The DTO-Bioflow Project: Bringing Biodiversity Data into the European Digital Twins of the Ocean

Klaas Deneudt[1] (klaas.deneudt@vliz.be), Laura Slaughter[2] (laura.slaughter@sintef.no), Arne Berre[2] (Arne.J.Berre@sintef.no), Syrine Ben Aziza[3] (syrine.benaziza@tno.nl), Carlota Muniz[1] (carlota.muniz@vliz.be), Tomas Martinovic[4] (tomas.martinovic@vsb.cz), Ute Brönner[2] (Ute.Broenner@sintef.no), Alain Arnaud[5] (aarnaud@mercator-ocean.fr), Jesse Harrison[6] (jesse.harrison@csc.fi), Francisco Hernandez[1] (francisco.hernandez@vliz.be)

[1]VLIZ, Flanders Marine Institute (Belgium)
[2]SINTEF (Norway)
[3]TNO (The Netherlands)
[4]IT4Innovations (Czech Republic)
[5]Mercator Ocean (France)
[6]CSC (Finland)

The Digital Twin of the Ocean (DTO)-BioFlow Project will transform access to ocean biodiversity data by enabling the sustainable integration of biodiversity monitoring data flows into the Digital Twin of the Ocean (EU DTO)[47]. The project unlocks the societal value of these data, thereby enriching digital tools and services in cooperation with the European Marine Observation and Data Network (EMOD)

EMODnet Project[48] and various DTO projects, including the European Digital Twin of the Ocean (EDITO)[49], EDITO Model[50], and the Iliad consortium[51]. To fully exploit the potential of these new and existing data flows, DTO-BioFlow will develop the biodiversity component of the DTO. The DTO fulfils the objectives of the EU Biodiversity Mission and Strategy by bringing together data, models, and new algorithms to support the development of policy-relevant tools and services for effective monitoring, restoration, and protection of marine biodiversity.

Figure 1, read left-to-right, shows how DTO-BioFlow provides support for the various steps in a digital twin pipeline. These steps move from data acquisition to data curation and representation, to models, and finally analysis. The twins support stakeholder requirements allowing the user to examine prediction-type queries to the twin, conduct what-if analyses, and interact with various types of visualisations. This will be developed through demonstrator use cases for the following seven areas: (1) invasive species management; (2) adaptive offshore construction and energy harvesting; (3) assessing pelagic biodiversity in relation to human impact; (4) spatial planning of sustainable mariculture; (5) ecosystem-based marine spatial planning and Marine Protected Areas (MPA) management; (6) low impact fisheries; and (7) biodiversity ecosystem services including carbon sequestration.

The key challenge is bringing new biodiversity data streams into the European DTO through a suitable interoperability architecture. New technology and data types (imaging, acoustics, DNA based, satellite) produce huge amounts of readings that require processing, aggregation, and translation into usable data products such as species occurrences, densities, biovolumes, species migration routes, etc. For many of these data types this hampers the flow of data from the biodiversity sensor networks to the biodiversity data integrators used to deal with more

[47] https://digitaltwinocean.mercator-ocean.eu
[48] https://emodnet.ec.europa.eu/en
[49] https://edito-infra.eu
[50] https://edito-modellab.eu
[51] https://ocean-twin.eu

traditional types of observation data. Sustainable ingestion procedures for these data towards the DTO will be provided, with the use of EMODnet Biology[52] as a primary focus. EMODnet is the portal that provides open and free access to interoperable data and data products on temporal and spatial distribution of marine species (angiosperms, benthos, birds, fish, macroalgae, mammals, reptiles, phyto- and zooplankton) from European regional seas. The infrastructure and data flow used within EMODnet Biology is based on the European Ocean Biogeographic Information System[53] (EurOBIS/OBIS), the Global Biodiversity Information Facility[54] (GBIF), and their underlying biodiversity ontologies. Other aggregators will also be considered when appropriate. The collected data will further be transformed to appropriate digital representations suitable for the data lakes of relevant digital twins with a transformation to analysis-ready data. Considered data formats for the digital twin models are Parquet/GeoParquet[55], SpatioTemporal Asset Catalogs[56] (STAC), and Zarr[57] drive, in particular for the analysis-ready data.

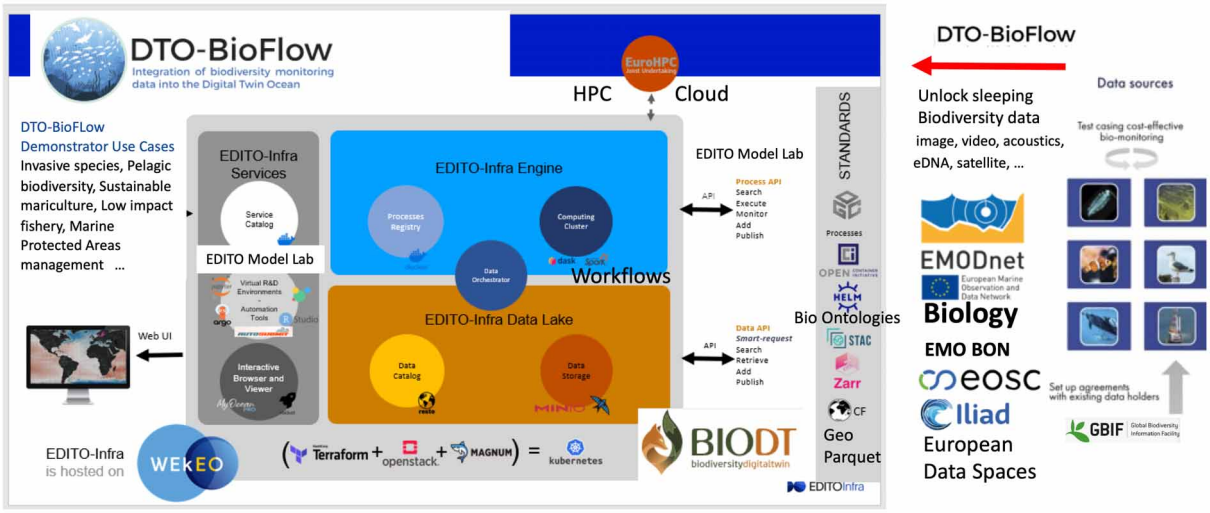

Figure 1 DTO-BioFlow Biodiversity data into Digital Twins of the Ocean.

The digital twin processing will be supported through the EDITOInfra data lake and the EDITO digital twin engine. Further support for High Performance Computing (HPC) model processing for digital twins is carried over from the Biodiversity Digital Twin (BioDT) Project[58], a land-based digital twin. Partners who are jointly involved in BioDT as well as DTO-BioFlow are responsible for transferring HPC-processing services into the marine biodiversity digital twins. This reuses existing HPC infrastructure services such as the LEXIS Project[59] combined with AI/Machine learning Graphics Processing Unit (GPU) clusters, workflow engines, and languages such as the common workflow language (CWL). Both projects take advantage of processing and service

[52] https://emodnet.ec.europa.eu/en/biology
[53] https://www.eurobis.org
[54] https://www.gbif.org
[55] https://geoparquet.org
[56] https://stacspec.org/en
[57] https://zarr.readthedocs.io/
[58] https://biodt.eu
[59] http://lexis-project.eu

support through the European Open Science Cloud (EOSC)[60]. Biodiversity digital twins will also be able to exploit the EDITO ModelLab services for various ocean models and relevant Virtual Research Environments like Jupyter Notebook Hubs[61] and the Galaxy platform[62].
The approach will further be harmonised with the interoperability architectures of supporting initiatives like the Iliad consortium, the United Nations Ocean Decade Initiative[63], DITTO (Digital Twins of the Ocean)[64] with the Turtle Interoperability Framework (TIF)[65]. The basis of the levels of legal, organisational, semantic, and technical interoperability of the European Interoperability Framework (EIF) will be followed.

Acknowledgements
We acknowledge the efforts of the entire DTO-Bioflow "Integration in DTO Infrastructure" team: Clément Villeviere, Elena Lazovik, and Jeroen Broekhuijsen from TNO (The Netherlands); Frederic Leclercq, and Bart Vanhoorne from VLIZ (Belgium); as well as Tuomas Rossi and Jarmo Makela from CSC-IT Center for Science LTD (Finland).

[60] https://eosc-portal.eu
[61] https://jupyter.org/hub
[62] https://usegalaxy.org
[63] https://oceandecade.org
[64] https://ditto-oceandecade.org
[65] https://doi.org/10.5194/egusphere-egu23-15620