European
MARINE BOARD
Advancing Seas & Ocean Science

7TH EUROPEAN MARINE BOARD FORUM

Big Data in Marine Science:

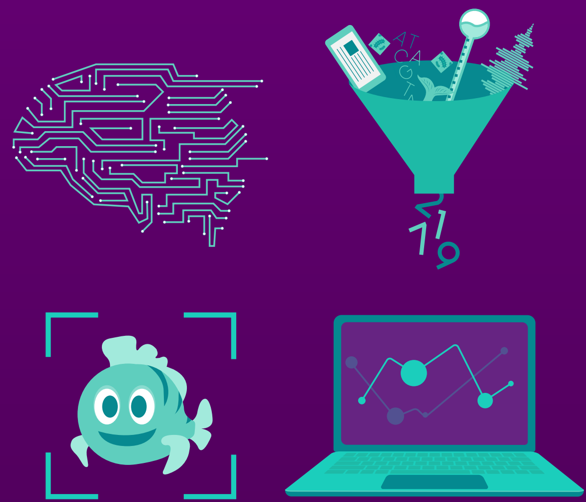
Supporting the European Green Deal, the EU Biodiversity Strategy and a Digital Twin Ocean

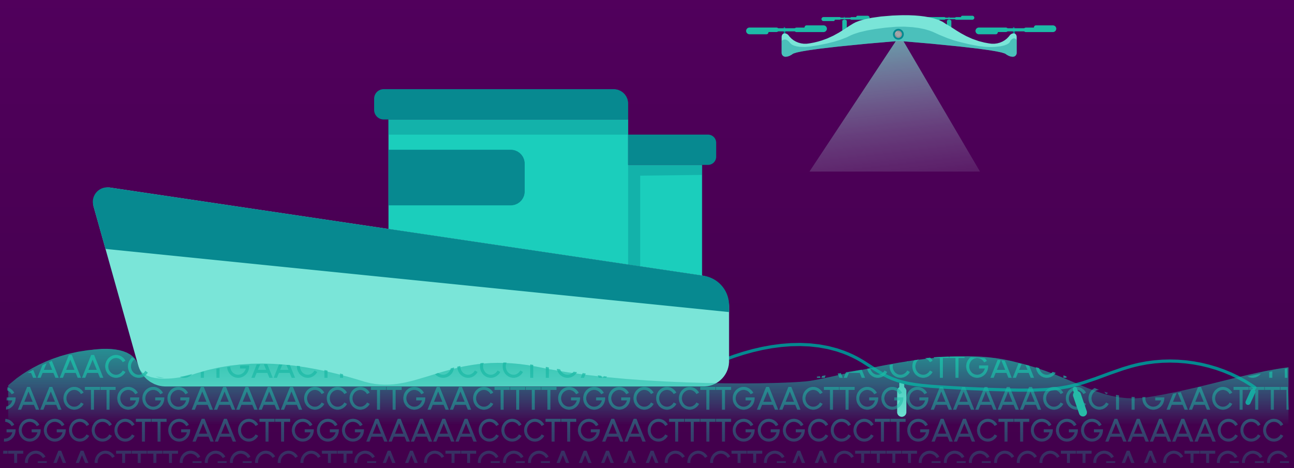
23rd October 2020, online

PROCEEDINGS

"Big data are the light through a new lens for us to observe, perceive, and hopefully better understand the Ocean and people."

Linwood Pendleton, Senior Vice-President for Science, C4IR Ocean, Norway & International Chair of Excellence, European Institute of Marine Studies, France





www.marineboard.eu

EUROPEAN MARINE BOARD

The European Marine Board provides a pan-European platform for its member organizations to develop common priorities, to advance marine research, and to bridge the gap between science and policy in order to meet future marine science challenges and opportunities.

The European Marine Board is an independent and self-sustaining science policy interface organization that currently represents 35 Member Organizations from 18 European countries. It was established in 1995 to facilitate enhanced cooperation between European marine science organizations towards the development of a common vision on the strategic research priorities for marine science in Europe. The EMB promotes and supports knowledge transfer for improved leadership in European marine research. Its membership includes major national marine or oceanographic institutes, research funding agencies and national consortia of universities with a strong marine research focus. Adopting a strategic role, the European Marine Board serves its member organizations by providing a forum within which marine research policy advice is developed and conveyed to national agencies and to the European Commission, with the objective of promoting the need for, and quality of, European marine research.

www.marineboard.eu

Edited by Britt Alexander, Paula Kellett, Joke Coopman, Ángel Muñiz Piniella & Sheila Heymans

This publication should be cited as: Alexander, B., Kellett, P., Coopman, J., Muñiz Piniella, Á., Heymans, J. J. (2021). 7th EMB Forum Proceedings. Big Data in Marine Science: Supporting the European Green Deal, EU Biodiversity Strategy, and a Digital Twin Ocean.
DOI: 10.5281/zenodo.4541335.

Cover infographics by Arteveldehogeschool students

TABLE OF CONTENTS

EUROPEAN MARINE BOARD FORUM SERIES

The European Marine Board Forum (EMB Forum) brings together European marine research stakeholders, representatives of the marine science community, funding agencies and national and European science institutions, to discuss research priorities and to promote marine science in Europe and globally. EMB Fora provide a platform for EMB members, partner organizations, individual scientists, and European and national policymakers to interact on a particular topic or theme of strategic importance to:

- Provide a focal meeting point for discussion;
- Facilitate the exchange of information and ideas and agree on a common position; and
- Enhance collaboration and reduce fragmentation and/or duplication in the European research effort.

The main messages, discussions, and decisions from EMB Fora are recorded and published as proceedings. Presentations and outputs of EMB Fora are available on the European Marine Board website: **www.marineboard.eu**

7th EMB Forum

The 7th EMB Forum was first planned as an in-person event on 29 April 2020. However, due to the COVID-19 pandemic, this meeting was adapted to an online launch event for the EMB Future Science Brief N°. 6 on 'Big Data in Marine Science'. The 7th EMB Forum was moved to 23 October 2020, with the hope that an in-person event would be possible. As the second wave of COVID-19 rapidly escalated in Belgium in the weeks leading up to the 7th EMB Forum, the event was moved entirely online. It provided an opportunity to further reflect on the role of big data in advancing marine science to support recent policy developments, namely the European Green Deal, the EU 2030 Biodiversity Strategy, and the development of a Digital Twin Ocean (DTO). 192 participants from Europe and further afield were present, representing academia, government, industry, and NGOs. Experts gave presentations and participated in panel discussions[1], with audience interaction. This document is a summary of the discussions, and full recordings of the sessions are available on the EMB YouTube Channel[2].

Credit: Adam Martinakis

Artwork included keynote in Linwood Pendleton's speech.

[1] Due to the October 23rd event initially being planned as hybrid, panelists were from Belgian organizations, and remote speakers from further afield.

[2] https://www.youtube.com/watch?v=rf2T4lEeh_k&list=PLXKTm_QGiR-hkKO6_aGo2t18TjMXY7S7D

7TH EUROPEAN MARINE BOARD FORUM
Programme

Moderator: Sheila Heymans, EMB Executive Director

Opening Session		
10:00 – 10.05	***Welcome***	**Gilles Lericolais**, EMB Chair & Ifremer CEO Advisor, France
10.05 – 10.15	***Opening Remarks***	**Andreea Strachinescu**, Head of Unit, Maritime Innovation, Marine Knowledge and Investments, DG MARE, EC
10.15 – 10.30	***EMB Future Science Brief N°. 6*** 'Big Data in Marine Science'	**Lionel Guidi**, Researcher at LOV-CNRS, France, Chair EMB Working Group 'Big Data in Marine Science'
10.30 – 10.50	***Keynote Address:*** 'Homo Digitus Oceanus: The Human Side of Big Ocean Data'	**Linwood Pendleton**, Senior Vice-President for Science, C4IR Ocean, Norway & International Chair of Excellence, European Institute of Marine Studies, France
10.50 – 11.00	***Youth Perspectives:*** 'Supporting the Ocean System with Integrated Big Data'	**EMB Young Ambassadors:** **Natalija Dunić,** Postdoctoral Researcher, IOR, Croatia **Alba González Vega,** PhD Student, IEO, Spain **Liam Lachs,** PhD Student, Newcastle University, UK
Session 1: Big Data and the European Green Deal		
11:30 – 11:45	***Opening Talk:*** Marine Spatial Planning Challenge Simulation Platform: Serious Gaming for Policy and Ecosystem Based Management	**Jeroen Steenbeek**, MSP Challenge & Senior Software Engineer, Ecopath International Initiative
11.45 – 12.30	***Panel Discussion***	**Felix Leinemann**, Head, Blue Economy Sectors, Aquaculture & Maritime Spatial Planning Unit, DG MARE, EC **Antonia Leroy**, Head of Ocean Policy, WWF Europe **Clea Parcerisas**, PhD Student, VLIZ, Belgium **Ondrej Socuvka**, Senior Policy Manager, Google Europe

Session 2: Big Data and the EU 2030 Biodiversity Strategy		
14:00 – 14:15	***Opening Talk:*** Big Data in Marine Science to Support the EU Biodiversity Strategy	**Sakina-Dorothée Ayata**, Associate Professor of Marine Ecology, LOV, Sorbonne University, France
14:15 – 15:00	***Panel Discussion***	**Anne Teller,** Policy Officer, DG ENV, EC **Patrick Roose**, Operational Director, RBINS, Belgium **Patricia Martin-Cabrera**, Data Manager, VLIZ, Belgium **Ward Appeltans**, Marine Biodiversity Programme Specialist, IOC-UNESCO & Project Manager, OBIS
Session 3: Big Data and a Digital Twin Ocean		
15:30 – 15:45	***Opening Talk:*** Digital Twin Ocean: Island Digital Ecosystem Avatars (IDEAS) for Sustainability	**Neil Davies**, Director, Gump Station on Moorea, French Polynesia & Research Affiliate, Berkeley Institute of Data Science, USA
15:45 – 16:45	***Panel Discussion***	**Fabienne Jacq**, Policy Officer, DG DEFIS, EC **Marilaure Grégoire**, Research Director, University of Liège, Belgium **Serge Scory**, Chair, SeaDataNet & Head, BMDC, Belgium **Kate Larkin**, Deputy Head, EMODnet Secretariat **Christian Kirchsteiger**, Policy Officer, DG Connect[3], EC
Closing Session		
16:45 – 16:55	***Closing Address***	**Ricardo Serrão Santos**, Portuguese Minister of the Sea
16:55 – 17.05	***Closing Statements***	**Sigi Gruber**, Head, Healthy Oceans & Seas Unit, DG RTD, EC

The full programme is available for download on the EMB website[4].

[3] Christian Kirchsteiger was unable to connect to the webinar.

[4] https://www.marineboard.eu/7th-marine-board-forum#:~:text=The%207thEMB%20Forum,of%20a%20Digital%20Twin%20Ocean.

KEY MESSAGES

Big data are rapidly transforming our society and leading to a paradigm shift in the way we face major challenges such as climate change and the sustainable management of marine resources. The expansion of the scale and scope of Ocean observing systems, and other data sources, generates a flood of highly complex Ocean data, often in near real-time. It is of utmost importance to integrate data so that it is readily available to support decision-making. The overarching vision and recommendations that resulted from the 7th EMB Forum discussions are presented below and are supported by the European Marine Board:

- **Develop a culture of transdisciplinary collaboration and trust.** This requires transparency and multicultural dialogues. Big data scientists, social scientists, ethicists, and others from academia and industry should be brought into the Ocean community to collaborate, put forward scientific hypotheses, and co-develop computational solutions that advance marine science and policy while building understanding and trust in big data and big-data derived knowledge. Decision-makers must work with scientists to determine the evidence required to plan for a sustainable future;

- **Develop trust in the conclusions drawn from big data.** It is important to question and test big data findings, being wary of artifacts. Funding and resources are needed to develop high quality ground-truthing datasets (i.e. data obtained from *in situ* observations) that can be used to train algorithms. This will help identify 'bad' data and minimize errors that could be misleading. Caution is also needed in the interpretation of real-time analysis of big data to avoid decision-making errors;

- **Advance open and transparent data through the accelerated adoption of the FAIR[5], CARE[6], and TRUST[7] data principles.** Quality control and assurance of data, as well as inclusion of appropriate metadata are important. Incentives and ethical, legal, and social issues should be addressed, and European marine data infrastructures need to work together to achieve these goals. Data sharing can be incentivized and improved by giving credit through Digital Object Identifiers (DOIs) and tracking the provenance of data;

- **Prioritize the federation and interoperability of Ocean data.** Progress is already being made: including through EMODnet[8], Copernicus[9], the Horizon 2020 BlueCloud[10] project, the Horizon 2020 Digital Twin Ocean call, and the International Oceanographic Data and Information Exchange (IODE[11])'s Ocean Data and Information System (ODIS[12]). Federated digital ecosystem avatars that focus on local-scale social-ecological foresight could form a web of digital twins that interact with each other and could be nested within digital twins for larger Earth systems.

- **Progress towards a paradigm shift using big data to engage the public through storytelling.** Big Ocean data should be used by society and educators to increase public awareness of the Ocean's role in a sustainable future. A Digital Twin Ocean, combined with engaging storytelling, will allow users to experience, play, visualize, and understand what is happening in the Ocean.

5 Findable, Accessible, Interoperable, Reusable (https://www.nature.com/articles/sdata201618)
6 Collective Benefit, Authority to Control, Responsibility, and Ethics (https://www.gida-global.org/care)
7 Transparency, Responsibility, User Focus, Sustainability, and Technology (https://www.nature.com/articles/s41597-020-0486-7)
8 https://emodnet.eu/en
9 https://www.copernicus.eu/en
10 https://www.blue-cloud.org/
11 https://www.iode.org/index.php?option=com_content&view=article&id=3&Itemid=33
12 https://www.iode.org/index.php?option=com_content&view=article&id=559&Itemid=100362

OPENING SESSION

Welcome

Gilles Lericolais, *Chair of the European Marine Board & Ifremer CEO Advisor, France*

Gilles Lericolais, Chair of the EMB, welcomed participants to the 7th EMB Forum, which was held under unusual circumstances. He introduced the EMB, and in particular the purpose of the EMB Fora in enabling discussion and knowledge exchange between stakeholders on both sides of the marine science-policy interface. He noted the timely conclusion of the EMB working group on 'Big Data in Marine Science', which ran from May 2019 – April 2020 and which resulted in the EMB Future Science Brief N°. 6[13]. He introduced the 7th EMB Forum as an opportunity to discuss the recommendations that were presented in the Future Science Brief, and how they could be taken forward to support the European Green Deal, and ultimately to better understand and protect our Ocean for future generations.

Opening Remarks

Andreea Strachinescu, *Head of Unit on Maritime Innovation, Marine Knowledge and Investments, DG Maritime Affairs and Fisheries (MARE), European Commission*

Andreea Strachinescu presented key opportunities in Europe to use big data to better understand the Ocean and its importance for humanity. She highlighted the value of Artificial Intelligence (AI) in the European Commission's 'European Data Strategy' and the importance of marine data in the 'Farm to Fork Strategy' of the European Green Deal. For example, data from Copernicus and the European Marine Observation Data Network (EMODnet) can reduce investment risks and facilitate sustainable practices in the fisheries and aquaculture sectors. EMODnet also contributes to Global Fishing Watch[14]: A global effort to visualize and track vessel data in near real-time via neural networks. She mentioned Horizon Europe as the main tool to support the digital transformation and the European Green Deal, as well as 'Mission Healthy Oceans, Seas, Coastal and Inland Waters'[15] and the 'Partnership for a Climate Neutral, Sustainable and Productive Blue Economy'[16]. She also highlighted the BlueCloud project[17] and the Digital Ocean Twin call as initiatives that aim to federate marine data to meet the goals of the European Green Deal.

[13] https://www.marineboard.eu/publications/big-data-marine-science
[14] https://globalfishingwatch.org/
[15] https://ec.europa.eu/info/publications/mission-starfish-2030-restore-our-ocean-and-waters_en
[16] https://ec.europa.eu/info/sites/info/files/research_and_innovation/funding/documents/ec_rtd_he-partnership-climate-neutral-sustainable-productive-blue-economy.pdf
[17] https://www.blue-cloud.org/

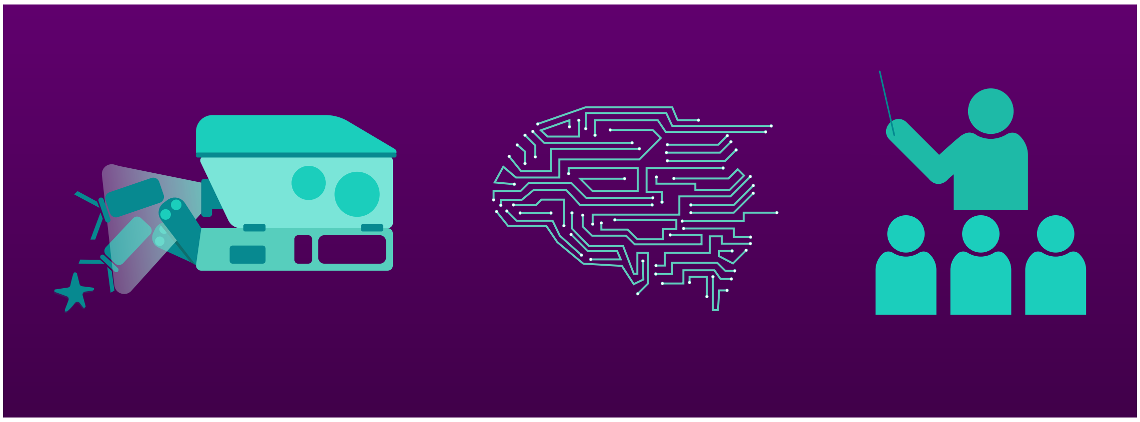
EMB Future Science Brief 'Big Data in Marine Science'

Lionel Guidi, *Researcher at Laboratoire d'Océanographie de Villefranche sur Mer (LOV), France & Chair EMB Working Group on 'Big Data in Marine Science'*

Lionel Guidi presented the main concepts within the EMB Future Science Brief N°. 6 on 'Big Data in Marine Science'. He explained that *"the highly dynamic and connected nature of the Ocean is a great challenge for marine biologists and computer scientists. This is why there is enormous value in bringing big data scientists into Ocean science because they can look at the Ocean as a computational problem and try to develop novel methods to find solutions"*. Big data are defined in the Future Science Brief using the 'Five Vs': 1) a large **V**olume of data; 2) a high **V**elocity and frequency of incoming data; 3) a high **V**ariety of complex heterogeneous data; 4) the **V**eracity and reliability has to be ensured; and 5) an inherent **V**alue to be able to extract meaning from the data and create products to develop actions. He explained that the biggest challenge marine science is facing in terms of big data is that of complexity, with a lot of different types of data of a different nature (e.g. different sources, formats, spatial and temporal scales). The Future Science Brief recommends improved integration of data science into marine science; continued development of automated data acquisition; advancement of data management (including the full implementation of the FAIR data principles); increased interoperability between computational infrastructures; development of standardized algorithms and community-maintained data sets; and improved training and collaboration.

"We are not yet there in terms of big data in marine science, but we have the means to get there."

- Lionel Guidi



Credit: Artevelde Hogeschool Students

Continued development of automated data acquisition, the development of standardized algorithms and community-maintained data sets, and improved training and collaboration are some of the key recommendations from the EMB Future Science Brief Nº. 6 on 'Big Data in Marine Science'.

Homo Digitus Oceanus: The Human Side of Big Ocean Data

Linwood Pendleton, *Senior Vice-President for Science, Centre for the 4th Industrial Revolution (C4IR), Norway & International* Chair *of Excellence, European Institute of Marine Studies, France*

Linwood Pendleton's keynote speech guided the audience through the 'human' challenges and opportunities of big Ocean data using artwork, philosophical parables, and poetry. He began by acknowledging the enormous technical challenges of big Ocean data, and the equally daunting and less well understood human challenges. He explained the rapid evolution of the big data world that we live in, which is driven by changes in how we observe the Earth and measure the Ocean. This includes dramatic increases in the number and types of sensors for monitoring the Ocean and a phase shift in the types of people and institutions that collect and share data. This explosion of ways of measuring the Ocean is also contributing to the democratization of data collection. He described that the biggest challenge to come is making sense of what we learn from big data. *"These are still just projections of the world, and we must always question and test what we see… the analytics of big data will reveal new insights into how the Ocean works, but not all of these new insights will be correct"* he explained. New patterns that are revealed through big data analysis may only be locally applicable, may reflect patterns and relationships from the past that won't hold in the future, or may simply be artifacts of when, where, and how we collected the data. It will be important to discern drivers from correlates in order to use findings from big data to inform policy. The International Accord for Open Data[18] calls for diligence in testing big data and the resulting findings and predictions, and self-correcting. There will also be a lot of heterogeneity in how big data are embraced, and it will not fit into all ways of knowing and thinking. He went on to discuss building trust in big Ocean data and what we learn from it, which requires acknowledgement of its limitations and biases. He explained that Ocean data are unrepresentative since we do not collect data everywhere or all the time, and we collect data to test our old hypotheses, which often reflect previous biases. This becomes a self-reinforcing view of the world making it is very difficult to exploit out-of-the-box insights from big data if the data are from inside the box. There is also potential for bias in human big data as data are generally collected from people that are digitally connected, creating bias against less connected people in a *"digital divide"*. He then described that the measure of data quality depends on the question to be answered. To ascertain the quality of big data, the data need to be considered within the context of their three-dimensional relationships e.g. time and human contexts, and combined with data about policies, economics, and human beings. Transparency is needed to build trust in data quality, and this requires tracking and showing the provenance of data. He highlighted the need to go beyond simple DOIs to the creation and use of tags that reflect key metadata for the individual datum within databases. With tags we can begin to create data impact factors to reward those who share their data. He concluded by stressing that many people are afraid that big Ocean data will mean we are buried in a sea of useless data as our ability to collect data may exceed our ability to understand and use it. To avoid this, we need collaboration between decision-makers and scientists to determine what evidence is required to plan for a sustainable future.

" The Ocean is big, data are big, but how much do we know, and how can that be useful?"

- Poem by Anna Zivian

[18] https://council.science/publications/open-data-in-a-big-data-world/

Youth Perspectives: Supporting the Ocean System with Integrated Big Data

Natalija Dunić, *Postdoctoral Researcher, Institute of Oceanography and Fisheries (IOR), Croatia*
Alba González Vega, *PhD Student, Spanish Institute of Oceanography (IEO), Spain*
Liam Lachs, *PhD Student, Newcastle University, UK*

EMB Young Ambassadors Natalija Dunić, Alba González Vega, and Liam Lachs presented their perspectives on the future of big data as a means to understand our dynamic living Ocean. They reflected on the current technological revolution that is allowing the exploration of previously unchartered territories in the Ocean, and leading to an explosion in the collection of Ocean data that has the power to assist us in reaching new levels of understanding of our Ocean and its connection to human life. Big data will assist scientists in taking a multi-dimensional approach to understanding the Ocean from micro- and local-scales, to regional and global processes. Data visualization techniques will need to adapt to the challenges of large volumes of heterogenous Ocean data, including social-economic data. They highlighted the importance of responsible data management and the role of AI in providing new frameworks to synthesize big data to allow better decision-making.

They contemplated the future of big Ocean data. *"As young scientists, we spend a lot of time trying to imagine what the future holds for us. When we think about big data, we imagine two blue worlds: the real-life Ocean that we all know, and a virtual Ocean that holds all the knowledge gathered by millions of scientists and machines around the world, where real-time and in situ data are constantly uploaded and used for calibration and validation. This virtual Ocean would be unified in format, collaborative, and of course open-access. This will, for example, allow us to quickly react to natural hazards, such as meteotsunamis, or to predict future oceanographic and ecosystem trends and climate change scenarios more accurately".*

They considered whether the development of AI would eventually put them out of work, but concluded that would not be the case. They explained that their philosophy is to leave to the machines what they are best at, such as collecting and uploading large amounts of data in real-time, unaffected by conditions that limit humans like a global pandemic. Humans should do what only humans can do, such as validating, analyzing, and interpreting data, formulating hypothesis, solving problems, and finding solutions. Their vision of the future is that of a symbiosis between humans and technology.

"Technology is not a competitor, but rather a collaborator."

- EMB Young Ambassadors

Credit: Adam Martinakis

Artwork included in Linwood Pendleton's speech.

Audience Questions & Answers

Lora Fleming (University of Exeter, UK): *How could you imagine the potential benefits of integrating human health and wellbeing data into the ongoing collection of "Ocean" data, both in terms of collection, analysis, and better integration of policy and training?*
Lionel Guidi: Collecting human data helps to demonstrate the human impacts of Ocean change.
Gilles Lericolais: The EMB Policy Brief N°. 8 'Policy Needs for Oceans and Human Health'[19] provides insight into the benefits of integrating human health data with Ocean data.

Raïssa Meyer (Helmholtz Centre for Polar and Marine Research, Germany): *Of the recommendations presented in the EMB 'Big Data in Marine Science' Future Science Brief, which would you say may pose the biggest challenge?*
Lionel Guidi: All of the recommendations presented are challenging, however those on the topic of data handling and management are pre-requisites to achieving many of the other challenges, so should be dealt with first. We are making progress, but there is still a lot to be done.

Onur Akdas (Dokuz Eylül University, Turkey): *Do you think that coastal residents should be target citizens for the collection of Ocean data by the public? Does it make a difference regarding the data?*
Linwood Pendleton: Citizens are an important source of data especially for places and things that cannot be monitored regularly, including many near-shore areas and estuaries, as well as biodiversity and plastic. Citizen science data are also important in understanding human uses of the Ocean. There is a lot of information in passive citizen science data i.e. images shared online with indications of water quality or biodiversity. Citizen science data will also be important for collecting data on human uses of the Ocean. In this latter category, it is important not to focus just on coastal residents, since marine ecosystem services are also enjoyed far from the sea, and opportunities to use citizen science far from coastal areas are important.

Anna Akimova (Thuenen Institute of Sea Fisheries, Germany): *What are the links or interactions between the EMB Big Data in Marine Science working group and the INSPIRE Directive?*
Britt Alexander (EMB Secretariat): INSPIRE is a directive on data standards, which is key to enabling data to be shared and integrated into infrastructure, so that it can be analyzed with big data methods (e.g. AI). The INSPIRE[20] directive is highlighted in the EMB Future Science Brief N°. 6.

[19] https://www.marineboard.eu/sites/marineboard.eu/files/public/publication/EMB_PB8_Policy_Needs_v4_web_0.pdf
[20] https://inspire.ec.europa.eu/

SESSION 1

Big Data and the European Green Deal

Marine Spatial Planning (MSP) Challenge Simulation Platform: Serious Gaming for Policy and Ecosystem-based Fisheries Management

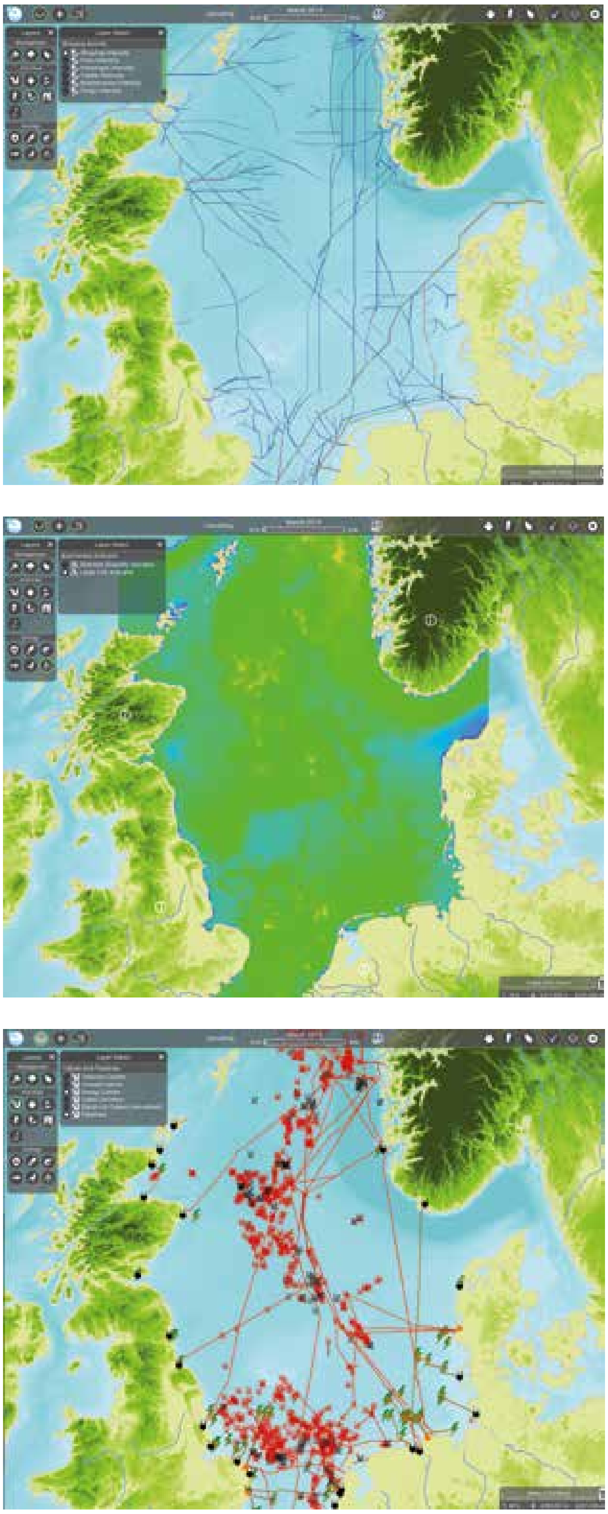
Jeroen Steenbeek, *MSP Challenge*
Senior Software Engineer, Ecopath International Initiative

Jeroen Steenbeek presented an overview of current and future uses of the MSP Challenge[21] simulation platform for policy and ecosystem-based fisheries management. Participants in the MSP Challenge are usually stakeholders, and they explore the marine spatial planning process through digital and social components, with multiple players participating in negotiating and approving spatial plans for a marine area that cover activities including shipping, offshore energy, mining, dredging, conservation, fisheries, and recreation. In a virtual environment, players experiment with, and learn about, the interplay between conflicting planning activities, society, and ecology. The games capture existing activities and ecology in an area of interest such as the North Sea and the Baltic Sea, for which games have already been developed. Participants take on the role of planner for a country, and are given goals that they should aim to achieve within 30-40 years. They then modify existing activities or plan new activities using a built-in knowledge-base. After the planning phase, there is a simulation phase where the spatial plans come into force using built-in simulation models for shipping, ecology, and energy. The outcomes are explored by participants for the next planning round where they can make changes. Jeroen explained that the MSP Challenge platform requires large volumes of data from a wide range of sources e.g. on human activities, policies, marine environment, ecology, climate forecasts, and interactions. The MSP Challenge could be considered a 'digital twin' where near real-time observational data are brought in to capture the full situation. He highlighted the huge potential for AI and smart analysis, particularly in quality checking real-time data. AI could also be used to better understand the decision-making processes. He concluded by highlighting that the current MSP Challenge platform is under development to bring in new data and soon all the code will be open source and the platform will be community-driven.

Credit: Breda University of Applies Sciences / Ecopath International Initiative

Impressions of an MSP Challenge session.

[21] https://www.mspchallenge.info/



Credit: Breda University of Applies Sciences / Ecopath International Initiative.

Examples of information presented to MSP Challenge participants during a North Sea session, reflecting shipping intensity (top), large fish ecological indicators (middle), and cables and pipelines (bottom).

Audience Questions & Answers

Anne-Katrien Lescrauwaet (Flanders Marine Institute, Belgium): *Bringing in near real-time data also brings in variability e.g. seasonal or spatial, into the decision-making process. How could the MSP Challenge deal with this?*
Jeroen Steenbeek: Models would need to be made more complex to be able to deal with near real-time data and its complexity. However, the models would need to remain simple enough for players to be able to play the full game within the limited time-frame of each session.

Pierre-Luigi Buttigeig (Max Planck Institute for Marine Microbiology, Germany): *Are you engaging expertise from psychology and decision-making in the design of the games?*
Jeroen Steenbeek: The games are primarily built by game developers, but there are a wide range of stakeholders involved in their design.

Anna Akimova (Thuenen Institute of Sea Fisheries, Germany): *Can you imagine feedback from the model to the data collection strategy to identify critical data gaps or resolution issues? Do you have an example of this?*
Jeroen Steenbeek: There are strategies in the MSP Challenge game to identify the data needed and the gaps. Data availability is one of the main bottlenecks to modelling. It is important to know if the data are available and in a usable format, and if they are in the public domain or not. Big data have amplified the challenge of getting the right data in the right format. Modelling is often used in general to show where the biggest data gaps are.

Carlos Tighe (Insight Centre for Data Analytics, Ireland): *What technologies are you using to build the games?*
Jeroen Steenbeek: All runtime tools are based on the Microsoft .NET framework e.g. Unity (C#) and Ecopath with Ecosim EwE (C#, VB.NET). A whole slew of Geographic Information System (GIS) and data analytical tools are used to define a MSP session. Propriety tools such as plug-ins for the EwE food web modelling are used to assist modelers in parameterizing and integrating peer-reviewed models.

Big Data and the European Green Deal

Panel Discussion

The panelists (Felix Leinemann, Antonia Leroy, Clea Parcerisas, and Ondrej Socuvka) were asked a series of questions by the moderator (Sheila Heymans), followed by a Q&A session with the audience.

Felix Leinemann, *Head, Blue Economy Sectors, Aquaculture & Maritime Spatial Planning Unit, DG Maritime Affairs and Fisheries (MARE) European Commission*

Question: ***What are the most pressing questions that need to be tackled to address the European Green Deal that big data can help answer?***

Felix Leinemann highlighted the need to scale-up marine renewable energy as part of the European Green Deal so we can increase its contribution to the EU's electricity production from the current share of 2% to 25% or 35% by 2050. This means that we will need to produce 25 times more energy from EU waters compared to today. *"We need to know the consequences of this increase for our maritime space so we can ensure that vulnerable marine ecosystems are adequately protected"* he explained. To achieve the goals of the EU 2030 Biodiversity Strategy for protected areas, we need a much better understanding of potential cumulative impacts of human activities on the marine environment and the interaction between offshore renewables and other activities at sea including fisheries, aquaculture, and shipping. He also highlighted the potential of offshore windfarms to provide space for the cultivation of shellfish and seaweed, which if scaled-up can relieve pressure on increasingly scarce land resources to feed the human population. He concluded that we can try to answer these marine spatial planning challenges with AI and big data.

Question: ***How can big data and AI help to overcome challenges related to the efficient use of maritime space?***

Felix Leinemann mentioned the need to be aware of, and to mitigate against, risks such as vessel collisions, discarded fishing gear, military activity, or dumped munitions and chemicals within areas that have the most potential for delivering offshore energy. Regional and sea-basin level planning is needed to be able to set up energy systems that will work across borders and keep ecosystems intact. He highlighted the EU's new offshore renewable energy strategy that lays out plans to scale-up cross-border and sea-basin cooperation.

Antonia Leroy, *Head of Ocean Policy, World Wildlife Fund (WWF), Europe*

Question: ***How can big data enable a more equitable way of working between marine industries for efficient use of marine space?***

Antonia Leroy began by mentioning the use of big data to assess and visualize what is happening at sea, including cumulative pressures of human activities, and its usefulness in helping to identify conflicts between users e.g. between wind farms and fisheries. She highlighted the importance of publicly accessible data to improve transparency and enable more informed and equitable decisions that account for economic and environmental perspectives. Big data can also create a more level playing field between stakeholders, for example by allowing the identification of illegal fishing, and can bring improvements to surveillance and monitoring at sea.

Question: ***How can big data contribute to the sustainable use of marine resources?***

Antonia Leroy cited ongoing discussions at EU level on the revision of the fisheries control regulations. The most common infringement in fisheries is not complying with the obligation to record and report catch and catch-related data, including the requirement for data to be transferred via satellite monitoring systems. She stressed the need to improve transparency and traceability along the seafood value chain. This could be helped through the implementation of electronic catch certificates, which would inform consumers about what is on their plate, and increase accountability. AI could be used to cross-check data from the catch certificates. She concluded by highlighting the need to ensure that proper regulatory structures are in place for AI to be used widely within the fisheries industry.

Clea Parcerisas, *PhD Student,*
Flanders Marine Institute (VLIZ), Belgium

Question: ***What do you see as the biggest challenge the European Green Deal will face that big data and AI can overcome?***

Clea Parcerisas began by introducing the use of big data and AI to make accurate predictions for scenarios and actions that have not happened yet and the importance of this for good management. Neural networks can model the interactions between complex systems and allow an understanding of the effects of certain actions beyond what humans can currently understand. She highlighted that data acquisition at sea remains the biggest challenge, however new technologies are becoming available to obtain more spatially distributed and continuous data, which can be used in modelling using AI. She concluded that data can always be bias and lead to errors, and highlighted the importance of continuing to test hypotheses.

Question: ***How can big data and AI contribute to addressing underwater noise levels for Good Environmental Status (GES) under the Marine Strategy Framework Directive (MSFD)?***

Clea Parcerisas reminded the audience that underwater noise is the 11th descriptor of Good Environmental Status (GES). To use noise to assess GES, it is necessary to understand the contribution of anthropogenic noise to underwater noise and to understand how this noise is affecting marine organisms. Models are required to understand both of these factors. She explained that terabytes of sound data feed into these models, together with other environmental and biological data, and these data help to determine sound levels that could be considered as acceptable for GES, and help to assess environmental status based on biodiversity levels and other ecological processes.

"Sound propagates much better than light underwater so marine organisms predominantly use sound to gather information about their environment. This is a useful way for us to measure marine biodiversity."

- Clea Parcerisas

Ondrej Socuvka, *Senior Policy Manager, Google Europe*

Question: ***What is the role of big tech in achieving the European Green Deal?***

Ondrej Socuvka expressed Google's support for the EU's ambitious plans on sustainability, and mentioned several of Google's sustainability initiatives using AI, including Global Fishing Watch[22] in which they are a partner. He also drew attention to Google's collaborations with researchers as part of their AI for Social Good Programme e.g. the National Oceanographic and Atmospheric Administration (NOAA)'s work using machine learning to characterize whale song[23].

Question: ***What lessons can the marine science community learn from the tech industry who are more advanced in using big data and AI?***

Ondrej Socuvka stressed the importance of interdisciplinary collaboration as a key to success. He highlighted Google's 'AI Impact Challenge', which provides grants and technical support to bring experts from different fields together to tackle climate change. Google Cloud AI Analytics[24] provides tools that are publicly available to be used. He mentioned the role of government in setting the regulatory environment for responsible AI, with a key objective of boosting public trust in AI, while recognizing the need to balance societal benefit and potential harm. Google stands for a risk-based approach whereby regulations should aim to achieve this balance, and should also keep in mind the cost of not using AI for tackling our biggest societal challenges. He advised academics to be aware of the risks and also the probability of harm of using AI, and urged them to consult Google's 'Responsible AI Practices'[25].

[22] https://globalfishingwatch.org/
[23] https://www.youtube.com/watch?v=JE3-LkMqBfM&feature=emb_title
[24] https://cloud.google.com/products/ai
[25] https://ai.google/responsibilities/responsible-ai-practices/

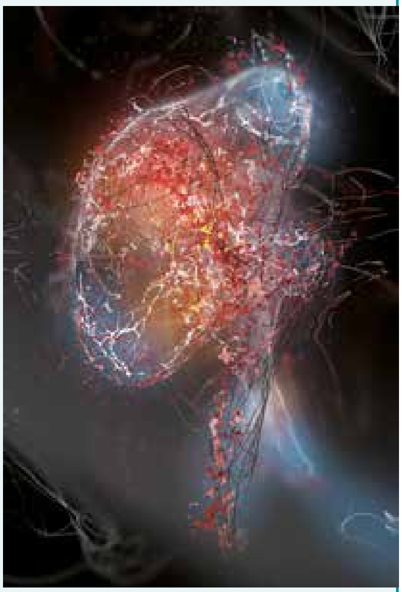
Audience Questions & Answers

Pierre-Luigi Buttigieg (Max Planck Institute for Marine Microbiology, Germany): *Machine learning can easily be poisoned by bad data. What strategies are being taken to make sure that data is informative and trustworthy? This is a major concern in our UN Ocean Decade digital ecosystem engineering work.*
Ondrej Socuvka: Google are working on data pattern recognition as part of the People and AI Research (PAIR)[26] initiative. These tools are openly available and allow researchers to easily spot trends and problematic areas in big data sets.

Felix Leinemann: *What are 'bad' data?*
Sheila Heymans (EMB): Bad data could be, for example, data that are based on preconceived ideas or biases.
Pierre-Luigi Buttigieg (Max Planck Institute for Marine Microbiology, Germany): They could be data originating from mis-calibrated instruments, spurious data points, or noise in the data. Algorithms can identify unreliable data, but these need to be trained with a large collection of consistent, comparable, and interoperable baseline "good" data. This is a priority area to direct funding. However, if you over-train algorithms, they may treat real signals that are "unusual" as mere errors and filter them out. Distinguishing between unstructured noise and potentially meaningful divergent signals that need human attention is key, and this is often tricky to develop.

Anna Akimova (Thuenen Institute of Sea Fisheries, Germany): *Can we use new technologies like AI to process long-term or historical data, which are required to understand climate related variability, and are often not digitized?*
Ondrej Socuvka: There are technological solutions for this. Google has a project on scanning books and analyzing their content using advanced visual text recognition. For large datasets, technological tools need to be deployed while respecting regulatory frameworks which still need to be defined and clarified, and these should clearly specify how research communities can use these datasets in a way that respects privacy.
Clea Parcerisas: The 'Text-to-knowledge (T2K) Research Group'[27] at the University of Ghent is working on extracting and digitizing biological data from old papers.

Andrej Abrmaic (Universidad de Las Palmas de Gran Canaria, Spain): *When does analysis become big data analysis? Are we already doing big data analysis through initiatives such as EMODnet, Copernicus, the EU Open Data Portal, and INSPIRE?*
Sheila Heymans (EMB): The EMB Future Science Brief N°. 6 on 'Big Data in Marine Science' outlines that while we have large quantities of data and highly variable data in marine science, it is not yet at the critical point where big data analytics (e.g. neural networks) are being widely deployed.



Credit: Adam Martinakis

Artwork included in Linwood Pendleton's speech.

[26] https://research.google/teams/brain/pair/
[27] https://ugentt2k.github.io/

SESSION 2
Big Data and the EU 2030 Biodiversity Strategy

Big Data in Marine Science: Supporting the EU Biodiversity Strategy

Sakina-Dorothée Ayata, *Associate Professor of Marine Ecology at Sorbonne University, Laboratoire d'Océanographie de Villefranche sur Mer (LOV), France*

Sakina-Dorothée Ayata began by highlighting the large social and economic cost of biodiversity loss for human societies, and the benefits of actions to protect biodiversity. For example, the conservation of fish stocks could increase the profits of the seafood industry by more than 49 billion Euros annually, and the protection of coastal wetlands could save the insurance industry about 50 billion Euros annually by reducing flood damage losses. She presented the EU 2030 Biodiversity Strategy and its objectives of promoting the monitoring, conservation, and restoration of biodiversity within and beyond Europe to lead the world towards an ambitious global biodiversity framework.

She presented two examples of how big data, digitalization and AI can contribute to achieving these objectives. She first described how big data could be used to define and monitor new marine protected areas (MPAs). Biodiversity hotspots can be identified by mapping plankton diversity using a high-throughput automated plankton imaging system combined with machine learning. Other types of marine data can be used to map the seabed and databases of species occurrences can then be exploited and combined with machine-learning based niche modelling to identify biodiversity hotspots that are most sensitive to climate change, and which would be suitable areas to designate as MPAs. She explained that data science can also be used to monitor biodiversity within newly defined MPAs to quantify their efficiency, for example using standardized observational methods, as well as new bioindicators based on the automatic acquisition of high-throughput data, particularly from images, genomics, satellites, or acoustics. The second example she gave was monitoring discarded bycatch from fishing trawlers using computer vision, i.e. automated analysis of video footage using AI[28]. Deep neural networks are used to develop a tool that is adapted to various video set-ups, and which is able to recognize individual fish and classify them into species even when some parts of the fish are missing from the images.

She concluded that an EU-wide methodology to map, assess and achieve good conditions in marine ecosystems should be developed to enable a new governance framework as part of the EU Biodiversity Strategy. She suggested that new bioindicators should be defined from machine-learning based analysis of high-throughput marine big data, particularly from imaging and metagenomics, and that the development of autonomous observation platforms could play an essential role in automation. This would provide data-driven support for stakeholders such as MPA managers, as well as governmental and European organizations, to select and prioritize species and habitats that need protection and monitoring.

[28] https://academic.oup.com/icesjms/article/77/4/1340/5542623

Credit: top left: GIS Posidonie; top right: Kyle McBurnie; bottom left: Sakina-Dorothée Ayata; bottom right: Nick Kamenos

Marine biodiversity encompasses a puzzling diversity of ecosystems, such as coralligenous habitats (top left), kelp forests (top right), intertidal reefs (bottom left), or mearl beds (bottom right).

Audience Questions & Answers

Sheila Heymans (EMB): *How good are fisheries surveillance systems at recognizing invertebrates, which are often brought up during trawling?*
Sakina-Dorothée Ayata: Invertebrates are often small so the resolution of the cameras may not be high enough to recognize them. Another issue may be the lack of ground-truthing data sets to recognize invertebrates using machine learning.

Ana Lara Lopez (EuroGOOS): *Are there examples of potential bioindicators to assess ecosystem condition?*
Sakina-Dorothée Ayata: There are none that are available now, but we can relate biodiversity information gained from images and genomics to environmental conditions using statistical models. Imaging can be used to identify plankton biodiversity, and metabarcoding can be used to identify a wider range of organisms. New indicators would need to be tested. Creating an indicator that can be used in all European seas is challenging as different species are needed to characterize GES in different sea basins. A coordinated study is needed to develop this type of bioindicator.

Big Data and the EU 2030 Biodiversity Strategy

Panel Discussion

The panelists (Anne Teller, Patrick Roose, Patricia Martin-Cabrera and Ward Appeltans) were asked a series of questions by the moderator (Sheila Heymans), followed by a Q&A session with the audience.

Anne Teller, *Policy Officer, DG Environment (ENV), European Commission*

Question: ***What are the most pressing questions for the EU Biodiversity Strategy, and how can big data and AI help to address them?***

Anne Teller drew attention to the first EU-wide ecosystem assessment published by the Europe Commission's Joint Research Centre (JRC)[29], which shows that while the marine environment is the largest ecosystem on Earth, it is also the least known. She mentioned that there are unique opportunities for big data, AI, and digitalization to improve our ability to assess GES, stressors in the marine environment, and costs of inaction. She highlighted the development of a legally binding instrument for restoration as part of the EU 2030 Biodiversity Strategy, which will fine-tune the definition of GES to emphasis critical habitats and flagship species. She also discussed the need for a strong monitoring and reporting system that will seek more accountability from Member States, and provide opportunities for corrective measures. The new 'Knowledge Centre for Biodiversity'[30] launched by the European Commission will be an important resource that will act as a one-stop-shop for key information about biodiversity research results and the impact of related policies to monitor progress under the EU 2030 Biodiversity Strategy.

Patrick Roose, *Operational Director, Royal Belgian Institute for Natural Science*

Question: ***What is the role of big data and AI in monitoring under the MSFD?***

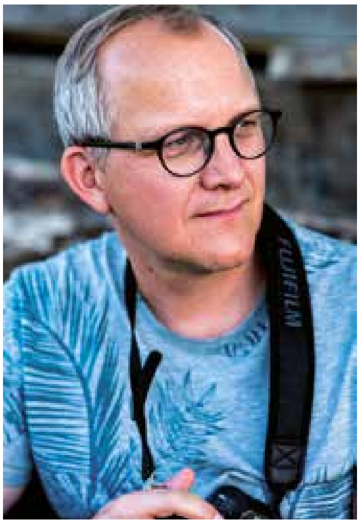
Patrick Roose explained that there is not much of a role for AI within the MSFD. However, there is huge potential, particularly in light of the large volumes of remote sensing and satellite data. Currently, there is an indicator-based approach, where samples are collected at sea from which ecosystem health and GES is extrapolated. Big data and automated observatories, in combination with in-depth sampling using research vessels, could allow a more holistic approach. They would allow large volumes of information to be gathered to establish patterns and a holistic overview of the environment in relation to changes in species, water column dynamics, currents, and temperature.

[29] https://publications.jrc.ec.europa.eu/repository/handle/JRC120383
[30] https://knowledge4policy.ec.europa.eu/biodiversity_en

Patricia Martin-Cabrera, *Data Manager, Flanders Marine Institute (VLIZ), Belgium; Partner in BlueCloud project*

Question: ***What are the main challenges in using biodiversity data to achieve the objectives of the EU Biodiversity Strategy and how can they be overcome using big data?***

Patricia Martin-Cabrera highlighted that there have been recent advances in Ocean observation technologies, FAIR data, and open science, meaning that marine data are easier to find and are more accessible. Imaging systems are used more often and are producing huge amounts of data. She explained that AI will be key to overcoming the challenges of large amounts of data e.g. by helping to identify and quantify species biomass and biodiversity. She gave the example of phytoplankton as essential drivers of biogeochemical processes and carbon fluxes. Phytoplankton abundance is classified as an Essential Ocean Variable (EOV) and an Essential Climate Variable (ECV) by the Global Ocean Observing System (GOOS). She stated that to protect at least 30% of European seas as part of the EU 2030 Biodiversity Strategy, big data will be important in improving knowledge and indicators by integrating and harmonizing data from different sources and from multidisciplinary fields. European research infrastructures and initiatives, including EMODnet, LifeWatch ERIC[31], and European Open Science Cloud (EOSC)[32] projects like BlueCloud[33], play an important role in making data interoperable and accessible.



Ward Appeltans, *Marine Biodiversity Programme Specialist, Intergovernmental Oceanographic Commission (IOC) of UNESCO, Ocean Biodiversity Information System (OBIS)*

Question: ***How do the objectives of the UN Decade of Ocean Science for Sustainable Development align with the EU 2030 Biodiversity Strategy?***

Ward Appeltans outlined the objectives of the Ocean Decade, i.e. to identify the science we need to build knowledge and increase the use of this knowledge to support a clean, healthy, resilient, productive, safe, and accessible Ocean. He expressed that many of the Ocean Decade challenges relate to the goals of the EU 2030 Biodiversity Strategy to reduce pollution; achieve GES; move to sustainable fisheries; develop a sustainable Ocean economy; understand the Ocean-climate nexus; develop solutions to fight climate change; limit invasive species; improve Ocean knowledge and observing systems; and enable transformative change through a new governance framework. The EU 2030 Biodiversity Strategy also refers to the Biodiversity Beyond National Jurisdiction (BBNJ) agreement, and actions under the Ocean Decade and by the Intergovernmental Oceanographic Commission (IOC) could support the proposed BBNJ treaty[34].

[31] https://www.lifewatch.eu/

[32] https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/european-open-science-cloud-eosc_en

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

[34] https://ioc.unesco.org/publications/non-paper-existing-and-potential-future-services-ioc-unesco-support-future-ilbi

Question: ***Given that the majority of species in the Ocean are not described, how can big data and AI help us to measure biodiversity?***

Sakina-Dorothée Ayata highlighted how plankton metagenomics can help measure biodiversity by measuring all the genes that are present in a drop of water. However, for approximately half of these genes it is not possible to identify its function or the organism they belong to. She explained that one way to access unknown biodiversity is to use Metagenome Assembled Genomes (MAGs), such as those that have been created using the Tara Oceans[35] metagenomics data, and which include both cultivatable and uncultivatable (as yet undescribed) organisms. Machine learning is used to identify the main drivers of MAG distribution, which is one way to better understand the distribution of biodiversity in the Ocean. She also highlighted that MAGs allow access to 'genomic dark matter', which is the part of the genome which we do not currently know the function of.

Question: ***How might big data, AI, and digitalization play a role in globally and regionally integrated data products for marine biodiversity indicators?***

Anne Teller highlighted the importance of ground-truthing data, and explained that this is an area where resources should be spent. She stated that DG Environment (ENV) is working with DG Research and Innovation (RTD) to support this financially. She highlighted the need to merge old and new techniques, and noted that policymakers are willing to take new scientific developments on board to improve monitoring, but scientists need to communicate rapid developments so that policymakers have the latest information. She concluded that to achieve a more integrated understanding of biodiversity we need to integrate models and data from different disciplines. Patricia Martin-Cabrera explained how BlueCloud is federating European marine data infrastructures to create data products that can be used in policymaking. This will provide discovery and access to multidisciplinary data, and the opportunity to perform high computation analyses.

Ward Appeltans contemplated a reality in which internet search engines could immediately answer questions about the marine environment, such as *'When was the last sperm whale sighting in the North Sea?'* or *'What was the average sea temperature in October off the coast of Scotland?'.* This could be possible through linking data systems like the Ocean Biodiversity Information System[36] (OBIS) or EMODnet, using high-quality semantically enhanced resources, and publishing the information as linked open data. Interconnected data nodes would allow us to ask more complex questions that require a combination of information from more than one source. He explained that this is the concept behind the 'Digital Ocean Ecosystem' that will be the foundation of the IOC's ODIS[37] for the Ocean Decade. Each data system can become a node in this distributed system and fulfil their own niche[38]. IOC's IODE is developing an interoperability layer to support ODIS and there is a technical work package to help regional digital stakeholders to share Ocean data globally[39]. He concluded that big data is not about building a bigger local database, but about making sure that AI algorithms can discover and pull data from the internet, and discover new marine indicators.

"Ultimately, asking Google about the state of the marine environment should be as simple as asking Google where the nearest pub is."

- Ward Appeltans

35 https://oceans.taraexpeditions.org/en/m/about-tara/les-expeditions/tara-oceans/

36 https://obis.org/

37 https://www.iode.org/index.php?option=com_content&view=article&id=559&Itemid=100362

38 https://www.youtube.com/watch?v=ZTK2nN2YR_Y

39 https://www.iode.org/index.php?option=com_oe&task=viewEventRecord&eventID=2755

Audience Questions & Answers

Anna Akimova (Thuenen Institute of Sea Fisheries, Germany): *What lessons can we learn from the protection of terrestrial ecosystems, which has had a longer history than the protection of marine ecosystems?*
Anne Teller: Biogeographical seminars, which are held to promote collaboration between scientists, policy-makers, and other stakeholders and to decide on priority areas for protection, may help. This process will be re-vamped in light of the EU 2030 Biodiversity Strategy, and should include marine ecosystems.

Justin Buck (NOC, UK): *Are FAIR and TRUST data principles sufficient to facilitate big data applications?*
Ward Appeltans: The CARE principles are also needed, as they consider respect for traditional knowledge.

Sheila Heymans (EMB): *How can we build trust within research communities to share data and properly acknowledge use of others data?*
Sakina-Dorothée Ayata: Publishing data in an open repository like Pangea[40] means you get a DOI and your data it citable. More scientists are already shifting their mindsets and sharing their data, particularly young scientists.

Nayra Pluma (Spain): *How can we easily find data sets from European data infrastructures?*
Patricia Martin-Cabrera: European data infrastructures have metadata catalogues that help you to find data. BlueCloud is working to federate all of the infrastructures together, and having everything in one space will make data discovery easier.
Ward Appeltans: As one of the first steps in building the ODIS, the IODE is building a catalogue of marine data and information sources (ODIS Cat)[41]. The aim is to build an index that can guide discovery of data sets and information sources to build the ODIS. The data infrastructures will be connected through semantically enhanced metadata, and will allow searching for specific data. Everyone is encouraged to add data sources here to make sure that the system we want to develop is harvesting as many data sources as possible.

Sheila Heymans: *How much human data are being incorporated into EOSC and ODIS?*
Ward Appeltans: We need to focus more on collecting and incorporating data on human uses of the Ocean. EMODnet is already working on this, and this is something that needs to be further addressed within the Ocean Decade.
Anne Teller: Integrated stories that link actions on the land to impact at sea are needed for policy action. For example, plastics legislation was put in place as a result of powerful images showing that the plastic bags people use while shopping end up in the Ocean and cause damage.

[40] https://www.pangaea.de/
[41] https://catalogue.odis.org/search

SESSION 3
Big Data and a Digital Twin Ocean (DTO)

Digital Twin Ocean: Island Digital Ecosystem Avatars (IDEA) for Sustainability

Neil Davies, *Director, Gump Station on Moorea, French Polynesia & Research Affiliate, Berkeley Institute of Data Science, USA*

Neil Davies presented on the concept of Island Digital Ecosystem Avatars (IDEAs) to provide social-ecological foresight through use-oriented computational simulations to support local decision-making towards sustainable development. He began by acknowledging that one of the major scientific challenges to achieving the UN 2030 Agenda is connecting large scale changes to local impacts as our understanding of complex ecosystems is still relatively primitive. He suggests that *"if humanity and the rest of life on Earth is to successfully navigate the next few decades, we will need to develop much greater capacity for social-ecological foresight…and this requires study of the Earth from genome up and planet down".* He highlighted the need to integrate data, models, and understanding across scales and disciplines, which requires coordination of standards, best practices, and knowledge infrastructures worldwide.

Neil Davies and his collaborators have developed a road map[42] for an IDEA in Moorea, French Polynesia, which uses sustainability science to address the complexity of this model social-ecological system. They began by sequencing the biodiversity on the island, and progressively developed understanding of the interactions, moving towards a digital model of the entire social-ecological system. He explained that this also involves understanding history, and simulating different scenarios for the future, and there is a need to connect the public, and particularly local communities, directly with the science to ensure it adds value to society. He continued by considering the potential for AI in advancing the IDEA concept, for example to guide management decisions. However, he highlighted that there are barriers to connecting computational models across scales and disciplines, and in gaining high quality open and FAIR data to train algorithms. He suggested that the social-ecological foresight and tools developed as part of the IDEA concept have the potential to be scaled-up to help move all geographic locations towards greater resilience, and should be the basis of a web of interacting digital Earth and Ocean twins. He highlighted some challenges of building realistic digital twins of social-ecological systems, and explained that it is unrealistic to think that we can make very precise predictions because the systems can be chaotic and the rules of the system are only partially understood. He concluded that avatars of social-ecological units should be combined with new methods of engaging in science-community dialogue and political processes such as citizens' assemblies. IDEAs could be developed for a number of different systems in Europe and some are already underway e.g., the avatar programme in Kosterpark, Sweden.

"One of the main scientific challenges to achieving the UN Agenda 2030 is being able to connect large-scale change to local impact."

- Neil Davies

[42] https://mooreaidea.ethz.ch/storage/pdfs/gigascience.2016.03.pdf

Credit : Russell J. Schmitt, Moorea Coral Reef LTER & IDEA Consortium.

A lagoon of Moorea, French Polynesia.

Big Data and a Digital Twin Ocean (DTO)

Panel Discussion

The panelists (Fabienne Jacq, Marilaure Gregoire, Serge Scory, and Kate Larkin) were asked a series of questions by the moderator (Sheila Heymans), followed by a Q&A session with the audience.

Fabienne Jacq, *Policy Officer, DG Defense, Industry, and Space (DEFIS), European Commission*

Question: ***What questions could we answer with a DTO that we cannot answer now?***

Fabienne Jacq highlighted that the European Commission is working on developing the basis for a DTO as part of the Copernicus Marine Environmental Monitoring Service (CMEMS) and Copernicus Climate Change Service (C3S). These services provide forecasting, and will be made available through the Destination Earth and DTO initiatives. Destination Earth aims to increase the use of AI, improve data interoperability, and link different scientific disciplines. She referred to the Moorea IDEA presented by Neil Davies as an interesting example that could be implemented as part of Destination Earth. A DTO will help to scale-up the downstream users of Copernicus and offer new opportunities for collaborative work. It will go beyond science to address what cannot be addressed by Copernicus, applying global thinking to high-resolution local solutions for coasts, beaches, islands, rivers, and land. It will also enable Ocean dimensions for which we currently have less data, such as marine biodiversity, to be more easily integrated into decision-making.

Marilaure Grégoire, *Research Director, University of Liège, Belgium & Member, Copernicus Marine Scientific and Technical Advisory Committee*

Marilaure Grégoire described how a DTO could answer many questions that link the changing Ocean to humans and could provide solutions to urgent problems. A DTO will be part of Destination Earth, and this will allow an Earth systems approach which will allow us to tackle problems that require understanding of the links between land, rivers, atmosphere, and the Ocean. The priority for CMEMS is to make these connections. She added that a DTO will also help to make decisions to not add extra human activities in highly impacted marine areas due to forecasted risk. Industries such as aquaculture, could use information from a DTO to anticipate the risk of hypoxia, harmful algal blooms, or jellyfish outbreaks so they can adapt in a timely manner. A DTO could also allow integration of our observing systems and models so that they can adapt their sampling strategy to the location of the most serious problem to solve. She concluded that a DTO could be used to map ecosystem services in high-resolution. This is particularly needed for biodiversity, and its links to human activities, as well as for climate regulation services, such as carbon dioxide sequestration by the Ocean. A multidisciplinary DTO needs to be ready to address unknown problems that will arise in the future.

Serge Scory, *Chair, SeaDataNet & Head, Belgian Marine Data Centre*

Serge Scory emphasized that a DTO should have a positive feedback loop to help inform what data still need to be collected so that data collection strategies can be optimized to increase the amount of pertinent information. Data collection is very resource-intensive and currently we often don't know whether we are collecting the right data, in the right place, at the right time, and in an optimal way. He added that a DTO would allow us to conduct large-scale experiments that are not possible in the field, and to perform detailed sensitivity analyses.

Kate Larkin, *Deputy Head, EMODnet Secretariat*

Kate Larkin highlighted a DTO as a good opportunity to do more complex and transformative science that we do not currently have the full capability to implement in terms of bringing together data, models, and simulations. There is already a good pipeline of marine data in Europe with data services and aggregators like EMODnet and Copernicus, which are single entry points to large volumes of *in situ* and satellite data. These services need to be taken forward in a DTO that connects marine data, human activities data, and social-economic data into a common workspace to enable science based on big data. She added that a DTO could provide continuous real-time monitoring that could engage citizens and provide operational tools and decision-making support for the blue economy that could reduce operating costs or save lives.

Question: ***What are the main challenges in developing the DTO and how can big data help them to be overcome?***

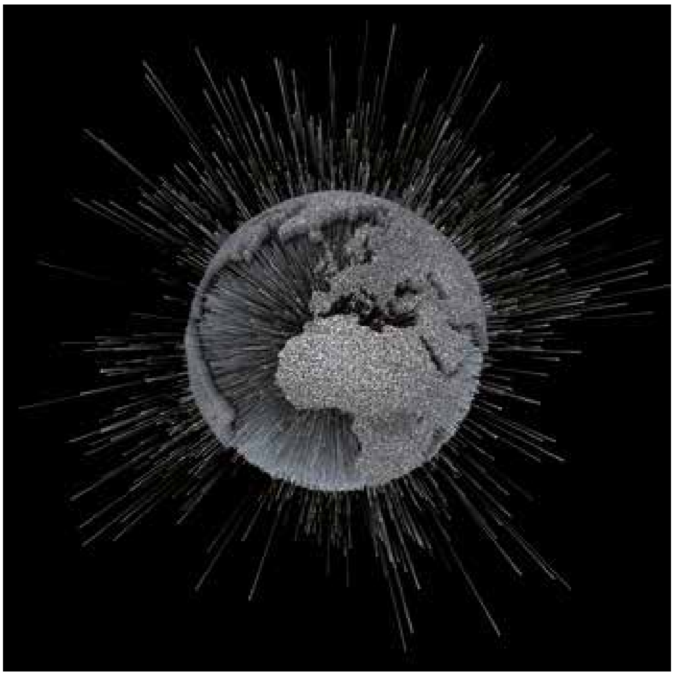
Fabienne Jacq emphasized that big data are a challenge in themselves, as well as a means to help solve challenges. The heterogeneity of marine data is a strength and a challenge. There is huge potential to learn from using technologies such as AI, from which information can be extracted such as signals that we cannot detect with numerical models. However, there is a lot of work to do before we truly have operational big data that are machine readable and ready to be processed by algorithms. In her view, the main challenge in achieving a DTO is non-technical, and relates to the complexity of implementing and deploying such an initiative in a short time. A lot of core science remains to be done, and advances in digital computing and systems integration are needed, as well as developing trust with users. The challenge is to combine both a top down and bottom up approach, building on precursors such as EMODnet, Copernicus, BlueCloud, and AtlantOS[43], etc. In parallel, a long-term vision and roadmap must be built, and champion users that are ready to change their behavior need to be identified.

[43] https://www.atlantos-h2020.eu/

Marilaure Grégoire mentioned that there are both technical challenges as well as those of connecting a DTO to the people, and developing trust with a diversity of users from private and public sectors and society. Access to high-performance computing will be needed to run AI algorithms on high-resolution data, and numerical codes will need to be adapted to new, low energy hardware that we will have in the future, such as the Graphics Processing Unit (GPU). Handling large volumes of heterogeneous data is also a challenge, and data have to be made interoperable and compliant with the FAIR principles. She reiterated the need to ensure data quality, as well as the quality of the information that we get from data.

Serge Scory expressed that one of the main challenges is the realistic modelling of the many complex interactions between physical, chemical, biological, and social-economical aspects of the Ocean. Scientists need to develop new ways of working by enhanced integration of disciplines. Big data can help overcome these challenges by using AI and data mining to identify patterns and enable more accurate modelling. He suggested that AI may also be used to facilitate and drastically improve quality control of data. He also highlighted that there are challenges remaining to be overcome for developing big data. Not all of the marine data networks are ready to handle the huge amount of data that are expected to come in the near future. SeaDataNet[44] and EMODnet need to support efforts like those of the ENVRI-FAIR[45] project to improve the FAIRness of data that are made available by the environmental research infrastructures.

Kate Larkin added that truly FAIR data, which is the benchmark for data interoperability, re-usability, and data provenance, is a challenge. Simulations and models are only as good as the data that are fed into them, and FAIR data are essential to reducing the uncertainties of predictions and for producing truly big data that are heterogeneous through interoperability with other digital twins from different domains. Data interoperability is important for tackling issues at the land-sea interface, in coastal areas, and at the Ocean-climate nexus, as well as for the international data ecosystem where data can be used to address international challenges. She emphasized that the scale-up towards automated data collection, processing, and analysis needs to be done in a step-wise, structured process to ensure data quality so that big data are fit-for-purpose.



Credit: Adam Martinakis

Artwork included in Linwood Pendleton's keynote speech.

[44] https://www.seadatanet.org/

[45] https://envri.eu/home-envri-fair/

Audience Questions & Answers

Dominique Durand (Norwegian Research Centre, Norway): *How can we make the Ocean community evolve and open up for new partnerships in order to be more creative and more efficient in addressing the challenge of building a DTO in partnership with industry, public authorities, and citizens?*
Kate Larkin: We can do this by developing dialogues with industry and citizens to show the win-win benefits of data sharing. For example, demonstrating to the aquaculture industry that sharing baseline environmental data around farms will enable the development of forecasting that benefits all farms in one area. Connections with computing and data science communities are also important.
Serge Scory: One way to improve cross-domain fertilization is to develop specific funding instruments that promote this. This is already the case in many Horizon 2020 projects.
Marilaure Grégoire: CMEMS has a regular call for service evolution and the development of downstream applications, including developing new tools like AI to handle big data.
Fabienne Jacq: CMEMS is becoming more open. The physics community now has links with the biogeochemistry community and together they are creating integrated models. They are now also working with the ecology community on marine ecosystem models. DTO and Destination Earth are open to social science and citizen science, which will force the Ocean community to become more open. Education is also a means to do this and the next generation of students may be less conservative, and can bring in new ideas for big data and AI.

Slawomir Sagan (The Institute of Oceanology of the Polish Academy of Sciences, Poland): *Many Ocean processes and their inter-linkages are still unknown, or described with high levels of uncertainty. Is it too early to create a DTO?*
Marilaure Grégoire: Now is a very good time to work on a DTO due to advances in Ocean modelling, observations, and data collection that are expected within the next decade. AI will allow us to better parameterize and constrain Ocean models.
Kate Larkin: A DTO would not be able to achieve everything immediately, but is a good focal point to empower a wider community to share data, and to enable access to more diverse data with higher resolution will ultimately reduce uncertainties. A wider variety of integrated data sets will also allow us to more easily identify gaps that can then inform future data collection efforts.

Anna Hermsen (TNO, the Netherlands): *How can we enable a paradigm shift to integrate science with different cultural disciplines such as art, music, humanities, philosophy etc. that will ultimately lead to innovation?*
Fabienne Jacq: The funding calls planned within ‘Mission Healthy Oceans, Seas, Coastal and Inland Waters’, and those already released for Destination Earth and the DTO have identified this as a goal, and we need alignment between funding sources and all stakeholders to make this happen.
Kate Larkin: A DTO could start with visualizing the big data sets that already exist in near real-time to help society see their link to the Ocean.
Sheila Heymans: This will also be important for linking Ocean models within a DTO to cultural ecosystem services.

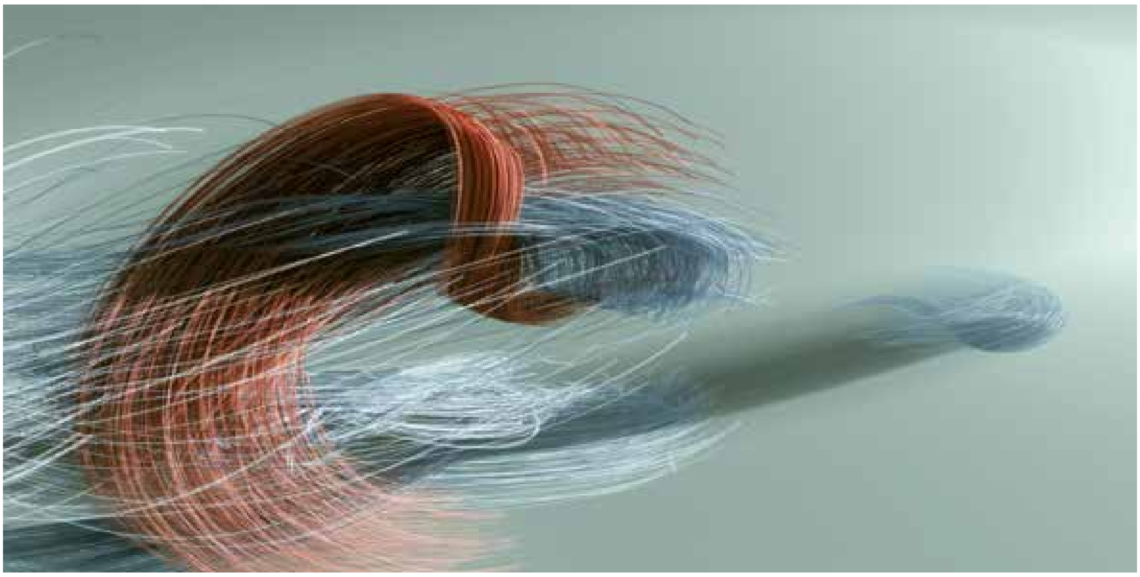
Closing Session

Ricardo Serrão Santos, *Portuguese Minister of the Sea*

Minister Santos provided perspectives on the day's discussions. He began by reminding the audience that science is fundamental to support good governance, and that this is particularly true in light of climate change. Oceanographers predicted global warming and the climate emergency in the mid-20th century, and this is not a question of believing or not, nor optimism versus pessimism. Ocean data comes from observations and monitoring from coastal zones to the deep sea. These allow us to clearly link the Ocean with climate change through understanding the Ocean as a stabilizing force of biogeophysical processes, and particularly in the regulation of the carbon cycle. He recalled the laborious work of Charles Keeling who collected time series on carbon dioxide in the atmosphere in Hawaii, which together with the work of Roger Revelle in the 1950s, has allowed us to understand the evolution of the state of the Ocean and the problems it is facing. These include acidification, deoxygenation, increasing temperature, changing currents, and coastal processes that impact fisheries, as well as plastics and other pollutants. He highlighted that more data does not necessarily mean more information and knowledge to face these problems. There are a large number of marine data sets that demand a large capacity for data processing and interoperability. If terrestrial, social, and economic data are also integrated the challenge is even bigger. The MSFD, the European Maritime and Fisheries Fund (EMFF), the Marine Spatial Planning Directive, the European Green Deal and the EU 2030 Biodiversity Strategy all aim to address the UN 2030 Agenda and its Sustainable Development Goals (SDGs) at the European level. These instruments call for very well-integrated, accessible, and up-to-date information on the marine environment. Europe has been doing a lot with EMODnet, CMEMS, and support by the European research infrastructures and major research and development projects to deploy Ocean observations and collect marine data. The DTO call is ambitious and is aligned with 'Mission Healthy Oceans, Seas, Coastal and Inland Waters' and has clear linkage with the UN Ocean Decade societal goal of a transparent and accessible Ocean with open access to data, information, and technology. He concluded that new knowledge and technologies alone cannot solve major threats to the Ocean, and that governance and new policies at regional and national levels, aligned with global and shared concerns are paramount. Rational and emotional linkages to the seas and Ocean are also much needed, and storytelling is an effective tool. Citizens informed about the challenges that the Ocean faces will allow for and promote social acceptance and social authorization of political measures based on science. He ended by emphasizing the need to reach the people and convey the most fundamental and powerful message expressed by Sylvia Earle "You should treat the Ocean as if your life depends on it, because it does".

Sigi Gruber, *Head, Healthy Oceans & Seas Unit, DG Research and Innovation (RTD) European Commission*

Sigi Gruber provided the final remarks and emphasized the key message that we need healthy waters and a healthy Ocean in the future to sustain life. Biodiversity has collapsed already and sea level is rising. She highlighted the ambition of the European Green Deal to make Europe carbon neutral by 2050 as the roadmap for action to achieve necessary environmental, social, and economic transformations, and that this is now the compass for recovery from the COVID-19 crisis. She emphasized the need for more data and more science to light the way through the 21st century. There is a need to instill confidence that data is bringing solutions back to the citizens through storytelling and 'Mission Healthy Oceans, Seas, Coastal and Inland Waters' has prioritized a systems approach that is need if we really want to make a difference. Data will be at the heart of everything. She drew attention to another important initiative: the 'European Partnership for a Climate Neutral, Sustainable and Productive Blue Economy', which will bring together all of the European Sea Basin strategies, where data collection will be one of the key pillars. As part of Horizon Europe, there are opportunities for new calls where the FAIR principles will have to be guiding. All these initiatives will underpin the UN Decade of Ocean Science for Sustainability and the UN Decade for Ecosystem Restoration. She concluded that Horizon Europe has an emphasis on co-creation, which should be implemented to involve citizens and bring them the solutions they need.



Credit: Adam Martinakis

Artwork included in Linwood Pendleton's keynote speech

ABBREVIATIONS

AI	Artificial Intelligence
AUV	Autonomous Underwater Vehicle
BBNJ	Biodiversity Beyond National Jurisdiction
CARE	Collective benefit, Authority to control, Responsibility and Ethics
CMEMS	Copernicus Marine Environmental Monitoring Service
COVID-19	Coronavirus disease of 2019
C3S	Copernicus Climate Change Service
DG ENV	Directorate General Environment
DG MARE	Directorate General for Maritime Affairs and Fisheries
DG RTD	Directorate General Research and Innnovation
DOI	Digital Object Identifier
DTO	Digital Twin Ocean
ECV	Essential Climate Variable
EC	European Commission
EMB	European Marine Board
EMODnet	European Marine Observation and Data Network
EMFF	European Maritime and Fisheries Fund
EOSC	European Open Science Cloud
EOV	Essential Ocean Variable
ERIC	European Research Infrastructure Consortium
EU	European Union
EwE	Ecopath with Ecosim
FAIR	Findable, Accessible, Interoperable, Reusable
GES	Good Environmental Status
GIS	Geographic Information System
GOOS	Global Ocean Observing System
GPU	Graphics Processing Unit
IDEA	Island Digital Ecosystem Avatar
IFREMER	The French Research Institute for the Exploitation of the Sea
INSPIRE	Infrastructure for Spatial Information in the European Community
IODE	International Oceanographic Data and Information Exchange
JRC	Joint Research Centre
MAGs	Metagenome Assembled Genomes
MPAs	Marine Protected Areas
MSFD	Marine Strategy Framework Directive
MSP	Marine Spatial Planning
NGO	Non-governmental Organization
NOAA	National Oceanographic and Atmospheric Administration, USA
OBIS	Ocean Biodiversity Information System
ODIS	Ocean Data and Information System
ODIS Cat	Ocean Data and Information System Catalogue
PAIR	People and AI Research
SDGs	Sustainable Development Goals
TRUST	Transparency, Responsibility, User focus, Sustainability, and Technology

European
MARINE BOARD
Advancing Seas & Ocean Science

European Marine Board IVZW
Wandelaarkaai 7 I 8400 Ostend I Belgium I Tel.: +32 (0)59 34 01 63 I E-mail: info@marineboard.eu I Web: www.marineboard.eu I Twitter: @EMarineBoard

Belgian register of legal entities: 0650.608.890