



## Digital Twin for Innovation in Oceanic Visualization and Exploration (DIVE)

### DATA MANAGEMENT AND INPUT DATA ACQUISITION

|             |                                                                                                                                |                           |
|-------------|--------------------------------------------------------------------------------------------------------------------------------|---------------------------|
| Prepared By | <br>_____<br>Nicolae Catalin Ristea (UNSTPB) | _____<br>Date: 20.11.2025 |
| Reviewed By | <br>_____<br>Cosmin Dănișor (UNSTPB)        | _____<br>Date: 30.11.2025 |
| Approved By | <br>_____<br>Mihai Datcu (UNSTPB)           | _____<br>Date: 30.11.2025 |

## AUTHORS

| No. | Name                   | Role in the project                    |
|-----|------------------------|----------------------------------------|
| 1.  | Nicolae Catalin Ristea | CS III, Machine learning researcher    |
| 2.  | Cosmin Dănișor         | CS III, EO Image processing researcher |
| 3.  | Corin Văduva           | CS II, EO Data analytics               |
| 4.  | Andrei Anghel          | CS I, EO Signal processing researcher  |
| 5.  | Răzvan Mateescu        | CS I, Coastal engineering              |
| 6.  | Mihai Datcu            | CS I, Principal Investigator           |

## DOCUMENT CHANGE RECORD

| Issue | Date       | Chapter | Change              |
|-------|------------|---------|---------------------|
| 1.1   | 2025-10-30 | All     | Document created    |
| 1.2   | 2025-12-05 | All     | Version 1 completed |

---

## TABLE OF CONTENTS

|    |                                 |   |
|----|---------------------------------|---|
| 1. | Introduction in Data Cube ..... | 4 |
|----|---------------------------------|---|

## 1. Introduction in Data Management and Input Data Acquisition

In this chapter, we focus on establishing a reliable and transparent framework for acquiring, managing, and distributing all input data required by the project. Treating input data as immutable assets, this task develops an open-source, cross-platform acquisition pipeline capable of handling both static datasets and live API streams. All datasets provided until now by stakeholders and WP3 experts will be documented in a lightweight DCAT-based metadata catalog, which records essential descriptors such as sensor, provider and storage location. Guided by this catalog, the pipeline will retrieve, fingerprint, and distribute each dataset to cloud or compute storage, ensuring traceability and reproducibility.

| Data name                               | Sensor                                                                                        | Provider                                                                        | Storage location                                                       |
|-----------------------------------------|-----------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|------------------------------------------------------------------------|
| Sentinel-1                              | C-band Synthetic Aperture Radar (SAR)                                                         | ESA / Copernicus                                                                | Copernicus Open Access Hub or DIAS platforms (e.g., WEKEO, CREODIAS)   |
| Sentinel-2                              | Multispectral optical imagery for land & coastal monitoring (13 spectral bands)               | ESA / Copernicus                                                                | Copernicus Open Access Hub, WEKEO, CREODIAS                            |
| Sentinel-3<br>OLCI                      | Ocean and Land Colour Instrument – ocean colour, chlorophyll, suspended matter, water quality | ESA / EUMETSAT (Copernicus)                                                     | EUMETSAT CODA, Copernicus Open Access Hub, WEKEO                       |
| AROME ECMWF                             | AROME high-resolution NWP model coupled with ECMWF boundary conditions                        | Météo-France and ECMWF                                                          | Météo-France data servers / ECMWF MARS or WEKEO (depending on license) |
| CMEMS                                   | Multi-sensor ocean observations + ocean circulation models                                    | Copernicus Marine Environment Monitoring Service                                | CMEMS data portal                                                      |
| ERA5                                    | ECMWF global atmospheric reanalysis system                                                    | ECMWF / Copernicus Climate Change Service (C3S)                                 | CDS (Climate Data Store)                                               |
| MET Norway<br>Norkyst800                | High-resolution coastal ocean model for Norwegian waters (800 m resolution)                   | MET Norway                                                                      | MET Norway Thredds server / Norwegian Meteorological Institute API     |
| TOPAZ                                   | TOPAZ4 Arctic ocean data assimilation system                                                  | Copernicus Marine (ARC MFC) / NERSC                                             | CMEMS portal (ARC MFC section)                                         |
| Oceanic AROME<br>weather<br>predictions | High-resolution AROME atmospheric model for ocean/coastal areas                               | Météo-France                                                                    | Météo-France servers (restricted access) / project-specific storage    |
| Individual Radar<br>Reflectivity Data   | Weather radar reflectivity sensors (C-band or X-band radars)                                  | National meteorological agencies (e.g., NOAA, DWD, Météo-France, ROMATSA, etc.) | National radar archives or project-specific data repository            |

Below, we provide an overview of the primary environmental and geophysical datasets that underpin the analytical and modelling activities within the project. Given the heterogeneity of the input sources, it is essential to characterize their origin, sensing principles, and thematic content. The following sections

therefore present brief descriptions of each dataset, highlighting their relevance, data characteristics, and role within the integrated data processing and modelling framework.

**Sentinel-1** consists of a constellation of C-band Synthetic Aperture Radar (SAR) satellites providing high-resolution, all-weather, day-and-night imaging of land and ocean surfaces. It supports applications such as sea-ice monitoring, surface roughness estimation, oil spill detection, and coastal dynamics.

**Sentinel-2** is a multispectral optical imaging mission within the Copernicus Programme designed for detailed observations of land and coastal environments. It delivers high-resolution imagery across 13 spectral bands, enabling the monitoring of water quality, turbidity, vegetation, shoreline changes, and suspended particulate matter. With frequent revisit times and radiometrically stable data, Sentinel-2 supports environmental mapping, change detection, and biogeophysical parameter retrieval essential for coastal and nearshore applications.

**Sentinel-3 OLCI** (Ocean and Land Colour Instrument) is a medium-resolution imaging spectrometer optimized for marine and oceanographic monitoring. It provides high-quality measurements of ocean colour, chlorophyll concentration, phytoplankton dynamics, suspended matter, and other water-quality indicators. Its consistent spectral calibration and wide swath allow for operational monitoring of ocean productivity, algal blooms, and ecological state assessments in both open-ocean and coastal regions. Sentinel-3 also integrates with complementary instruments measuring sea-surface temperature and topography for comprehensive ocean analysis.

**AROME**, a convection-permitting numerical weather prediction model operated by Météo-France, is often run with boundary conditions from ECMWF's global forecast system. It delivers fine-scale atmospheric fields (e.g., wind, precipitation, temperature) essential for regional and coastal forecasting.

**CMEMS** delivers operational oceanographic products derived from satellite observations, in-situ measurements, and numerical ocean models. It provides global and regional fields of currents, sea surface height, sea temperature, salinity, sea ice, and biogeochemical indicators used for ocean monitoring and forecasting.

**ERA5** is a global atmospheric reanalysis product produced by ECMWF, integrating observations from satellites, ground stations, and atmospheric models. It offers hourly, gridded fields of atmospheric, land, and ocean variables from 1950 onwards, ensuring temporal consistency for climate and environmental analyses.

**TOPAZ** is a high-resolution Arctic Ocean forecasting and reanalysis system run by the Nansen Environmental and Remote Sensing Center (NERSC) within CMEMS. It provides 3D ocean and sea-ice variables, including temperature, salinity, currents, ice concentration, and ice thickness, with advanced data

**Norkyst800** is a high-resolution coastal ocean forecasting system developed by MET Norway to model the hydrodynamics of the Norwegian coastal and fjord regions. With a spatial resolution of approximately 800 meters, it simulates 3D ocean currents, temperature, salinity, and sea-level variations with fine detail suitable for complex coastal topography. The system is forced by atmospheric models such as AROME and assimilates observational data to provide accurate short-term forecasts supporting marine operations, ecosystem monitoring, and coastal management.

**Oceanic AROME** is a configuration of the AROME high-resolution atmospheric model specifically optimized for marine and coastal environments. It produces detailed short-range forecasts of wind fields, precipitation patterns, humidity, and boundary-layer processes relevant for ocean-atmosphere interactions.

**Radar reflectivity datasets** originate from ground-based weather radars that measure the returned electromagnetic signal from precipitation particles. They provide high-resolution spatiotemporal fields characterizing rainfall intensity, storm structure, and convective development, supporting nowcasting and hydrometeorological applications.