

# MarDATA @ OSCM 2024

### Project Week "Digital Twin Cabo Verde"

<u>Arrival OSCM (Cabo Verde):</u>	10.03.2024
<u> Departure OSCM (Cabo Verde):</u>	19.03.2024

#### Summary:

MarDATA doctoral researchers and supervisors visit the Ocean Science Center Mindelo (OSCM). Together with the Master Students of the WASCAL Master Program "CLIMATE CHANGE AND MARINE SCIENCES", data-related research questions will be addressed during a project week on the island.

The project week was designed to allow MarDATA doctoral researchers to apply their data science skills and expertise outside their own scientific questions. At the same time, the close involvement of WASCAL master's students should enable them to gain new insights into data-driven marine science and analysis methods.

The aim of the challenges dealt with during the week is divided into two parts. Firstly, they should give the participants the opportunity to apply the marine data science expertise they have acquired in real-life use cases. In addition, the week will serve to decisively support the efforts to develop a "Cabo Verde" digital twin.

Accordingly, the focus of the research tasks (aka challenges) is on data processing, analysis or/and visualization. A total of three challenges were defined in advance. These cover (1) user-focused, (2) science-focused and (3) conceptual aspects of a digital twin "Cabo Verde" and are described in more detail below.

Since WASCAL master's students are likely to have only a basic knowledge (if any) of data science methodologies, we will be running the project week with accompanying training. MarDATA doctoral researchers and supervisors will provide lectures and tutorials for WASCAL students, both prior to the start of the hackathon and accompanying it.

In addition, an "Introduction to Python" course is offered as part of the WASCAL master's program, which covers most of the methods required during the project week. This course will be given by staff of the Data Science Unit (GEOMAR) in February 2024.

#### Teams & preliminary schedule:

Depending on the number of participants, about 4 to 5 participants will form a team. There should be at least 2 WASCAL students per team.

	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday
	09.03.24	10.03.24	11.03.24	12.03.24	13.03.24	14.03.24	15.03.24	16.03.24	17.03.24	18.03.24	19.03.24	20.03
08:00			Welcome	MarDATAs train WASCALs	Lecture	Lecture	Lecture		Freizeit	Presentation of results		
10:00			Tour OSCM		Hackathon Hackathon		n Hackathon	Social (Field-			Abreise	
11:00												
12:00		Anreise				Hackathon						
13:00												
14:00												
15:00								, is an a comp				
16:00	Anreise		Challenge Intros	Hackathon		ENT	WUR	F				
17:00												
18:00												
19:00	Welcome So	Welcome Social							Farewell-Social			
20:00												
21:00												
22:00												
23:00												
00-00												
_ 0.00												

#### <u>Challenges:</u>

## Challenge 1 - "User Applications: Fishermen's App"

<u>Keywords</u>: citizen/community science, smartphone app, socio-economic data, AI-based image annotation, near-realtime data

Mentors (on site): Björn Fiedler, Ivanice Monteiro (OSCM), N.N. (UTA)

available data/materials:

- incoming app data reports (approx. 700 reports so far, numerical, \*.csv)
- incoming pictures (\*.png)
- App source code (R)

#### Scope:

The goal of this challenge is to practice the sifting, cleaning, merging and visualization of already collected and frequently incoming marine data and to enable the further use of the data currently collected in near-realtime with the help of a smartphone app used by artisanal fishermen. Data include socio-economic parameters (e.g., costs for fishing effort), parameters on catches (e.g., weight) as well as biodiversity data (pictures of species).

Solutions ideally also include a roadmap for optimized integration of new/future data and options to provide immediate information to stakeholders. Further development of the app is not an explicit goal, but would be technically possible.

Potential fields of activity during the project week:

- Transforming fish catch images into information (AI / Annotation / Classification)
  - Could ease the information collection for fishers
  - Could improve the quality of the provided annotations
  - Could result in recommendations to collect "better" images
- Software-Engineering Designing and prototyping an architecture for actual use of the data (Database, Backend / API, Frontend(s))
  - Feedback portal for fishermen (e.g., stats on expenses spent on fishing, catches per week/month/year)
  - Develop a pilot marketplace to inform buyers of available catch (like this: https://www.fischvomkutter.de/moeltenort.html or https://abalobi.org/)
  - Link to challenge 2

# Challenge 2 - "Scientific Applications: Marine Data Fusion"

Keywords: data fusion, FAIR data, prediction, imputation, uncertainty, visualization

Mentors (on site): Willi Rath, Nuno Vieira (?), Lucileida Ramos (UTA) (?). Matthias Schaber (?)

<u>available data:</u>

- hydroacoustic biomass data (XY campaigns)
- bathymetry data
- CV00 time series data (biogeochemical data (ship), currents/temp/sal/o2 (mooring))
- autonomous surface vehicle (Saildrone/Wave Glider) data: temp/sal/bgc
- gridded atmospheric, hydrographic and biogeochemical <u>Copernicus Marine Data</u> products
- gridded high-resolution ocean-circulation data VIKING20X

#### Scope:

The goal of this challenge is to practice the integration of existing marine data sets (in-situ point data, in-situ time-series, gridded fields, along-track data) that allows for joint analysis. The study region around the Cape Verde Islands serves as a specific example used in the hackathon but should not limit the scope of solutions to only those applicable in the tropical oceans. Solution can include visualization with existing infrastructure (e.g. Digital Earth Viewer, BELUGA), as well as further analysis using, e.g., machine learning methods. Particular attention should be paid to linking existing biomass data, hydroacoustic and (simulated) ocean circulation data. It would also be desirable to develop a blueprint for the integration of new/further data.

Potential fields of activity during the project week:

- Integration of as many as possible of the <u>Copernicus Marine Data products</u> for the Cape-Verde region as context for local in-situ observations
- ML/AI-driven Biomass prediction based on location, hydrographic parameters (temperature / salinity) and bio-geo-chemical parameters (oxygen, nutrients, ...)
- ML/AI-driven imputation of data
- Uncertainty estimation based on data from different sources

## Challenge 3 - "Conceptual Framework: Digital Twin Cabo Verde"

<u>Keywords:</u> Stakeholder engagement, Digital Twin Framework, Narrating the integrative Cabo Verde data story, Establishing operational feedback loops

Mentors (on site): Timm Schoening, DSU GEOMAR, Estanislau Lima (UTA)

#### Scope:

All digital twins need observation data, interactive models and simulation data, dynamic visualization interfaces as well as feedback mechanisms to transfer knowledge generated to stakeholders to put into action for the benefit of society and nature.

The goal of this challenge is to design an overarching framework for the implementation of future digital twins of Cabo Verde. These sustainably built twins shall be able to access all required data to address specific what-if questions around the archipelago. Infrastructure and data workflow requirements are important here. Yet, the main focus shall lie on establishing an effective feedback loop among data interpreters and decision makers.

Potential fields of activity during the project week:

- Infrastructure requirements
  - Which key infrastructures are required?
  - Which infrastructures exist locally?
  - Which infrastructures from elsewhere are available?
  - Which infrastructures need to be implemented?
  - Who has access to which infrastructures?
- Requirements for sustainable monitoring activities around Cabo Verde
  - Enabling uncertainty quantification of existing data sources
  - Providing the means to create necessary data (where to sample next? What to sample next?)
- Establishing a narrative, pitching the Digital Twin vision to regional stakeholders:
  - Collection of what-if questions to be addressed by DTs
  - Collection of scientific, societal and industrial gains
  - Highlighting the DT potential of informing the necessary stakeholders about risks, uncertainties and potential actions
- Facilitating an operational feedback loop with society, industry and government:
  - Who are the local stakeholders?
  - What are their roles (legislation, oversight, monitoring, data provisioning)?
  - How do they interact?
  - Do they have means to operate a twin or ingest twin results?
  - Are stakeholders missing (agencies, NGOs, interest groups)?
- Your ideas of actions towards a Cabo Verde Digital Twin framework:
  - Go beyond the bullets above
  - Think in interactions
  - Embrace the effectivity and efficiency of DTs for the benefit of the Cabo Verde archipelago and its people