BlueBRIDGE

Understanding how ecosystems of e-infrastructures can support Blue Growth

11 January 2017, Brussels, Belgium

Executive Summary

The BlueBRIDGE H2020 project[1] recently convened a one-day workshop in Brussels entitled "Understanding how ecosystems of e-infrastructures can support Blue Growth"[2].

The aim of the event was to showcase how the Blue Growth sector is benefitting from services that the BlueBRIDGE project has to offer. The latter primarily regard the combination of existing e-infrastructure services with an easy to use interface. The workshop also provided the opportunity to further discuss opportunities related to data management with e-Infrastructures.

The workshop brought together stakeholders operating in the Blue Growth sector, all with data management challenges needing integrated solutions.

The stakeholders ranged from international fisheries organisations, to organisations providing scientific advice and recommendations to policy makers and governments on marine protected areas and fish stocks, to private enterprises providing support to aquafarms in evaluating their economic performance, or in detecting areas to locate cages, to academics and scientists supporting international fisheries commissions in analysing marine species or performing environmental, societal and nutritional research. Representatives from the European Commission (DG CONNECT, DG JRC, DG MARE, DG ENVIRONMENT, DG RTD) were also present. The full list of participants is available at the following link http://www.bluebridge-vres.eu/participants.

The Workshop discussion focused on three main themes:

1. How data-driven science is shaping Blue Growth practices: Practical Insights from BlueBRIDGE Users
2. How European e-infrastructures can support Blue Growth related data science and innovation
3. How the Blue Growth sector can benefit from the European Open Science Cloud.

The stakeholders operating in the Blue Growth sector opened the workshop by highlighting how it can become unpractical for individual organisations to implement data management solutions.

1 BlueBRIDGE (Building Research environments fostering Innovation, Decision making, Governance and Education to support Blue growth) receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 675680, www.bluebridge-vres.eu.

2 http://www.bluebridge-vres.eu/agenda-bluebridge-workshop-11-january-2017-brussels-belgium

The rapid developments of on-line data services and exponential growth of data quantity require **collaborative and computing intensive solutions** that can only be available through e-Infrastructures. A characteristic of e-Infrastructures is the re-use of existing solutions. **Re-use of existing services is an effective way to save time and money**. Indeed, this is the approach adopted by the BlueBRIDGE project and the thread of the presentations of the event.

BlueBRIDGE delivers specialised data services for aquaculture farm management, the ecosystem approach to fisheries, spatial data analysis, and marine spatial planning. These services are operated through Virtual Research Environments[3] built on top of a hybrid-data infrastructure[4]. This hybrid-data infrastructure is capable of dynamically federating computing and storage resources coming from third party providers[5], datasets belonging to heterogeneous data sources, analytical tools developed by different organisations, and to offer all of these services in a unique collaborative environment.

Several users have already adopted BlueBRIDGE services to solve their data management issues and have provided really positive feedback.

WHAT THEY SAY ABOUT US

Anton Ellenbroek, IT Specialist at the Food and Agriculture Organization (FAO) of the United Nations: *"BlueBRIDGE is a comprehensive platform that can help you to manage users, data, and software in a single environment. It provides dedicated data services for data access, management, and analysis; services related to stock assessment and analysis of geospatial features that can be combined in adaptable products"*

Scott Large, Professional Officer at the International Council for the Exploration of the Sea (ICES): "*Integrating the local computing environment with e-infrastructure enables users to remain in a familiar space, while still having access to services that make e-infrastructures so attractive and useful*"

Miles Mac-Millan Lawler, Programme Leader at GRID Arendal "*BlueBRIDGE is helping us with the collation and access to relevant data and metadata; it has speed up our processing time - seconds and minutes instead of hours and days! … and most importantly the BlueBRIDGE infrastructure can be accessed by external projects (e.g. JRC BIOPAMA Project)*"

Kostas Seferis, CEO at I2S "*BlueBRIDGE can help aquafarming companies in building powerful instruments for estimating their financial performances, allowing them to acquire the necessary knowledge to take decisions on future investments*"

Paul Taconet, Engineer at the Institut de recherche pour le développement (IRD) "*The main benefit that we see in the usage of BlueBRIDGE is its capability to providing a single environment where different researchers and scientists can collaborate, access data and computing resources and share their research results in a transparent way. The possibility of integrating the BlueBRIDGE services with our partner websites (such as the IRD website or the tuna RFMOs websites) is also a great advantage. We can re-use existing services without developing them from scratch*"

Currently, BlueBRIDGE is the only European initiative providing the Blue Growth community with a unique solution to cover the full data management workflow (from data collection to publication).

3 http://www.bluebridge-vres.eu/news/vre-three-letters-enhance-cooperation-and-collaboration-amongst-researchers-modern-science

4 D4Science, www.d4science.org

5 BlueBRIDGE has already established fruitful collaborations with existing European e-infrastructures service providers [such as EGI.eu, the e-infrastructure that provides advanced computing services for research, and EUDAT, the European Collaborative Data Infrastructure that provides a common model and service infrastructure for managing data spanning all European research data centres and community data repositories] and is seeking more collaborations.

The services developed by BlueBRIDGE are thus uniquely positioned to constitute a cornerstone for the establishment of the upcoming European Open Science Cloud (EOSC)[6].

EOSC was introduced to the workshop by Victoria Tsoukala, Program Officer-Seconded National Expert e-Infrastructures and Open Science Cloud at European Commission, DG CONNECT. *"EOSC is part of the Digital Single Market strategy and therefore it contributes to the Blue Growth strategy[7], which is largely focused on exploiting marine and aquaculture data to boost the economy and the marine sector. EOSC will enable data interoperability and synergy of existing resources and services. It will provide cloud-based resources for SMEs, researchers, entities and innovators involved in the Blue Growth sector. It will enable a single environment that will allow regions to have access to different data and information. In this respect,* ***I can see BlueBRIDGE and its services very much aligned with the vision towards the EOSC and I can see how, in collaboration with other members of the e-infrastructure community, BlueBRIDGE could offer some of its services as part of the EOSC list of services****"*.

Further considerations on the role of data-driven science and e-infrastructures for the Blue Growth sector and on the benefits brought to the sector by EOSC were discussed in the event's final panel, the findings of which are summarized in the following:

The role of data-driven science and innovation for the Blue Growth sector

1. At every step of the knowledge production chain, reliable data are needed.
2. Society demands more and more detailed information on environmental, societal, and nutritional issues: this requires careful and integrated data management solutions.
3. Innovation in the data chain is often most effective when SMEs are engaged at an early stage
4. Scientific advice and recommendations can only be based on data coming from the Blue Growth sector if stakeholders can and will open their data collections
5. There needs to be a concerted effort to raise awareness of both the short and the long-term benefits that Open Science can bring these key stakeholders. This is especially true in the aquaculture sector, where the main players are SMEs that are not accustomed to data sharing and open science practices.

How can European e-infrastructures support Blue Growth related data science and innovation

1. Horizontal e-infrastructures are essential, but by themselves are not sufficient. They need to be complemented by thematic infrastructures that serve the specific needs of communities. This is the approach adopted by BlueBRIDGE. Today, BlueBRIDGE is the only initiative serving the Blue Growth community with data management services covering the full data production chain.
2. Federating resources for communities requires the development of easy to use interfaces that hide integration technicalities. This is particularly needed in the Blue Growth sector where users are very heterogeneous and include SMEs, policy advisors and marine experts who do not necessarily have coding skills.
3. Provenance metadata is becoming more and more important especially when providing

6 The European Open Science Cloud (EOSC) is the initiative launched by the European Commission with the aim to make "science more efficient and productive and let millions of researchers share and analyse research data in a trusted environment across technologies, disciplines and borders". http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

7 https://ec.europa.eu/commission/priorities/digital-single-market_en

fisheries policy advice and recommendations: understanding which data, methods and algorithms were used for the fisheries data analysis is key for the transparency of the results and for the repeatability of experiments.

4. An effective data collection system needs to be in place to ensure data availability.
5. Too many e-infrastructure services are the result of short-term funded projects and are not fully sustained beyond the lifetime of the project. There is a growing need for such projects to provide robust sustainability plans, which envisage the establishment of legal entities for potential financial transactions and for exchange of expertise to begin to happen.

How can the Blue Growth sector benefit from the European Open Science Cloud (EOSC)[8]

1. The European Commission plans to set up a "Blue Cloud pilot" within EOSC that will rely on three technological layers:
   a. Layer 1: Thematic infrastructures such as Copernicus and EMODnet must be at the foundations of such a pilot, as they will provide the data to fuel the process.
   b. Layer 2: Horizontal e-infrastructures providing basic services will be the engine to compute and store data.
   c. Layer 3: Thematic e-infrastructures built on top of the horizontal e-infrastructures, like the one underlying BlueBRIDGE, must be part of such a pilot to provide the users with customized services and user friendly interfaces. These are particularly fundamental in the Blue Growth sector, where scientists, researchers, policy advisors and SMEs need customized services to solve specific problems and do not always have the skills to directly use the horizontal infrastructure services.
2. The challenges that EOSC is going to deal with are too complex to be addressed by a small group of people. There is a strong need for collaboration.
3. A healthy ecosystem of e-infrastructures is exactly what EOSC aims to achieve. To meet this objective, the complementarities of each e-infrastructure need to be identified.

8 http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

Workshop report

The Workshop Report is structured as follows:

- Section 1 presents the BlueBRIDGE offer and approach.
- Section 2 describes how users from the Blue Growth sector are benefitting from the adoption of the BlueBRIDGE services.
- Section 3 presents the main findings of the final panel discussion focusing on the role that data-driven science and innovation plays in the Blue Growth sector, on how European e-infrastructures concretely facilitate this role and on how the Blue Growth sector will benefit from the European Open Science Cloud.

1. BlueBRIDGE in a nutshell

The BlueBRIDGE event showed data services in support of the entire **knowledge production chain** from data collection to data aggregation, including the analysis and the production of indicators and their publication.

The stakeholders that BlueBRIDGE addresses are very heterogeneous: they span from large European and international institutions, whose mission is to provide advice on sustainable use of marine resources to relevant commissions and member states, to local academic institutions and to small and medium private enterprises (SMEs) that deal with data management issues on a daily basis.

The core of the BlueBRIDGE approach is to open the "silos" where data and results produced in one domain are owned by that domain and are difficult to share with others. The Blue Growth sector has stakeholders belonging to different sectors (research, education and business). The idea of BlueBRIDGE is to **facilitate the sharing of data and results as much as possible and to share resources made available by one sector with others to avoid duplication of effort.**

BlueBRIDGE aims to fill the gap between the e-infrastructure specialists and communities. Over the past years, the e-infrastructure sector has focused on building horizontal data access and processing solutions. **BlueBRIDGE wants to build data management services on top of the existing e-infrastructure solutions in order to serve specific needs of the Blue Growth community.**

Solving the data management issues of the Blue Growth community usually requires the involvement of skills from different domains. This is why BlueBRIDGE builds upon the concept of the "**collaborative environment**".

BlueBRIDGE provides users with collaborative Virtual Research Environments (the VREs) where they can access resources and services (data, computing & storage resources, analytics tools) that are part of the underlying BlueBRIDGE hybrid data infrastructure, D4Science. These resources and services are offered to the users in a dynamic fashion (users can customize their environment) and through a user-friendly interface that hides the complexity related to the management of an infrastructure. In this way, users can focus on the problem that they have to solve.

D4Science is a self-sustained[9] hybrid data infrastructure executing around 60,000 models & algorithms per month and providing access to over a billion records hosted in more than 50 worldwide repositories. Currently, D4Science serves over 2,700 users from multiple scientific domains (e.g. fisheries, biodiversity, ocean observation, etc.). As mentioned before, the added-value of D4Science is that it is a framework in which infrastructure resources (e.g. data and services) made available by different data infrastructures can be dynamically packaged to serve the needs associated with particular scientific or societal questions. All of this is completely transparent for the user.

Currently, the infrastructure can access data from Copernicus, SeaDataNet, EMODnet, OBIS (Ocean Biogeographic Information System) and many other data providers and relies on the computational resources provided by EGI. Besides using all these resources, the users can also easily integrate their own software, models and algorithms in the infrastructure to perform particular tasks. In addition, they have the possibility to make available these resources and services available through their own portals.

9 D4Science is operated and maintained by the Italian National Council of Research (ISTI-CNR).

All these features make BlueBRIDGE unique. BlueBRIDGE is the demonstration that concrete services can be built through ecosystems of e-infrastructures.

The picture below gives an overview of what BlueBRIDGE has to offer.

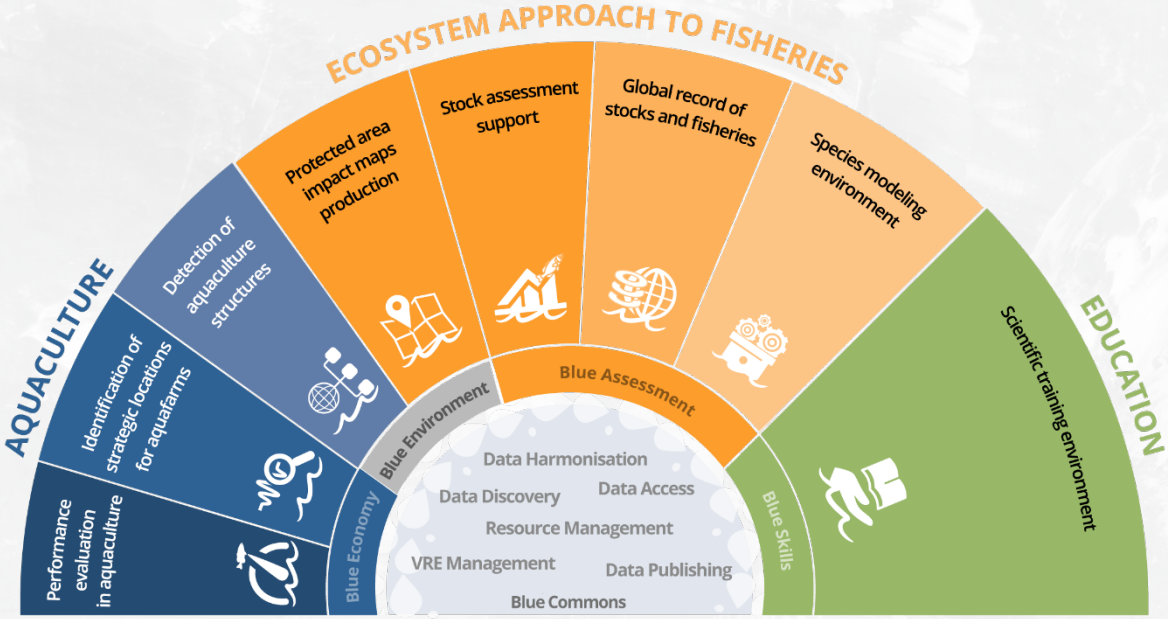


Figure 1: BlueBRIDGE Service Portfolio

2. Data-driven science shaping Blue Growth practices: Practical Insights from the BlueBRIDGE Users

The **target users of the BlueBRIDGE services** are international organisations providing scientific advice and recommendations to policy makers and governments on marine protected areas and fish stocks, small private enterprises supporting aquafarms in evaluating their techno-economic performances or in detecting strategic areas where to locate their cages, academics and scientists supporting international fisher commissions analysing marine species trends, or performing environmental, societal and nutritional research.

They all need integrated data production workflows providing:

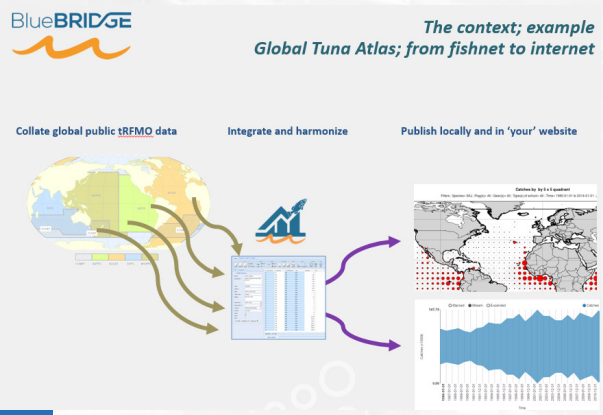
- **advanced data harmonization tools**, for example, to map reference schemes to Master Data including geospatial references
- **time series & spatial analysis services** allowing them, for example, to host and expose R/WPS based models
- **higher powered computing resources**, for example, to run ecosystem models, mixed-fisheries models, single-stock forecasts and management strategy evaluations for the provision of policy advice
- **reporting, communication and dissemination facilities** able to provide a bridge across different data contributors and across different publication formats.

BlueBRIDGE users recognise that these improved analytics and processing of data can improve the delivery of research results, products and services leading to greater economic growth and sustainability.

Five use cases illustrating how BlueBRIDGE can support the needs of the Blue Growth community were presented in the workshop and they are reported in this section.

Data Driven Science and BlueBRIDGE, a match made in the cloud, Anton Ellenbroek, IT Specialist, Food and Agriculture Organisation (FAO) of the United Nations[10]

The Fisheries Statistics and Information Branch of the Food and Agricultural Organizations of the United Nations (FIAS-FAO of the UN) intends to use the BlueBRIDGE project to promote the Ecosystems Approach to Fisheries and Aquaculture. It aims to deliver improved data services for collection and collation of statistics, and to have a collaborative environment to manage the entire data chain for fisheries statistics; from the fishnet to the internet. This includes services for fisheries data harmonization using FAO and other Master Data Management services, on-line and multi-user stock assessment services, visualization of fisheries data, and the inclusion of this data in a joint Global Record of Stocks and Fisheries, where a global knowledge base is being developed. For aquaculture inventories, FAO works with BlueBRIDGE to develop inventories of aquaculture sites using the semi-automated detection of features from EU and global satellite data.



This image describes how in the final solution, data are collected

10 http://www.slideshare.net/BlueBridgeVREs/03-bb-advantages-to-bg-ae

from, for example, globally distributed fisheries organizations, and harmonized using shared and re-usable data services to, for example, align with global reference data for gear and country classifications. After data are harmonized, BlueBRIDGE services can then offer analytical services such as CMSY or SS3 for stock assessment, or use the data in a rich data visualization portal. Both service types are not exclusive to FAO, but also of value to other project partners such as IRD and ICES. The data are also not internal to BlueBRIDGE, but with the service oriented approach, are easily visualised in web-portals that can be embedded in 'any' site.

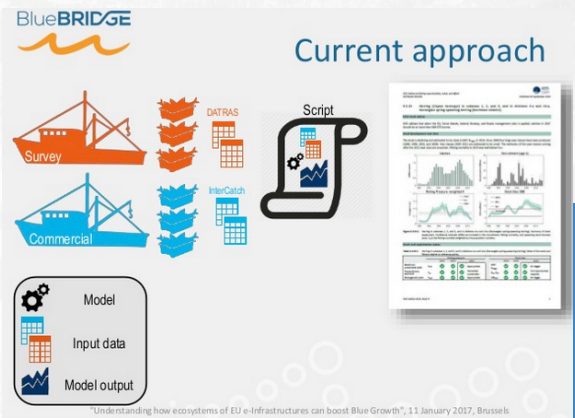
FAO is further contributing with data and expertise to the development of the Global Record on Stocks and Fisheries that will provide a dynamic and interactive collection of existing information from three currently separated information systems. The data services are embedded in the D4Science infrastructure, and are published in a catalogue. This opens perspectives to enrich the catalogue with, for example, information gathered in the data collation and stock assessment tasks; information on catch and landings in harmonized formats, and stock analysis results are currently very difficult to link to stock status assessment reports, and it requires a lot of effort to reproduce an assessment result, if possible at all. The GRSF, together with the stock assessment data services can provide this unique integrated service.

For the detection of often remote and temporary aquaculture sites, the use of semi-automated detection algorithms mining features from large EU and global satellite datasets is developed to improve FAO's understanding of the location and structure of aquaculture. The tool that is being developed not only detects sites, but also calculates the number of cages and their coverage. This reduces the time of the analysis significantly, and promises to be re-applicable to other typologies since the ontology to detect the features can be trained to detect other features as well.

How ICES uses BlueBRIDGE tools to overcome their challenges, Scott Large, Professional Officer, International Council for the Exploration of the Sea (ICES)[11]

The mission of the International Council of the Exploration of the Sea (ICES) is to provide advice on the sustainable use of living marine resources and protection of the marine environment. In particular, ICES provides stock fisheries advice (they currently have 263 fish stocks across the northeast Atlantic) and ecosystems and fisheries overviews exploiting a set of scientific tools such as ecosystem models, mixed fisheries models, single stock forecasts and management strategy evaluation. All of this requires lot of different models, usually developed by different scientists, and also the computational capacity to run these models.

Currently, to provide single stock fish advice, ICES collects data from different sources: for survey data they use the DATRAS database and for commercial data they get data from regional data providers that then goes into the InterCatch database. From these databases they get a script for each of their 263 different stocks that are elaborated and transformed into a suitable format for the stock assessment model. Once the stock assessment model has ran, a report is produced as an output. This process is very compartmentalised and therefore requires a lot of manipulation from scientists and as such is not as fully transparent as ICES would like it to be.



11 http://www.slideshare.net/BlueBridgeVREs/04-ices-brusselsworkshoplarge

The optimal process would be to have input data pass to a single script in order to generate the output in a transparent way. ICES is currently working on a product called the ICES Transparent Assessment Framework[12] that aims to make the full process collaborative, easier, flexible and integrated and able to manage provenance.

To achieve the desired workflow ICES is exploiting the BlueBRIDGE capabilities: the different models that ICES is using for the single stock assessment are now part of the BlueBRIDGE infrastructure and therefore they can be ran in parallel, optimizing some time. Additionally the R code that the scientists used to run on their desktops is now submitted to the BlueBRIDGE e-infrastructure that is able to process it and send back the results to the scientists who no longer have to deal with the code and all its related technicalities. Integrating R with BlueBRIDGE brings many advantages to the scientists: first of all, it allows them to manage provenance; to enable extant R scripts to be easily modified and to reduce the overhead for the scientists to fully comply with service and generalization requirements.

Porting the ICES stock assessment workflow on the BlueBRIDGE e-infrastructure allows ICES to streamline the full process and to execute the work in a smarter way. Having a transparent process in place, where the technological complexity is completely managed by BlueBRIDGE is a great advantage: it allows ICES to reduce the training time of its scientists and improve their productivity, as they can really only focus on scientific development instead of the technological complexity of the process.

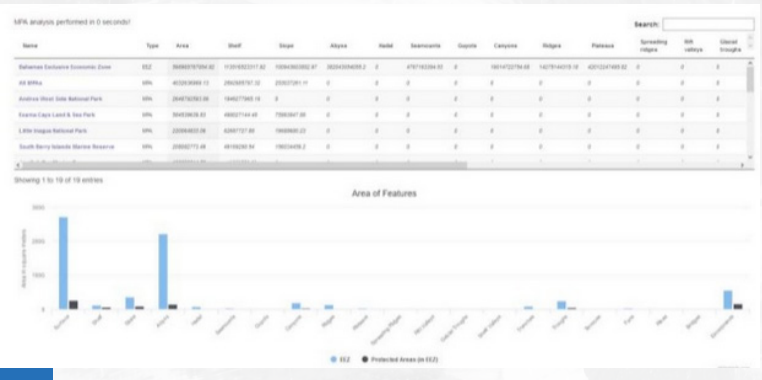
BlueBRIDGE: Supporting Maritime spatial planning through provision of data and analysis, Miles Macmillan-Lawler, Programme Leader, GRID-Arendal[13]

Organisations and private companies dealing with maritime spatial planning require more and more information on the location of activities for human uses and how these relate to the environment in order to inform decision makers properly. Some of the key challenges include incomplete or non-existent and not accessible (in one place) information on the location of human uses (for example aquaculture facilities) and time consuming processes to collate and analyse the relationships between human uses, management and environment.

GRID-Arendal is a centre collaborating with the United Nations Environment Programme (UNEP). GRID-Arendal together with CLS (Collecte Localisation Satellite, France) joined BlueBRIDGE to address two main challenges:

1. to develop rapid inventories of aquaculture facilities (Aquaculture Atlas for Greece and Indonesia);
2. to develop tools to assess which ecological features are in marine managed areas (for example Marine Protected Areas, MPAs, of seagrass beds, mangroves, corals, but also canyons and sea mountains) which is very important to contribute to the achievement of the Convention on Biological Diversity Aichi target 11[14].

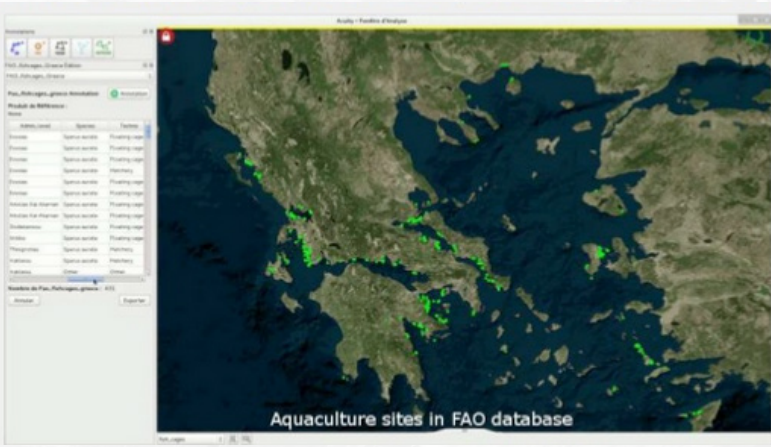
For the first challenge, CLS is trying to answer the following question: where are all the aquaculture cages in Greece? CLS set up a BlueBRIDGE VRE that gave them the possibility to easily access earth observation imagery taken from Google Maps or Bing imageries to derive the aquaculture cage locations; to make available the processing algorithm directly in the VRE and to set up editing facilities to adjust the final results after the algorithm computation.



Exploiting the data from the FAO database on 431 farms, CLS was able to assess 253 farms by mapping 8699 fish cages for a total of 1,931,043 m^2. Thanks to the computational resources made available by BlueBRIDGE, execution of the algorithm takes only few seconds and the human adjustments, afterwards, are done in a much easier way. As a further step, CLS is planning to exploit

12 taf.ices.dk
13 http://www.slideshare.net/BlueBridgeVREs/05-blue-environment-public-day-miles-macmillan-lawler
14 https://www.cbd.int/sp/targets/rationale/target-11/

the BlueBRIDGE VRE to detect aquaculture ponds on land and mangrove protected areas in Indonesia.


Aquaculture sites in FAO database

The impact of this VRE is very clear: with such little overhead countries can identify the location of aquaculture facilities to feed into the Maritime Spatial Planning strategy.

For the second challenge, GRID-Arendal wanted to answer the following question: What is protected in the Bahamas protected area network?

GRID-Arendal developed an algorithm that uses global scale data on marine protected areas from the Protected Planet database, data on exclusive economic zones (Marine Regions), including data on ecoregions provided by Marine Ecoregions of the World (MEOW) and by the Pelagic Provinces of the World (PPOW), data on seagrass, mangroves and coral reefs provided by the World Conservation Monitoring Centre (WCMC) and on seafloor geomorphology directly owned by GRID-Arendal.

The algorithm takes six minutes to run on the BlueBRIDGE e-infrastructure to do this full analysis and this picture is an example of the outputs of the BlueBRIDGE VRE: it is a table where for each feature of each MPA, it tells how much is in there. This is the basis to start developing some dynamic charting of the protected area.

BlueBRIDGE allowed GRID-Arendal to collect and access relevant data and metadata in a unique environment (instead of having them fragmented on diverse sources), increased the algorithm processing time (seconds and minutes instead of hours and days), made the full process repeatable and comparable and provided them with standardized reporting tools.

Additionally, as GRID-Arendal is collaborating with the Joint Research Centre (*JRC*) of the European Commission on the BIOPAMA project[15], thanks to BlueBRIDGE, GRID-Arendal was able to make the above algorithm available accessible from the BIOPAMA project portal.

Thanks to this VRE countries can report on their protected area networks to determine how well they are representing a standard range of ecologically relevant features (Aichi target 11).

Data Science and e-infrastructures can help Aquaculture to improve performance and sustainability, Kostas Seferis, CEO, I2S SA[16]

Aquaculture is the fastest growing food industry. It now accounts for nearly 50 percent of the world's fish that is used for food. Global production is forecast to increase from 45 million tons in 2014 to 85 million tons by 2030. Producers are looking into ways to increase their production in a sustainable way.

Aquafarming companies therefore need advanced tools that allow them to improve their

15 The Biodiversity and Protected Areas Management Programme (BIOPAMA) aims to address threats to biodiversity in African, Caribbean and Pacific (ACP) countries, while reducing poverty in communities in and around protected areas. Specifically, the programme will enhance existing institutions and networks by making the best available science and knowledge available for building capacity to improve policies and better decision-making on biodiversity conservation, protected areas management and access and benefit sharing. In particular, BIOPAMA works on a protected areas component, that includes capacity building for regional and national institutions, technical personnel, and protected areas managers and improved access and availability of biodiversity data through the establishment of regional observatories and information systems to improve decision making.

16 http://www.slideshare.net/BlueBridgeVREs/06-blue-bridgeworkshopkostasseferisv3

performance in terms of cost, feed conversion, growth rate and mortality and at the same time, reduce the impact on the environment. Aquaculture investors also need tools to better understand where to invest, in order to achieve the best fish growth and, at the same time, less diseases and mortalities.

I2S is a Greek private enterprise that provides data management services to over 2000 aquafarms worldwide. I2S developed the aquaManager tool[17], an IT system for production planning, management, cost analysis and financial forecast for aquaculture. Understanding the advantages that BlueBRIDGE can bring to them in terms of easy accessibility to additional datasets, computational resources and the capability of dynamically set up VREs on demand, I2S has integrated its aquaManager tool in a BlueBRIDGE VRE to offer extra services and facilities to its customer needs. The VRE allows aquafarming SMEs to generate optimized feeding and growth strategies by taking historical data as input and by producing models for feeding, growth and feed conversion. It provides "what if analysis" and also supports benchmarking. All of this was made possible thanks to the capability of easily porting on the BlueBRIDGE infrastructure, the analysis and simulation algorithms to calculate aquaculture performance in terms of main production KPIs and the feeding and growth models. Additionally, through the VRE, I2S is able to provide its customers with a performance benchmarking service that allows aquafarms to benchmark their performance by comparing their results to "global" models for selected species and regions.

VREs to improve fisheries science: The case of the Global Tuna fisheries, Paul Taconet, Engineer, Institut de recherche pour le développement (IRD)[18]

IRD is the French National Institute for Sustainable Development (IRD) specialised in providing scientific advice on fisheries and tuna fisheries. Currently, IRD is supporting the Global Tuna Atlas project to build a global tuna atlas database and related tools to access it. IRD is in charge of collecting data on tuna fisheries and of analysing it to provide scientific advice for the decision makers to manage better the tuna stocks.

The picture below shows the map of the distribution of catches of tuna, from 2010 to 2014.

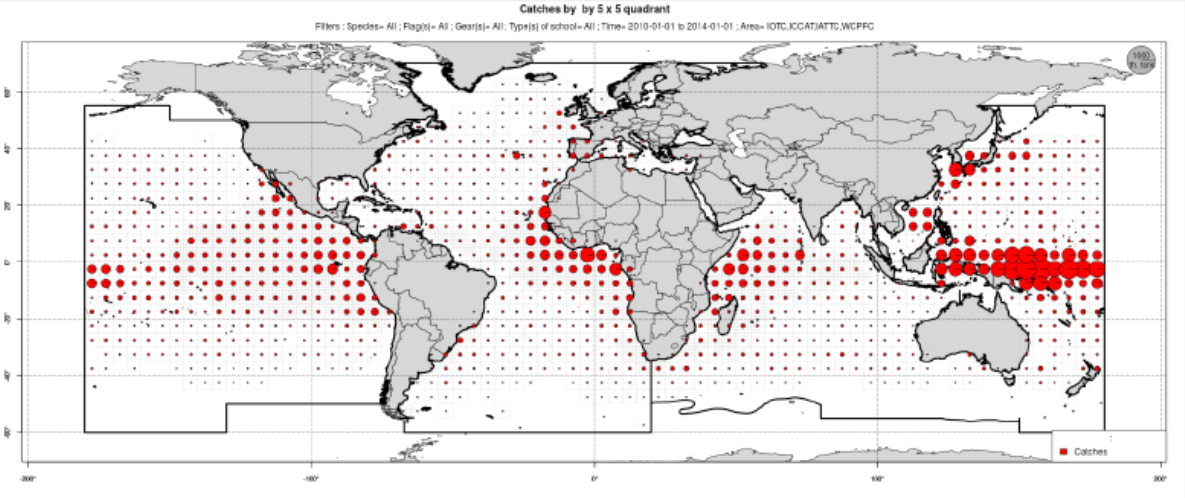


Figure : Distribution of the global catches of tuna and tuna-like species for the period 2010 - 2014
Source : SARDARA database using data collated from IOTC, ICCAT, IATTC, WCPFC

The map has been built using data collected from different regional fisheries management organisations

17 www.aquamanager.com
18 https://dl.dropboxusercontent.com/u/9472381/07_tunaatlas_ird_presentation.pdf

(RFMOs) and then manipulated by different scientists (that for example have converted some data expressed in catches to the number of fish into weight of fish catch). Clearly, this non-standard and non-transparent process does not allow other scientists to be able to reproduce the same map and makes difficult to understand the assumptions made to generate it.

To make this process transparent and reproducible only two solutions are available:

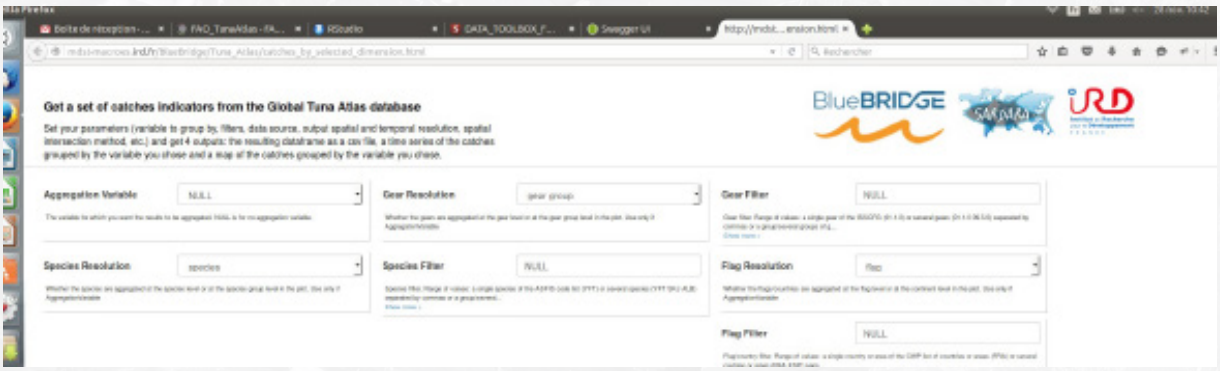
1. to propose and promote standards for tuna fisheries data (for formats, code lists and for description of the data). This clearly can only be implemented at a policy level;
2. to propose and promote efficient solutions and guidelines to handle the data, able to easily discover, access, process and report data. This is where BlueBRIDGE can make the difference.

IRD has joined BlueBRIDGE and has built the Tuna Atlas VRE. The VRE enables the users to:

- access existing data in several formats (not all the users are interested in the same formats),
- store fisheries time series data,
- discover available data (metadata catalogue),
- process data (to create new data and to compute fisheries indicators),
- reproduce or modify a workflow that creates a dataset
- prepare data for reporting and describe newly produced datasets (through metadata).

The VRE allows IRD to access all the data and tools needed to produce such a map in a single collaborative environment. Being collaborative, the environment can be enriched over time as different scientists can add new data and new processes.

Additionally, as many of the users of the map are policy makers and NGOs that do not necessarily have the skills to access and use a VRE, the catches indicators of the Global Tuna Atlas database have been made available on an external portal through a easy to use interface, as shown in the picture below.



IRD will continue enriching the tuna atlas database through BlueBRIDGE and is willing to move from tuna fisheries to any fisheries, as all the issues mentioned below are common to any fisheries. This is a clear example of how BlueBRIDGE can support informed decision making.

3. The role that data-driven science and innovation plays in the Blue Growth sector and how European e-infrastructures concretely facilitate this role

This section summarizes the final panel debate discussion. Renowned representatives from the Blue Growth and the e-infrastructure sectors discussed how the increased availability of data and data management services is changing science and bringing innovation to the Blue Growth sector and how the European e-infrastructures are accelerating this process.

The panel was chaired by Marc Taconet, FAO-FI, Chief Fisheries Statistics and Information Branch (FIPS) & BlueBRIDGE External Advisory Board Chair. Panellists included:

» Licia Florio, Project Development Officer at GÉANT;

» Neil Holdsworth, Head of Data and Information at the International Council for the Exploration of the Sea (ICES);

» Wainer Lusoli, Policy Officer at the European Commission - DG Research & Innovation;

» Per Öster, Research Infrastructures Director at CSC - IT Center for Science & EUDAT & EOSC pilot representative;

» Marco Weydert, Marine Resources Unit, DG-Research & Innovation, EU Commission.

Figure 2: Panel members

Main findings from the final panel discussion

The role of data-driven science and innovation for the Blue Growth sector

1. Generating value at the different stages of the data value chain is at the centre of the future Blue economy. At every step in the knowledge production chain, **reliable data are needed**.
2. **Society demands more and detailed information on environmental, societal, and**

nutritional issues: **this requires careful and integrated data management solutions.** Through improved analytics and processing of data the productivity of the Blue Growth sector can further improve.

3. **Innovation in the data chain is often most effective when SMEs are engaged at an early stage**
4. Scientific advice and recommendations can only be based on data coming from the Blue Growth sector if stakeholder **can and will open their data collections**
5. Open Science can undoubtedly bring many benefits to specific communities: data is free and public, and the findings and methods can be portrayed in a variety of web-based mediums. However, these new possibilities have not yet been fully exploited because almost all datasets, models, papers, and statistics produced are usually available exclusively within the community that has generated them. But why is this? The main reason is that exposing data products to other domains or actors requires additional effort and investment that does not always bring an immediate return of investment for the community itself. **There needs to be a concerted effort to raise awareness of both the short and long-term benefits that Open Science can bring to these key stakeholders. This is especially true in the aquaculture sector where the main players are SMEs that are not accustomed to data sharing and open science practices.**

How European e-infrastructures can support the Blue Growth related data science and innovation

1. **Horizontal e-infrastructures** (such as EGI or EUDAT) are essential, but they are not sufficient by themselves. They **need to be complemented by thematic infrastructures able to serve the specific needs of the communities**. Customised services should be identified by clustering the needs of the different communities and be developed on top of the horizontal services already provided by the e-infrastructures. **This is the approach adopted by BlueBRIDGE. Today, BlueBRIDGE is the only initiative serving the Blue Growth community with data management services covering the full data production chain. The BlueBRIDGE hybrid data infrastructure is a quintessential example** of a collaborative infrastructure that federates existing e-infrastructure services and resources and manages them securely and efficiently.
2. Federating resources is not just a matter of discovery and access: users need easy to use interfaces. Establishing a federation of e-infrastructures is not only a matter of discovering resources geographically located somewhere else, or accessing resources maintained by someone else. **Federating resources means offering them to users with easy to use interfaces where the integration technicalities are hidden**, like in the case presented by IRD for the Global Tuna Atlas. This is particularly needed in the Blue Growth sector, where the users are very heterogeneous including SMEs, policy advisors and marine experts many of whom do not necessarily have coding skills. **The possibility of making the BlueBRIDGE services available via third party web-portals through customized interfaces** is one of the ways to hide the infrastructure complexity to the end users.
3. **Provenance is becoming more and more important especially when providing fisheries policy advice and recommendations**: understanding which data, methods and algorithms were used for the fisheries data analysis is key for the transparency of the results and for the repeatability of experiments. Having metadata automatically generated and stored together with the result of the computation today is a must. BlueBRIDGE services have been designed with this in mind.

4. **An effective data collection needs to be in place to ensure data availability**. Just like many other communities, this is still a challenge that the marine community faces. The lack of standard formats, common policies and licenses are still barriers for wider adoption of services, especially when it comes to ecosystems of infrastructures where different organisations have to share and re-use services. How the provision of these services is regulated is not always clear and this instils mistrust in the users. A clear governance is needed.
5. Too many e-infrastructure services are the results of short-term funded projects and are not fully sustained beyond the lifetime of the project. There is a growing need for such projects to provide robust sustainability plans which envisage the establishment of legal entities potential financial transactions and exchange of expertise can begin to happen. At the basis of this lack of long term vision there is the wrong funding model applied so far who has focussed only on scientific excellence rather than sustainability. This is now being addressed: **the idea behind the upcoming European Open Science Cloud should force all the providers to think sustainable.**

How the Blue Growth sector can benefit from the European Open Science Cloud (EOSC)

1. The European Commission plans to set up a "**Blue Cloud pilot**" within EOSC. This pilot will analyse the needs of the Blue Growth community and provide them with the right services. The players that will be part of the Blue Cloud Pilot must cover the full data work-flow starting from data collection, data access and harmonisation up to data analysis and publication. Thematic infrastructures such as **Copernicus** and **EMODnet**, the European Marine Observation and Data Network, must be at the foundations of such a Blue Cloud Pilot as they will provide the data to fuel the process. **Horizontal e-infrastructures providing basic services will be the engine to compute and store data**. One element is missing and that is skills. Scientists, researchers, policy advisors and SMEs operating in the Blue Growth sector need customized services to solve specific problems and they do not always have the skills to directly use the horizontal infrastructure services. **Another layer on top of the horizontal e-infrastructures is needed and is represented by the thematic e-infrastructures, like the one underlying BlueBRIDGE**, that provides dedicated community services and acts as a "mediator" between the e-infrastructure world and the end users.
2. It is clear that with EOSC a framework is in place. There will be competition among different players to be part of it. All are potentially good candidates but those that will succeed will be those able to understand who they are, who the others are, the different services provided, the skills and the capacities. **The challenges that EOSC is going to deal with are too complex to be addressed by a small group of people. There is a strong need for collaboration.**
3. Just as in the marine ecosystem, each organism can live with its own resources but growth and development depends on the health and equilibrium of the overall ecosystem, the same dynamic can be applied to e-infrastructures which serve the scientific community. **A healthy ecosystem of e-infrastructures is exactly what EOSC is looking to achieve.** Each e-infrastructure providing its own services and a dependence on services provided by other infrastructures in order to build new services. **A balanced governance model is essential to ensure that such inter-dependency is successful. To reach this vision the complementarities of each e-infrastructures need to be identified.**

We at BlueBRIDGE, want to thank all the participants that joined us in Brussels and we would like to invite all of them to contact us to establish new fruitful collaborations. We are planning to organize a new event later in early 2018 involving all the major infrastructure leaders for Blue Growth in Europe. If you like this idea get in touch with us.

Contact us at info@bluebridge-vres.eu

SPECIAL THANKS TO

Donatella Castelli, ISTI-CNR & BlueBRIDGE Coordinator

Anton Ellenbroek, IT Specialist, Food and Agriculture Organisation (FAO) of the United Nations

Licia Florio, Project Development Officer, GÉANT

Sara Garavelli, Project Manager, Trust-IT Services

Neil Holdsworth, Head of Data and Information, International Council for the Exploration of the Sea (ICES)

Scott Large, Professional Officer, International Council for the Exploration of the Sea (ICES)

Wainer Lusoli, Policy Officer, DG Research & Innovation, European Commission

Miles Macmillan-Lawler, Programme Leader, GRID-Arendal

Per Öster, Research Infrastructures Director, CSC - IT Center for Science & EUDAT & EOSC pilot

Pasquale Pagano, ISTI-CNR & BlueBRIDGE Technical Coordinator

Kostas Seferis, CEO, I2S SA

Marc Taconet, FAO-FI, Chief Fisheries Statistics and Information Branch (FIPS) & BlueBRIDGE External Advisory Board Chair

Paul Taconet, Engineer, Institut de recherche pour le développement (IRD)

Victoria Tsoukala, Program Officer-Seconded National Expert- eInfrastructures and Open Science Cloud at European Commission, DG CONNECT

Marco Weydert, Marine Resources Unit, DG-Research & Innovation, EU Commission

…and to all the workshop participants

Find out more about BlueBRIDGE www.bluebridge-vres.eu

Follow us on Twitter @BlueBridgeVREs

Join our LinkedIn network

BlueBRIDGE Consortium Members

Coordinator: **Italian National Research Council - CNR** (IT)

Partners: **ERCIM** (FR); **Engineering** (IT); **University of Athens** (GR); **The Food and Agriculture Organisation of the United Nations – FAO** (IT); **The International Council for the Exploration of the Sea – ICES** (DK); **Institut de recherche pour le developpement – IRD** (FR); **Foundation for Research and Technology Hellas – FORTH** (GR); **Trust-IT Services Ltd** (UK); **I2S** (GR); **CITE** (GR), **Collecte Localisation Satellites SA – CLS** (FR); **GRID-Arendal** (NO); **Pole Mer Bretagne Atlantique** (FR).

European Commission

BlueBRIDGE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 675680