all the versions of all components.</p> <p>Generated/Collected: Generated</p> <p>Origin(s): It is generated from the source code (cf. Sec. 3.5.2) of gCube components. The code is compiled, tested, packaged and verified and the resulting artefacts are finally published in this dataset.</p> <p>Nature: It contains binary artefacts of the various gCube software components. For each artefact also source code</p>

documentation packages and source code packages are available.

Scale: It contains about 10,000 software artefacts plus, approximatively, the same number of documentation artefacts. The current growth rate is about 600 new entries for each gCube release.

Potential use: It is required by anybody that needs to create or update a gCube infrastructure.

Scientific publication: N/A

Data and Metadata Standards

Data and metadata representation standards: gCube software is written in Java and managed with the Apache Maven build tool. The format of software artefacts is standard byte-compiled Java code packaged in one of the standard formats jar, war, gar (depending on the nature of the artefact). The documentation artefacts are standard html pages packaged in jar files. Finally, the metadata follows the Maven Repository Metadata specifications (<http://maven.apache.org/ref/3.2.5/maven-repository-metadata/>).

Dataset Sharing

Access Procedure:

- Apache Maven tool
- Web at <http://maven.research-infrastructures.eu/nexus/content/repositories/gcube-releases/>
- gCube Website: <https://www.gcube-system.org/software-releases>

Embargo Periods: N/A

Dissemination Mechanisms: Any software updated (at each gCube release) is announced in the official gCube website with the publication of release notes and released artefacts.

Software and tools for re-use: the main tool that can be used to consume the it is Apache Maven.

Access rights: It is publicly available for consultation and download. Only authorized people have rights to add new entries.

Licence: gCube code is released under the EUPL (<http://ec.europa.eu/idabc/eupl>) licence.

Repository: It is published in a Maven repository called "gcube-releases" hosted by a Nexus server at: <http://maven.research-infrastructures.eu/nexus/content/repositories/gcube-releases/>

Dataset archive and preservation

Preservation strategy: N/A

Preservation tool: N/A

3.5.2 GCUBE SOURCE CODE

Dataset Description

The source code of the gCube system is organized per component. It supports versioning and all versions of the code are maintained.

Generated/Collected: Generated

Origin(s): It is generated by developers of the gCube system.

Nature: It contains the source code of the various gCube software components. Majority of dataset is made by Java source code files and configuration files. For web components also image files are included.

Scale: It maintains source code of about 500 active components. All changes are maintained for each component. The

<p>current size of all gCube active components is about 1.5GB of source code.</p> <p>Potential use: It can be used by anybody that needs to work with gCube source code, for example to inspect the code, fix defects, maintain and evolve the gCube components.</p> <p>Scientific publication: no publications are available at the moment. To support future publications, the source code will be made available in the Zenodo platform to easily reference to it.</p>
Data and Metadata Standards
<p>Data and metadata representation standards: All gCube software is written in Java and managed with the Apache Maven build tool. Therefore, format of source code is for the vast majority Java source files. In addition, xml configuration files and images are the other two predominant formats. Concerning metadata, each component has a project descriptor that follows the Maven POM specifications https://maven.apache.org/pom.html.</p>
Dataset Sharing
<p>Access Procedure:</p> <ul style="list-style-type: none"> • web http://svn.research-infrastructures.eu/public/d4science/gcube/ • the SVN (or equivalent) tool <p>Embargo Periods: N/A</p> <p>Dissemination Mechanisms: When the dataset is updated (at each gCube release) it is announced in the official gCube website with the publication of release notes and links to the source code.</p> <p>Software and tools for re-use: main tool to be used to use the data is Subversion.</p> <p>Access rights: The dataset is publicly available for consultation and download. Only authorized people have rights to write to the dataset.</p> <p>Licence: EUPL http://ec.europa.eu/idabc/eupl</p> <p>Repository: This dataset is published in a SVN repository hosted at http://svn.research-infrastructures.eu/public/d4science/gcube/</p>
Dataset archive and preservation
<p>Preservation strategy: The plan to preserve the gCube source code is to migrate it to the GitHub.com platform.</p> <p>Preservation tool: N/A</p>

4 CONCLUSION AND NEXT STEPS

This deliverable is the first version of the BlueBRIDGE Data Management Plan (the second one, D2.2, will be produced at Month 18 and the third one, D2.3, at Month 27). It contains an initial description of the datasets collected, processed or generated by the project and an initial plan on how the sharing, archiving and preservation of these datasets will be guaranteed.

For the sake of the Data Management description the BlueBRIDGE data space has been clustered into:

Basic Datasets, i.e., datasets worth to have in the BlueBRIDGE data space because of their potential use when building one or more Virtual Research Environment;

Datasets Resulting from Activities performed within a VRE, i.e., datasets resulting from the operation of a Virtual Research Environment;

Project Output, i.e., datasets resulting from the operation of the BlueBRIDGE project and produced by the BlueBRIDGE consortium;

Infrastructure Operation Datasets, i.e., datasets resulting from the operation of the BlueBRIDGE supporting infrastructure, datasets underlying infrastructure services; and

Software, i.e., datasets resulting from the software enabling BlueBRIDGE.

The presented plan will be refined and extended in the next months also taking into account the technological progresses done in WP9 concerning Data Publishing. In the context of this activity efforts will be done to facilitate data discoverability, accessibility, and re-use.

REFERENCES

- [1] Assante, M., Candela, L., Castelli, D. and Tani, A. (2016) Are Scientific Data Repositories Coping with Research Data Publishing? *Data Science Journal*, 15:6, pp. 1–24, DOI: 10.5334/dsj-2016-006
- [2] Candela, L., Castelli, D., Michel Assoumou, J., Pagano, P. and Zoppi, F. (2015) Quality Plan. BlueDRIDGE Project Deliverable. D2.1
- [3] Genova, F., Hanahoe, H., Laaksonen, L., Morais-Pires, C., Wittenburg, P. and Wood, J. (2014) The Data Harvest Report: How sharing research data can yield knowledge, jobs and growth. RDA Europe Report
- [4] Piwowar, H.A., Vision, T.J. and Whitlock, M.C. (2011) Data archiving is a good investment. *Nature*, Vol. 473, 285/285; 10.1038/473285a
- [5] Piwowar, H.A., Day, R.S. and Fridsma, D.B. (2007) Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2(3): e308; 10.1371/journal.pone.0000308
- [6] Renear, A.H., Sacchi, S. and Wickett, K.M. (2010) Definitions of dataset in the scientific and technical literature. *Proceedings of the American Society for Information Science and Technology*, 47(1): 1–4. DOI: 10.1002/meet.14504701240