Draft Strategic Roadmap (Release 1)

Work Package	WP6 – Blue Cloud Roadmap, Exploitation & Sustainability Measures, and trans-European Liaisons
Lead Partner	Seascape Belgium (SSBE)
Lead Author (Org)	SSBE
Contributing Author(s)	Julia Vera, Kate Larkin, Xiaoyu Fang (SSBE), Sara Pittonet (Trust-IT), Dick Schaap (MARIS), Pasquale Pagano (IST-CNR)
Reviewers	Cécile Nys (IFREMER), Nathalie Tonné (SSBE), Silvana Muscella, Sara Pittonet, Federico Drago (Trust-IT), Anton Ellenbroek, Marc Taconet (FAO), Dick M.A. Schaap (MARIS), Pasquale Pagano (ISTI-CNR), Patricia Cabrera (VLIZ)
Date	10.06.2021
Version	1.3

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Blue-Cloud - Piloting Innovative services for Marine Research & the Blue Economy - has received funding from the European Union's Horizon programme call BG-07-2019-2020, topic: [A] 2019 - Blue Cloud services, Grant Agreement n. 862409.

DISCLAIMER

"Blue-Cloud, Piloting Innovative services for Marine Research & the Blue Economy" has received funding from the European Union's Horizon programme call BG-07-2019-2020, topic: [A] 2019 - Blue Cloud services, Grant Agreement n.862409.

This document contains information on Blue-Cloud core activities. 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 Blue-Cloud Consortium, 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 in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

COPYRIGHT NOTICE

Table of Contents

Executive summary: Blue-Cloud in a snapshot & the road ahead

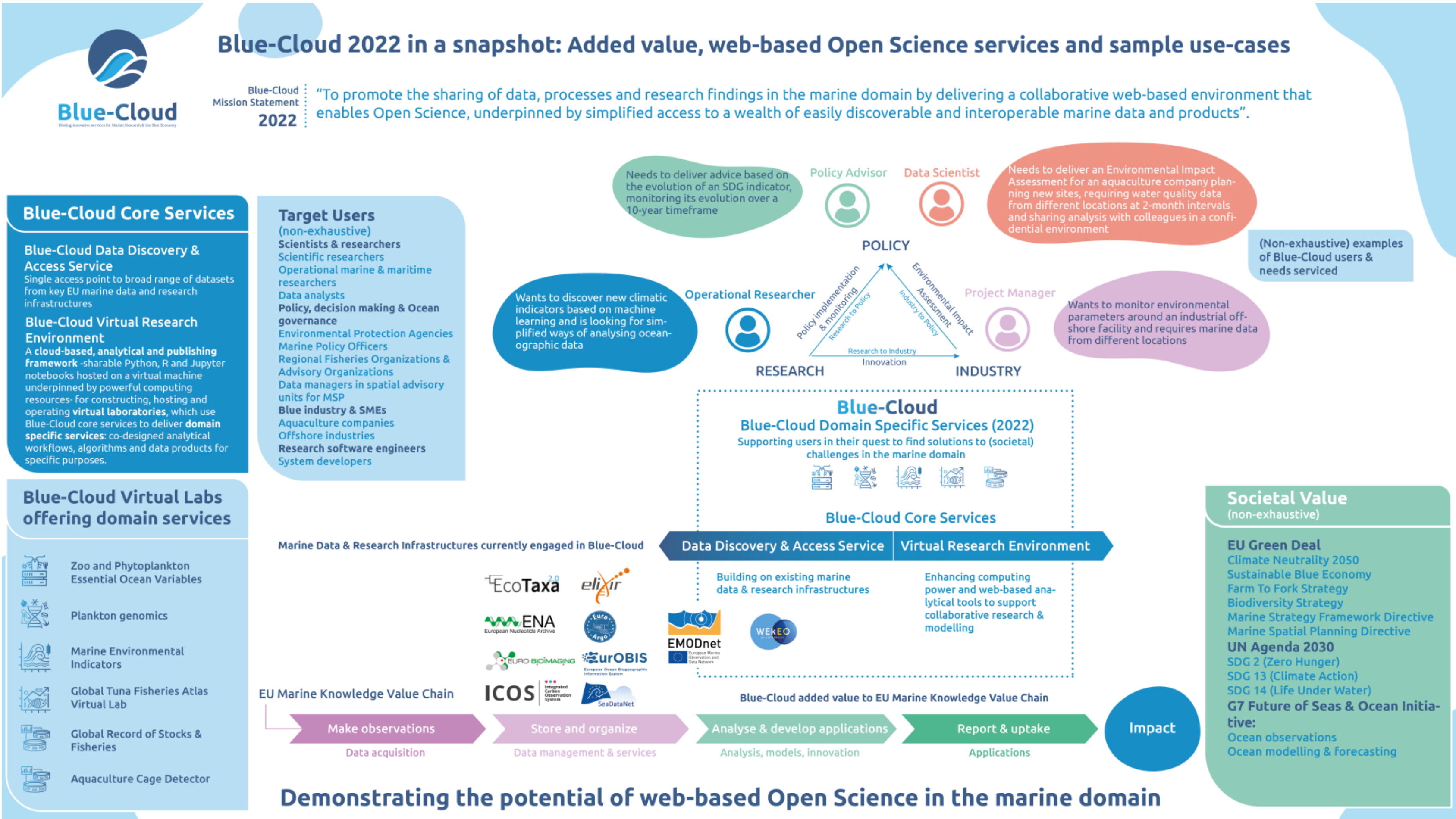

Figure 1: Blue-Cloud 2022 in a snapshot: Added value, web-based Open Science services and sample use-cases. **Credit**: Seascape Belgium

Blue-Cloud - Piloting Innovative services for Marine Research & the Blue Economy - has received funding from the European Union's Horizon programme call BG-07-2019-2020, topic: [A] 2019 - Blue Cloud services, Grant Agreement n. 862409.

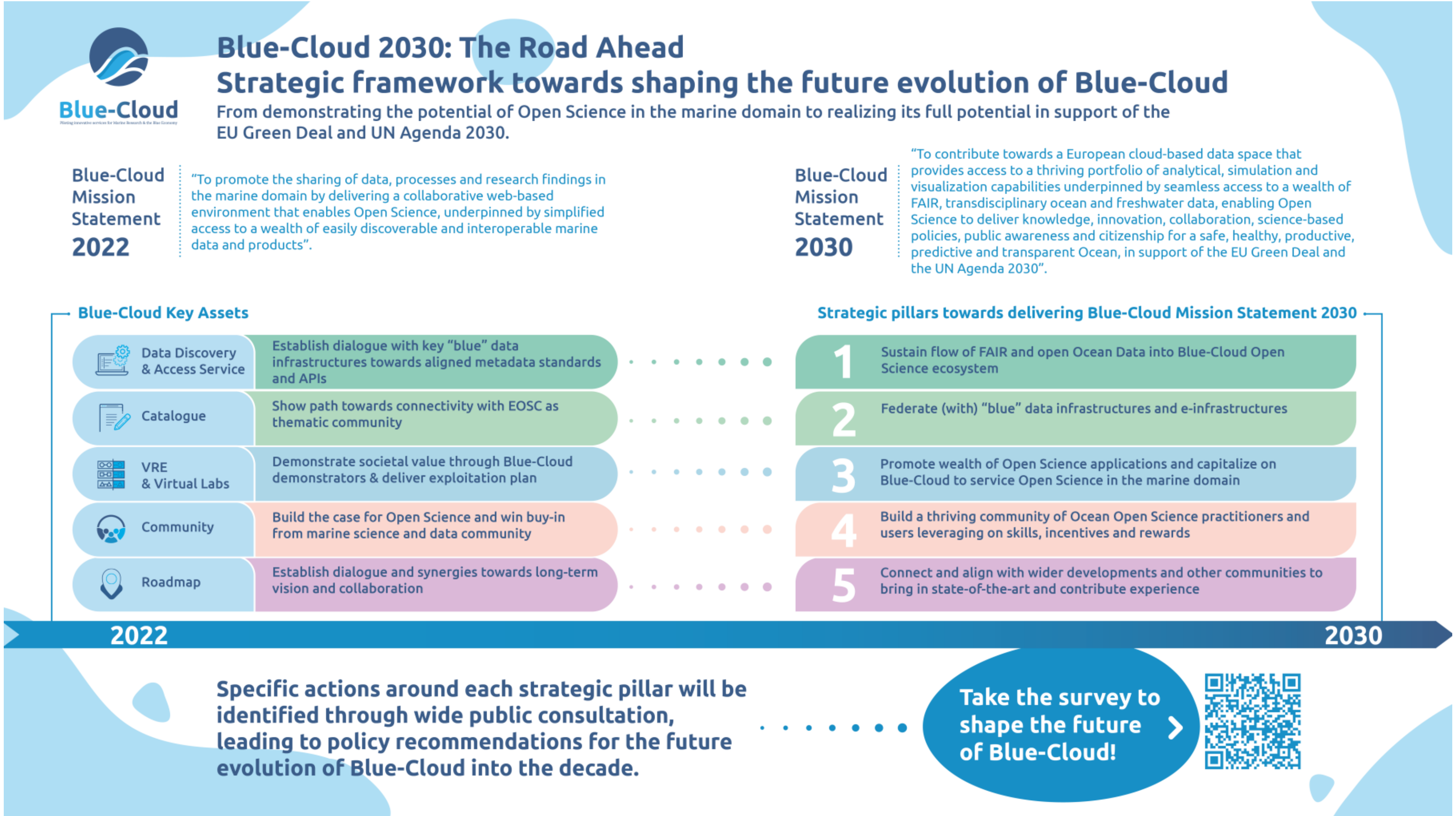

Figure 2: Blue-Cloud 2030: The Road Ahead. Strategic framework towards shaping the future evolution of Blue-Cloud. **Credit**: Seascape Belgium

Executive summary

The **Horizon 2020 Blue-Cloud (B-C) project** launched in October 2019, aiming to **demonstrate the potential of web-based Open Science in the marine domain**. To deliver on this objective, it is **piloting** the development of a **web-based cyber platform** that will provide marine scientists with enhanced **analytical capabilities**. It will facilitate their engagement in **collaborative research** and will provide them with access to powerful **cloud-computing resources**, a range of **analytical tools** and **simplified access to multi-disciplinary data** from *in situ* and satellite-derived observations to model outputs. B-C is co-designed by, and builds on, existing European capability, including trusted data services EMODnet, CMEMS and other key research and data infrastructures and e-infrastructures. In the **short-term**, the project is building this cyber platform by means of a smart federation of selected, multidisciplinary data repositories, analytical tools and computing facilities. The added-value of B-C will be demonstrated by five specific "demonstrators" or use-cases. In the **medium-** and **long-term future**, B-C aspires to upscale this cyber platform, its resources, services and applications, together with a thriving community of Open Science service providers and users. B-C will evolve to further align with wider developments at European level to catalyse **transformative solutions** to priority **societal challenges** and unravel **new opportunities** for **innovation**, in support of the **EU Green Deal** and **UN Agenda 2030**. To guide the long-term capitalization and further development of this ambition into the future, the Blue-Cloud Project is producing a **roadmap** to **2030**, which is being developed as a **co-designed**, **community-oriented policy document** with substantial stakeholder consultation and input. This document is a first **B-C Draft Strategic Roadmap,** which is being launched for public consultation from June to September 2021. For this reason, this document should not be seen as a **preliminary blueprint**, but rather as an intermediary step towards gathering **input**, **feedback** and **insight** from the **B-C Community** towards its evolution and grounding, benefiting from wide stakeholder input.

Section 1 presents the process followed towards the development of the roadmap. **Section 2** reflects on the **policy context** that motivates the **Blue-Cloud**'s efforts, exploring the emergence of **Open Science** and the opportunities it brings in **support** of the **European Green Deal** and the **United Nations Agenda 2030 for Sustainable Development.** It also introduces the **added value** that **Blue-Cloud** will contribute towards seizing these opportunities through its **core assets** and **services**. **Section 3** reflects on an overarching **vision** that could guide the **future capitalization** and further development of these assets, working around five **strategic pillars** of action to support successful Open Science in the marine domain and inviting **feedback** and **contributions** towards grounding **B-C**'s **Strategic Roadmap to 2030**.

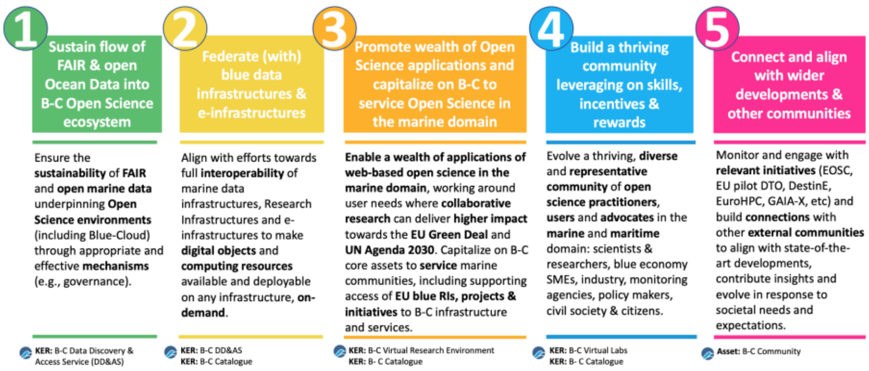

Figure 3: B-C Key Exploitable Results (KER) & strategic pillars of action towards evolving Blue-Cloud efforts into the future

Blue-Cloud - Piloting Innovative services for Marine Research & the Blue Economy - has received funding from the European Union's Horizon programme call BG-07-2019-2020, topic: [A] 2019 - Blue Cloud services, Grant Agreement n. 862409.

A **Technical Annex** is provided as additional reading for stakeholders who are less familiar with Blue-Cloud's landscape and its offering, providing additional insight into Europe's marine knowledge value chain and the **services** and **technological** and **demonstrative assets** that the Blue-Cloud project will bring as added value to this landscape by 2022.

As the **first early draft "roadmap"**, this document sets the stage for, but **does not yet include**, **policy recommendations**, which will be the core content of the final roadmap. This current format is used for a better understanding of the Blue-Cloud efforts, **inviting feedback** and **contributions** from the **B-C Project Consortium**, the **B-C External Stakeholder Expert Board**, **users** of B-C services and **related projects** and **initiatives** with whom the project is in dialogue, but also from organizations and professionals **not yet directly engaged** in Blue-Cloud, but who could be interested in joining its efforts and/or benefitting from the use of its **assets** in the future. The **policy recommendations** for the final B-C Roadmap to 2030 will be drafted from community responses to the public consultation, in particular to the following key questions:

- How should **Blue-Cloud** evolve to support a **thriving ecosystem** for web-based **Open Science** and **Open data** in the **marine domain**?
- What additional **mechanisms** (e.g., governance) could add value to the existing EU long term marine data services to further ensure the **sustainability** of **FAIR Ocean data** underpinning B-C's Open Science environment and services? What additional mechanisms could best contribute to a broad **alignment** of the **marine community** towards future **EU pilot DTO** and **DestinE** developments?
- What **applications** of **Open Science** in **the marine domain** could have a higher probability of success, given current availability of data, models and actors willing to engage in collaborative science, across a broad range of topics? Which of such applications should be prioritized towards addressing current **user needs** and delivering **highest societal impact**?
- How should B-C's Open Science environment and services evolve to be fit-for-use not only for scientists, but also for other Open Science users such as **policy makers** and **blue economy SMEs** and **industry**? What needs do these users have that B-C could evolve to address? What **skills**, **incentives** and **rewards** can contribute to bring Open Science practitioners on board?
- How can B-C evolve to **further connect** with **marine data infrastructures** and **research infrastructures** to deliver full interoperability of marine data through the **B-C Data Discovery & Access Service**, aligning and in collaboration with other **international efforts**?
- What actions would be required to enable **B-C's Catalogue** of **analytical methods**, **algorithms** and **applications** to **be deployed in EOSC,** but also in other infrastructures **-closer to data-** or across **supercomputing platforms** in Europe?
- How should **B-C's assets** evolve to align with future **EU pilot DTO** and **DestinE** developments and other relevant, wider initiatives?

Join us in shaping the future of Blue-Cloud:

Bring **your views** and **opinion** on these questions and have your say on shaping **strategic policy recommendations** towards the future development of Blue-Cloud, so it evolves to deliver on your **needs** and **expectations**. Take the survey and provide feedback to the end of **September 2021**:

https://www.blue-cloud.org/form/blue-cloud-roadmap-2030-2nd-online-consultation

1 About the Blue-Cloud Roadmap to 2030

The **Horizon 2020 pilot Blue-Cloud project**[1] is the component of the "**The Future of Seas and Oceans Flagship Initiative"** aiming to demonstrate the potential of web-based **Open Science in the marine domain.** The project is **piloting** the development of a **web-based cyber platform** to provide marine researchers with enhanced **analytical capabilities** to enable **Open Science**. It will provide them with powerful **computing facilities**, a range of **analytical tools** and **simplified access to data** from in-situ and satellite observations, data products and model outputs available across different blue data infrastructures and marine domains. Blue-Cloud will form a **marine thematic community** interacting and integrating with developments of the **European Open Science Cloud** (EOSC).

The Blue-Cloud project will deliver **innovative, demonstrative assets**, which will have to be taken up, sustained and evolved to capitalize on its results. The **Blue-Cloud Roadmap to 2030** seeks to provide guidance for the future evolution of these efforts, by providing a **vision** towards their development beyond project-end and into the future (2030). The Roadmap will be the result of a **collective reflection** on how the results of the Blue-Cloud project could be exploited and evolved into the future to maximize their impact, catering to a much wider user base, aligning with wider developments at European and global level and creating new opportunities for **Ocean**, **science-based innovation** in support of the **EU Green Deal** and the **UN Agenda 2030**.

The **Blue-Cloud Roadmap to 2030** is being developed as a **co-designed**, **community-oriented policy document**, specifically aimed at further, long-term strategic development of the Blue-Cloud's efforts. It is being developed over the course of the three-year pilot Blue-Cloud project, with substantial stakeholder consultation and input. In 2020, building from an initial Concept Note[2], the project saw the start of an **initial phase** of **stakeholder dialogue**[3] that extended into early 2021, including discussion amongst Project Partners[4], workshops with **B-C External Stakeholder & Expert Board**[5] (ESEB), targeted conversations with representatives of the **European Commission** and consultation to key stakeholder communities (marine researchers, marine data infrastructures, research infrastructures and e-infrastructures).

This document is the first **draft release** of the **Blue-Cloud Roadmap to 2030**. However, it should not be seen as a preliminary blueprint, but rather as an **intermediary step** towards gathering **additional input**, **feedback** and **insight** from the **Blue-Cloud Community** towards its evolution and grounding, benefiting from wide stakeholder consultation. This draft release will open for public consultation in **June 2021** and contributions will be welcomed throughout the **Summer** of **2021**. This consultation will aim to gather as many contributions as possible from the **wider marine & maritime community** in Europe and internationally and other actors across all **key stakeholder groups** conforming the **Blue-Cloud Community**.

[1] https://www.blue-cloud.org/
[2] Blue-Cloud strategic Roadmap Concept Note (Milestone MS41)
[3] Summary of Key Messages & Recommendations of the 1st Phase of Consultations with the Blue-Cloud Community (SSBE, 2020)
[4] https://www.blue-cloud.org/partners
[5] https://www.blue-cloud.org/eseb

2 Leveraging web-based Open Science in the marine domain: Blue-Cloud's added value

2.1 Opportunities for web-based Open Science in the marine domain

The motivations and goals behind the **Blue-Cloud**'s efforts are shaped by the context of our times, driven by a confluence of **global change** -large scale, Planetary changes in the Earth system and in society- and the **digital revolution**. The Blue-Cloud project is part of EU action in response to this context, which on the one hand presents unprecedented challenges -connected to **climate change**, **pollution**, **waste** and the **loss of biodiversity** and **ecosystem integrity** at a planetary scale- while on the other is delivering new emerging opportunities enabled by the dawn of the **digital age**. Digital technologies are making it possible to accelerate the development of **science-based solutions** to these challenges and fostering **digital environmental cooperation** at an also unprecedented global scale, thereby making it more plausible to coordinate a more effective global response. This response, in return, has the potential to create **new jobs, wealth** and **wellbeing** by **advancing global sustainability** through environmental stewardship. Europe is embracing this context of challenges and opportunities through the **EU Green Deal**[6] and the **EU Strategy for Data**[7], further contributing to effective international action by aligning the objectives of these ambitious policy frameworks with those of the **United Nations (UN) Agenda 2030 for Sustainable Development**. From a marine perspective, the EU has launched **Mission Starfish**[8], which sets clear goals to support the EU Green Deal through specific actions around **healthy Oceans**, **seas, coastal** and **inland waters** and is working through the **G7** to align action through the **G7 Future of Seas and Oceans Initiative**. A common denominator across these EU and international policy efforts is the recognition of the role that **science** can play as a **catalyst of action** in support of **sustainable development**, as well as the need to ensure that our planet's largest ecosystem -the Ocean- can continue to deliver the valuable life supporting services underpinning such development. Also building on this notion, the UN has called for the decade spanning between 2020 and 2030 to be the **Decade of Ocean Science for Sustainable Development**. Central to the Ocean Decade is the goal of triggering a revolution in Ocean science, seeking to promote and facilitate "***science that is shared openly and available for reuse***". **Open data** and **Open Science** are recognized as central to achieve the objectives of the **EU Green Deal** and the associated, systemic and transformative solutions for "healthy oceans, seas, coastal and inland waters" proposed by **Mission Starfish** and the **UN Ocean Decade**, in support of the **UN Agenda 2030**.

There are different ways in which **web-based Open Science in the marine domain** can contribute to support an inclusive economic and ecological transition, opening opportunities to improve our **knowledge** and **understanding** of the **Ocean**, informing better **science-based policies**, **triggering innovation** across the blue economy and supporting the development of an **Ocean citizenship**. One obvious way is contributing to improve our **knowledge** and understanding of **marine ecosystems** by supporting collaborative research, but also to deploying **more efficient, effective and cost-saving**

[6] https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en
[7] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en
[8] https://ec.europa.eu/info/horizon-europe/missions-horizon-europe/healthy-oceans-seas-coastal-and-inland-waters_en

environmental **monitoring systems** to **measure progress and compliance** with **policy targets**, positively impacting all cross-cutting priorities shared by the **EU Green Deal** and **UN Agenda 2030**.

Another one is enhancing our **Ocean modelling** and **prediction** capabilities, in support of timely, adaptative crisis management. Europe's pledge to be climate neutral by 2050 will require a quantum leap in our capacity to observe, understand and predict complex and interconnected natural and anthropogenic processes occurring at different spatial and temporal scales. Novel and emerging technologies -including **cloud** and **High-Performance Computing** (HPC), **Artificial Intelligence**, novel in-situ data gathering capabilities (such as **Internet of Things** or **citizen science**) or advanced and **high-resolution modelling technologies**- are empowering scientists in their capacity to understand and predict such complex interactions. The EU is responding to this opportunity by proposing a "**Digital Twin of the Ocean**" (DTO) as a high precision digital model representing a consistent high-resolution, multi-dimensional and (nearly) real-time description of the marine realm. **Open Science**, underpinned by **FAIR** (**f**indable, **a**ccessible, **i**nteroperable and **r**eusable) and **open data**, will be the driving force behind this DTO, which will rely on **analytical models** built by **thematic experts** in the **marine domain** and open the window to new, richer interactions with other communities and stakeholders, including ICT innovators, industry, NGOs and citizens.

What is web-based Open Science?

The evolution of ICT, and specially web-based technology, has brought unprecedented changes and possibilities as it enables not only **sharing** and **processing** large amounts of data, but also the **results** of **its analysis**, feeding a virtuous circle of **knowledge development**. As a result, and just like any other sector in society, **science** is also undergoing a profound transformation, as **scientific data**, **methods** and **results** are shared across teams, geographical borders and disciplines, opening unprecedented, new collaboration opportunities underpinned by game changers like **cloud computing**, **Artificial Intelligence (AI), big data analysis** and **machine learning**. The associated concept of "**Open Science**" –"*an approach to the scientific process that focuses on spreading knowledge as soon as it is available by using web-based digital and collaborative technologies"*- is shifting the prime focus of researchers away from "publishing" towards **knowledge sharing**[9]. This shift is timely and aligns with society's expectations on how to govern our collective use of data. In a recent consultation launched by the European Commission on the European Strategy for Data[10], 91.5% of respondents agreed that "*more data should be available for the common good, for example for improving mobility, delivering personalised medicine, reducing energy consumption and making our society greener*", signalling broad societal consensus that data should be used for tackling societal, climate and environment-related challenges, contributing to healthier, more prosperous and sustainable societies.

Another important way is contributing to trigger new opportunities for **innovation and economic development**, establishing **closer links** and **collaboration** between **marine stakeholders** and the **maritime industry** for a **cleaner** and **circular economy**, particularly as the benefits of sharing data through trusted marine data infrastructures such as EMODnet[11] become obvious. Improving knowledge, reducing operational risks or adding value to data through integration to deliver new, added value data products tailored to business needs are some of obvious advantages of collaborative, Open Science that are beginning to trigger a cultural shift in data sharing amongst industry players.

Web-based Open Science can also contribute to expanding our current body of Ocean knowledge by enabling and facilitating transboundary and interdisciplinary science. Building Open Science

[9] J.C Burgelman et al (2019). *Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century* https://www.frontiersin.org/articles/10.3389/fdata.2019.00043/full
[10] European Commission (2020) *Summary Report of the Public Consultation on the European Strategy for Data*
[11] https://www.emodnet.eu/emodnet-business-brochure

capabilities in the marine domain will prepare the marine community to engage in collaborative research and science across a wealth of other disciplines, capitalizing on the opportunities brought by the single European data space introduced by the European Data Strategy, which will become articulated with the **9 sectoral data spaces** foreseen by the strategy, building a competitive data and knowledge economy in Europe endorsing Open Science[12] and supported by the **European Open Science Cloud** (EOSC) as an "*internet for science*". This will allow the marine community to build on previous strategies of **Marine Knowledge 2020**[13], a key pillar of the 2011 Commission Communication on Blue Growth[14] and to further research and broaden the body of evidence of our interdependencies with the Ocean by engaging with other disciplines, for example advancing knowledge in the meta-discipline of **Ocean and Human Health (OHH)[15]**, linking different priorities across the **EU Green Deal.**

Last but not least, web-based Open Science can contribute to **boosting Ocean literacy** and leading to **better informed, attentive and engaged citizens in Ocean** matters, contributing to **develop** a sense of **Ocean citizenship.** It not only contributes to breaking down barriers to data access, which is the first step towards data democratization, but it also unfolds new, creative ways of using marine data to support Ocean literacy efforts, leveraging on new technologies. Another way in which Open Science can contribute to Ocean citizenship by further empowering **citizen science**, which in turn can contribute to feed the marine knowledge value chain by encouraging and enabling the generation of "non-scientific" data streams that complement and refine "traditional" methods of observation.

Figure 4 summarizes the window of opportunity that **web-based Open Science** brings towards delivering on the societal objectives laid out by the **EU Green Deal**, the **UN Agenda 2030** and the **G7 Future of Oceans Initiative**, which are the driving force behind Blue-Cloud's efforts.

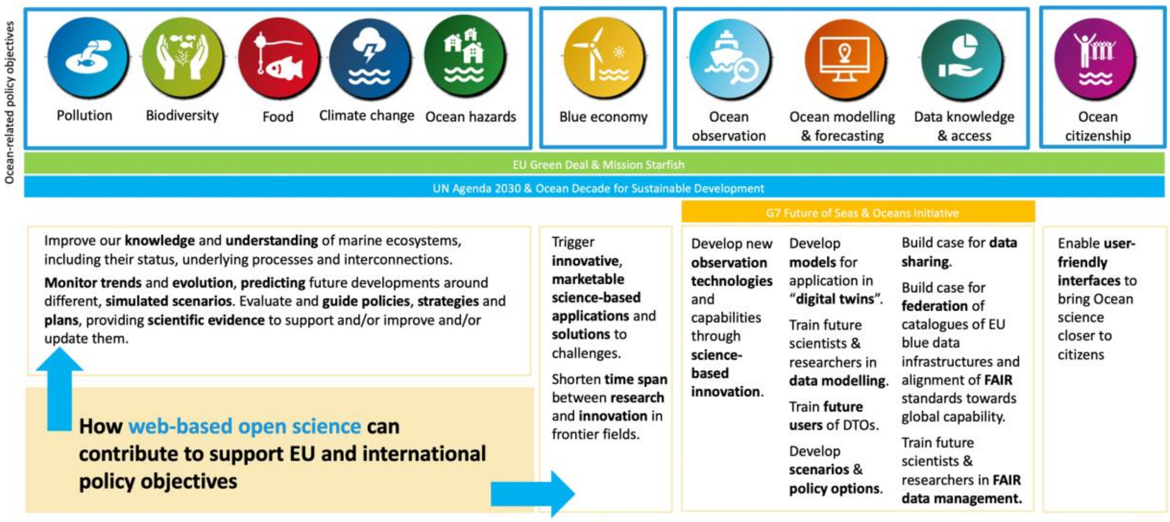

Figure 4: Web-based Open Science's contribution to relevant EU and global policy objectives. **Credit**: Seascape Belgium

2.2 Blue-Cloud's added value to Europe's marine knowledge value chain

Over the past decades, Europe has developed an impressive capability for marine observation, anticipating the need to support and advance our knowledge of the Ocean. These efforts have

[12] https://ec.europa.eu/digital-single-market/en/european-open-science-cloud
[13] https://ec.europa.eu/maritimeaffairs/sites/maritimeaffairs/files/docs/body/marine-knowledge-2020-green-paper_en.pdf
[14] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=COM:2014:254:REV1&from=EN
[15] https://sophie2020.eu/

contributed to the emergence of an advanced **European "marine knowledge value chain"** which brings together multiple European actors and assets.

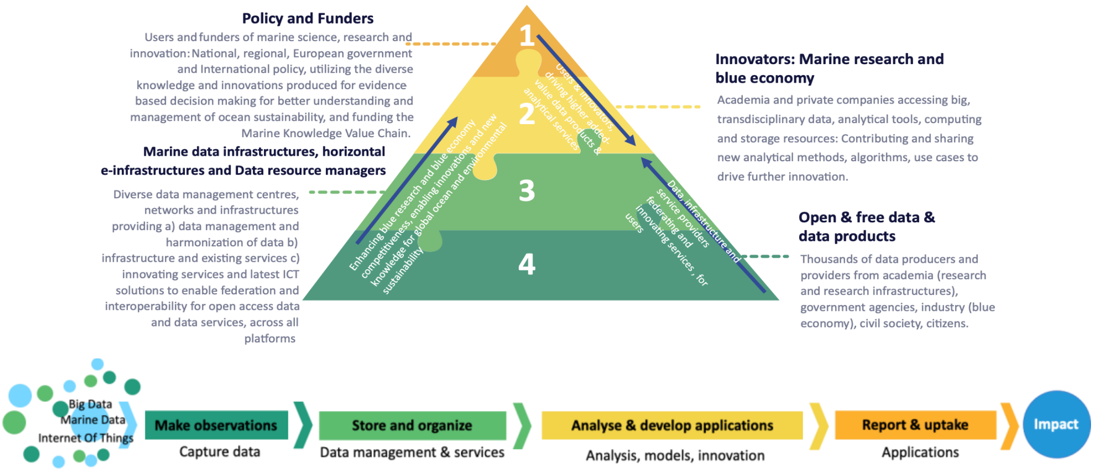

Figure 5: The European marine knowledge value chain: Data, infrastructure and service providers and users, including marine research and blue economy, policy makers and funders. Credit: MARIS and Seascape Belgium

Europe's efforts to support increased knowledge of our seas[16], aligned with action at national level, has led to a remarkable increase of abundance of **marine data**, stemming from academia, major Research & Development projects and initiatives, Research Infrastructures (RIs), industry and civil society. The emergence of citizen science and technological developments such as the "Internet of Things (IoT)" anticipates unprecedented potential to exponentially grow such abundance.

A resilient, existing European network of distributed **infrastructures** for Ocean observing and data gathering, handling and sharing is making it possible to **capture**, **manage**, **enhance** and **channel** this wealth of data into an additional layer of actors in the marine and maritime research and blue economy communities. Marine data infrastructures and services, such as the **European Marine Observation and Data Network** (**EMODnet**) and the **Copernicus Marine Environmental Monitoring Service** (**CMEMS**), together with **Research Infrastructures** and **horizontal e-infrastructures**, play a key, strategic role in making data **findable**, **accessible**, **interoperable** and **reusable** (FAIR) to **marine data user communities**. Thanks to all these efforts, marine data is now recognized as a **public good** in Europe and multi-resolution maps of all Europe's seas and oceans are available, together with large catalogues of **open** and **free data products** that span over multiple disciplinary themes.

Efforts along the marine knowledge value chain have also extended into applying technology towards enhancing the **analytical capabilities** of the marine community. The EU has supported the development of **e-infrastructures** as the pathway towards integrating **data services** with **analytical tools** in a seamless way, increasingly at large scale. Piloting initiatives such as the i-Marine project, the BlueBridge project or the SeaDataCloud VRE development have shown how comprehensive data management solutions can support the application of a science-based ecosystem approach to specific domains (e.g. fisheries or marine spatial planning), as well as providing attractive capacity building environments for interdisciplinary research communities involved in increasing our knowledge of the marine environment. Thanks to these efforts, a growing **catalogue** of **virtual laboratories** ("**virtual**

[16] COM(2014) 254 final/2 Innovation in the Blue Economy: Realising the potential of our seas and oceans for jobs and growth https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=COM:2014:254:REV1&from=EN

labs") is making web-based **Open Science** more **accessible** to researchers in the marine community, providing them with greater **computational resources** and **tools**.

What is a virtual lab?

A virtual laboratory is a web-based, problem solving environment that enables users across different physical locations and expertise to collaborate efficiently in an ongoing way, providing them with tools and services to share data, analytical methods and results towards solving complex problems.

In spite of its impressive capabilities, Europe's marine knowledge value chain still faces **challenges** towards further enhancing its performance and unlocking the full potential of Open Science, for the benefit of all engaged actors and ultimately, European citizens. On the one hand, it requires "lubrication" at different points of its underpinning gear, including **advancing** and **harmonizing data management practices towards FAIR principles** across blue data infrastructures to support greater **interoperability**; making more of the currently available data **open** (to the extent possible); and **connecting to other data spaces** to support interdisciplinary research. But alongside responding to those challenges, Europe's marine knowledge value chain can benefit from developing more robust **frameworks** for **analysing data**. Harnessing the power of **web-based, "cloud" technologies** can boost the value chain's **productivity**, bringing a huge leap forward in the way that **marine data** -as well as the **processes** and **results** linked to **exploiting** and **analysing** such data- is **used**, **shared**, **enriched** and **applied**. Previous experiences have led the way towards testing and advancing technological infrastructures, methods and tools towards this end, introducing opportunities to pilot supporting frameworks for Open Science along the marine knowledge value chain. However, they were limited in their scope to influence changes both upstream ("data access") and downstream ("sharing outputs" across different communities and gearing efforts towards bridging the "science-policy" gap). The **Blue-Cloud project** is a step forward in that direction, **evolving existing computing** and **analytical capabilities** and underpinning them with **easier access to data** delivered across key blue data infrastructures through the use of **common languages** (standardized metadata), seeking to signpost potential solutions towards **federating their data services**, while showcasing how **Open Science** can deliver **impact** on key, **priority policy areas** catering to a **broader range of users**.

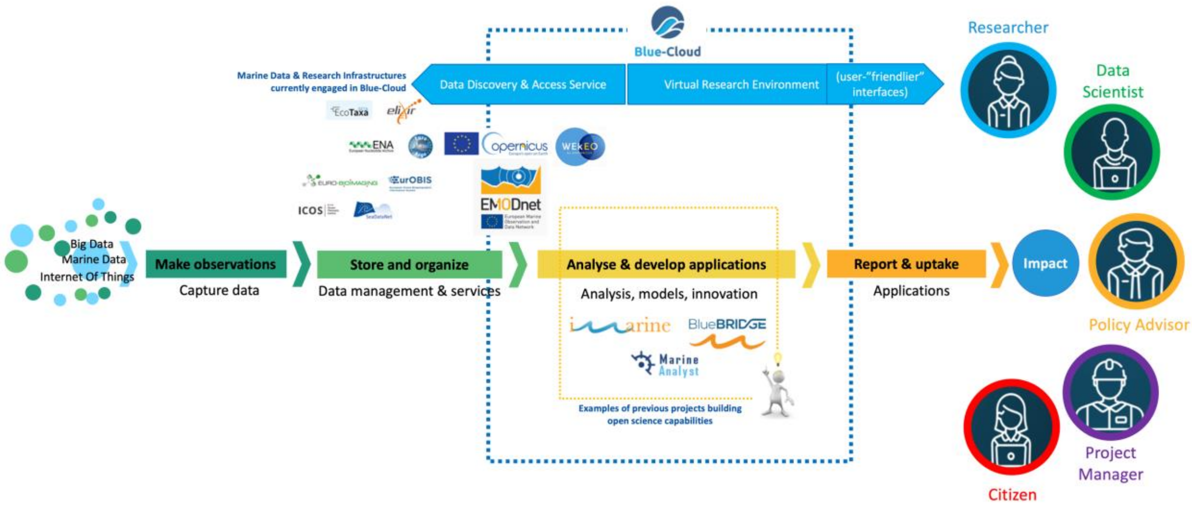

Figure 6: Blue-Cloud's contribution to Europe's marine knowledge value chain. **Credit**: Seascape Belgium

Blue-Cloud's **underpinning technology** and innovative concept builds on existing EU capability and has been designed to deliver **core services** which are deployed through smart federation of leading marine data and e-infrastructures, namely:

- A **Data Discovery & Access Service (B-C DD&AS)**: Offering quick **discovery** and **access** to a range of **marine data** by federating and increasing interoperability between key European data management infrastructures (EMODnet, CMEMs, SeaDataNet, EurOBIS, Euro-Argo and Argo GDAC, ELIXIR-ENA, EuroBioImaging, EcoTaxa, and ICOS-Marine) to facilitate users in finding and retrieving multi-disciplinary datasets from multiple repositories.
- A **Virtual Research Environment (VRE):** Providing a **cloud-based, analytical** and **publishing framework** as a federation of computing platforms and analytical services for constructing, hosting and operating **virtual laboratories ("labs")** for specific applications.
- **Blue-Cloud Virtual Labs**: Configured with specific analytical workflows to serve as **demonstrators** showcasing the potential of open science in the marine domain, demonstrating how providing researchers with **web-based**, **analytical tools** and **greater cloud computing power**, underpinned by **open access to data** available across different infrastructures, can lead to a more efficient development of **innovative data products** that contribute to **improve knowledge** and **understanding** of the Ocean. While virtual labs can be adopted and adapted for inputs and analyses to advance knowledge and innovation across different marine disciplines, five **B-C demonstrators** currently deal with analytical workflows for specific applications and are feeding Open Science resources into **B-C**'s **Catalogue**.

Blue-Cloud's **added value** to Europe's marine knowledge value chain and to society is clearly conveyed through the **B-C** "**demonstrators**", which provide practical examples showcasing how the B-C core assets enable web-based **Open Science** in support of **greater societal ambitions**, contributing to policy objectives of the **EU Green Deal** and the **UN Agenda 2030**. These use-cases all have one thing in common: they demonstrate how providing researchers with **web-based**, **analytical tools** and **greater cloud computing power** -allowing them to perform heavy computational analyses otherwise not possible or too timely to perform on local computers-, underpinned by **open access to data** available across different European blue data infrastructures, can lead to a more effective and efficient development of **innovative data products** that contribute to **improve knowledge** and **understanding** of the Ocean, and thus the overall performance of Europe's marine knowledge ecosystem. Furthermore, as Blue-Cloud is built towards connection with **EOSC**, it will both contribute a "role model" for the development of other thematic communities within this "system of systems" and benefit from its large-scale deployment, which will open opportunities for marine knowledge to inform and feed from other, multidisciplinary knowledge.

The Blue-Cloud's **core services** and **demonstrators**, including the **assets** and **domain specific services** that the project will deliver by **2022** are further described in the **Technical Annex** to this document.

3 The road ahead: Vision & paths of action to 2030

The **Blue-Cloud project** is **demonstrating the potential of web-based Open Science** in the **marine domain**. As a three-year project, the **assets** delivered through Blue-Cloud's efforts by 2022 will require resources for their **future capitalization** and further development. What should be the **overarching vision** for the future evolution of these efforts to 2025? To 2030? What **actions** can be articulated to guide progress towards such a vision? **Who** would have to **join forces** in this **mission**? What **resources** will need to be mobilized? In this section we explore the **road ahead**.

3.1 Methodological framework: Setting a strategic course of action

Defining a **vision** and identifying the **strategic course of action** that should guide the future development of the Blue-Cloud's efforts requires **aligning** with a wealth of stakeholders who are currently knitting and building the underlying foundations for a thriving Open Science community in the marine domain. While in times of Covid-19 convening a large community of stakeholders to invite a long conversation to build such a shared vision has become a difficult task, intensive efforts towards initiating a remote conversation have delivered considerable input and insight. Feeding from initial dialogue with key stakeholders (as introduced in Section 1), we have produced this early draft of the Blue-Cloud Roadmap to 2030 following a "*backcasting*" approach, consisting of working backwards from a very broad "**vision of success**" to deliver a comprehensive overview of the challenges ahead and to reflect on the different ways in which the Blue-Cloud efforts could evolve to contribute to the realization of such vision. It seems logical to presume that if the **Blue-Cloud Vision 2022** has been built towards "***demonstrating the potential of Open Science in the marine domain***", the **Blue-Cloud Vision 2030** should be articulated around the idea of "***realizing the full potential of Open Science in the marine domain, in support of the EU Green Deal and the UN Agenda 2030***". We suggest using this exercise as the starting point for a conversation that will evolve throughout 2021 and into 2022 to narrow the scope of this vision and underpin it with more specific objectives, feeding from the comments and contributions of the Blue-Cloud Community to co-create a shared, targeted **Vision 2030**, including a set of **policy** and **strategy recommendations** to **2030** for the future evolution of Blue-Cloud's assets.

B-C Vision 2030
"Big picture" of what B-C wants to achieve
"Unlocking the potential of web-based Open Science in the marine domain in support of the EU Green Deal and UN Agenda 2030 (including Mission Starfish & Decade of OSSD) by 2025 and realizing the full potential of Open Science by 2030 to guide the next decade of challenges".

From a **policy perspective**, and for the purpose of our efforts, we depart from the assumption that Open Science in the marine domain will have cusped its momentum when it is recognized as driving the **science-based knowledge**, **transformative solutions** and **innovation** required to restore our Ocean and waters by 2030 and to deliver on the objectives of the **EU Green Deal**, contributing with European leadership, concrete actions and results towards meeting the **UN Agenda 2030 SDGs**.

Feeding from different initiatives and expert insight[17], it can be argued that **successful web-based Open Science in the marine domain will require**:

- **Open and FAIR data:** A wealth of marine observations and data that fully capture the state of the Ocean and seas around the world that is openly available to anyone with an interest in exploiting it. This wealth of data should be FAIR (findable, accessible, interoperable and reusable) and to the extent possible open and free, but not necessarily so (in full alignment with EOSC recommendation regarding the protection of security and property rights[18]).
- **Analytical resources:** A wealth of cloud computing resources and supporting web-based analytical tools allowing anyone interested in transforming data into knowledge to be able to attempt to do so, in collaboration with others, at a reasonable cost.
- **Infrastructures:** Basic physical and organizational structures and facilities needed to host, support and operate web-based, open data and analytical resources.
- **People:** A critical mass of participants across the marine and maritime community and other interdisciplinary communities, requiring incentives, rewards and metrics in support of a cultural change that recognizes FAIR and OPEN practice; training; demonstrators of added value of FAIR and open data; and fit-for-purpose interfaces that facilitate, encourage and support wide participation and collaborative science.
- **Policies:** Principles of action for web-based Open Science, widely accepted by the broad marine and maritime community, within the overarching principles provided by EOSC[19].

How should Blue-Cloud's efforts evolve to service a wealth of users and support key stakeholders towards achieving progress in the delivery of successful web-based marine Open Science by 2025? By 2030? This is the key question that we are seeking to respond through this Roadmap.

Table 1 below brings together our projected "vision of success" for web-based Open Science in the marine domain, the different components of such vision, the results that B-C will contribute by 2022 through its **key exploitable results** (KER) and **potential paths of action** that could be laid out for the **future** evolution of these assets **(2025-2030)**. Combined together and taking into consideration the SWOT (*s*trengths, *w*eaknesses, *o*pportunities and *c*hallenges) analyses conducted for B-C, these elements provide a **strategic framework** that can be useful to then inspire a travel plan or **mission** guiding the future evolution of Blue-Cloud's efforts, as a starting point guiding the conversation. Through the early draft **Roadmap** and the consultation, the community will be invited to convey their thoughts on this shared journey, welcoming all contributions. Our efforts will be carefully designed to invite as many travel companions on board as possible, to go as far as possible, together.

[17] Gagliardi, Dimitri & Cox, Deborah & Li, Yanchao. (2014). *What are the factors driving and hindering the adoption of Open Science? An exploratory study.* Manchester Institute of Innovation Research working paper series. 10.13140/2.1.3016.9602
[18] EOSC Strategic Research and Innovation Agenda Version 1.0 https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf
[19] EOSC Strategic Research and Innovation Agenda Version 1.0 https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf

Table 1: Strategic Framework Towards Developing a Vision 2030

The vision of success for each "Open Science supporting component" stems from the relevant policies motivating the B-C efforts (EU Green Deal & Mission Starfish · UN Agenda 2030 & Decade of Ocean Science for Sustainable Development)

Vision of Success 2030 What does successful Open Science in the marine domain look like in Europe by 2030?		Blue-Cloud 2022	Into the future 2025 - 2030
Open science drives **science-based knowledge**, **transformative solutions** and **innovation** to deliver "**Mission Starfish**" objective of restoring our Ocean and waters by 2030 and to support the objectives of the **EU Green Deal**, contributing with European leadership and concrete actions towards meeting the **UN Agenda 2030** SDGs.		**Demonstrate value of Open Science to EU Green Deal and UN Agenda 2030**	**Unlock and realize potential of Open Science in support of the EU Green Deal and UN Agenda 2030**
What does Open Science need to support this vision?		**Possible Paths of Action**	
Open & FAIR data		**What will the B-C deliver? How should the B-C efforts evolve into the future to support these objectives?**	
Ocean observations	• European seabed fully & coherently mapped in high-resolution by 2030. Expand the global Ocean observation system.	• Support FAIR through standard metadata, a Data Discovery & Access Service (DD&AS) and an EOSC compliant, scalable catalogue of "blue" research outputs. **KER:** DD&AS, underpinned by metadata standard for marine data sets. **KER:** B-C Catalogue, underpinned by metadata scheme for research outputs.	• Achieve wide adoption of common metadata standards across blue data infrastructures and marine community. • Support international efforts towards global Ocean data coverage and EU quest for FAIR data in the marine domain by gaining buy-in from the public and private sectors towards FAIR data sharing making a strong case for Open Science and innovation.
Open access to FAIR data	• All public marine data collected by Member States & EU is pooled centrally and openly available and all private marine data is FAIR. A wealth of FAIR data sustainably feeds EU & global marine knowledge value chain.		
Standards for interoperability	• EU standards for marine metadata are available and aligned with international standards, enabling full machine-to-machine discovery and interoperability.		
FAIR data & research objects	• All EU marine data & research infrastructures are federated and have their full catalogues available in an integrated, open access EU Data lake that supports a wealth of Open Science services and disciplines through EOSC.		
Federated infrastructures & analytical resources			
Marine data infrastructures, RIs, e-infrastructures	• Marine data, research infrastructures and e-infrastructures are federated into a single data space (flexible and scalable "system of systems") allowing users to access and exploit **data**, **computing** and **analytical resources seamlessly**, pooling their hosting, HPC and cloud computing power together to service the needs millions of researchers in Europe.	• Showcase approach to interoperability of infrastructures and signpost outstanding challenges. **KER:** DD&AS • B-C VRE offers users a virtual environment with a broad range of analytical services, underpinned by access to data across blue infrastructures and metadata standard for research outputs. **KER:** Virtual Research Environment	• Achieve wide adoption of DD&AS in support of federation of European blue data infrastructures. • Align with technical interoperability efforts towards standards and protocols in support of federation of data AND e-infrastructures. • Support simulation capabilities of European Digital Twin of the Ocean with analytical models, what-if scenarios and other tools developed in virtual labs.
Virtual Research Environments			
Cloud computing resources			
Standards for interoperability	• Access and interoperability standards and protocols between computing infrastructures and workflows (i.e., methods & algorithms) are available, allowing them to be deployed on demand, closer to data.		
Digital Twin Ocean	• EU global digital twin of all oceans and waters is operational.		
A diverse, representative and thriving community of practitioners, users and advocates			
Critical mass of participants **Incentives** **Rewards** **Training** **User-interfaces**	• Millions of researchers and professionals engage in Open Science in the marine domain, boosting research productivity and innovation along the marine knowledge value chain and developing a wealth of user-centric applications of Open Science that trigger transformative solutions and science-based policies towards meeting the EU Green Deal & UN Agenda 2030. • Millions of students use available Open Science capabilities to support their academic objectives. • Each European is a citizen of our ocean and waters.	• Build the case for Open Science in the marine domain. • Nurture and promote wider user uptake of B-C services, capitalizing on B-C assets to service EU RIs, initiatives and projects. **KER:** Virtual Research Environment **KER:** Virtual Labs **Asset:** Community	• Mainstream and service Open Science in the marine domain, servicing a wealth of users through multidisciplinary, multipurpose virtual labs. • Achieve a diverse and representative community with skills, incentives, demonstrators, rewards and fit-for-purpose user interfaces to bring all "blue" stakeholders onboard.
Policies			
Shared policies and rules of participation	• Transparent and consistent terms for participation in Open Science and for use of marine data commons are established with wide stakeholder engagement, achieving trust and confidence for broad participation and a sustainable operational and financial framework for Open Science.	• Alignment with policies established by EOSC	• Support EOSC efforts for wide adoption of policies for Open Science in the marine domain and provide assurance of sustainability, quality of and trust in Open Science.

Extracting from the **strategic framework** above, the different **paths of action** identified have been structured around **five strategic pillars** that Blue-Cloud's future evolution should strive to shore up as part of its long-term vision, namely:

- **Strategic pillar 1:** Encourage efforts towards appropriate and effective **mechanisms** (e.g., governance) along Europe's marine knowledge value chain to ensure the sustainability of **FAIR Ocean data** underpinning (B-C and other) **web-based Open Science environments**, towards long-term sustainability of Open Science services.
- **Strategic pillar 2:** Align with efforts towards full **interoperability** of marine data infrastructures, Research Infrastructures and e-infrastructures to make **digital objects** and **computing resources** available and deployable on any infrastructure, **on-demand**, -supporting open and standardized approaches, where possible.
- **Strategic pillar 3:** Enable a wealth of **applications of Open Science in the marine domain**, delivering on **user needs** by supporting **collaborative research** around topics with a relatively higher degree of **feasibility** and **impact** towards the **EU Green Deal** and **UN Agenda 2030**. Capitalize on B-C core assets to **service** the **needs** of the marine and maritime communities for web-based Open Science -including supporting access of **EU blue RIs, projects & initiatives** to B-C infrastructure and services-, cost-effectively contributing to a wider community of Open Science practitioners and to user-centric applications.
- **Strategic pillar 4:** Evolve a thriving, diverse and representative community of Open Science **practitioners**, **users** and **advocates** in the marine domain -including scientists & researchers, blue economy SMEs, industry, monitoring agencies, policy makers, civil society & citizens- by supporting **skill development** and creating **incentives** and **rewards** for Open Science.
- **Strategic pillar 5:** Continuously monitor and engage with **relevant initiatives** (such as EOSC, EU pilot DTO, DestinE, EuroHPC or GAIA-X, amongst others) and with other **external communities** towards aligning with state-of-the-art technological developments, contributing insights and lessons learned from marine & maritime communities and evolving in response to societal needs and expectations.

By working around these **strategic pillars**, the future evolution of Blue-Cloud will benefit from generating and exploiting **positive feedbacks**:

- Supporting those applications of **web-based Open Science** in the **marine domain** that have a higher chance of materializing will help to catalyse the development of a **growing community of Open Science practitioners** and **users**, who in return can contribute to make collaboration on other research and innovation topics **more feasible**, removing barriers to Open Science.
- An expanding community of Open Science practitioners will contribute to win **buy-in** for **collaborative research**, building a stronger case for **data sharing** and opening new opportunities for **data ingestion** for blue data infrastructures, which in turn will contribute towards FAIRification of available data and **bring new users** and **advocates** to the community.
- Advancing **FAIR data** will make analytical environments more attractive and reliable, **driving more users to e-infrastructures** to seize their computing resources to combine, process and run analysis on growing catalogues of interoperable data sets. Federating e-infrastructures will in turn contribute to driving more users to those catalogues, creating a **win-win scenario for collaboration** between blue data infrastructures and e-infrastructures.

- Aligning with **wider developments** and **external communities** will contribute to broaden Blue-Cloud data space and ecosystem to service **multidisciplinary** research and a **wider user base**, creating the conditions for increasingly connecting to and shaping the EU's future, sectoral **Green Deal data space** as one of the cornerstones of the **European Data Space**.

3.2 Turning strategy into action: Evolving B-C Key Exploitable Results

Considering the strategic framework outlined, Blue-Cloud's efforts will need to evolve with a clear definition of the added value that it can contribute towards delivering on the long-term aspirations for Open Science in the marine domain, building on its **current assets** and **key exploitable results** (as described in the **Technical Annex** to this document). The direction of such contribution is captured in the following, tentative **Mission Statement**:

B-C Mission Statement 2030

General statement on how B-C will achieve our Vision 2030

*"To contribute towards a **European cloud-based data space** that provides access to a thriving portfolio of **analytical**, **simulation** and **visualization capabilities** underpinned by **seamless access** to a **wealth of FAIR**, transdisciplinary **ocean** and **freshwater data**, enabling **Open Science** to deliver knowledge, innovation, science-based policies, collaboration, public awareness and citizenship for a safe, healthy, productive, predictive and transparent Ocean, in support of the **EU Green Deal** and **UN Agenda 2030**".*

Blue-Cloud's **Key Exploitable Results** (B-C Data Discovery & Access Service; B-C Virtual Research Environment, B-C Virtual Labs and B-C Catalogue) should evolve to realize this mission, delivering positive impact around the five **strategic pillars** of action in support of Open Science in the marine domain. In this section we identify potential ways in which such evolution could materialize. Figure 7 lays out the connection between B-C KERs and each of these strategic pillars.

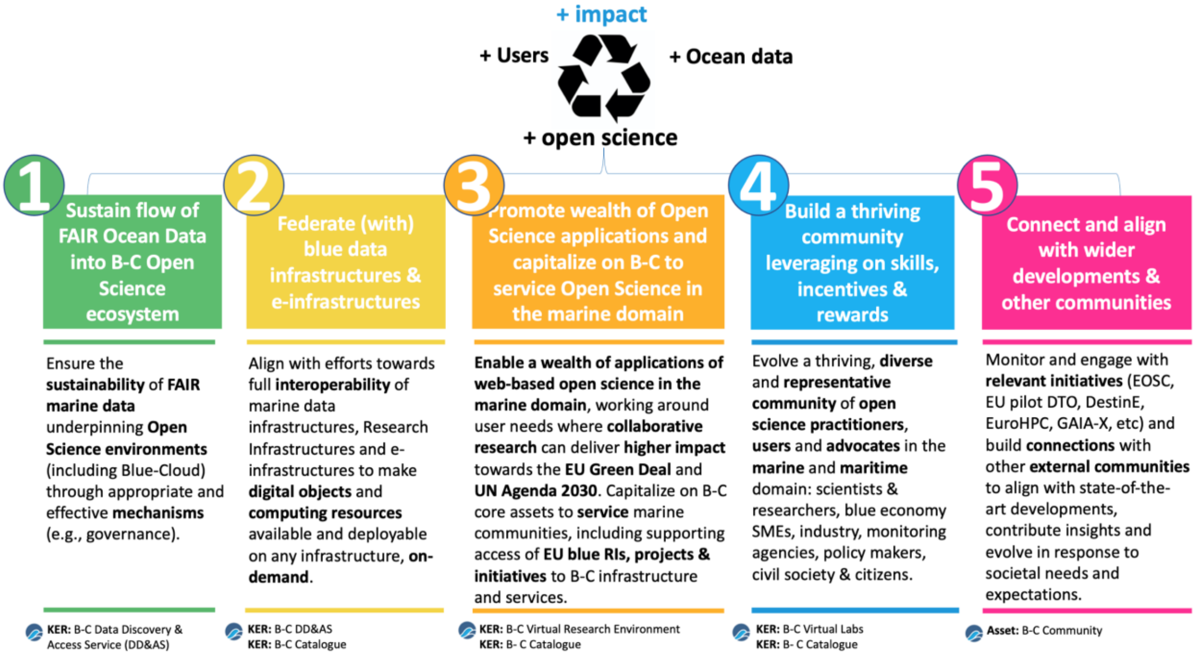

Figure 7: How B-C KERs connect to strategic pillars of action for future evolution of B-C servicing Open Science in the marine domain

3.2.1 B-C DD&AS: Medium- (2025) and long-term (2030) potential and evolution

Medium-term (2025) potential and evolution

The initial pilot version of **B-C Data Discovery & Access Service** is being deployed by interacting with existing web services and APIs from participating blue data infrastructures. However, this means that both the central services and also its users will still be confronted with differences in metadata formats and in supporting vocabularies. In the medium term, this could be further streamlined by upgrading and increasingly harmonising the underlying metadata catalogues. For example, in practice, each of the blue data infrastructures adopting and supporting more and multiple key elements of **ISO 19115 -19139 metadata model** (which is also in use in **INSPIRE**[20]).

In addition, **semantic interoperability** could be deployed, allowing different approaches. One approach could be encouraging blue data infrastructures to adopt **SeaDataNet**'s **Common Vocabularies,** which are leading in the marine data community in Europe and beyond and which currently consist of more than 110 different lists of topics, including variables, units, regions, platforms, instruments, organisations, instrument manufacturers, processing methods, biota, etc. and which are maintained as a crowd source with a clear governance structure, following requirements of user communities. Another option for semantic interoperability could be a **semantic mapping and brokerage at central level**, taking into account the different vocabularies in use by the blue data infrastructure, if any. This way, the metadata profiles and support by controlled vocabularies at each of the blue data infrastructures might be upgraded. In this process, blue data infrastructures might also consider upgrading and/or changing types of protocols for their web services and API's, as currently there is a wide range of protocols being applied and several deployments are lacking functionalities, proper documentation and reliability, in particular in an operational setting.

In a comparable way, also a **further harmonisation of data formats** might be pushed forward, e.g., by adopting two data container models, such as an **ASCII model** and a **NetCDF model**, including **support for controlled vocabularies**, which would allow to formulate dedicated profiles for each of the types of data sets and products that are currently provided by the blue data infrastructures. This would allow to harmonise data models for possibly 90% of all data and products, while still allowing other data model types for specific data types, like e.g., SEG-Y for seismic, FASTQ for genomic data, and others, which are supported by often extensive and precious software packages in specific scientific domains. These upgrading activities should fit perfectly in the strive of blue data infrastructures for improved **FAIRness** and for which several infrastructures are already underway, through projects like **ENVRI-FAIR** and other **EOSC** related projects. Such an upgrading of the underlying blue data infrastructures would contribute to optimising the functioning of the overarching query mechanism, possibly introducing a unified interface also at the second query level and facilitating delivery of data sets from several infrastructures in data formats following the common data models and vocabularies, improving their users' experience.

In the medium term, the range of blue data infrastructures in B-C Data Discovery & Access Service could be **expanded** with **several other relevant marine repositories**. To this effect, a clear distinction should be made between well established and operational blue data infrastructures, which are European aggregators, and project repositories. The recommendation would be for the latter to

[20] https://inspire.ec.europa.eu/

preferably connect and become a provider node for the former, and in this way to become integrated into B-C Data Discovery & Access Service without extra effort on their side. The current B-C Data Discovery & Access Service has a focus on **delayed mode data sets** and **data products**. In the medium term, this might be expanded towards the **Near Real-Time and Real-Time metadata and data brokerage**, however in a close cooperation with **EMODnet Physics**. In the transition from initial to medium-term, also the **Technology Readiness Level (TRL)** of B-C Data Discovery & Access Service and each of its underlying BDI's should increase from the **current average TRL7** (system prototype demonstration in operational environment) **towards TRL8** (system complete and qualified) or **better TRL9** (actual system proven in operational environment). This will require further attention for robustness of services and for an overall monitoring system, considering availability and performances of all services and service chains.

<u>Long-term (2030) potential and evolution of B-C DD&AS</u>

In the longer term, considering the earlier medium-term activities for **harmonising metadata** and **data formats** and **associated delivery mechanisms** of each of the blue data infrastructures and the central **Blue-Cloud Data Discovery & Access Service**, a next step will be the **development of a central Data Lake**, which would serve as a **central data cache** for a major subset of all available data and data products. The selection should be made on **user requirements** of much needed data sets for various applications. The Data Lake could be dynamically maintained through the **Blue-Cloud brokerage system**, interacting with each of the blue data infrastructures, and following their updating and increased population. In addition, several direct APIs with other major systems such as **WEkEO** can be foreseen. The function of the Data Lake would be to build and maintain a service ready **Big Data Collection** which could directly interact with various demanding applications in **B-C Virtual Research Environment**. These applications could be **statistical algorithms**, **data mining**, or **on-demand numerical models**. For that purpose, the Data Lake would not only store big amounts of data, but also transform those into other computer formats which are more suited for fast data throughput and interaction with numerical models and other data intensive applications. Use could be made of **Object Storage** technologies to ensure **high scalability**, **availability** and **durability**. The Data Lake could support cloud-optimised formats (COG/ZARR), including STAC for metadata. As the range of connected blue data infrastructures expands, probably more repositories will get connected to the European infrastructures, thus becoming available in Blue-Cloud Data Discovery & Access Service and Data Lake. New Data Lake components should also strive to achieve **Technology Readiness Level – TRL9** (actual system proven in operational environment) and as part of this, the overall Blue-Cloud monitoring system should be expanded, considering availability and performances of all services and service chains. Considering the higher data throughput and expected large increases in number and weight of user transactions, considerable upscaling of computing and storage resources and associated bandwidths would be necessary.

3.2.2 B-C VRE: Medium- (2025) and long-term (2030) potential and evolution

<u>Medium-term (2025) potential and evolution</u>

With its generic functionality and its wide connectivity options, **B-C VRE** offers researchers major opportunities to undertake **world-class science**. The Virtual Labs come to demonstrate this and serve as a way to promote and market B-C VRE, reaching out and attracting the interest of many marine scientists and researchers wishing to join Blue-Cloud to access its services. In the **medium term**, this should result in an expanding number of users and uses. The modular architecture of Blue-Cloud VRE is scalable and sustainable, being fit for **connecting additional infrastructures**, implementing more and advanced **blue analytical services**, configuring more dedicated **Virtual Labs**, and targeting broader (groups of) **users**. This architecture will be expanded to integrate cluster of GPUs to empower the cloud computing platform with native capacities to handle **analytical methods** based on **Artificial Intelligence**. Those resources will become accessible upon request, within the limit assigned to each Virtual Lab. New way will be explored for adopting additional **cloud storage**, **cloud computing**, **deep learning** and **neural networks** for supporting **big data processes** for **validation**, **extraction**, **interpolation** and **products generation**. Moreover, there will be challenges to serve existing and potential new users with improved and new functionalities for analysing and processing data sets as part of research and for generating data products and knowledge. Data scientists will be empowered to dig into large data sets with new classes of algorithms such as **Artificial Intelligence**, **Machine learning**, and **Data Mining**. This will be implemented and validated through the Blue-Cloud demonstrators. The overall **Technology Readiness Level** of **B-C VRE** should be increased from **TRL7** (system prototype demonstration in operational environment) to **TRL8** (system complete and qualified), which requires all components to be upgraded and made more operationally robust.

Long-term (2030) potential and evolution of B-C VRE

On the longer term, the so far "**data centric approach**" of B-C VRE could be complemented with **mathematical modelling** of **processes**. This could be partly realized in cooperation with **CMEMS**, which is developing and running **large scale ocean mathematical models**. This could be complemented with a set of **on-demand models**, allowing to set up and run **what-if scenarios** for the ocean and marine environment. This way, B-C VRE could develop into a **Digital Twin of the Ocean**, capable of mirroring real-life processes, measures and impacts from within its virtual environment.

Progressing in this direction will require further efforts, including developing and expanding **visualization techniques**, being able to run **what-if scenarios** as interactive games in 4-D (gamification of decision-making processes). This could be very intuitive and useful for decision processes, such as Marine Spatial Planning, developing climate scenarios or for gaining new scientific insights. Techniques from the gaming and animation industry should be studied and brought in for this purpose, also demanding **High Performance Computing** facilities and availability of **GPUs** becoming in the same order of magnitude (tens of thousand) of the **classic CPUs**. To lower the threshold for using new models and visualization techniques by researchers, dashboards should be designed and implemented into B-C VRE to **ease interfacing**. Moreover, **visualization techniques** are expected to become part of the publishing of scientific results, in order to bring science closer to the **general public** in a more effective way. At the same time, use of **Artificial intelligence** and **Data Mining** will be amplified, making use of progressing state-of-the-art and increasing computing resources and capabilities.

B-C VRE could potentially attract many thousands of researchers that will develop and deploy hundreds of **Virtual Labs**, making use of and interacting with models of major processes. Also, B-C VRE could become a platform that is used by many EU Research Projects for **specific workflows** and **products**. To this end, the overall **Technology Readiness Level** of the core Blue-Cloud VRE should be increased from **TRL8** (system complete and qualified) to **TRL9** (actual system proven in operational environment), which requires further upgrading of all components and operations.

3.2.3 B-C Virtual Labs & Catalogue: Medium- (2025) and long-term (2030) potential and evolution

The B-C demonstrators showcase the value that Open Science can deliver towards supporting the realization of the overarching goals of the **EU Green Deal** and **UN Agenda 2030**. Its associated **Virtual Labs** are designed to welcome new practitioners and users on board, broadening their communities to support and enrich their collaborative research with input and contributions from other Open Science practitioners. In the first years of its development, Blue-Cloud is expected to mainly service **scientists & researchers** of the **marine community**. However, its future ambition is to **serve users all along Europe**'s **marine knowledge value chain**, evolving its capabilities to service **policy makers** and **blue economy SMEs** and **industries**, and ultimately **citizens**, **closing the gap** between **scientists** and **society** to **advance science-based solutions** and **align collective action**. Its services could be used to develop tailored applications to different societal challenges, including helping to track pollution, climate change or other impacts of human activities in real-time. With the technology improvements described above, new services developed through future, upscaled Virtual Labs could be open for use by all economic actors, public authorities and civil society to monitor, observe and enable responses to **climate change**, enhance **maritime spatial planning**, support **safe navigation** or promote **citizen participation** in **ocean** and **water governance** and **policy making**. They could serve as **risk-management tools** for the financial and insurance sectors to assess sustainability credentials of investments. The possibilities are **endless.** In the future, **clear guidance** and **incentives** for the development of Open Science in the marine domain could contribute to a wealth of applications. This is important as, in order to deliver the transformative solutions sought by the **EU Green Deal** and **UN Agenda 2030**, Open Science will have to be deployed at a **much larger scale**, well beyond Blue-Cloud's initial demonstrators. Unfortunately, resources are not endless but **scarce**, which is why efforts should evolve towards prioritizing those areas where **collaborative research** has a higher degree of **feasibility** and can deliver relative **higher impact** towards targeted policy objectives (see Figure 8). What does high "**feasibility**" and "**impact**" mean in the **marine domain**? While some general criteria can be outlined, the response will be drafted from community responses to the public consultation:

- **Feasibility:** Abundant **data** and **knowledge** already exist around the topic; plenty of **actors** are already working on the field and making progress; and actors could **benefit** from sharing efforts, with potential **win-win scenarios** for **collaboration**.
- **Impact:** Collaboration in a virtual, web-based environment can deliver **significant added value** (i.e., societal benefits) along the priorities of the **EU Green Deal & Mission Starfish** and **UN Agenda 2030 & Ocean Decade for Sustainable Development**.

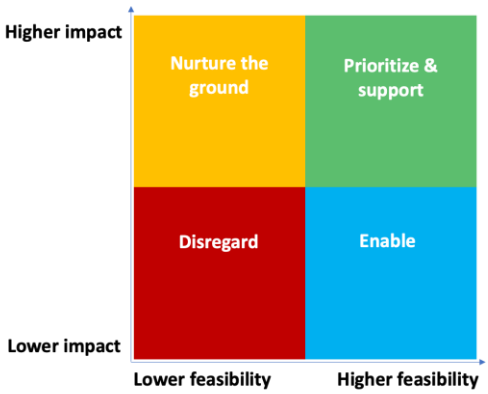

Figure 8: Tentative criteria for the development of future demonstrators. Credit: Adapted by Seascape Belgium

The consultation launched on **B-C Strategic Roadmap to 2030** will seek to gain insight into thematic areas where new applications of B-C services could potentially have a relatively higher chance of success and that could potentially deliver a relatively higher societal impact, based on the experience of B-C community. It will also strive to gather ideas for **potential demonstrators** that could possibly satisfy them, so as to articulate recommendations for their further support and development, including access to **B-C**'s **core services**.

Initial stakeholder dialogue has already provided insight for additional applications of Open Science or "**use-cases**" that could exploit the Blue-Cloud services to deliver societal value. Proposals stemming from the Blue-Cloud community have signposted ideas for **virtual labs** across a broad range of topics, including collaborative efforts towards **Marine Spatial Planning (MSP)**; **monitoring biodiversity**; designating and monitoring **Marine Protected Areas**; artificially intelligent systems to map indicators and predict **coastal flooding** and **erosion**; but also transdisciplinary demonstrators that **integrate social science** and **economic data** or **health data** and **wellbeing data** to inform future decisions or to influence and support **behavioural change**. Blue-Cloud is in dialogue with several EU projects to identify synergies and opportunities towards using Blue-Cloud infrastructure and services to support on-going marine research and innovation initiatives. The outcomes of this dialogue, together with feedback gathered through public consultation on the **B-C Roadmap to 2030** will seek to support the identification of future applications of web-based open science in the marine domain around specific user needs, such as the one showcased as example in the box below.

Open Science applications in support of Marine Spatial Planning

Credit & innovators: DG Joint Research Centre · Cogea (https://www.cogea.it) · CETMAR (https://cetmar.org/?lang=en)

MSP -the process of ensuring that human activities at sea take place in an efficient, safe and sustainable way- has been gaining momentum in Europe over the past decade, as it encourages compatible uses of marine space and resources, reducing tensions around conflicting uses and balancing sustainable use and marine conservation. Striking a balance among the ecological, biological, socioeconomic and institutional aspects within ecoregions is one of the pillars of **ecosystem-based management** (EBM). The increasing availability of data on **human activities** in **European coastal regions** and **seas** (readily accessible through **EMODnet**) opens an opportunity for **Open Science applications** deploying **analytical models** to support better and more timely **planning decisions**, contributing to optimizing the sustainable development of **EU blue economy**.

Supporting the implementation of the EU Marine Spatial Planning Directive

The **MSP Directive** will reach an important critical point in 2021, as the deadline for the establishment of **maritime spatial plans** by EU Member States is met. To date, most maritime spatial planning efforts have taken place at local and national level[21]. With the new spatial plans, setting up a **collaborative**, **EU-wide platform** that could support their aggregation, providing the context and analytical tools to ensure spatial and ecological coherence across borders and sectors and suggesting eventual adaptations would not only be **feasible**, but contribute considerable **added value** towards the policy objectives of the

Modelling and assessing the socioeconomic impact of Marine Spatial Planning on coastal communities

In the last years, different initiatives have developed **models** to quantify the **socio-economic effects** of **MSP** on coastal communities[22]. Evolving and finetuning these models could further contribute to use existing data on **human activities** in coastal areas and at sea to develop different **use scenarios** of **available marine space**, providing policy makers with tools to test different planning options in real time. Besides gaining insight into uses that contribute **higher social** and **economic value**, with a **lower environmental footprint**, collaborating in an EU platform would allow them to factor in how **neighbouring countries** are planning their ocean space, enabling for example **cost-efficient,**

[21] https://www.msp-platform.eu/msp-practice/msp-projects
[22] Surís-Regueiro J.C. *et al* (2021) An applied framework to estimate the direct economic impact of MSP. Marine Policy 127

directive and creating the conditions for a more **sustainable use** of our seas and Ocean.	**collaborative investment decisions** (i.e., on deployment of shared Ocean energy grid infrastructures).
Potential users: ● **Public authorities:** To ensure and demonstrate a transparent and sustainable use of marine space. ● **Blue economy entrepreneurs, SMEs & industry:** To identify and evolve business opportunities benefitting from a level-playing field and sound economic planning in the marine space. ● **Civil society:** To monitor and influence the use of marine space according to societal needs. ● **Citizens:** To actively monitor and participate in policy decisions and Ocean governance.	

In the future, the **B-C Catalogue** will be instrumental in making new applications of Open Science, including their underpinning data, analytical models and results, easily discoverable, accessible and reusable by supporting their dissemination through **EOSC**.

3.3 Service exploitation and sustainability: Responding to user needs

The Blue-Cloud Project Partnership is made up of 20 beneficiaries[23] who are involved in different aspects of the development of the Blue-Cloud project and in delivery of its planned **Key Exploitable Results**, either directly or in other supporting activities, such as management, promotion, marketing, business planning or long-term strategy. Throughout the next months, Project Partners will convene who will continue as active partners in medium-term exploitation of the KERs and who will team up in the search for long-term sustainability. The future roles of individual partners will have to be determined with regard to each KER and the different activities connected to their future exploitation. A first draft **Service Exploitation & Sustainability Plan**[24] (SE&SP) has been produced to guide these discussions and the outcome will be available to inform the final version of this Roadmap. While Blue-Cloud's SE&SP will be geared at identifying and servicing the needs of **current users** of Blue-Cloud services, addressing the needs of **different**, **future users** will depend on a range of factors that strongly influence Blue-Cloud's ability to successfully deliver on their expectations (see Table 2).

Table 2: B-C's user needs & expectations and influencing factors (selected examples)

User profile	User needs	User expectations	Key Success Factor	Line of Action
Policy advisor at public agency	Needs to deliver advice based on the evolution of an SDG indicator, monitoring its evolution over a 10-year timeframe	Confidence that data will continue to be available over a 10-year period to enable comparability	Sustainable flow of marine data into B-C and commitment of data infrastructures to support & update APIs	**Sustainability of FAIR Ocean data**
IT manager at NGO	Needs to work on a 2-year research project in a Marine Protected Area, working with field data collectors, citizen scientists and satellite experts	A reliable infrastructure offering a sound environment and services for open, collaborative research	Sound technical solution offering state-of-the-art computing resources and user-friendly analytical tools	**Service quality and sustainability**
Project manager at industrial	Needs to monitor environmental parameters around an industrial	One-time service offering easy access to marine data	Flexible service with flexible pricing,	**Competitive offering**

[23] https://www.blue-cloud.org/partners
[24] D6.3 Blue-Cloud Services Exploitation and Sustainability Plan (Release 1)

maritime facility	offshore facility and requires marine data from different locations	available across different infrastructures	Service Level Agreement	
Data scientist at consulting company	Needs to deliver an Environmental Impact Assessment for an aquaculture company planning new sites, requiring water quality data from different locations at 2-month intervals and sharing analysis in a confidential environment	A service offering comparable water quality data from different locations, analytical tools and confidence in applied data management protocols and access rights	FAIR Data management standards for marine data and sound data management protocols	**Continuous adaptation of environment to adopt and adapt to state-of-the-art FAIR data management standards**

Some of these factors are external to B-C (depending on developments in its landscape), while others are internal (depending on developments within its overall management and operational framework). The **B-C Roadmap** will seek to deliver recommendations towards addressing those external factors that fall outside of the scope of B-C's influence and thus require action at wider policy level. **B-C Service & Exploitation Plan** will focus on establishing a sound management framework to address those factors that are key to delivering on user expectations and can be addressed with sound technical, operational and commercial policies, taking into consideration B-C's **surrounding landscape**.

3.4 Identifying policy recommendations to 2030

The SWOT analysis undertaken for B-C, together with the vision, the strategic framework, the potential evolution of B-C KERs and the service exploitation and sustainability considerations introduced in this section seek to inform the **identification** of **policy recommendations** for the future evolution of Blue-Cloud's efforts, including **key actions** underpinning their implementation. Articulating such recommendations requires gathering responses to key questions, which will be addressed to the Blue-Cloud community through the planned, public consultation:

- How should **Blue-Cloud** evolve to support a **thriving ecosystem** for web-based **Open Science** and **Open data** in the **marine domain**?
- What additional **mechanisms** (e.g., governance) could add value to the existing EU long term marine data services to further ensure the **sustainability** of **FAIR Ocean data** underpinning B-C's Open Science environment and services? Which could best contribute to a broad **alignment** of the **marine community** towards future **EU pilot DTO** and **DestinE** developments?
- What **applications** of **Open Science** in **the marine domain** could have a higher probability of success, given current availability of data, models and actors willing to engage in collaborative science, across a broad range of topics? Which of such applications should be prioritized towards addressing current **user needs** and delivering **highest societal impact**?
- How should B-C's Open Science environment and services evolve to be fit-for-use not only for operational researchers & scientists, but also for other Open Science users such as **policy makers** and **blue economy SMEs** and **industry**? What needs do these users have that B-C could evolve to address? What **skills**, **incentives** and **rewards** can contribute to bring Open Science practitioners on board?

- How can B-C evolve to **further connect** with **marine data infrastructures** and **research infrastructures** to deliver full interoperability of marine data through the **B-C Data Discovery & Access Service**, aligning and in collaboration with other **international efforts**?
- What actions would be required to enable **B-C's Catalogue** of **analytical methods**, **algorithms** and **applications** to **be deployed in EOSC,** but also in other infrastructures **-closer to data-** or across **supercomputing platforms** in Europe?
- How should **B-C's assets** evolve to align with future **EU pilot DTO** and **DestinE** developments and other relevant, wider initiatives?

Join us in shaping the future of Blue-Cloud:

Bring **your views** and **opinion** on these questions and have your say on shaping **strategic policy recommendations** towards the future development of Blue-Cloud, so it evolves to deliver on your **needs** and **expectations**. Take the survey and provide feedback to the end of **September 2021**:

https://www.blue-cloud.org/form/blue-cloud-roadmap-2030-2nd-online-consultation

Conclusions

This early draft of the **Blue-Cloud Roadmap to 2030** introduces the policy context, opportunities and challenges motivating **Blue-Cloud**'s efforts, exploring the emergence of **Open Science** in the marine domain in the context of the digital age and how it can contribute to **support** the **European Green Deal** and the **United Nations Agenda 2030 for Sustainable Development**. It analyses relevant developments shaping Europe's **marine knowledge value chain** and describes the **added value** that **Blue-Cloud** will bring to this landscape by **2022**, in the form of **Key Exploitable Results**.

Progress achieved so far in the development of Blue-Cloud's assets confirms the key role of **Open Science** in the marine domain in **delivering science-based knowledge**, **solutions** and **innovation** to address societal challenges. B-C's pilot demonstrators are showcasing value to **specific policy objectives**, identified in **Mission Starfish** and **SDGs 2, 13** and **14**. But also, through its services, Blue-Cloud is building on the solid foundations of Europe's marine knowledge value chain to continue making progress towards delivering on the needs of **successful open science in the marine domain**, fostering the development of science that is shared openly and available for re-use. Through its **B-C Data Discovery & Access Service**, it is simplifying access to data available across different blue infrastructures from within its **Virtual Research Environment**, which provides a strong analytical framework for different open science practitioners to meet and work together. **B-C**'s **Catalogue** will further make a wealth of analytical resources available for re-use. Last but not least, an **emerging community** of **early open science practitioners** is gathering around its existing **Virtual Labs**, which will inspire other users and practitioners to join and build a larger, thriving Open Science community in the marine domain, into the future.

Capitalizing and evolving the Blue-Cloud's efforts into the decade will be key to **unlock the full potential of Open Science in the marine domain**, enabling the development of a wealth of Open Science applications that can deliver science-based solutions and innovation to address societal ambitions and related policy objectives at a proper scale. This will require aligning action with a wealth of stakeholders who are currently knitting and building the underlying foundations for a thriving Ocean Open Data and Open Science community in Europe. This early draft of the B-C Roadmap to 2030 introduces an overarching **vision** to guide the future capitalization and further development of Blue-Cloud's **KERs**, proposing strategic, high-level **paths of action** that could be activated to guide progress towards such a vision.

As the **first early draft "roadmap"**, this document **does not yet include policy recommendations**, but sets the scene for **wide stakeholder consultation** towards co-designing a **community vision** and identifying **concrete actions** for the future evolution of Blue-Cloud's **added value** into the decade, delivering insight into Blue-Cloud's current **strengths** and **weaknesses** and into the **opportunities** and **challenges** ahead. Its content will be released in June 2021 for consultation, welcoming contributions throughout the Summer. The **feedback** and **contributions** gathered through this public consultation, together with additional dialogue with key stakeholder communities throughout 2021 and 2022 will shape the final, community-driven **Blue-Cloud Roadmap to 2030** which will be delivered to the **European Commission** by **July 2022** (as Project deliverable D6.4).

Technical Annex

TA1 Europe's marine knowledge value chain

Over the past decades, Europe has developed an impressive capability for marine observation, anticipating the need to support and advance our knowledge of the Ocean. These efforts have contributed to the emergence of an advanced **European "marine knowledge value chain"** which brings together multiple European actors and assets (see Figure 9).

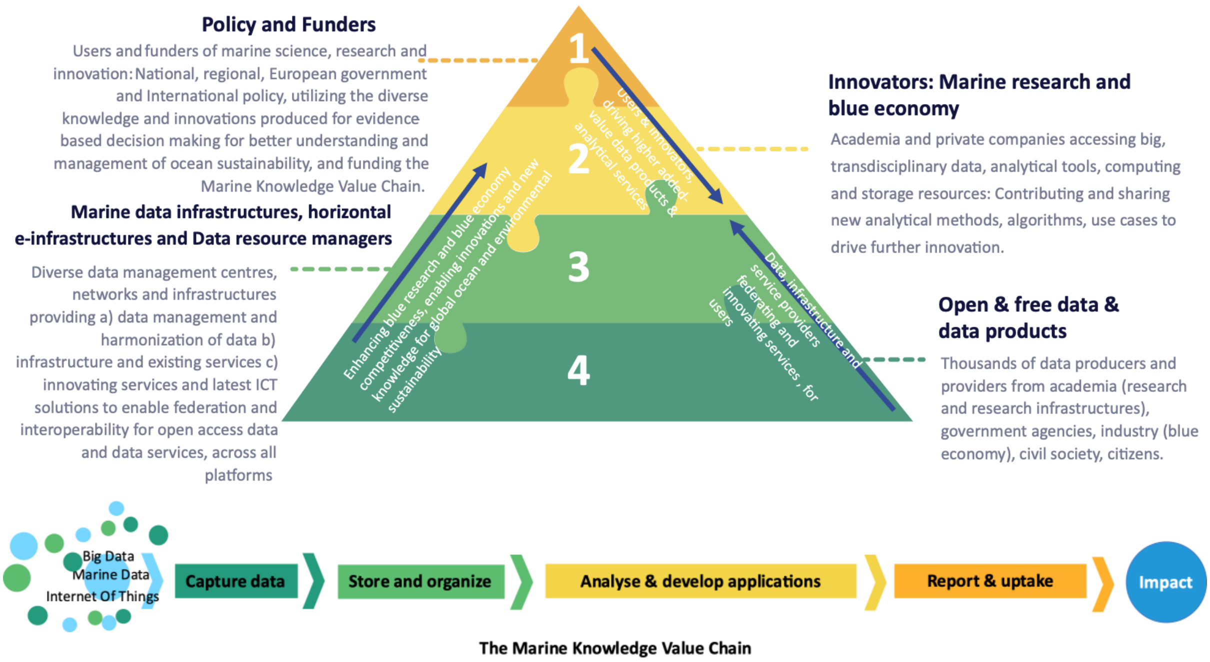

Figure 9: The European marine knowledge value chain: Data, infrastructure and service providers and users, including marine research and blue economy, policy makers and funders. Credit: MARIS and Seascape Belgium

Europe's efforts to support increased knowledge of our seas[25], aligned with action at national level, has led to a remarkable increase of abundance of **marine data**, stemming from academia, major Research & Development projects and initiatives, Research Infrastructures (RIs), industry and civil society. The emergence of citizen science and technological developments such as the "Internet of Things (IoT)" anticipates unprecedented potential to exponentially grow such abundance. A resilient, existing European network of distributed **infrastructures** for data gathering, handling and sharing is making it possible to **manage**, **enhance** and **channel** this wealth of data into an additional layer of actors in the marine and maritime research and blue economy communities. Marine data infrastructures and services, such as the **European Marine Observation and Data Network** (**EMODnet**) and the **Copernicus Marine Environmental Monitoring Service** (**CMEMS**), together with **Research Infrastructures** and **horizontal e-infrastructures**, play a key, strategic role in making data **discoverable**, **accessible**, **interoperable** and **reusable** to **marine data user communities**. Through their efforts, these communities are able to apply new **analytical** methods to **transform available data into knowledge**, adding modelling and forecasting capabilities and opening new opportunities to drive **innovation**. The resulting knowledge is key to inform better **science-based policies**, using existing evidence to drive a **sustainable use and conservation of the Ocean** through sound management and governance.

[25] COM(2014) 254 final/2 Innovation in the Blue Economy: Realising the potential of our seas and oceans for jobs and growth https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=COM:2014:254:REV1&from=EN

Further insight into how this **marine knowledge value chain** has shaped throughout the years, the **progress** achieved so far and the **challenges** that it is currently facing, is key to understanding the **added value** and **momentum** that the Blue-Cloud's efforts are striving to deliver.

In 30 years, Europe has developed an impressive Ocean observing capability

Fit for purpose and sustained ocean observations are an essential part of worldwide efforts to understand marine social-ecological systems. Observations can be samples collected on ships, measurements from instruments on fixed platforms, autonomous and drifting systems, submersible platforms, ships at sea or remote observing systems such as satellites and aircrafts. In Europe, oceanographic and marine data are collected by several thousands of **research institutes**, **governmental organizations**, and **private companies**. Various heterogeneous observing sensors are installed on research vessels, submarines, aircraft, and moorings, drifting buoys, gliders, floats, fixed platforms, and satellites. These sensors measure physical, chemical, biological, geological and geophysical parameters, with further data resulting from the analysis of water and sediment samples for a wide variety of parameters. Governments deploy monitoring programs at national and regional level to support national marine governance goals and European policy commitments. Research institutes in Europe operate research vessels and many other platforms and instruments for gathering in-situ observations and samples in European waters, but also on a global scale. Marine data are also collected by private organizations in support of economic activities. In the past 30 years, the **EU** has also developed a **world-leading position** in **ocean observation** from **space**. In 2014, the EU launched **Copernicus**, the European Union's Earth observation and monitoring programme, served by a set of dedicated satellites and in-situ systems. Today, it produces a wealth of data and information on the Earth's sub-systems (land, atmosphere, oceans and inland waters) and cross-cutting processes (climate change, disaster management and security). It is fair to say that thanks to these efforts and through close collaboration with its Member States, the **European Union** boasts an unrivalled **marine observation** and **forecasting infrastructure** that is increasing our understanding of the oceans.

Europe has made huge progress advancing FAIR and open data in the marine community

Only a few decades ago, marine data from all those observations described above were difficult to find, hardly accessible and extremely cumbersome to put together because of different standards, nomenclature and baselines. As a response to the need to develop a more coherent and interconnected system towards exploiting data from observations, since the early nineties, a wide range of initiatives funded and/or supported by the EU have not only contributed to further develop Europe's capabilities for collecting and managing marine in-situ and remote sensing data, but also to develop and advance **data centres** and **data management systems** servicing the marine domain. As a result, great progress has been achieved in terms of developing **data standards** and **data management services**, through the establishment of dedicated infrastructures.

EMODnet, CMEMS and the Data Collection Framework collectively implement the **Marine Knowledge 2020 Strategy**. Over the past decade there has been significant development towards **FAIR** and **open data** in the **marine domain** integrating data from various sources, harmonising and standardising these to EU e.g., INSPIRE (and increasingly international) standards and making more information freely available as interoperable data layers and data products following the principle of "*measure once, use many times*".

These EU data brokers and services have many operational partnerships with other, well established, European **marine data management infrastructures.** For instance, the EU infrastructure SeaDataNet is a network of national ocean data centres that are a central part of the EMODnet data flow for a number of thematic areas. And EurOBIS (marine biodiversity focus) is the EU counterpart of the global OBIS data infrastructure, with very close links and data flow with EMODnet Biology. Other thematic research and distributed infrastructures include **ELIXIR-ENA** (life sciences) and **ICOS-Ocean** (ocean carbon component of ICOS), amongst many others, are funded by, and operated by EU Member States.

These blue infrastructures have already started working together to continue supporting, expanding and mainstreaming the application of FAIR data management practices, collaborating through EU funded initiatives such as **ENVRI-FAIR**, which is working towards better analysing and improving progress towards FAIR in four environmental subdomains, including the marine domain, as well as aligning requirements towards **EOSC**. Adding to these collaborations, blue infrastructures are working in close interaction with international initiatives led by the **United Nations Intergovernmental Oceanographic Commission** (IOC), the **World Meteorological Organization** (WMO), the **Food and Agricultural Organization** (FAO), the **Group on Earth Observations** (GEO) and the **International Council for the Exploration of the Sea** (ICES), amongst others.

Thanks to all these efforts, marine data is now recognized as a **public good** in Europe and multi-resolution maps of all Europe's seas and oceans are available, together with large catalogues of **open** and **free data products** that span over multiple disciplinary themes. European efforts to promote and apply the principles of free and **open access** and **interoperability** in the marine domain are demonstrating clear value, through long-term data brokers and services e.g., **EMODnet**, and community research and development funded through **European Research Framework Programmes**, most recently **Horizon 2020** and the upcoming **Horizon Europe.**

<u>The marine community has responded positively to pilot analytical frameworks seeking to support web-based Open Science in the marine domain</u>

Efforts along the marine knowledge value chain have also extended into applying technology towards enhancing the **analytical capabilities** of the marine community. Data availability is not a sufficient condition to support systemic and transformative solutions for healthy oceans, seas, coastal and inland waters. As **Mission Starfish** points out, "*while the effects of pollution, human activities and climate change are documented and observable, significant gaps in our understanding and knowledge of the hydrosphere and the challenges it faces remain*". Gaps in our understanding of how to address these challenges effectively and holistically, also remain. As mentioned before, the emergence of IoT and citizen science may bring unprecedented potential to exponentially grow the abundance of Ocean data. The question is, are we ready to make the most out of this wealth of data?

Through its Research and Innovation (R&I) programme, the EU has supported the development of **e-infrastructures** as the pathway towards integrating **data services** with **analytical tools** in a seamless way, increasingly at large scale. Piloting initiatives such as the i-Marine project, the BlueBridge project, Marine Analyst or the SeaDataCloud VRE development have shown how comprehensive data management solutions can support the application of a science-based ecosystem approach to specific domains (e.g. fisheries or marine spatial planning), as well as providing attractive capacity building environments for interdisciplinary research communities involved in increasing our knowledge of the marine environment. Thanks to these efforts, a growing **catalogue** of **virtual laboratories** (“**virtual labs**”) is making **Open Science** more **accessible** to researchers in the marine community, providing them with greater **computational resources** and **tools**. The e-infrastructure D4science[26] has serviced more than 2,500 users in the marine domain through its over 60 “blue” virtual labs, as a solid proof of the growing awareness and interest from the marine research community of the opportunities to thrive with the use of new digital technologies. D4Science currently services 7,000 user from 44 countries.

What is a virtual lab?

A virtual laboratory is a web-based, problem solving environment that enables users across different physical locations and expertise to collaborate efficiently in an ongoing way, providing them with tools and services to share data, analytical methods and results towards solving complex problems.

Blue data infrastructures are also increasingly involved in generating models and new, knowledge-oriented data products which are run by their teams or made available as services for external users from research, government and industry. In 2018, the EU launched **WEkEO**[27] as one of the five Copernicus Discovery and Access Service (DIAS) to provide a single access point to all Copernicus **data** and information, alongside **processing resources**, **tools** and other relevant data. The overarching objective of DIAS is to enhance access to Copernicus data and information for further use in an efficient **computing environment** implementing the paradigm of “bringing the user to the data”, as one condition for unlocking the potential value of Copernicus for innovation, science, entrepreneurship, business and economic growth and science-based policies. WEkEO is the service for **marine environmental data**, **virtual environments** for data processing, and skilled **user support**.

But a number of challenges are yet impeding progress to unlock the full potential of web-based Open Science in the marine domain, in support of the EU Green Deal and the UN Agenda 2030

In spite of this extraordinary progress, a number of barriers are still impeding the marine community from embracing **Open Science** and unlocking its full potential in the marine domain, in support of wider societal objectives. In 2020, the Blue-Cloud project launched an initial phase of dialogue and consultations with different stakeholders across the Blue-Cloud community[28]. Although the dialogue was focused on better identifying their needs and expectations with regard to the Blue-Cloud project, it was also instrumental to gain insight into some of the **challenges** that are currently hindering Open Science in the marine domain, and therefore a more productive marine knowledge value chain in Europe:

- Through the stakeholder consultation in 2020, a survey was launched towards identifying the most pressing “pain-points” for the community in its dealings with marine data. A **large majority** of respondents who identified as **users of marine data** signalled as “significant” or “very significant” problems: “*harmonizing data from different sources*” (86%); “*performing computing and analytical processes across data sets*” (78%); “*having to search for data in different portals*”

[26] https://services.d4science.org/explore
[27] https://wekeo.eu/web/guest/home
[28] Summary of Key Messages & Recommendations of the 1st Phase of Consultations with the Blue-Cloud Community (Seascape Belgium, 2020)

(77%); "*sharing and processing large data sets from observations*" (75%) and "*finding the data for which I am searching*" (72%). A **majority** of respondents who identified as **producers of marine data** signalled as "significant" or "very significant" problem "*transforming datasets into higher added value products with specific applications*" (69%). Some of these results come to support **Mission Starfish**'s claim that additional effort is required to further federate data infrastructures, to continue improving the **discovery**, **access** and **interoperability** of Ocean related data. Initial stakeholder dialogue with the **Blue-Cloud External Stakeholder & Expert Board** (ESEB)[29] has further identified the lack of **standards** for **metadata** as one of the main challenges blocking progress towards FAIR in the marine community. A recent report of the EOSC Secretariat supports the idea that **properly structured metadata** to aid findability, along with provision of services via uniform and compatible encodings using community-adopted standards, will be required to support better discovery and access to data but also machine-based processing of data flows in the marine domain, in support of interoperability[30]. The Blue-Cloud project, as described in the next section, is seeking to contribute with effective steps towards larger on-going efforts for better discovery, access and interoperability in the marine domain, facilitating dialogue amongst blue data infrastructures, followed by concrete actions.

- However, pooling data centrally will not solve some of the most compelling needs of the marine community regarding their need to **process large data sets** from observations and performing **computing** and **analytical processes** across data sets. Nor will it immediately inspire them to transform datasets into higher added value products. Transforming **data** into **knowledge** and into **innovative applications**, **products** and ultimately **solutions** to current societal challenges will demand **computing resources**, **stronger analytical frameworks** for Ocean related data and **closer public-private collaboration** between scientists, researchers, innovators, entrepreneurs and business developers. Providing broader access to **user-friendly collaborative tools** and **services** towards **transforming Ocean data** into **societal knowledge** is one of the actions proposed by Mission Starfish towards the objective of "healthy oceans, seas, coastal and inland waters". While the cost of computing may have been a limiting factor for the success of Open Science environments in the past, the emergence of new, **"pay-on-demand" services** for computing resources (allowing users to only pay for the computing resources they use, rather than having to pay for more costly monthly or yearly subscriptions or to invest in computing infrastructure themselves) opens new opportunities for the development of collaborative tools and schemes. The private sector is making computing services increasingly affordable, providing users with collaborative environments (such as Amazon's SageMaker, Google's Colab or Microsoft's Azure Notebooks) catering to their data computing needs. Signalling the relevance of these developments, NOAA has put its World Ocean Database (WOD) onto Amazon's servers, bringing the data and the computing power closer together. Users of Amazon's cloud services now have faster access to WOD, thus creating better conditions for data scientists and new discoveries. The data is open and can be downloaded (without requiring user authentication).
- For these collaborative environments and/or schemes to deliver results, however, they require engaging a **critical mass of players**. At the moment, as initial dialogue with the Blue-Cloud ESEB

[29] https://blue-cloud.org/eseb
[30] European Commission - Directorate-General for Research and Innovation (2020) Six Recommendations For Implementation of FAIR Practice https://ec.europa.eu/info/publications/six-recommendations-implementation-fair-practice_en

has highlighted, Open Science is not yet mainstreamed in the marine domain -neither across public nor private organizations. In general, and in spite of the early signals of a significant cultural change, **uptake** of sound (FAIR) **data management** practices -one of the underlying enablers of Open Science at a broad scale- is still low across the academic sector. So is the uptake of **Open Science tools** and **environments**. With 1.7 million researchers in Europe, and given the scale of the challenges ahead, a penetration of 7,000 users in a successful Virtual Research Environment (VRE) signals a long road ahead for the scientific community. As the aforementioned report of the EOSC Secretariat also points out, it is widely seen that researchers do not see sufficient benefits of **FAIR** data, and therefore are not willing to put in the efforts in implementing FAIR data management practices. A similar reasoning could be made for **Open Science**. The report points at the need to **fund** FAIR **awareness-raising**, **training** and **education** in academic and research institutions, as well as **funding**, **rewarding** and **recognising improvements** of **FAIR practice** and developing and **monitoring** adequate policies for FAIR data and research objects. It also highlights that while a successful strategy towards mainstreaming wide adoption of FAIR will benefit from top-down policies, FAIR seems to evolve naturally and more effectively when motivated by **need** and **common benefits**. The report signals "communities" as being key to advance towards agreed formats for data, common vocabularies, metadata standards and procedures for how, when and where data is shared.

Therefore, making the case for both **FAIR data management** and **Open Science** in the **marine domain** by **showcasing** its **benefits** is key to **achieve buy-in** from the marine, academic, scientific and research communities. Building an initial community of **early Open Science advocates and practitioners** can be a powerful contributor/catalysator to that objective. **Virtual Research Environments** provide a good ground for researchers to experiment with available **tools** and **facilities** and to grasp the benefits of Open Science, acting as a **bottom-up community building tool**. However, through its Report, the EOSC Secretariat also alerts that "*solving findability and accessibility of data within a discipline by bringing the data together in a virtual research environment can result in a larger silo of data that no longer interoperates with other disciplines*". While virtual labs have a clear, demonstrative value in the short to medium term, building them to **anticipate easy integration into larger environments** that allow their **exploitation** by **larger, interdisciplinary audiences** can be a way of mitigating this risk.

FAIR data practices are equally challenged across **public agencies**, the **private sector** and **NGOs**. Showing the benefits of FAIR data management practices is equally important to progress public-private collaboration efforts. Initiatives like EMODnet's **data ingestion service** are an important step in the right direction, as they encourage all types of actors across the public sector, private sector and civil society to contribute and share datasets for further processing as **FAIR data** and in most cases towards publishing as **open data**. Like amongst the research community, making the **benefits** of this **win-win collaboration visible** is critical to achieve buy-in from the private sector and civil society. However, again, FAIR data is a necessary but not a sufficient condition for Open Science. The **drivers** and **motivations** of private companies and NGOs to engage in sharing of data, processes and analysis results are different from those of the academic and the scientific community. Understanding these drivers and motivations and creating the conditions to respond to their needs and expectations will be key to bring these actors into collaborative Open Science schemes geared at delivering solutions to societal challenges. On-going initiatives powered by the private sector (e.g. Kaggle, featured in the box below) can provide inspiration towards developing **collaborative Open Science environments** that provide win-win results to all stakeholders involved, by generating value that aligns with their distinctive motivations.

Kaggle: An example of win-win collaboration in support of data science in the private sector

Recently acquired by **Google**, online data science and machine learning community Kaggle is home to over **one million users** ranging from computer science Ph.D. holders conducting cutting edge research to absolute beginners. Kaggle is best known for its data science competitions that offer (substantial) cash prizes, but it also serves as an educational tool for autodidacts as well as a place to present one's portfolio of related work. By connecting talented data scientists with tough problems, motivating them through lucrative cash prizes, and assisting their professional development through educational and portfolio resources, Kaggle creates substantial value for its users. Private partner organizations can similarly develop custom-built solutions for their business challenges while identifying the best talent to recruit. Finally, Google earns revenue from these partners while at the same time building its credibility in the arena of data science, familiarizing users and building customer loyalty towards their tools and lowering barriers for the uptake of other related commercial services[31].

In summary, in spite of its impressive capabilities, Europe's marine knowledge value chain still faces challenges towards further enhancing its performance and unlocking the full potential of Open Science, for the benefit of all engaged actors and ultimately, European citizens. On the one hand, it requires "lubrication" at different points of its underpinning gear, including **advancing data management practices towards FAIR principles** across blue data infrastructures; making more of the currently available data **open** (and to the extent possible, **free**); and **connecting to other data spaces** (as described in section 2.2) to support interdisciplinary research. But alongside responding to those challenges, Europe's marine knowledge value chain can benefit from developing more robust **frameworks** for **analysing data**. Harnessing the power of **web-based, "cloud" technologies** can boost the value chain's **productivity**, bringing a huge leap forward in the way that **marine data** -as well as the **processes** and **results** linked to **exploiting** and **analysing** such data- is **used**, **shared**, **enriched** and **applied**. Previous experiences have led the way towards testing and advancing technological infrastructures, methods and tools towards this end, introducing opportunities to pilot supporting frameworks for Open Science along the marine knowledge value chain. However, they were limited in their scope to influence changes both upstream ("data access") and downstream ("sharing outputs" across different communities and gearing efforts towards bridging the "science-policy" gap). The **Blue-Cloud project** is a step forward in that direction, **evolving existing computing** and **analytical capabilities** and underpinning them with **easier access to data** delivered across key blue data infrastructures through the use of **common languages** (standardized metadata), seeking to signpost potential solutions towards **federating their data services**, while showcasing how **Open Science** can deliver **impact** on key, **priority policy areas** catering to a **broader range of users**. In the next section, we look deeper into the Blue-Cloud's efforts and how they are, and will contribute to, **unlocking the potential of web-based Open Science** along Europe's **marine knowledge value chain**, in support of the **European Green Deal** and the **UN Agenda 2030.**

[31] Kaggle: Building a Market for Data Science (and Scientists) https://digital.hbs.edu/platform-digit/submission/kaggle-building-a-market-for-data-science-and-scientists/#:~:text=The%20online%20data%20science%20and,edge%20research%20to%20absolute%20beginners.

TA2 Blue-Cloud's added value and core services

The Blue-Cloud project launched in October 2019 aiming to **demonstrate the potential of Open Science in the marine domain**. This is the **vision** that captures what Blue-Cloud is seeking to achieve by 2022. To achieve its vision, the Blue-Cloud project is developing a **web-based, cyber platform** that provides marine scientists and researchers with enhanced **analytical capabilities** that enable **Open Science**, including powerful cloud-computing resources, a range of analytical tools and simplified access to data from in-situ and satellite observations, data products and model outputs available across different blue data infrastructures. Initial stakeholder dialogue and consultations in 2019-2021 have led to the articulation of a short-term **mission** to guide the project through its implementation, towards delivering on its vision:

Blue-Cloud Mission 2022

*"To promote the sharing of data, processes and research findings in the marine domain by delivering a **collaborative web-based environment** that enables **Open Science**, underpinned by simplified access to a wealth of easily discoverable and interoperable marine data and products".*

The Blue-Cloud project will work around accomplishing **five overarching, strategic goals** by 2022:

1. Build the case for Open Science in the marine domain.
2. Establish dialogue and showcase approach towards improving interoperability of marine data infrastructures, research infrastructures and e-infrastructures.
3. Bring together a thriving community of early practitioners of Open Science that attracts engagement and participation of different stakeholders.
4. Promote wider user uptake of Blue-Cloud's Open Science services.
5. Demonstrate value to EU Green Deal and UN Agenda 2030.

In the **medium-** and **long-term future**, the Blue-Cloud aspires to upscale the results of this effort, aligning with wider developments at European level to unravel new opportunities for **innovation**.

TA2.1 Building on existing assets

To deliver on its **mission**, the Blue-Cloud project has brought together for the first time some of the key **blue data infrastructures**, **Research Infrastructures** and **e-infrastructures** currently servicing the **marine knowledge value** chain in **Europe**, namely (see Figure 10):

- **Blue (marine data & research) infrastructures:**
  - Copernicus & CMEMS (climate and ocean analysis and forecasting)
  - EMODnet (bathymetry, biology, chemistry, geology, physics, seabed habitats and human activities)
  - ELIXIR-ENA (life sciences, biogenomics)
  - Euro-Argo and Argo GDAC (ocean physics and marine biogeochemistry)
  - EuroBioImaging (microscopy)
  - EurOBIS (marine biodiversity)
  - ICOS-Marine (carbon)
  - SeaDataNet (marine environment)

- **e-Infrastructures:**
    - D4Science
    - EUDAT
    - WEkEO

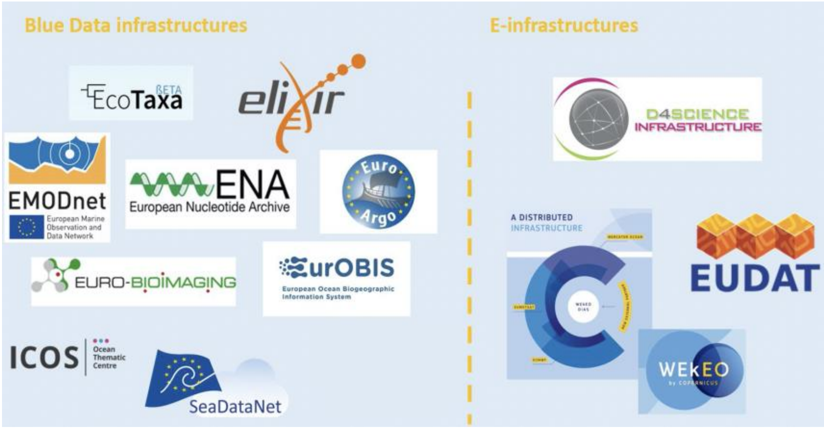

Figure 10: Leading infrastructures bundling their forces for the pilot Blue-Cloud project. **Credit**: MARIS

The Blue-Cloud project has opened a dialogue amongst these key players to explore a common approach towards making the wealth of data available across these different infrastructures more easily **discoverable** and **accessible** to users, but also more easily **combined** and **exploited** by them to support their **analytical research needs** and **objectives**. While the Blue-Cloud project will only be working with selected data catalogues and not with the full offer available across the participating blue data infrastructures, this effort will deliver a guiding reference towards further broadening its scope with additional catalogues from these and potentially other European (and international) blue data infrastructures and towards integration with **EOSC**. The challenges around data are closely related to the recent discussions around Data Space concepts[32] aligned with EOSC Core and EOSC Exchange of delivering services. New rules from the European Commission in a proposal for a Regulation on European Data Governance issued in November 2020 will pave the way for data to be harnessed and for sectoral European data spaces to benefit society, citizens and businesses. The selected data catalogues within Blue-Cloud could be part of the Data Spaces' concepts offering an environment for hosting and for processing research data in support of European science.

TA2.2 Delivering value to stakeholders and building a thriving community

Within the **marine knowledge value chain**, the **Blue-Cloud** project will contribute to bridge and connect different communities, bringing value to **six stakeholder groups** that are central to its implementation, evolution and sustainability, namely:

- The **marine & maritime research community**, which produces and analyses data to create knowledge of the Ocean.
- The "**blue economy**" **entrepreneurs**, **SMEs** and **industry** that make use of the Ocean and of available knowledge to deliver products and services that satisfy society's needs.

[32] https://ec.europa.eu/digital-single-market/en/news/commission-proposes-measures-boost-data-sharing-and-support-european-data-spaces

- The **Policy, decision-making** and **governance institutions** that provide the legal and administrative frameworks to manage and preserve the Ocean.
- The **marine data, RIs** and **e-infrastructures** that are in dialogue with the Blue-Cloud but also other existing ones that are contributing towards the abundance of FAIR and open data in the marine domain and towards providing computing resources and digital services.
- The **European Open Science Cloud**, which the Blue-Cloud seeks to connect to, to enable trans-disciplinary and transformative research and innovation.
- The **ICT sector** that is driving new breakthroughs in the use of artificial intelligence, big data and machine learning.

These key groups are central to the **Blue-Cloud** as they represent: the Blue-Cloud's key "**target users**" (marine & maritime researchers; "blue economy" entrepreneurs, SMEs & industry; Ocean monitoring, management and governance institutions); the Blue-Cloud "**strategic allies**" (partners along the marine knowledge value chain, including blue data infrastructures and e-infrastructures; AI, ML & Big Data and data service providers; EOSC); and the Blue-Cloud "**enablers**" (policy makers and international funding institutions). The Blue-Cloud seeks to deliver **added-value** to all these communities, aligning a **user-centric approach** towards the creation of such added value:

- **Target users:** Enhancing their capabilities to perform **Open Science** in the marine domain, leading to improved knowledge of the Ocean and of our collective ability to sustain its productivity and good health through science-based management, innovation and policies.
- **Strategic allies:** Providing a strong case and a framework for **closer, long-term collaboration** along the marine knowledge value chain by enhancing value to shared user communities.
- **Enablers:** Boosting the **productivity** of **European infrastructures** and **assets** in the marine knowledge domain by building synergies around their value to users, in support of the EU Green Deal and of European stewardship in the global Ocean governance arena.

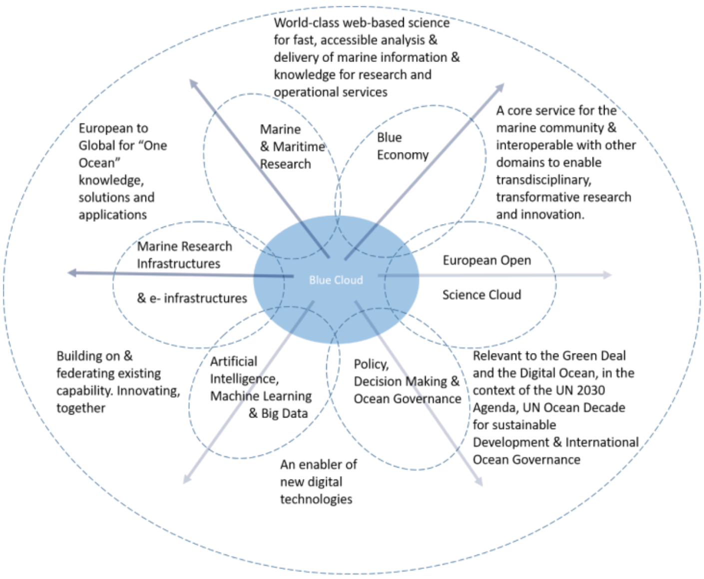

Figure 11: The Blue-Cloud Stakeholder Communities. Credit: Seascape Belgium

In Figure 11, the arrows show the direction of expansion of the Blue-Cloud community, from a core group of stakeholders in the centre, to engagement of wider stakeholder groups and sectors spanning the full spectrum of the marine and maritime domain, and wider sectors in new digital technologies. Whilst the focus is on Europe, Blue-Cloud is set in a global context and sets out to deliver a world-class capability for web-based Open Science.

The B-C Project has established different **mechanisms** to invite these targeted **stakeholder communities** to engage with its efforts. These mechanisms span from basic **communication activities**, including web-based news items, newsletters, webinars and wide social media outreach, to **targeted dialogue** and **consultation**, to **motivation mechanisms** articulated around the development of **synergies** with other initiatives. From a **communication perspective**, the Blue-Cloud project's website[33] plays a central role in communicating with **stakeholders** but also with **users**, being the entry point to the **B-C web-based Open Science environment (**or **Virtual Research Environment**). It also provides access to key assets, such as the **B-C Catalogue** and different **user-oriented training materials**. It showcases progress and results achieved by the **B-C demonstrators** (introduced in **Section 4**), including **webinars** and **interviews** with their teams and with representatives of the blue data infrastructures teaming up with their efforts. At the time of writing this roadmap, the website has reached well over **11,000 visitors**.

From a **motivational perspective**, as the B-C seeks to evolve towards shaping up as a **thematic service** for **EOSC**, interacting with peer projects that could potentially benefit from its services (e.g., FNS Food Cloud, amongst many others) and teaming up with other efforts (e.g., EOSC-FUTURE project, where science clusters such as ENVRI-FAIR, EOSC-Life, ESCAPE, PANOSC or SSHOC will connect and interact with EOSC core services) is key to better understand how the B-C services should evolve to address their needs and expectations, and build on complementarities, generating **synergies**. An overview of synergies established up to date is available here: https://blue-cloud.org/synergies.

TA2.3 Leveraging on innovation to enable Open Science: B-C's core services

Blue-Cloud's **underpinning technology** and **innovative concept** (see Figure 10) has been designed to deliver **core services** which are deployed through smart federation of leading marine data and e-infrastructures. In the **short-term**, the project will deliver the following **services**, which are considered its **core assets**:

- A **Data Discovery & Access Service (B-C DD&AS)**: Offering quick **discovery** and **access** to a range of **marine data** by federating and increasing interoperability between key European data management infrastructures (EMODnet, CMEMs, SeaDataNet, EurOBIS, Euro-Argo and

[33] https://blue-cloud.org/

Argo GDAC, ELIXIR-ENA, EuroBioImaging, EcoTaxa, and ICOS-Marine) to facilitate users in finding and retrieving multi-disciplinary datasets from multiple repositories.

- A **Virtual Research Environment (VRE):** Providing a **cloud-based, analytical** and **publishing framework** as a federation of computing platforms and analytical services for constructing, hosting and operating **virtual laboratories ("labs")** for specific applications.
- **Blue-Cloud Virtual Labs**: Configured with specific analytical workflows to serve as **demonstrators** showcasing the potential of open science in the marine domain, demonstrating how providing researchers with **web-based**, **analytical tools** and **greater cloud computing power**, underpinned by **open access to data** available across different infrastructures, can lead to a more efficient development of **innovative data products** that contribute to **improve knowledge** and **understanding** of the Ocean. While virtual labs can be adopted and adapted for inputs and analyses to advance knowledge and innovation across different marine disciplines, B-C demonstrators currently deal with analytical workflows for:
    - Zoo- and Phytoplankton EOV products.
    - Plankton Genomics.
    - Marine Environmental Indicators.
    - Fisheries data analytics.
    - Aquaculture Monitor.

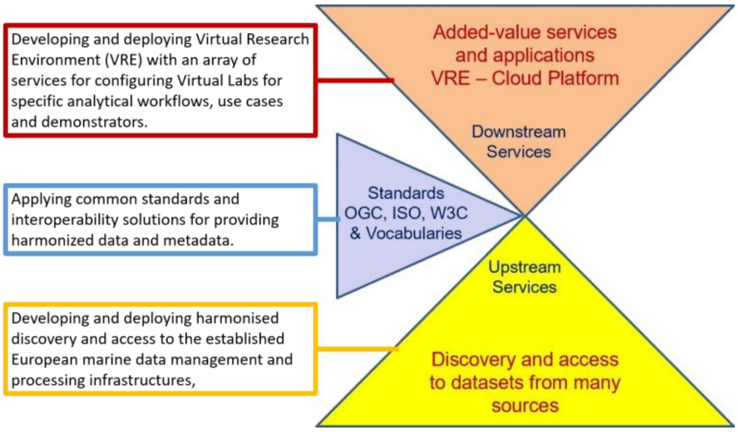

Figure 12: Leading concept and services for the Blue-Cloud development

In this section we look deeper into Blue-Cloud's **core services**. The B-C Virtual Labs and their **domain specific services** are further described in Section 4, in the framework of the Blue-Cloud demonstrators.

<u>**Blue-Cloud Data Discovery & Access Service**</u>

Blue-Cloud's **Data Discovery & Access Service** delivers a one-stop shop towards leading European **blue data infrastructures** for exploring and retrieving their joint offer of data sets and data products. An overarching **discovery interface** and **shopping mechanism** is being structured to allow the interfacing with each of the participating **blue data infrastructures**. This will allow users and machines to find and retrieve data sets from an impressive and highly diversified array of key infrastructures dealing with **physics**, **biology**, **biodiversity**, **chemistry**, and **bio genomics** data and data products. The **discovery** functionality is implemented with a two-step approach:

- The first step is a **user query** on a few **universal criteria** and at a collection level. For this purpose, use is made of the existing **GEODAB metadata brokerage solution**, which supports dynamic harvesting of a set of metadata from each blue data infrastructure. It allows transforming and ingesting these into a harmonised **Blue-Cloud metadata catalogue** at collection level, which makes it easier for users to identify which of the blue data infrastructures might have interesting data or data products and in how many collections.
- The second step is a **user query** which allows to **formulate additional search criteria** and to interact with each blue data infrastructure, whereby the search profiles are driven by the characteristics of each of the infrastructures and their specific contents and services. In this second step there is no central caching, but direct use of **machine-to-machine services** to support **dynamic queries** per infrastructure at data granule level, thereby identifying **individual data sets**. The user can include selected data sets from multiple infrastructures in a **shopping basket** which is managed through Blue-Cloud's **data broker service**. This component uses the shopping list and centrally stored business knowledge about each blue data infrastructure to build a list of data requests, which are then deployed to each of the relevant infrastructures to **retrieve** the associated data sets. These are temporarily stored in a **central cloud cache**, from which users can **download** them. All data and data products are openly delivered, so the shopping process does not require further negotiation with data providers, but only depends on processing time by the Blue-Cloud broker and each of the blue data infrastructures. To support this process, Blue-Cloud service has to interact with web services and/or APIs of each of the infrastructures, which in practice requires dealing with a range of **different service protocols** and with different **metadata** and **data formats**.

The core added value of **B-C Data Discovery & Access Service** to the user is overcoming not only differences in source metadata profiles across different blue data infrastructures, but also in machine-to-machine protocols for queries, aggregation levels of local data resources, delivery mechanisms, and local configurations. It makes it easier for users to **interact with multiple infrastructures** without first having to study and learn to operate with each of them. Moreover, it provides an innovative **B-C shopping ledger** for users and providers to keep track of their transactions, their shopping profiles, and allowing for monitoring and analysis of key indicators regarding users' **shopping behaviour**.

Blue-Cloud Virtual Research Environment

The initial **Blue-Cloud Virtual Research Environment (B-C VRE)** is based upon **D4Science**, an e-infrastructure allowing the Blue-Cloud community to make optimal use of major experience gained in earlier projects undertaken towards developing and operating a generic VRE with many core services, including building and running multiple **virtual labs**, each dedicated to specific research targets. D4Science provides proven solutions for connecting to external computing platforms and means for orchestrating distributed services, which in practice proves to be instrumental for further developing the initial B-C VRE and its smart federation with the **EUDAT**, **DIAS**, **EGI**, and **EOSC e-infrastructures**.

B-C VRE facilitates collaborative research using a variety of **data sets** and **analytical tools**, complemented by generic services such as **sub-setting**, **pre-processing**, **harmonizing**, **publishing** and **visualization**. In the framework of the Blue-Cloud project, a number of demonstrators are being developed (as described in **Section 4**). Each demonstrator is enacting a family of **analytical workflows** (or pipelines) that consist of a series of applications and make use of selected datasets as input. Results of analytical workflows can be documented with **provenance information** for **reproducibility**, provided with a **Digital Object Identifier (DOI)** for **citation**, **published**, and **visualized**. For each demonstrator, a **Virtual Lab** has been deployed within B-C VRE, with their researchers in group accounts. Multi-disciplinary datasets can be retrieved by means of **B-C Data Discovery & Access Service** and loaded into the **VRE Data Pool**. Also, direct APIs are made possible, such as the one used for DIAS WEkEO service, facilitating regular and automatic provision of new data sets for specific Virtual Labs, following pre-set requirements. Moreover, users can introduce their own data sets, and ingest data sets retrieved and ingested from other major data portals and resources, like **NOAA World Ocean Database**[34], **FAO** databases, and many others. Analytical services for stocking the overall VRE toolbox and equipping virtual labs and their analytical workflows are already available in B-C VRE, are partly added by the developers of the demonstrators, and are running on associate computing platforms, which are federated to become part of B-C VRE. For instance, the **WEkEO DIAS infrastructure** provides access to the **Sentinel** satellite images and also has several tools to analyse and process these, on the same computing platform, close to the images. Results of such processing can be ported into B-C virtual labs.

B-C VRE has a **common dashboard** shared by all virtual labs supporting collaborative research. The dashboard includes **common facilities** for accessing a **shared workspace**, a **social networking** area, a **data analytics platform** and a **publishing platform**. These facilities are all correlated to each other and build a system where (i) datasets can seamlessly flow across the various components to be easily shared among users and openly discussed by social networking practices; (ii) generated datasets and products are automatically enriched and enhanced with metadata capturing their entire lifecycle, their versions, and the detailed list of authors and tasks performed leading to the development status and shapes. The dashboard allows also to configure new workflows as combinations of services and for future expanding of B-C VRE with more connected computing platforms and more analytical and generic services. One important functionality is providing support for a federation of identities between different AAI systems that the engaged computing platforms might operate. Interoperability is arranged with social media login (OpenID, OAuth 2.0, and SAML), the authoritative eduGAIN network, and others. **OGC standards**[35] for interacting with the services and the datasets and resulting data products are used as much as possible.

All B-C VRE provided facilities are exploitable as web applications accessible either through the Blue-Cloud VRE **gateway menu** for common users or through the **APIs** for developers. The data analytics platform offers several types of facilities to users of B-C VRE:

- **Jupyter notebooks for expert users:** This allows users to execute code fragments leading to the generation of the results. The notebooks are very suitable as they combine code fragments, text cells, and figures or animations that illustrate different steps in the process. Also, notebooks can be exported to HTML and PDF and are easy to share, improving reproducibility and peer-reviews. Jupyter notebooks are designed for a single user, while a multiple-user instance has been established using Jupyterhub.

[34] https://www.ncei.noaa.gov/products/world-ocean-database
[35] https://www.ogc.org/docs/is

- **RStudio**: Offered to deliver a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management;
- **A software importer tool:** For analytical methods implemented in **Java**, **Python**, **R**, **C** and other languages. A high throughput computing engine is also available for expert users wishing to execute their software code in a distributed infrastructure. Any imported software becomes accessible via a standard Web Processing Interface (WPS). Any execution automatically generates provenance information making results sharable, re-usable and reproducible.

The **B-C VRE** provides an **innovative computing platform** capable to hide the complexity of the infrastructure, while enabling easy integration and deployment of analytical algorithms and software tools. Additional services can be deployed as Docker containers, without configuration efforts. It also delivers an innovative approach for publishing results respecting ownership, provenance, and controlled access, while supporting collaborative research from the formulation of a new analytical approach to its experimentation, validation and delivery. In terms of **Technology Readiness Levels (TRLs)**, **B-C VRE common facilities** can be qualified as **TRL8** (*system complete and qualified*) with the ambition to achieve at least **TRL7** level for all the **virtual labs** operated by the B-C VRE (as described in Section 4) within the lifetime of the project.

So far, the B-C services have been operating on a "testing" mode, restricted to the B-C Project's circle of "early users". Over 300 users have accessed these services. In the **second semester** of **2021**, B-C Demonstrators will start opening their **virtual labs** to the public, inviting other **scientists** and **researchers** to **test** and **experiment** with their many functionalities, enriching them with new data, suggesting improvements of their algorithms, developing new data products, models, hypotheses and/or results, including potential new, innovative applications.

TA2.4 Aligning with wider developments

The Blue-Cloud services are being deployed with attention to other, wider developments that in the medium term will positively impact Europe's **marine knowledge value chain**, such as the upcoming **EU pilot Digital Twin of the Ocean** referred to in **Section 2**; the **EuroHPC**[36] joint undertaking, which will develop a world class, supercomputing ecosystem in Europe; or the Gaia-X project[37], building the requirements for an EU data infrastructure. Through this roadmap, the B-C project will reflect on ways in which its **key exploitable results** could be **further capitalized** by and **benefit** from these initiatives.

TA2.5 Blue-Cloud as a model thematic service for EOSC

> *"Europe is going to co-create a framework to allow the use of data. It should consist of a trusted pool of non-personalized data that governments, businesses and other stakeholders can contribute to. And thereby, we open up data as a resource for innovation and bring new solutions to the market. And our scientists are already beginning to do this. We are creating a European Open Science Cloud (EOSC) now. It is a trusted space for researchers to store their data and to access data from researchers from other disciplines. We will create a pool of information leading to a web of research insight." Ursula von der Leyen, President of the European Commission*

[36] https://eurohpc-ju.europa.eu/
[37] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

During the annual meeting of the **World Economic Forum** held in 2020, the President of the European Commission pointed out the value of **EOSC** and the absolute importance of data as a key resource for innovation. The B-C project is built on this notion, working to bring the following elements to EOSC:

- The experience of a **thematic EOSC** that can serve as a role model for the development of other thematic clouds within EOSC. The cyber-platform built by the B-C project will enable collaborative research in support of the **EU Green Deal** and **UN Agenda 2030**.
- A set of **domain-specific services** deployed through B-C pilot demonstrators which address specific research and innovation needs for **Oceans**, **seas** and **freshwater bodies**. These services will support marine ecosystem research, conservation and forecasting, as well as enabling innovation in the Blue Economy. The **Blue-Cloud Catalogue** -Blue-Cloud's thematic catalogue of services made available through its Virtual Research Environment- is being designed and developed for interoperability with **EOSC**, with the purpose of making B-C's assets and services findable and accessible via its **Portal Catalogue & Marketplace**[38], creating opportunities for new applications and workflows. The included service metadata will include APIs to access specific services with thematic functionalities. The **B-C Catalogue** is one of the ways in which the B-C project will be contributing to cluster **Open Science marine thematic services** through EOSC by 2022, for easy discovery, access, interoperability and re-use by a much wider community. In the medium to long term, EOSC will be instrumental in deploying B-C's strategy of promoting wider user uptake of Open Science, supporting the transformative science and innovation needed to deliver the **EU Green Deal** and **UN Agenda 2030.**
- A mechanism to **easily access** and **discover blue data**. The blue data infrastructures federated in B-C Data Discovery & Access Service manage important volumes of data, which are made easily discoverable and accessible. The metadata on the B-C data resources will be shared and included in EOSC metadata catalogue service with deep links to Blue-Cloud Data Discovery & Access service to facilitate the actual data retrieval by EOSC user communities.
- A proven methodology showing how researchers interacting with e-infrastructure developers can establish a cyber platform with tools and services, which support many scientific challenges and are fit-for-purpose. Meanwhile building a thriving community and synergies between IT developers and blue domain researchers, which can provide a learning curve for the further deployment and exploitation of EOSC into a successful initiative.
- Bringing in the Blue-Cloud community towards exploring the added-value and opportunities of EOSC and taking an active role in making EOSC more attractive and dynamic.

[38] https://marketplace.eosc-portal.eu

TA3 Demonstrating value to society: B-C Demonstrators

The Blue-Cloud's "**demonstrators**" are case studies showcasing how scientists and researchers are using the technology and services offered by **Blue-Cloud**'s **web based, Open Science environment** described in the previous section. They are also practical examples showcasing how **Open Science** can enhance the **marine knowledge value chain** in support of greater societal objectives. They all have one thing in common: they demonstrate how providing researchers with **web-based**, **analytical tools** and **greater cloud computing power** -allowing them to perform heavy computational analyses otherwise not possible or too timely to perform on local computers-, underpinned by **open access to data** available across different European blue data infrastructures, can lead to a more effective and efficient development of **innovative data products** that contribute to **improve knowledge** and **understanding** of the Ocean, and thus the overall performance of Europe's marine knowledge ecosystem. Below we introduce their scientific objectives and how they are leveraging on Blue-Cloud's core services to deliver societal impact.

TA3.1 Blue-Cloud's demonstrators: Showcasing uses in the marine domain

<u>**Demonstrator 1: "Zoo- and Phytoplankton Essential Ocean Variables (EOVs)"**</u>

The **Zoo- and Phytoplankton Essential Ocean Variables (EOVs)** demonstrator is working on the development of **innovative data products** that help to estimate the **abundance** and **concentration** of **plankton** (which is an Essential Ocean Variable - EOV). Using the Blue-Cloud Open Science environment and feeding from different data sources, it has developed an **online service** (**Virtual Lab**) that provides users with built-in, **open, transparent methodologies** to estimate **plankton abundance**, **distribution** and **dynamics**, based on **big data analysis** and **machine learning** (e.g. neural networks).

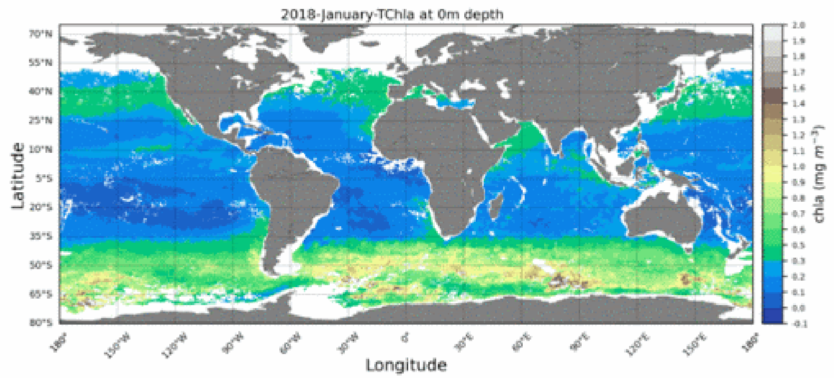

Figure 13: Main output of the model to quantify the relative contributions of **phytoplankton abundances** in one station. The model allows the integration of different EOV variables, not only to display data-driven trends, but also to understand interactions in a mechanistic way.

Any interested **user** is able to join this Virtual Lab and take up these available **EOV methodologies** to easily re-run analysis with different datasets, as well as sharing and comparing results. Users of the **Zoo-phytoplankton Essential Ocean Variables (EOV)** demonstrator will be able to capitalize on the offered reusable methodologies and tools to:

- Generate **maps of zooplankton abundances**.

- Produce **global ocean 3D key phytoplankton products of chlorophyll-a (Chla) concentration**, as a proxy for total **phytoplankton biomass** and phytoplankton functional types, as a proxy for **phytoplankton diversity**.
- Quantify the relative contributions of the top-down and bottom-up drivers in **phytoplankton dynamics**, applying a mechanistic model using near real-time data.

The resulting **EOV products** contribute to **improve knowledge** and vastly **reduce uncertainty** on the present state of **marine plankton ecosystems** and their response to ongoing and future **climate change**, providing valuable information for the **modelling**, **assessment** and **management** of marine ecosystems. **Plankton** is considered the **cornerstone** of marine ecosystems and of the food supply throughout the Ocean, as well as being essential in the global carbon and oxygen cycles producing an estimated 80% of the Planet's oxygen. This resulting knowledge can contribute to inform the **EU Biodiversity Strategy for 2030** and **"Farm To Fork" Strategy** in support of the **EU Green Deal**, as well as the **UN Agenda 2030**'s **Sustainable Development Goals 2** (Zero Hunger) and **14** (Life Under Water).

Virtual Lab: Zoo- and Phytoplankton Essential Ocean Variables (EOVs)

Lead Partner: VLIZ · **Scientific Partners:** VLIZ, Sorbonne University, CNRS-LOV, University of Liege

Potential users and applications	Policy Impact
• **Fisheries advisory organisations:** To study the availability of food resources for fish stocks and assess the effects on fish stocks. • **Marine policy officers:** To address threats such as food insecurity, foreseen under the EU Biodiversity Strategy for 2030. • **Researchers** (e.g. from environmental agencies): To contribute to the understanding of environmental conditions and factors at new scales of observations.	Knowledge generated through this Open Science ecosystem can help marine policy officers to **address threats** related to **food security**, as foreseen under the **EU Biodiversity Strategy for 2030** and the **"Farm to Fork" Strategy** underpinning the **EU Green Deal**, but also the **UN SDGs 2** (Zero Hunger) and **14** (Life Under Water). Moreover, the proposed EOV products are of interest for fundamental research, contributing to modelling of marine ecosystems. These models will be available to feed the future EU pilot **Digital Twin of the Ocean**, which will unlock new, innovative applications of these Open Science outputs.

Users of this **Virtual Lab** will be exploiting and benefiting from open access to the following data sources:

- **Argo GDAC:** Salinity, oxygen, chlorophyll data
- **CMEMS:** Satellite-derived reflectance, sea level anomaly, PAR, physical data
- **EurOBIS:** Zooplankton abundances
- **EMODnet Biology:** Zooplankton (from MBA CPR survey) & phytoplankton abundance
- **GEBCO:** Bathymetry
- **GlobColour:** Satellite derived Photosynthetically Available Radiation
- **GSFC, NASA:** Distance to nearest coastline
- **LifeWatch:** Nutrients, PAR and temperature
- **SeaDataCloud**: Temperature and salinity
- **World Ocean Atlas 2018:** Nitrate, Silicate and Phosphate

Demonstrator 2: "Plankton Genomics"

Using the Blue-Cloud Open Science environment -underpinned by **access** to **biomolecular, imaging** and **environmental data** available across different European blue data infrastructures and cloud computing- the "**Plankton Genomics"** demonstrator has developed an **online service** (**Virtual Lab**) that provides users with **analytical methodologies** and **tools** ("**workflows**") to extract new **"plankton" data products** from existing data, in particular from genomic samples collected by the "Tara Oceans" project[39]. These analytical tools are relevant to study **plankton diversity** at basin and global scales when observations are scarce, which is of particular interest for researchers in the fields of **plankton biogeography**, **marine biogeochemistry**, and **ecosystem health**.

Users of the **Plankton Genomics** demonstrator will be able to capitalize this **analytical framework** to:

- Generate georeferenced tables of **occurrence of known** and **unknown plankton taxa & genes**, together with environmental variables, allowing **species** and **functions discovery** from the **genomic dark matter** of Tara Oceans data.
- Produce **global maps** of potential **distribution** of the selected **plankton taxa** and **genes**, through machine-learning based regression on environmental variables.

Scientific researchers, in particular taxonomists, computational ecologists and bioinformaticians in quest of the **identification** of **unknown sequences** in the oceanic environment will be able to 1) **obtain lists** of unknown taxonomic and functional units, 2) **correlate** these unknowns with **environmental variables**, 3) **project** the **results** of this correlation over the **world's ocean** and 4) **visualize** these bio-geographies as **maps**.

<table>
<tr><td colspan="2">Plankton Genomics
Lead Partner: EMBL · Scientific Partners: Sorbonne Université, CNRS, VLIZ</td></tr>
<tr><td>Potential users and applications
● Scientific researchers (in particular taxonomists, computational ecologists and bioinformaticians): To pursue the identification of unknown taxa and genes in the ocean, predict other locations where the sequences are likely to occur, and overall, to advance research on plankton biogeography, marine biogeochemistry and ecosystem health.</td><td>Policy Impact
Knowledge generated through this Open Science ecosystem can contribute to advance scientific progress in the fields of ecosystem health, informing progress towards the EU Biodiversity Strategy 2030. It can also contribute to advance sequencing of the DNA of our ocean and waters by learning more about those sequences, which is one of the objectives of Mission Starfish and offers unprecedented opportunities to discover and use new molecules and biotechnologies.</td></tr>
<tr><td colspan="2">Users of this Virtual Lab will be exploiting and benefiting from open access to the following data sources:
● ELIXIR-ENA: Genomics data, in particular that of Tara Oceans and Arctic (global ocean, plankton)
● CMEMS: Ocean colour, altimetry, temperature and salinity field data, ocean and climate variables
● EcoTaxa: Quantitative plankton images
● Tentatively: Data from other (external) sources, such as NOAA's World Ocean Atlas (climatologies)</td></tr>
</table>

The resulting new data products can contribute to generate **new hypotheses** and **guide research** in the fields of **plankton biogeography**, **marine biogeochemistry** and **ecosystem health**, leading to new, deeper assessments of **plankton distributions**, **dynamics** and **fine-grained diversity** (to a molecular resolution). The fine scale genomic data used gives unprecedented ability to **detect species** and unknown, but coherent, **genetic entities**, which is particularly relevant in scenarios where collection of new samples is difficult or impossible. This resulting knowledge can contribute to inform the **EU Biodiversity Strategy for 2030** in support of the **EU Green Deal**, as well as the **UN Agenda 2030**'s **Sustainable Development Goal 14** (Life Under Water). More importantly, sequencing and learning about the **DNA** of our ocean and waters, as a **common public good**, is one of the objectives of **Mission**

[39] https://oceans.taraexpeditions.org/en/m/about-tara/les-expeditions/tara-oceans/

Starfish and will offer immense opportunities to discover and use new molecules and biotechnologies for the **wellbeing** and **health** of **European citizens**.

Demonstrator 3: "Marine Environmental Indicators"

The "**Marine Environmental Indicator"** demonstrator is showcasing how using **analytical tools** and greater computing power underpinned by access to data available across different European blue data infrastructures can ease the way towards developing dynamic **Marine Environmental Indicators** (MEI) that inform the **environmental quality** of the **Ocean**. As a scoping effort focusing on the **Mediterranean Sea**, the demonstrator is using the Blue-Cloud Open Science environment to set up an online service (**Virtual Lab**) allowing users to:

- Calculate and share **marine environmental indicators** feeding from data sets available across different sources.
- Obtain **new, added-value data** applying **big data analysis** and **machine learning methods** on existing data sets.
- Perform online and "on the fly" operations, such as **selecting** a portion of data for a specific area and period of time, **performing analytics** with several methodologies on the selected variables or **displaying** the available indicators on **tables**, **maps** and other **graphic visualizations**.

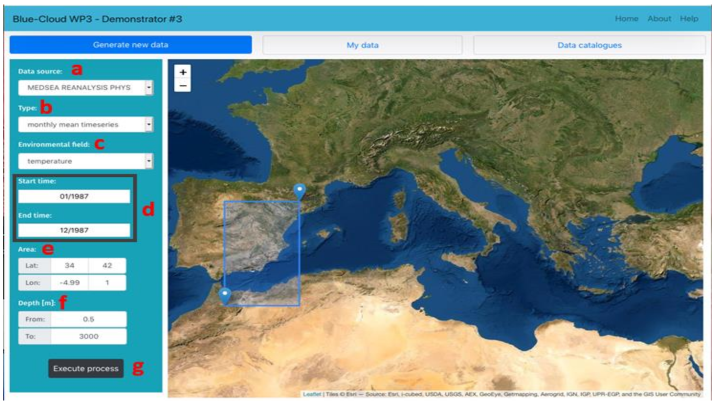

Figure 14: Demonstrator offers a user-friendly interface for execution of WPS methods allowing section of: Output type [b] [c] Lon/lat area [e] & depth layer [f]

To enable these capabilities, the Virtual Lab offers:

- **Prototype Marine Environmental Indicators (MEI) applications**, which allows users to display available data to generate new added- value data, and to apply methods to deliver new **marine environmental indicators**.
- **Prototype of methods**, adopting a big data approach to extracting information from data sources and generating new marine environmental indicators.
- **Data processing workflows** delivered in notebooks for specific research on Ocean patterns and Ocean regimes, delivering **marine physical indicators** based on machine learning. Unsupervised classification algorithms are applied on vertical profiles ("patterns") or time series ("regimes").
- A **Storm Severity Index (SSI)** providing insight into changes of the storm climate of a sea region and into the impact of such changes on sea circulation patterns, delivering relevant **marine meteorological indicators**.

- A prototype for recalculating missing sub-variables of the "Essential Ocean Variable" **Inorganic Carbon** and (potentially) providing calculation of uncertainties, to produce fit-for-purpose data products catering different user segments.

<table>
<tr><td colspan="2">Marine Environmental Indicators
Lead Partner: CMCC · Scientific Partners: IFREMER, MOi, KNMI, University of Bergen</td></tr>
<tr><td>Potential users and applications</td><td>Policy Impact</td></tr>
<tr><td>● Governmental & Environmental Protection Agencies: To perform statistical analyses of the quality and characteristics of the marine environment (Mediterranean Sea) towards MSFD compliance. Potential to scale up in the next version to the global Ocean.
● IOC UNESCO: To measure progress towards SDG 14.3 (Ocean acidification).
● Researchers: To facilitate the discovery of new climatic indicators based on machine learning, and a simplified way to analyse oceanographic data.</td><td>Knowledge generated through this Open Science ecosystem can contribute to measure progress towards the objectives set forth by the EU MSFD, as well as to inform progress towards SDG target 14.3 on Ocean acidification. In the future, it could be instrumental to deliver a more dynamic, timely understanding of the impact of climate change mitigation measures, potentially contributing towards more effective carbon pricing throughout the economy. With more adequate interfaces, it could evolve as an effective educational and informational tool to the general public.</td></tr>
<tr><td colspan="2">Users of this Virtual Lab will be exploiting and benefiting from open access to the following data sources:
● CMEMS & C3S (DIAS-WEkEO): Ocean and climate model data (subset of MEDSEA_REANALYSIS_PHYS_006_004 and GLOBAL_REANALYSIS_PHY_001_030 CMEMS, ERA5 data) and observations
● ICOS-Marine and ENVRI-FAIR: Inorganic carbon data.
Throughout the project, direct access to the following data sources might also be contemplated:
● EMODnet: Physics, biology and chemistry data.
● Euro-Argo and Argo GDAC: Salinity, oxygen, chlorophyll data.
● SeaDataNet: Biogeochemistry, physics, biology, environmental data</td></tr>
</table>

There are different ways in which this demonstrator can contribute to **enhance marine knowledge**, in support of the EU Green Deal and the UN Sustainable Development Goals (SDGs):

- Sustaining **Good Environmental Status** of EU marine waters by progressing towards the targets set out in the **European Marine Strategy Framework Directive** (MSFD) will remain central to EU action on Ocean matters. As anticipated in Section 2, the **MEI Virtual Lab** can provide practical ways for Open Science to ease the way towards **monitoring progress** towards policy objectives, providing a more accurate picture of the status of the Ocean, including the visualization of long-term series to better assess **environmental trends**. Other specific applications might include, for example, better insight into the impact of toxic algae blooms in the environmental quality and productivity of marine ecosystems.

- As also discussed in Section 2, the EC has also set out a clear ambition to achieve **climate neutrality** by **2050**. To deliver on this objective, the EC is intending to review all relevant climate-related policy instruments, comprising the Emissions Trading System; Member State targets to reduce emissions in sectors outside the Emissions Trading System; and the regulation on land use, land use change and forestry. The role of oceans in mitigating and adapting to climate change is increasingly recognised, as lasting solutions to climate change will require greater attention to **nature-based solutions**, including **healthy** and **resilient seas** and **oceans**. Although marine ecosystems are yet hardly, formally considered within the scope of climate-related policy instruments at international nor European level, their role in carbon sequestration is increasingly recognized[40]. Whereas their formal recognition in governance mechanisms might remain a long-term aspiration, MEIs can already start delivering a more dynamic, timely understanding of the **impact** of **climate change mitigation measures**, which currently take years to assess. In the future, the information derived could potentially contribute towards more effective **carbon pricing** throughout the economy, which is one of the key objectives of upcoming climate change related European policy reforms. Although the mechanisms to bring this information into governance mechanisms and pricing systems need to be further matured, the value of MEI's for climate change policy is already starting to be explored and progress can be catalysed through the **MEIs Virtual Lab**, which could guide and provide feedback to EU national agencies and ministries on their national commitments to reduce GHG emissions. The virtual lab could also inform **UNESCO**'s **Intergovernmental Oceanographic Commission (IOC-UNESCO)**, as the custodian agency for the **SDG target 14.3** on **Ocean acidification**.

- As it is the case with demonstrators 4 and 5, in the long term, this **Virtual Lab** has the added value potential of opening a window of opportunity to be used as a **powerful educational** and **informational tool** to the **general public**, for a better-informed civil society, but also so that citizens become more aware of the connections between our use of the Ocean and its impact on marine ecosystems and in turn, on our own health. Although the Virtual Lab would have to develop new interfaces to be fit for this purpose, this fact brings yet another potential, future contribution of **Open Science** to the marine knowledge value chain closer to its realization.

Demonstrator 4 "Fish, A Matter of Scales"

The "**Fish, A Matter of Scales**" demonstrator is working on **innovative ways** of **delivering knowledge on fisheries** to **improve fisheries management** and **monitor trends** in **fisheries indicators**. It aims to implement a fully FAIR data management approach across findable and accessible fisheries data and, in combination with interoperable (meta)data across different blue data infrastructures, deliver reproducible **data analytics** through a variety of products.

Using the Blue-Cloud Open Science environment, and building on previous efforts, it has set up two different **Virtual Labs**. The first one allows users to search for reliable and fact-checked fisheries content, with features ranging from **global fisheries maps**, **catch statistics** and **overviews**, high resolution bathymetry and **aggregate maps** of **major world fisheries**. The second one provides users with a suite of **analytical services**, including indicators, interactive maps, models and methods (currently under development) based on widely published software to serve a community of **fisheries data analysts**, as well as expanded information from other sources to bring context and perspective to the fisheries maps and for **approved status assessments of fisheries**:

- **Global Tuna Fisheries Atlas Virtual Lab:** Updated for 2019, this **virtual lab** provides a complete overview of **FAO statistics** (global tuna and billfish fisheries), showcasing a **workflow** that facilitates the visualization of these statistics in combination with other maps (e.g. **CMEMS** and **EMODNet** layers) in unprecedented detail and precision. Through a **user interface**, it

[40] Climate action requires new accounting guidance and governance frameworks to manage carbon in shelf seas (T. Luisetti et al, 2020) https://www.nature.com/articles/s41467-020-18242-w

offers global thematic maps showing the **Earth's fisheries**, their **production** and **trade**, as well as **fish distributions maps** and **ecological zones**; added maps created through the full benefit of CMEMS satellite products based on Copernicus data; and measurements and other accurate analytical techniques to place and analyse fisheries in a wider environmental context.

- **Global Record of Stocks & Fisheries Virtual Lab:** This **virtual lab** provides a reference repository of fish stocks and fisheries combining harmonized data from three global data providers, offering different working environments to **validate** and **harmonize new data sources** (not public) for **approved status assessments of fisheries** and to publish data (making it public). It uses **semantic technology** for data harmonization.

<table>
<tr><th colspan="2">Fish A Matter of Scales
Lead Partner: FAO · Scientific Partners: FAO, FORTH, IRD</th></tr>
<tr><th>Potential users and applications</th><th>Policy Impact</th></tr>
<tr><td>● Regional fisheries data analysts: To help them assess how fisheries develop over time in their area of interest and forecast future evolution under different scenarios, informing fisheries managers for decision making processes. To train data analysts in the use of algorithms in support of SDG 14.4.1 analysis.
● Developers: To access a “boiler plate” solution for the management of fisheries time-series on catch and effort that brings collated statistical data into a data harmonization and quantitative analysis process.
● Other communities: To benefit from a FAIR compliant, data management solution spanning statistical and geospatial data workflows in the fisheries, aquaculture and related aquatic and land-based domains.</td><td>Knowledge generated through this Open Science ecosystem can inform and support decision making across different EU policies, including the Common Fisheries Policy, the EU “Farm-To-Fork” Strategy and the EU Biodiversity Strategy by enhancing policy monitoring capabilities. It can also inform fisheries management decisions at a regional level and support capacity building to monitor progress towards the UN Sustainable Development Goals (SDGs), more specifically SDG 2 (Zero Hunger), SDG13 (Climate Action) and SDG 14 (Life Under Water).
In the long-term, it could evolve as an educational and informative tool targeting the general public, opening an opportunity for citizens to gain insight into fish stocks and fisheries and the impact of their consumption choices, with easy access through web or QR codes.</td></tr>
<tr><td colspan="2">Users of this Virtual Lab will be exploiting and benefiting from open access to the following data sources:
● FAO: Global stocks and fisheries data; FAO global project data and fish tagging data
● Global Effort Maps of Fisheries
● FNS Cloud: Data on fish composition
● Any ISO/OGC dataset that is FAIR compliant, including CMEMS and EMODnet</td></tr>
</table>

From a technical perspective, this demonstrator shows how **FAIR data management** practices allows to **harmonize** and **combine** data, which can be made available for generic analytical services. From a policy perspective, this demonstrator shows how the availability of **user-friendly analytical frameworks** can support the marine knowledge value chain, servicing **intermediate users** (such as fisheries data analysts) but also informing **policy makers**, with a potential to evolve towards delivering useful resources that bring the issues at stake closer to the **general public**. Again, its contribution to delivering on the objectives of the EU Green Deal is worth analysing not only from a short-term perspective, but also with a view to the potential evolution of the services and added value delivered by this demonstrator:

- With the launch of the **EU Biodiversity Strategy**, all EU policies will be expected to contribute to preserving and restoring Europe’s **natural capital**. The **EU “Farm to Fork” Strategy** will continue to work under the common fisheries policy to reduce the **adverse impacts** that fishing can have on **ecosystems**, especially in sensitive areas. **Combining state-of-the-art data** on **fish stocks** and **fisheries** with other **marine environmental indicators** is essential to understand the relations between fisheries and marine ecosystems, informing policy making processes.

- The **"Farm to Fork" Strategy** will also strive to stimulate sustainable food consumption and promote affordable healthy food for all, making European food the global standard for sustainability. The EC will explore new ways to give consumers **better information**, including by digital means, on details such as where the food comes from, its nutritional value, and its **environmental footprint**. **Fish provenance** and **traceability** are relevant information that can contribute towards better assessing the environmental footprint of fisheries, as well as informing more sustainable consumption choices. In the short and medium term, this virtual lab can contribute to support **fisheries monitoring** to inform more sustainable **fisheries management** decisions. In the future, this virtual lab further opens a window of opportunity to evolve to be used as a powerful **educational and informational tool** to the general public, so that consumers become more aware and are able to connect the **impact** of their **consuming behaviour** on available fish stocks.
- In an international context, the "**Fish A Matter of Scales**" demonstrator provides **data analysts** from **Regional Fisheries Management Organizations** (RFMOs) with fisheries analytical models and tools to monitor how fisheries in their area of interest develop over time, also in relation to environmental variables and other ancillary data (for example, correlating fish stocks with environmental data to assess the impact of climate change on fish stocks behaviour). This will contribute to **inform** fisheries management decisions, but also support **capacity building** to monitor progress towards the **UN Sustainable Development Goals** (SDGs), and more specifically **SDG 2** (Zero Hunger), **SDG 13** (Climate Action) and **SDG 14** (Life Under Water).

Last but not least, the **Fish A Matter of Scales** demonstrator is currently exploring synergies with the "**Food Nutrition Security Cloud**"[41], with the potential to broaden the analytical capabilities developed to include nutrition considerations when addressing fisheries considerations.

Demonstrator 5 "Aquaculture Monitor"

The "**Aquaculture Monitor**" demonstrator showcases how **artificial intelligence**, underpinned by **open access** to **multi-source data**, can deliver **remote sensing capabilities** for **aquaculture cage detection**, **monitoring** and **classification** (coastal/pond; land-type), supporting **aquaculture planning** in coastal areas. Using the Blue-Cloud Open Science environment, the demonstrator has set up an **"Aquaculture Atlas Generation" Virtual Lab** delivering a range of products and capabilities to users:

- **Aquaculture Cage Atlas:** The Atlas offers an online overview of satellite data derived **maps of cages** and **cage clusters**, allowing users to visualize geospatial data and to access information at farm level (including feeding systems, farm material, farm design and fish species). Users can overlay farms with satellite images and perform different analysis, like comparing the footprint of coastal ponds and rice paddy fields and check their evolution over a period of 20

[41] https://www.fns-cloud.eu/

years. The products are delivered through a map viewer, allowing registered users to edit features of the detected cages and cage clusters and to automatically map across feature sets to enrich maps. Estimates of **cage activity** over a production season can be produced, provided there is a large enough data sample available. "Super users" can also edit farm attributes and validate them at farm level, publishing the results in a geo-network catalogue.

- **Aquaculture Ponds Atlas:** This test-service uses some of the same data sources and similar analytical processes as the Cage Atlas to deliver a **coastal land-use classification map** of **aquaculture ponds**, fully based on Copernicus data for its remote sensing component.

Aquaculture Monitor **Lead Partner:** FAO · **Scientific Partners:** FAO, CLS	
Potential users and applications	**Policy Impact**
• **Aquaculture data analysts and managers:** To monitor how aquaculture develops over time in their area of interest and inform management decision-making processes. • **Data managers in spatial advisory units for maritime spatial monitoring and planning:** To monitor and inform decision-making processes related to maritime spatial planning. • **System developers:** To access a "template" solution for the management of sentinel and other satellites data access and processing in WeKEO. The demonstrator example workflow on cage monitoring can be adapted to other data and analytical WeKEO processes. • **Other communities** (e.g. aquaculture industry and SMEs)**:** To benefit from a FAIR compliant data management solution spanning statistical and geospatial data workflow in the fisheries, aquaculture and related aquatic and land-based domains.	Knowledge generated through this **Open Science ecosystem** can contribute to support the **EU** "**Farm to Fork Strategy**" in its objective to develop the potential of **sustainable seafood** as a **source of low-carbon food**, but also as the source of new sources of **protein** that can relieve pressure on agricultural land. It can specifically support better **aquaculture monitoring**, better **impact assessments** of fish and plant farming on marine ecosystems and inform more **sustainable aquaculture management** decisions. It can likewise support monitoring progress towards the UN SDGs, more specifically **SDG 2** (Zero Hunger), **SDG 13** (Climate Action) and **SDG 14** (Life Under Water). In the future, it could be scaled up and improved with **more user-friendly interfaces** and with new application interfaces (APIs) to provide information to the **general public** on aquaculture locations, production and tracking throughout different web portals.
Users of this **Virtual Lab** will be exploiting and benefiting from open access to the following data sources: • **WekEO:** Sentinel 1 (S1) and Sentinel 2 (S2) • **CLS:** VHR images • **FIRMS**	

To support users in the deployment of these capabilities, the Virtual Lab currently provides use cases of fish farm detection for **Greece** and **Malta**; and coastal ponds and rice fields detection for **Sulawesi** and **Indonesia**. Advanced users will be supported in 2021 with facilities to run **additional analytical services** using Copernicus and other data, while FAO will provide a **mapping service** to connect to **local statistical datasets**. The long-term ambition of this effort is to deliver a full-fledged service providing overviews of **national aquaculture sectors**, using OGC compliant data services and available capabilities across different Virtual Labs available in the Blue-Cloud Open Science environment.

In a similar way to demonstrator 4, the **Aquaculture Monitor** demonstrator shows the potential of servicing **intermediate users** (such as fisheries data analysts) with a view to also informing **policy makers**, potentially evolving in the future towards engaging the **general public**, with wider policy implications:

- In the short- to medium-term, this demonstrator can contribute to support the "**Farm to Fork Strategy**", which is seeking to develop the potential of **sustainable seafood** as a **source of low-carbon food**, but also as the source of new sources of **protein** that can relieve pressure on agricultural land. **Aquaculture** has a crucial role in meeting the challenge of how to feed a growing population while reducing the environmental impacts of food production. It has already overtaken fisheries as the main source of fish for human consumption and accounted for 52% of global production in 2018. FAO forecasts that aquaculture's share of production will increase to 59% (109 million tonnes) by 2030. While fish is less resource-intensive to farm and has a lower carbon footprint than meat, the decline of wild-fish stocks at biologically sustainable levels will continue to drive demand for Ocean aquaculture upwards, requiring **closer monitoring** to keep its **impact** on **marine ecosystems** in **balance**. As the **Aquaculture Atlas Generation Virtual Lab** allows mixing in-situ data (farm inventories) with satellite data, it can support EU and global international monitoring efforts, by for example providing **early warning of diseases** (combining aquaculture data with CMEMS data on currents) or facilitating environmental impact assessments (combining aquaculture data with CMEMS ocean colour and seagrass maps). In short, combining current data on aquaculture with other marine environmental variables, indicators and other ancillary data such as site-inventories can contribute to support better **aquaculture monitoring**, to better **assess the impact** of **fish and plant farming** on **marine ecosystems** and **informing more sustainable aquaculture management decisions** and **relevant policy making processes**.

As with the "**Fish A Matter of Scale**" demonstrator, in the future, this virtual lab also opens a window of opportunity to be used as a sound **educational tool** to inform the general public, so that consumers become more aware and are able to connect the impact of their **food consuming behaviour** on **marine ecosystems**.

TA3.2 Lessons learned from the Blue-Cloud demonstrators so far

At the moment of writing this document, the Blue-Cloud demonstrators have been working on closed environments. However, their full potential will unleash as their **Virtual Labs** are open to the public, inviting users to use and test the different **analytical frameworks** they provide to **service** and **support Open Science** in the **marine domain**. Over the next year, the Blue-Cloud project will have the opportunity to gather feedback from all early users of B-C Open Science environment to finetune its infrastructure and its related services.

As early adopters of the Blue-Cloud's core services, the B-C demonstrators provide useful insight into some of the emerging, (direct and indirect) positive impacts of the project, but also the challenges still faced to fully cater to their needs. Some of these challenges will be resolved through the project, with the resources available. Others signal greater challenges beyond the scope of the project, which need to be properly identified and understood, in order to tackle them through future efforts.

*"Thanks to the Blue-Cloud project, our "Zoo-and Phyto Plankton EOVs" demonstrator has gained interest from **major biodiversity initiatives**. For instance, we are currently exploring how the demonstrator can contribute to the development of the post-2020 global biodiversity framework in the **Convention on Biological Diversity**". - Patricia Cabrera (VLIZ)*

From a positive perspective, the value of weaving a multi-stakeholder, Open Science community in the marine domain is showing clear benefits. Bringing the **multidisciplinary teams** engaged in **blue data infrastructures**, **e-infrastructures** and **scientific research together** is demonstrating value towards a much-needed dialogue and interaction that contributes to a better understanding of "user needs" and interoperability challenges, creating momentum for closer collaboration. But the value of the community extends beyond this "technical" dialogue and into the realm of signalling potential, specific applications of the data products developed for specific blue economy industries and into advancing science-based policies. For example, the **Zoo- and Phytoplankton EOVs** demonstrator's innovative model has gained interest from major biodiversity initiatives, like the **Convention on Biological Diversity**[42]. On the more challenging side, issues connected to **data harmonization** and **interoperability** (due to the lack of common standards across blue data infrastructures, especially for in-situ observations), but also to the **interoperability between blue data infrastructures and computing infrastructures** for the execution of processes still remain, limiting the possibilities of Open Science practitioners. Potential approaches to address these challenges are being explored. For example, one important technical consideration emerging from the internal dialogue established amongst infrastructures and researchers is the fact that **standardized**, **harmonized** and **scalable data access systems** are a necessary condition for the deployment of **scalable algorithms** (i.e., the capability to effectively exploit **Big Data**). Interoperability and data harmonization standards are the way for cost-efficient and sustainable developments, both for scientific research and applications. Related to this is the need to achieve separation between **computing infrastructures** and **algorithms/analytical methods**. Ideally, a **FAIR catalogue of algorithms** should make it possible to **deploy any desired, available algorithm on any computing infrastructure**, **on-demand**, **closer to** the required "input" **data**. Making this possible requires the definition of a **protocol for interoperability** of **algorithms** across **computing infrastructures**. This would open new opportunities, enabling the possibility of implementing a more reliable, distributed "system of systems" (underpinned by all federated infrastructures) that could serve a wealth of input data into analytical applications (which in the case of big data is a necessary condition for successful discovery and exploitation).

From a wider, societal perspective, B-C domain specific services -such as those offered by Demonstrators 4 and 5- are proving valuable towards bridging the gap between the "last sea-mile" of cloud services and ground-based communities, opening a window of opportunity to serve resource-poor communities with **FAIR marine data services**. Adapting to EU environmental standards and/or regulations often requires resources that many communities lack, requiring sustained (international) governmental support towards alignment or compliance. B-C virtual labs are demonstrating that they can contribute to service community needs in a cost-effective manner.

Feeding from comprehensive input received through initial stakeholder dialogue throughout 2019-2021, we have summarised Blue-Cloud's "**strengths**" and "**weaknesses**" together with the landscape of "**opportunities**" and "**threats**" surrounding its development to produce a **SWOT analysis** that can be useful to guide a **long-term vision** for the evolution of its efforts into the future (see Table 3).

[42] https://www.cbd.int/conferences/post2020

Table 3: Current outlook of B-C Strengths, Weaknesses, Opportunities and Threats (SWOT)

<table>
<tr><th>Strengths</th><th>Weaknesses</th></tr>
<tr><td colspan="2">Internal: Specific to the Blue-Cloud's efforts</td></tr>
<tr><td>
Brings together representative Blue data and e-infrastructures for further dialogue and concrete actions to further address interoperability issues.
Improved, stronger analytical framework (Virtual Research Environment) developed building on previous efforts (i-Marine, Blue-Bridge project) and tested for different uses across different domains.
Builds on existing European initiatives and enhances them, not replaces them nor reproduces them.
Enables and supports implementation of FAIR principles in the marine domain (through use of standardized metadata).
Will provide feedback to improve and consolidate good practices towards FAIRness of research outputs (concerning metadata and publication protocols).
Supports a culture change towards Open Science by providing an example of best practice.
Familiarizes and trains users in the deployment of Open Science tools in the marine domain, in anticipation of future developments (EOSC, DTO).
Geared at showcasing the value of Open Science towards EU Green Deal and UN Agenda 2030.
</td><td>
Engaged blue data and e-infrastructures and teams need financial resources to continue supporting services "free of charge" to users beyond the project, ensuring sustainability of efforts.
Data harmonization and interoperability across participating blue data infrastructures not yet solved.
Lack of interoperability protocols between blue data infrastructures and e-infrastructures.
Limitations of "data duplication" model (i.e. data sets are "copied" into analytical framework). Hosting = expensive.
APIs developed to connect with blue data infrastructures depend on infrastructures maintaining them updated with technical specifications (or else they stop functioning).
Small user base, limiting potential towards triggering ideas for innovative applications.
Working in an "enclosed" analytical environment could potentially raise barriers to reach larger pools of potential users of marine data.
</td></tr>
<tr><th>Opportunities</th><th>Threats</th></tr>
<tr><td colspan="2">External: Framing the context in which the Blue-Cloud project is developing</td></tr>
<tr><td>
Current technological developments anticipate an unprecedented growth in abundance of marine data.
The price of computing resources is lowering for users as "pay-on-demand" for cloud-computing is introduced.
Strong interest in the marine community in sharing marine data, processes and results, creating momentum towards advancing metadata standards and to agreeing on basic principles for data sharing.
GreenData4All space opens the opportunity for linkages with other domains/sectors (i.e "human health", other economic sectors) and multidisciplinary research.
EOSC will expose existing blue data infrastructures and researchers to much wider audiences, providing more "reasons" for effective collaboration.
The global transition towards data science will make a wealth of scientific methods, processes, algorithms and systems available to extract knowledge and insights from big data. Emerging initiatives such as Digital Twin of the Ocean will welcome common languages in support of data interoperability and interrogation and the integration of analytical resources to support their simulation capabilities.
New <u>developments in user interfaces</u> allow engagement of non-expert users in digital platforms.
International developments (UN DOSS, G7 FSOI) open opportunities for global collaboration beyond EU.
</td><td>
Competitors delivering cloud services (<u>Amazon</u>, <u>Google</u>, <u>Microsoft</u>) offering comparable solutions.
Lack of FAIR data management (not fully mainstreamed) in the marine domain is impeding progress towards seizing the potential of Open Science.
Open science currently relies on individual scientists with heavy workloads, lack of time/resources/recognition for data management and balanced priorities towards delivering academic publications (favouring data embargoes).
Resources required for FAIR data management (est. 10% of research cost) to enable Open Science are not currently embedded in research budgets.
Lack of agreed metadata schemes in marine domain.
Virtual Labs servicing marine researchers are widely set up across different e-infrastructures, with a fragmented offer resembling the "old days" of marine data.
Data management efforts are seen as diluting resources away from marine observing systems and marine research, which need continuity of long time series to support the system.
Unproductive duplication of efforts and sustainability caused by lack of incentives pushing for interoperability amongst existing blue data infrastructures.
Lack of standard protocols of interoperability between blue data and e-infrastructures limits potential.
Data ownership and the road to FAIR data vs open data (differences in data policies and access rights).
</td></tr>
</table>

List of Figures

List of Tables

Blue-Cloud

Glossary

Word	Definition
AI	Artificial Intelligence
Argo	International programme on ocean observation using Argo floats
API	Application Programming Interface
B-C	Blue-Cloud
BlueBridge	H2020 project Supporting Blue Growth with innovative applications based on EU e-infrastructures
C3S	Copernicus Climate Change Service. CFP
CAMS	Copernicus Atmosphere Monitoring Service
CMCC	Euro-Mediterranean Centre on Climate Change
CMEMS	Copernicus Marine Environment Monitoring Service
CNR	Italian National Research Council
D4Science Infrastructure	Data Infrastructure promoting Open Science (managed by CNR)
DG	Directorate-General (of the EC)
DG DEFIS	Directorate-General for Defence Industry and Space (formerly DG GROW)
DG RTD	Directorate-General for Research and Innovation
DG MARE	Directorate-General for Maritime Affairs and Fisheries
DG GROW	Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs (succeeded by DG DEFIS)
DG ENV	Directorate-General for Environment
DG CONNECT	Directorate-General for Communications Networks, Content and Technology
DG DEVCO	Directorate-General for International Cooperation and Development
DIAS	Data and Information Access Service (funded by EC COPERNICUS programme)
DTO	Digital Twin of the Ocean
EASME	Executive Agency for Small and Medium-sized Enterprises
EC	European Commission
EcoTaxa	Web application dedicated to the visual exploration and the taxonomic annotation of images focused on planktonic biodiversity
EMBL	European Molecular Biology Laboratory
EMBL-EBI	European Bioinformatics Institute
EMBRC	European Marine Biological Resource Centre
EMODnet	European Marine Observation and Data Network
ENA	European Nucleotide Archive
EOSC	European Open Science Cloud
ESEB	External Stakeholder & Expert Board (of Blue-Cloud)
EU	European Union
EUDAT	Pan-European network and collaborative data infrastructure, consisting of a network of academic computing centres
EuroBioimaging	European Research Infrastructure for Imaging Technologies in Biological and Biomedical Sciences
EurOBIS	European Ocean Biodiversity Information System
FAIR	Findable, Accessible, Interoperable and Reusable

FAO	Fisheries and Agriculture Organisation of the United Nations
G7 FSOI	G7 Future of Seas and Oceans Initiative
GPU	Graphics Processing Unit
H2020	Horizon 2020 EU Framework Programme
HPC	High-Performance Computing
ICOS	Integrated Carbon Observation System
ICT	Information and communications technology
Ifremer	The French Research Institute for Exploitation of the Sea
iMarine	Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources
ISC	International Science Council
IoT	Internet of Things
JRC	Joint Research Centre
KERs	Key Exploitable Results
KPIs	Key Performance Indicators
MARIS	SME expert in European marine data management infrastructures (Blue-Cloud technical coordinator)
ML	Machine Learning
MOi	Mercator Ocean International
MSFD	Marine Strategy Framework Directive
MSP	Marine Spatial Planning
MSPD	Marine Spatial Planning Directive
NOAA	National Oceanic and Atmosphere Administration (United States of America)
OGC	Open Geospatial Consortium
SDG	Sustainable Development Goals
SeaDataNet	A pan-European infrastructure to ease the access to marine data measured by the countries bordering the European seas (network of NODCs)
SeaDataCloud	EU HORIZON 2020 project aiming at considerably advancing SeaDataNet Services and increasing their usage, adopting cloud and High Performance Computing technology for better performance.
SME	Small-Medium Enterprises
SSBE	SME specialised in marine policy (Roadmap coordinator) - Seascape Belgium
Trust-IT	SME specialised in market and technical research analyses in the field of ICT (Blue-Cloud coordinator)
UN	United Nations
UN 2030	The 2030 Agenda for Sustainable Development
VLIZ	Flanders Marine Institute
VRE	Virtual Research Environment
Web-based Open Science	An approach to the scientific process that focuses on spreading knowledge as soon as it is available using web-based digital, collaborative technologies. In this document, “web-based Open Science” and “Open Science” are considered synonyms, but “web-based Open Science” is used to highlight the key role of web-based technologies in enabling Open Science.
WEkEO DIAS	WEkEO is one of the 5 Copernicus DIAS, bringing in the CMEMS, C3S and CAMS