MDMC
Master in Data Management and Curation

AREA SCIENCE PARK

SISSA

Implementation of FAIR Principles for Marine Data at OGS

Alessandra Lanzoni, Andrea Corbo, Sebastian Plehan, Elena Partescano
National Institute of Oceanography and Applied Geophysics - OGS

2025-12-03

Overview

- The Trieste Marine Observatories system and E2M3A buoy
- The “Buoy Controller” Hardware & Software
- Instrument maintenance
- Data flow
- Data publication ERDDAP and NODC portal
- ERDDAP and FAIR principles
- Metadata and Standard Vocabularies

The Gulf of Trieste and the Northern Adriatic Sea

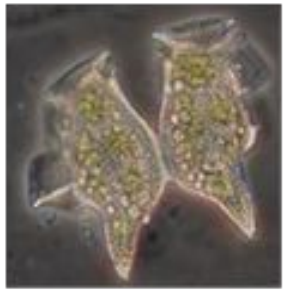
- Morphology
- Physical forcing
- Marked seasonal and interannual variability
- Biogeochemical characteristics
- Anthropogenic impacts
- Extreme events
- Extensive observational sites

Red tides

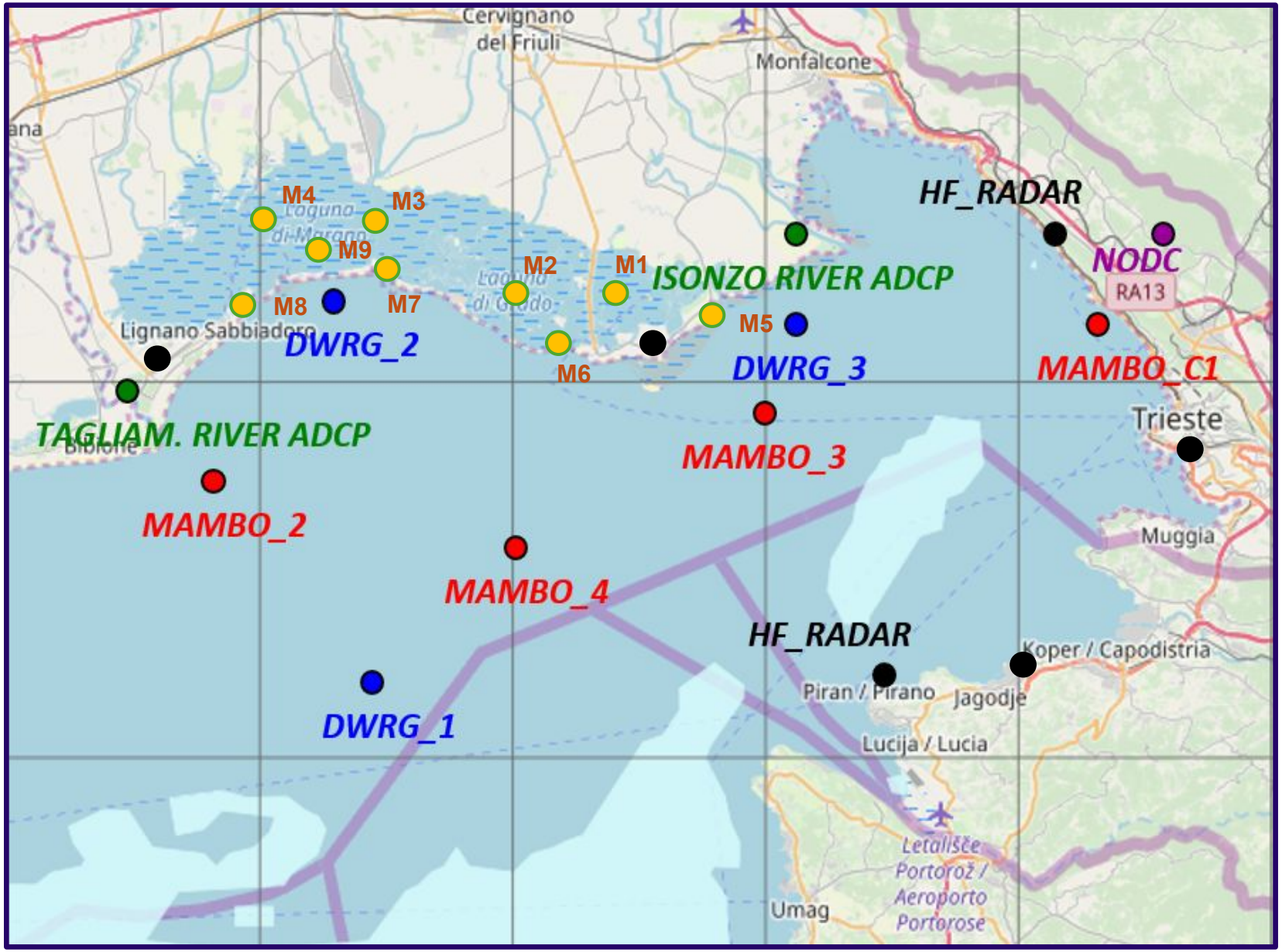
Mucilage

Jellyfish

Toxic algae

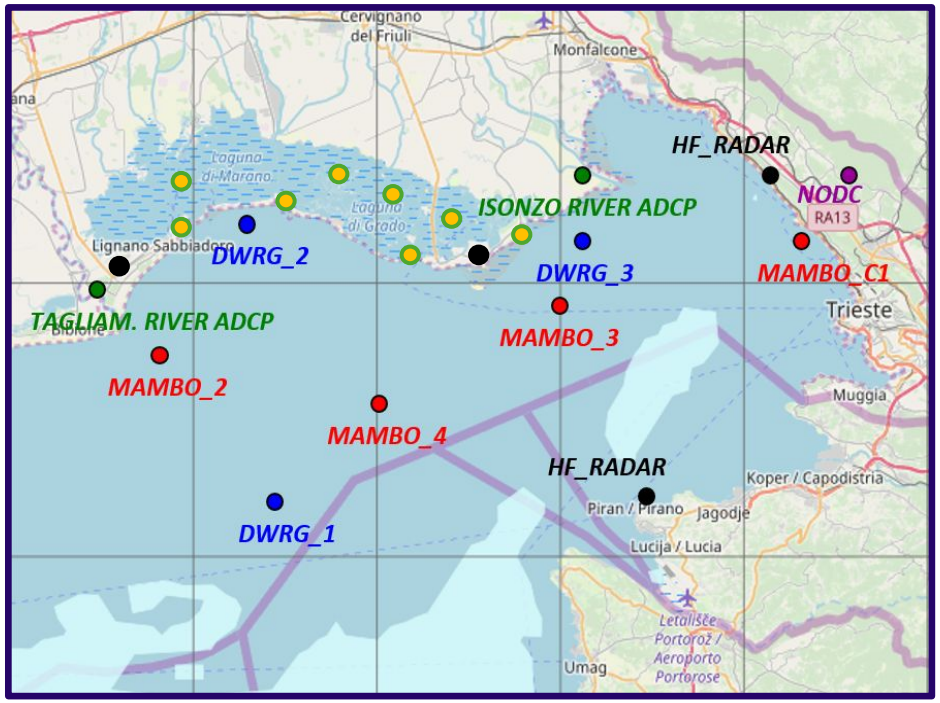


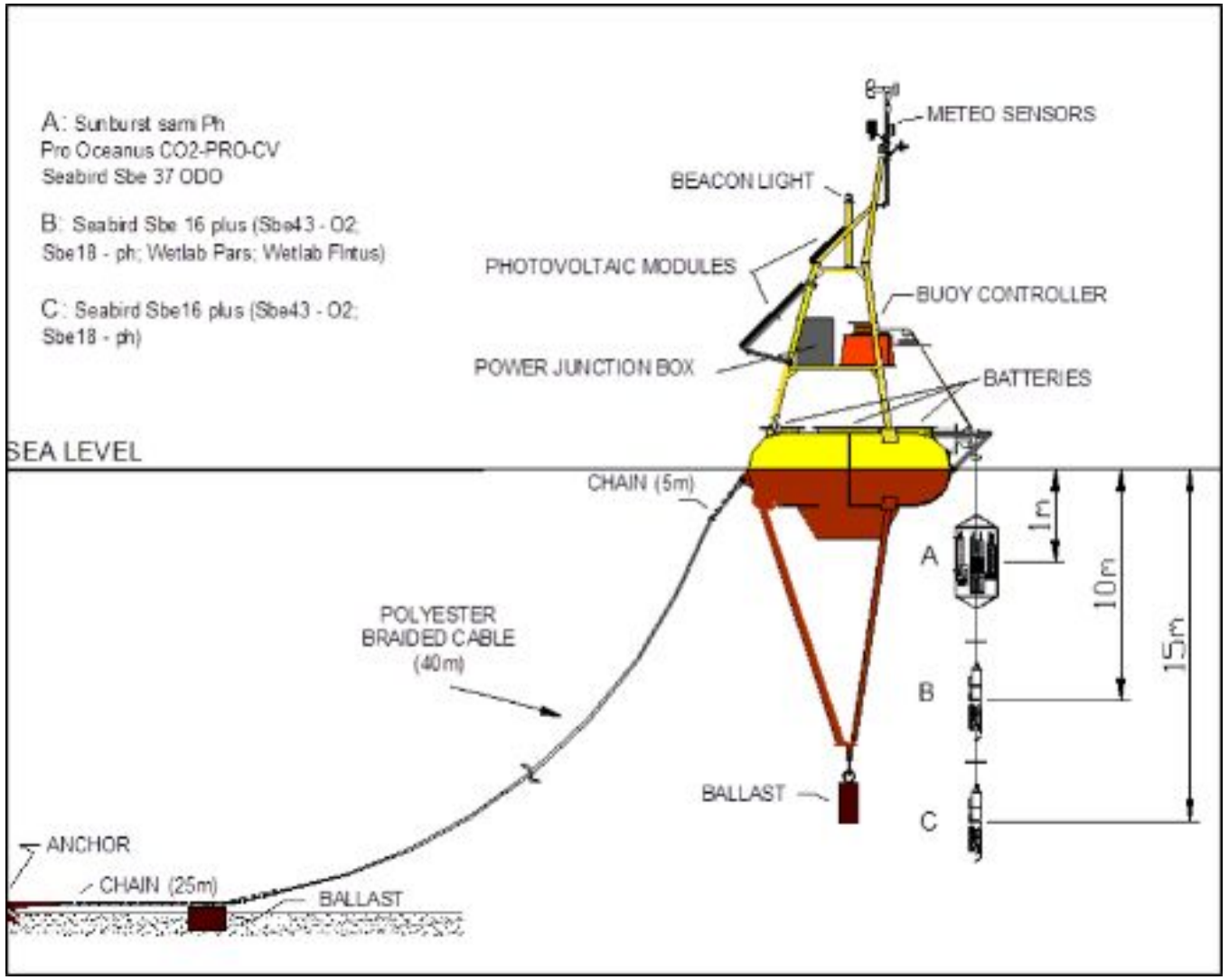
Trieste marine observatory system



- Civil Protection ADCP
- Radar
- Elastic Beacon
- Directional Wave Buoy
- NODC Data Center
- Danubius –RI (ERIC)

Meteo – Oceanographic buoys MAMBO_miramare





Meteo – Oceanographic buoys MAMBO_miramare

Meteorological Station:

- Air temperature, humidity
- Atmospheric pressure
- Wind speed and direction

Probes at -2m:

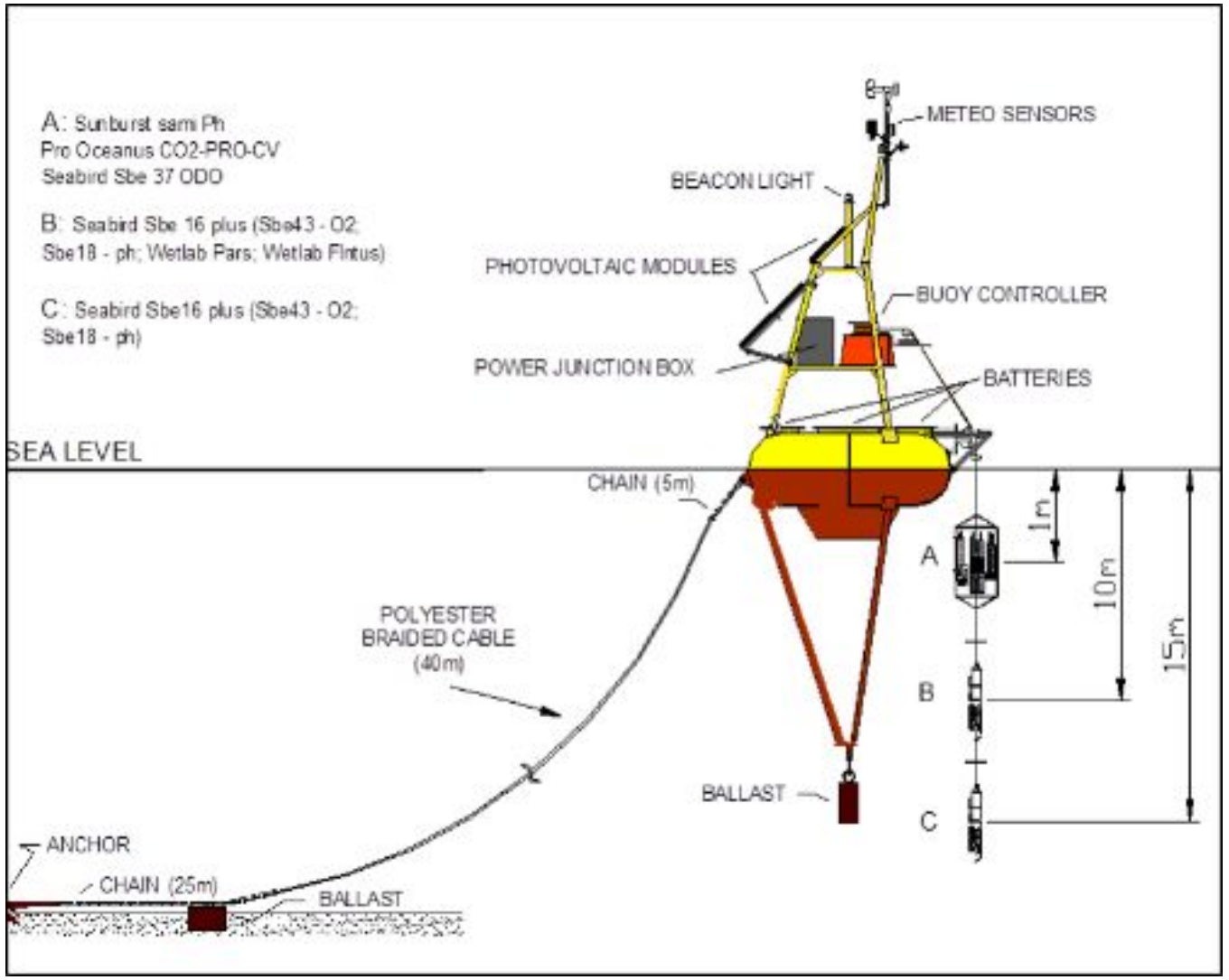
- Temperature.
- Conductivity/Salinity.
- Dissolved oxygen.
- pH and pCO2

Probe at -10m:

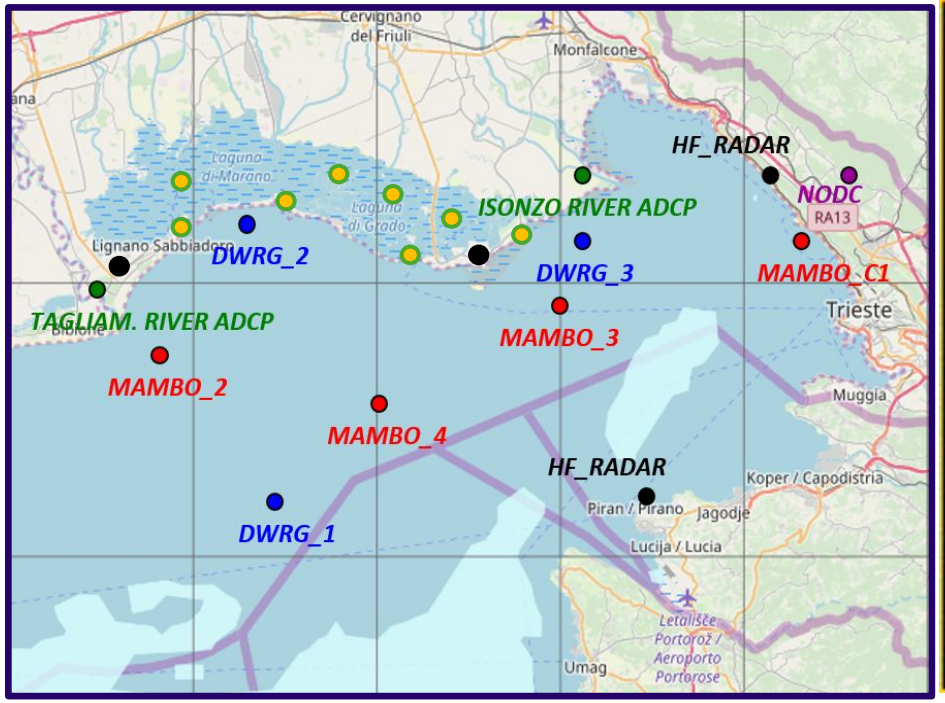
- Temperature.
- Conductivity/Salinity.
- Pressure.
- Dissolved Oxygen.
- Turbidity,

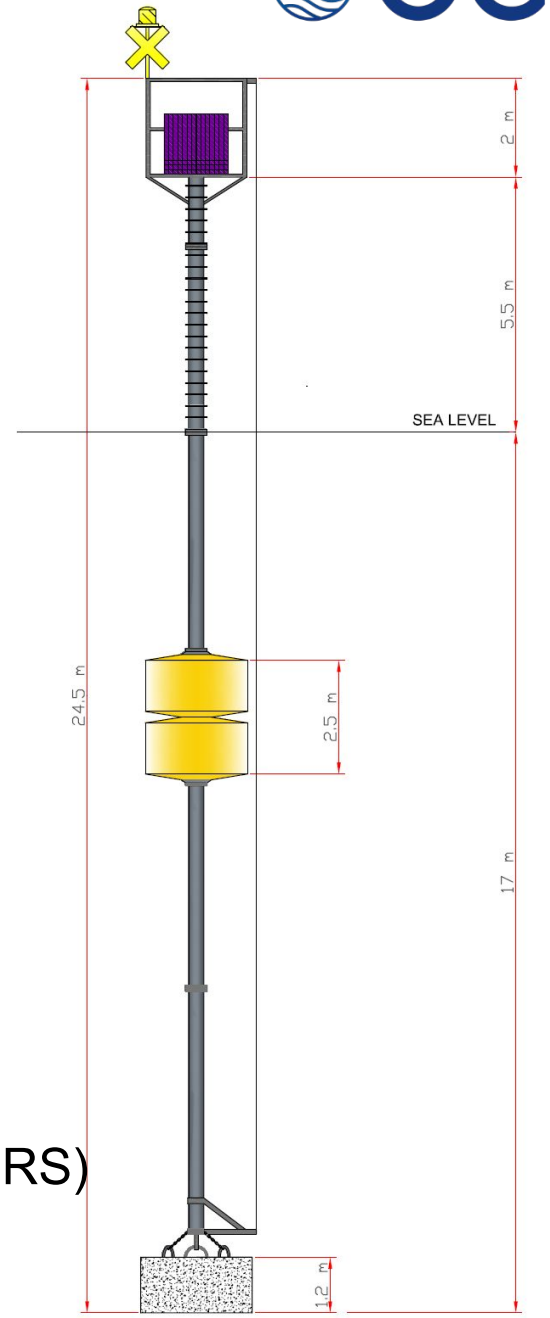
Probe at -15m:

- Temperature.
- Conductivity/Salinity.
- pH.
- Dissolved oxygen.



Oceanographic buoys MAMBO 2,3,4





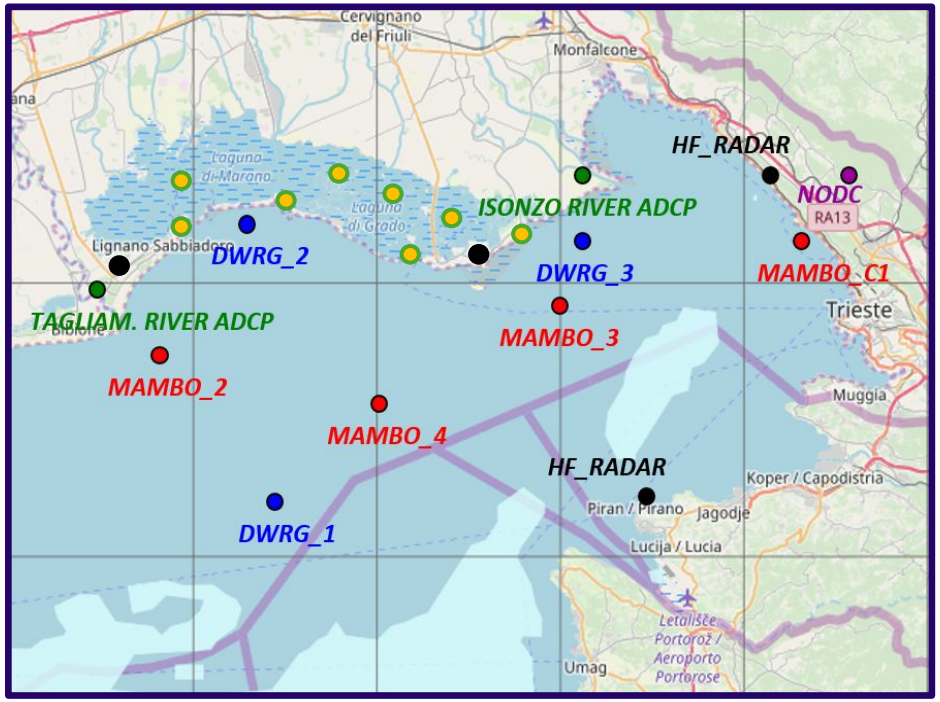
Meteorological Station:

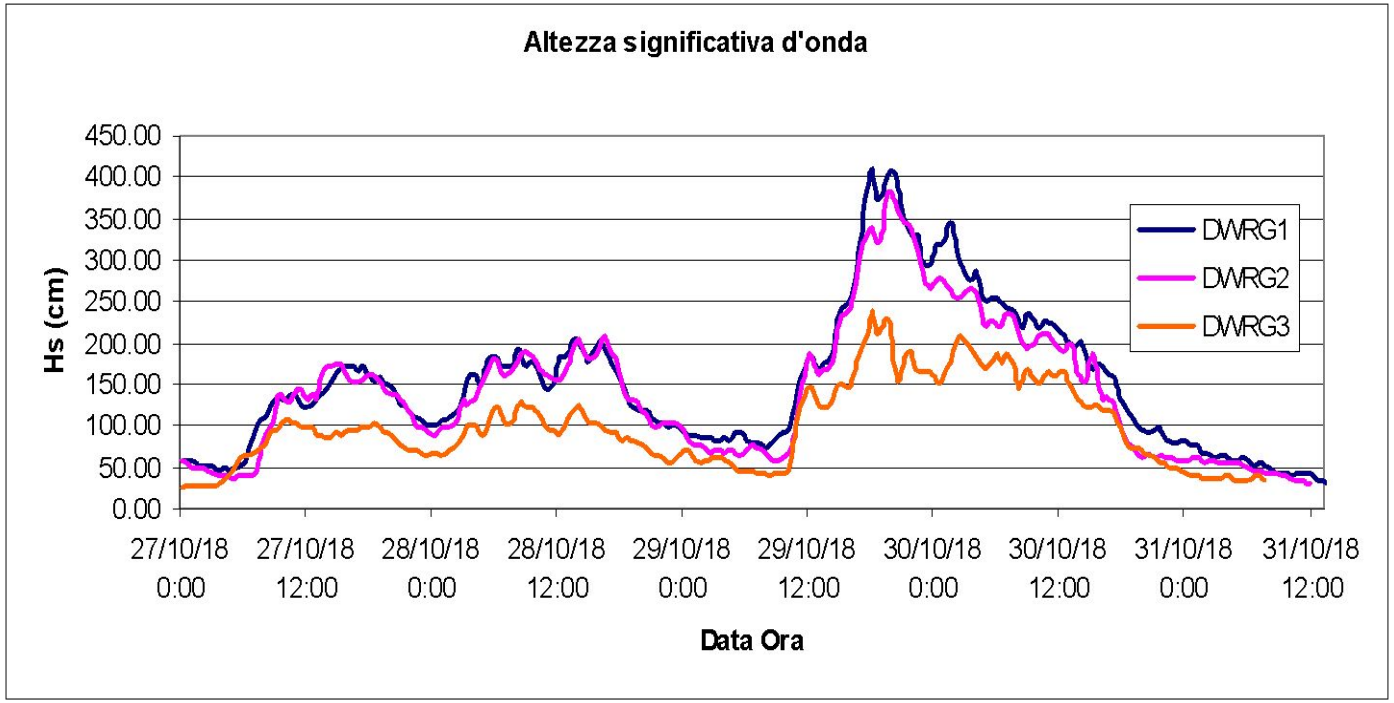
- Air temperature and humidity
- Atmospheric pressure
- Wind speed and direction

Probes:

- Temperature.
- Conductivity/Salinity.
- Dissolved oxygen.
- ADCP current profiler
- OBS sismografo di fondo (CRS)

Directional Wave buoys

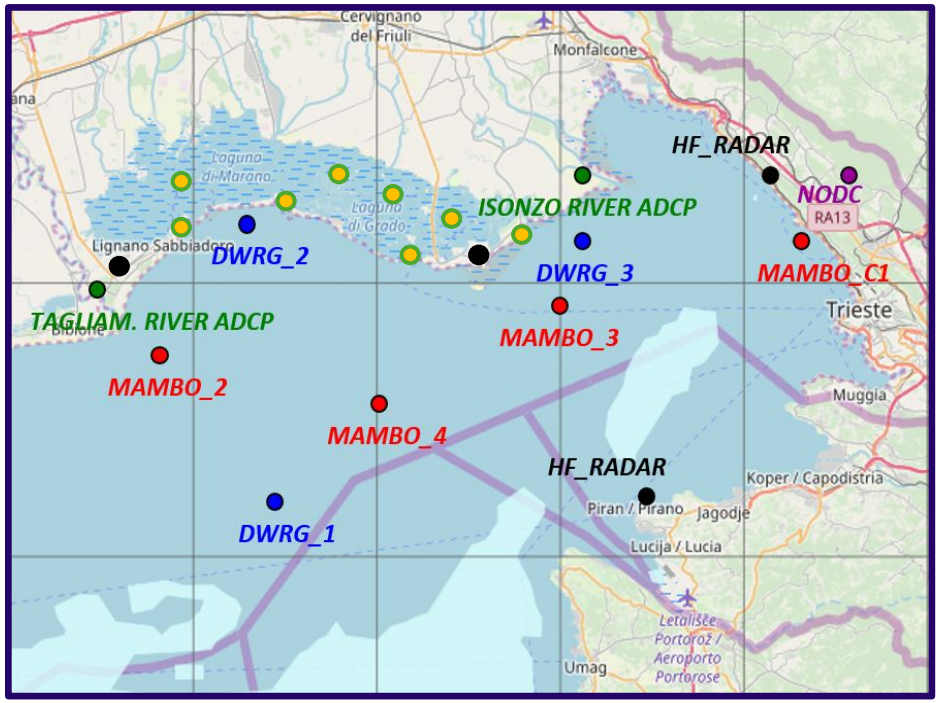




Function:

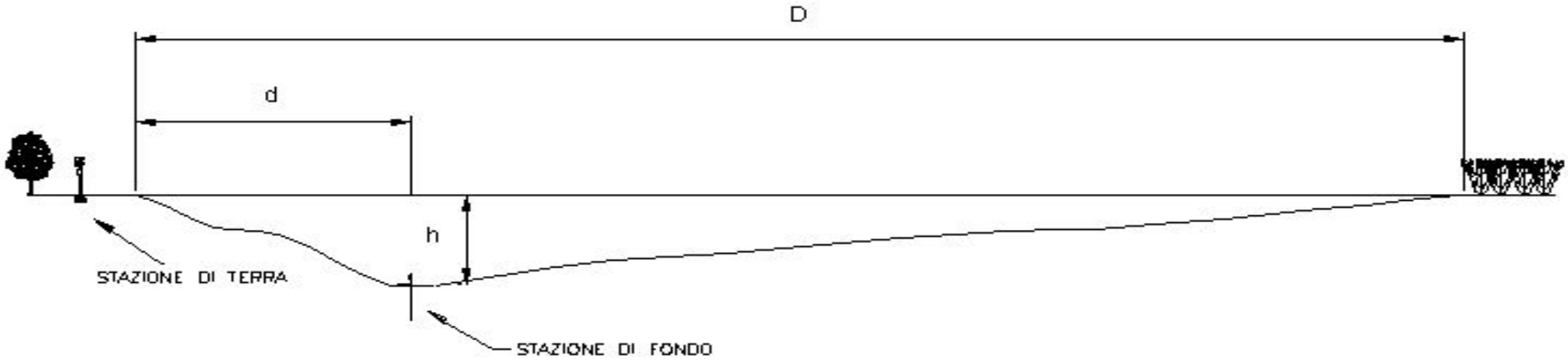
- Acquisition every 30 minutes
- Directional and source waves
- Maximum wave height
- Significant wave height

Isonzo River ADCP

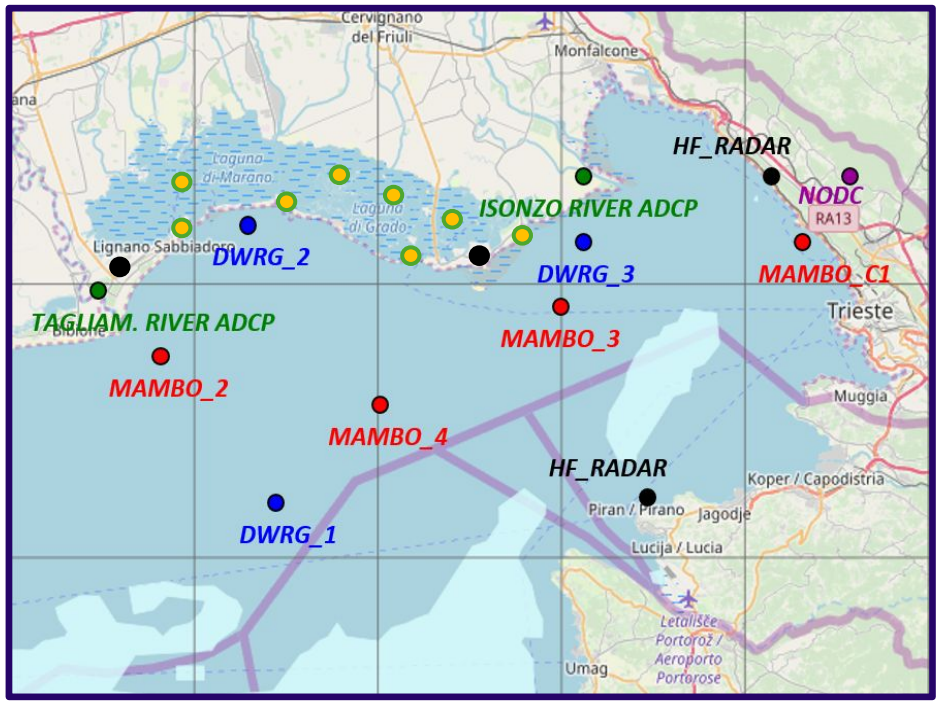


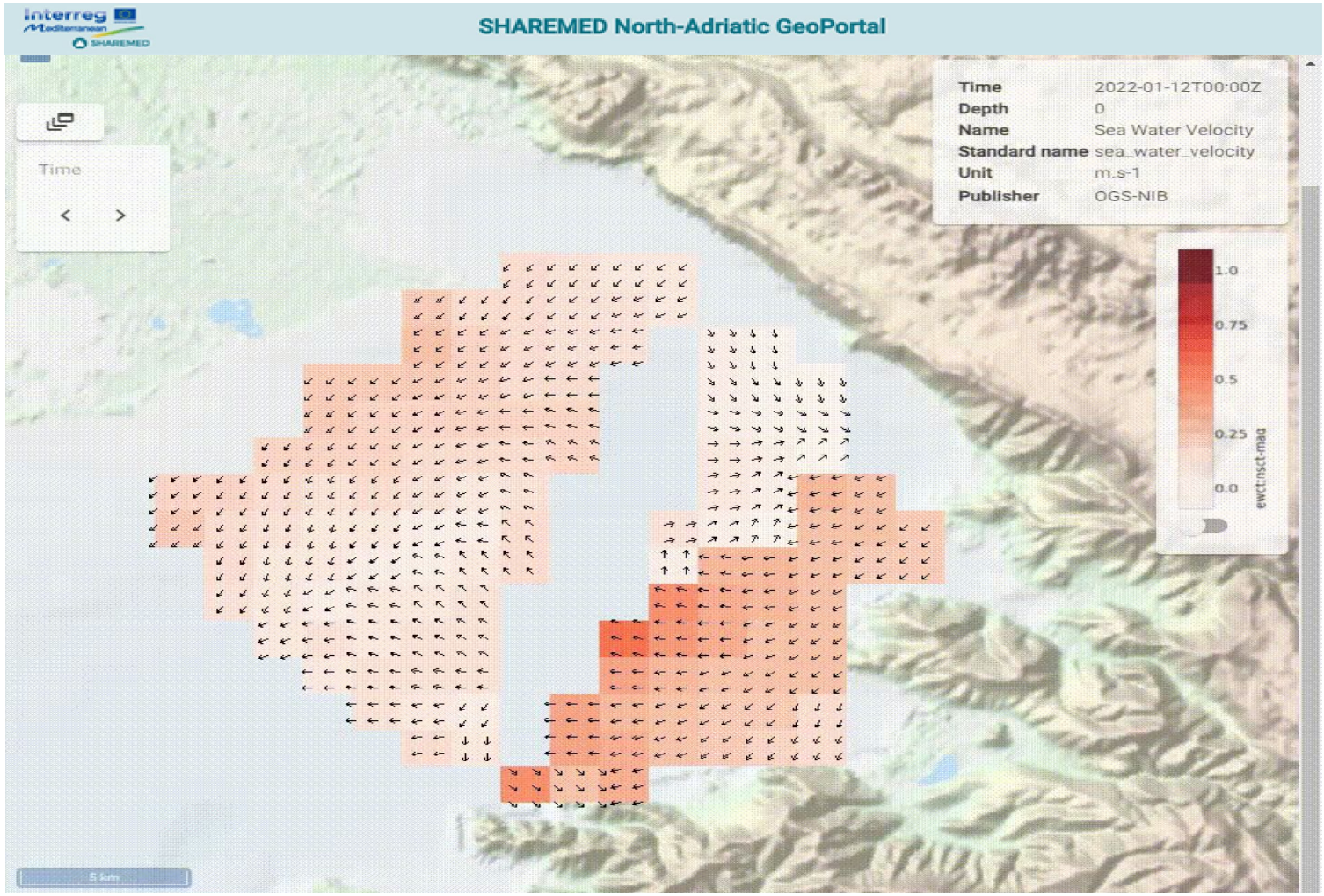
The monitoring stations is equipped with a Nortek 1000 kHz Aquadopp profilers fixed at the river bottom by means of an immobile stainless-steel structure.

Acquisition every 10 minutes (cell size = 0.5 m) over the entire water column (12 -13 meters)



HF_Radar

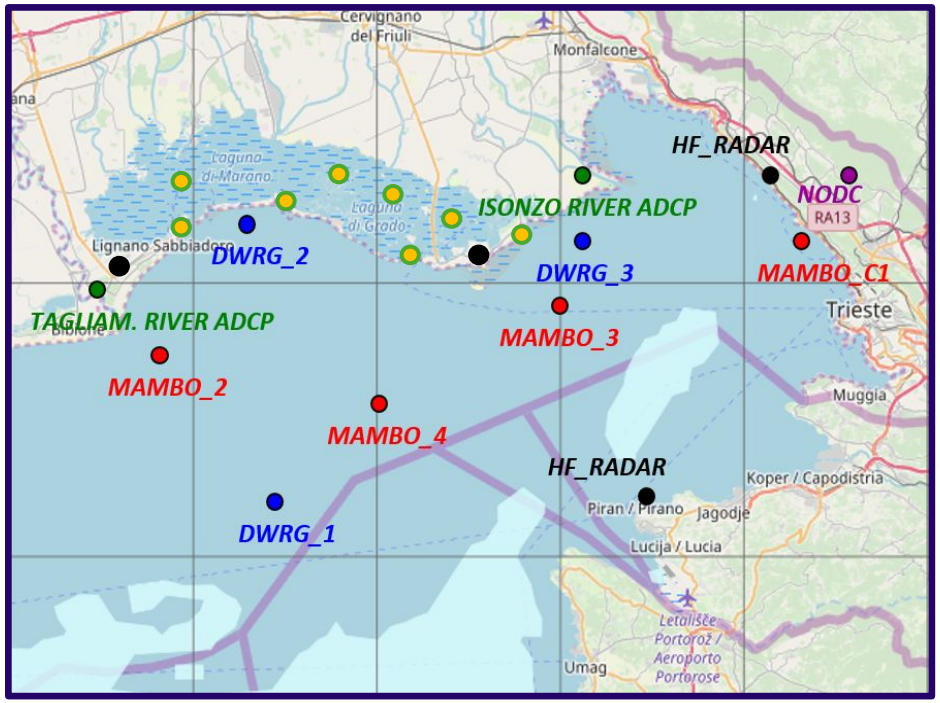


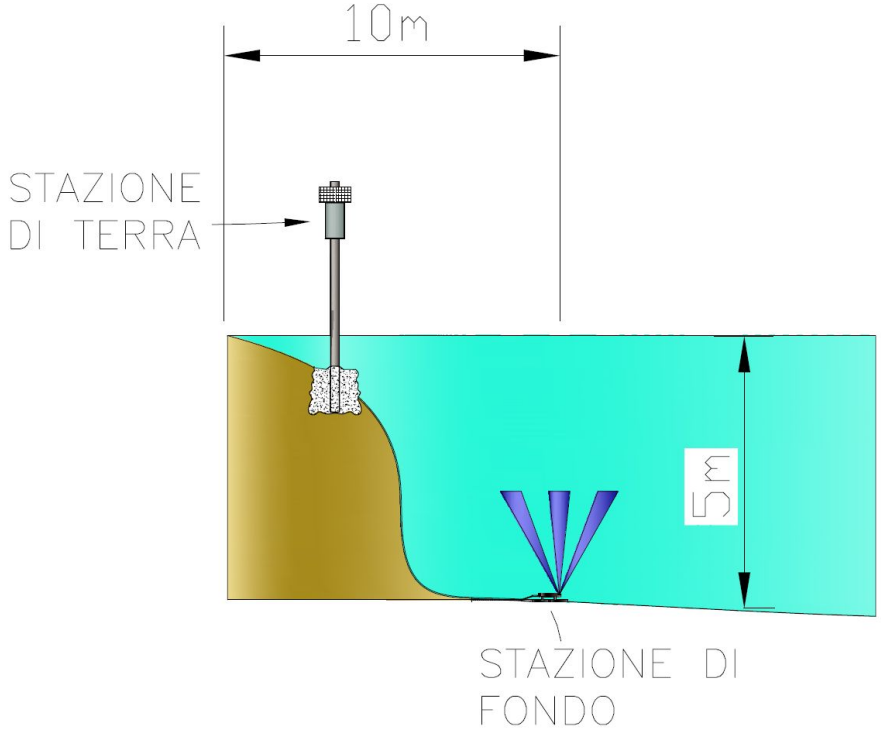


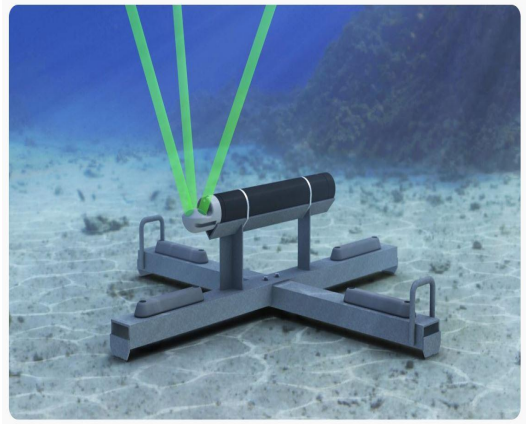
CODER ARRAY:

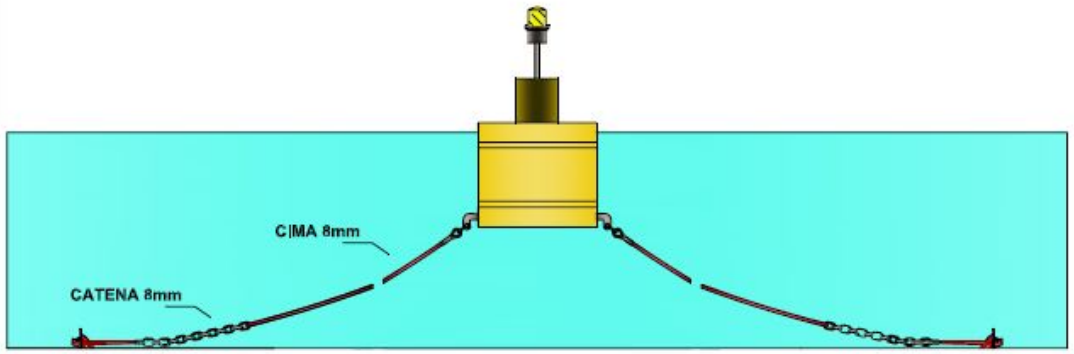
With the 2 new radar improvement of the resolution in front of the Istrian Coast

Danubius-RI implementation in the Marano-Grado lagoon







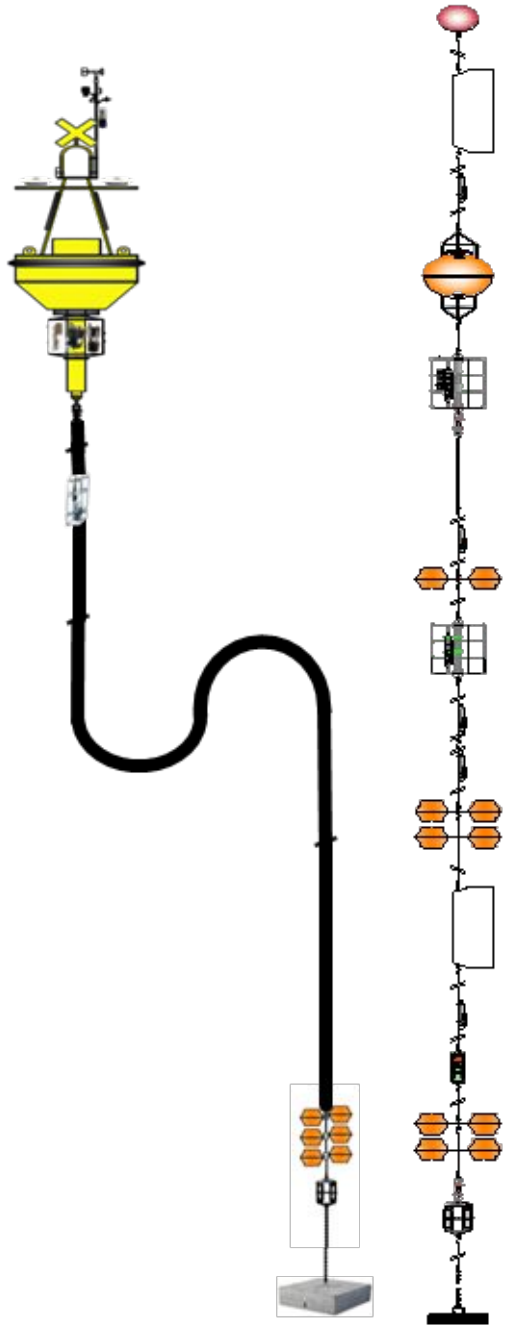


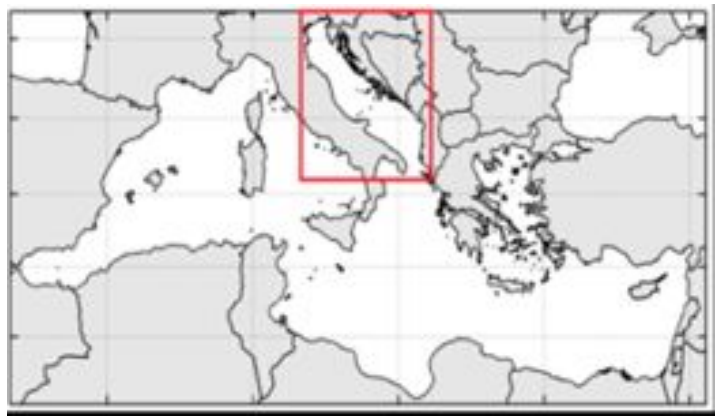
Probes buoys:

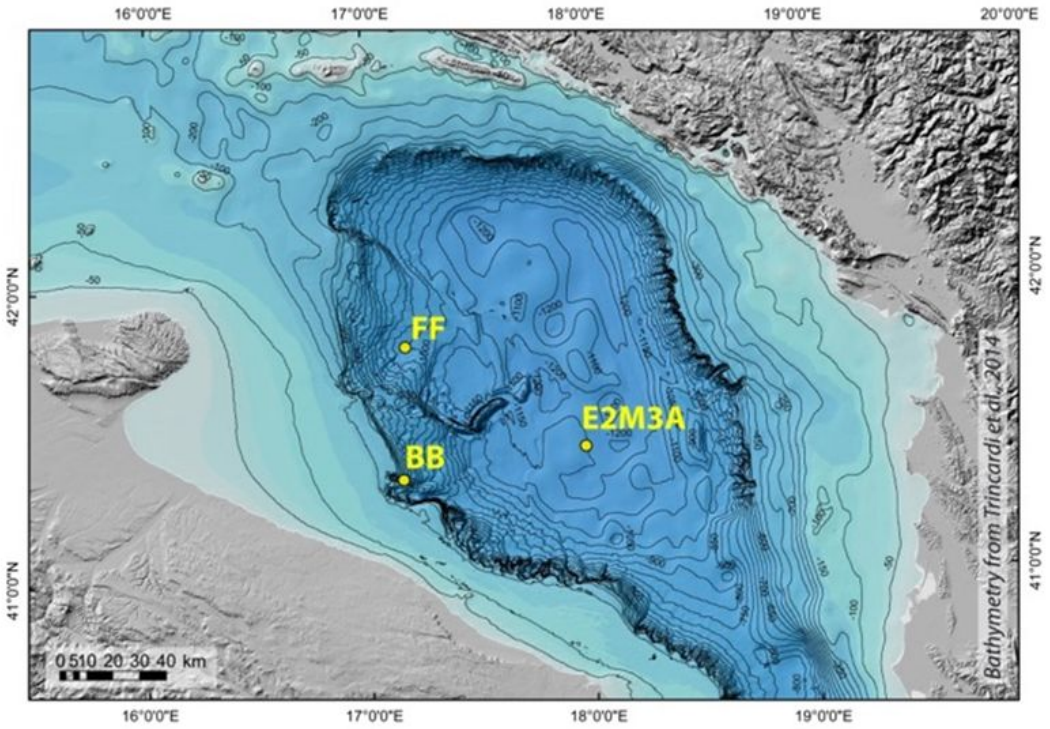
- Temperature.
- Conductivity/Salinity
- Dissolved oxygen
- Chlorophyll A
- SUNA Nitrates analyzer

E2M3A - South Adriatic Sea

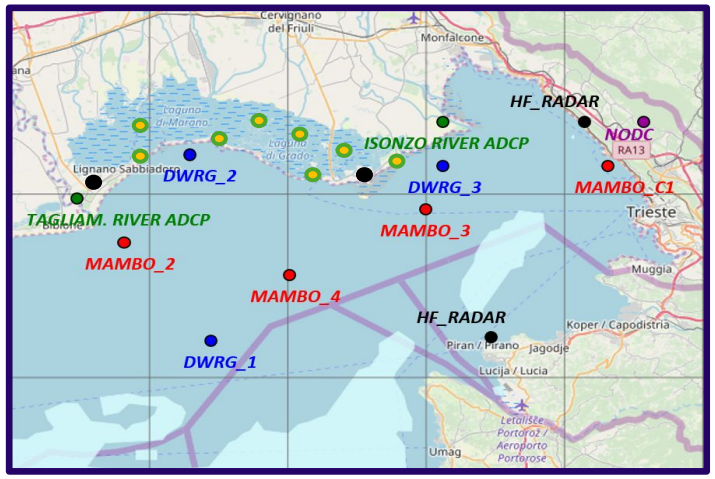
Parameters measured during LAST deployment	Depth
Temperature and salinity	120, 591, 889, 989, 1156, 1165 m
Temperature and salinity	332, 770 m
Pressure	591, 889, 989, 1156, 1165 m
Pressure	332,770 m
Turbidity	1165 m
Dissolved Oxygen	770, 889, 989, 1156, 1165 m
Dissolved Oxygen	332, 770 m
Trasmittance	332, 770 m
Currents (profiling)	170-318 m
Currents	1165 m
Particulate	1153 m







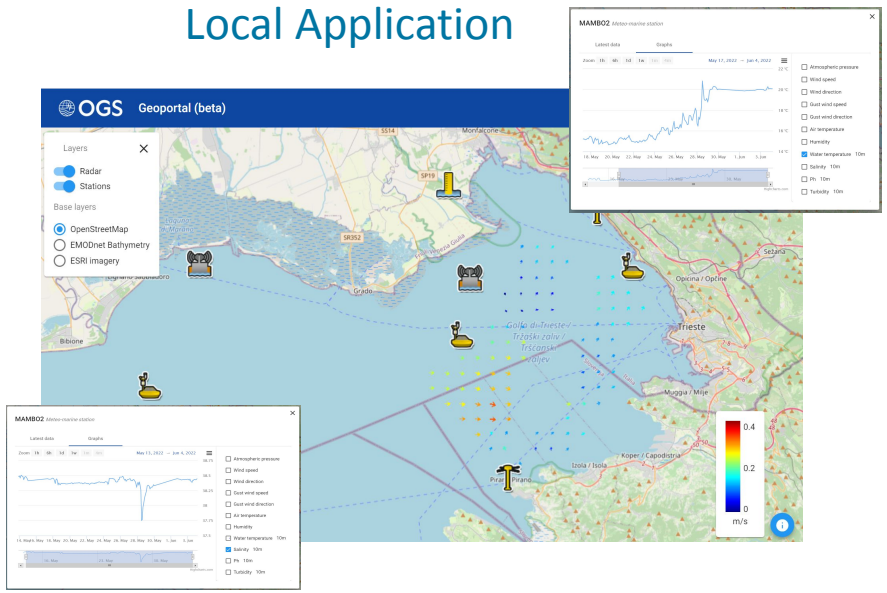
Data flow and products



Realtime and delayed mode data

NODC
National
Oceanographic
Data Centre at OGS

Local Application



International Infrastructures as:

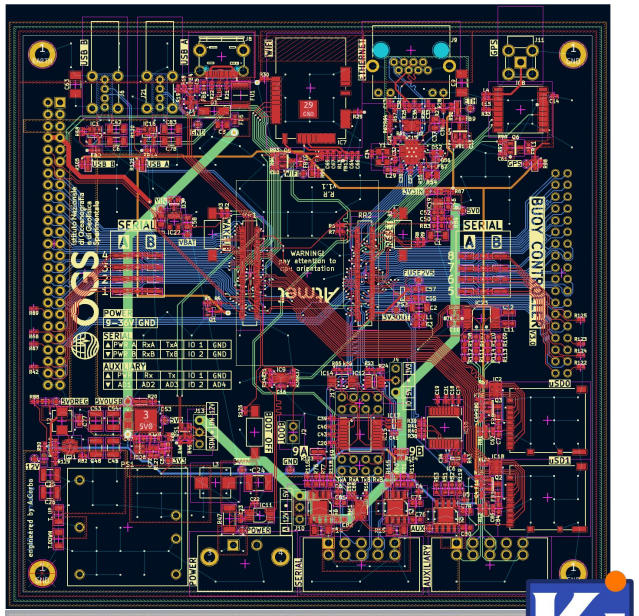
EMODnet European Marine Observation and Data Network
LTER Italia
EMBRC EUROPEAN MARINE BIOLOGICAL RESOURCE CENTRE
SeaDataNet
JERICO RI SCIENCE - SERVICES - SUSTAINABILITY
ICOS Ocean Thematic Centre
DANUBIUS-RI
GOA-ON Global Ocean Acidification Observing Network
IMBER Integrated Marine Biogeochemistry and Ecosystem Research

Regional services:

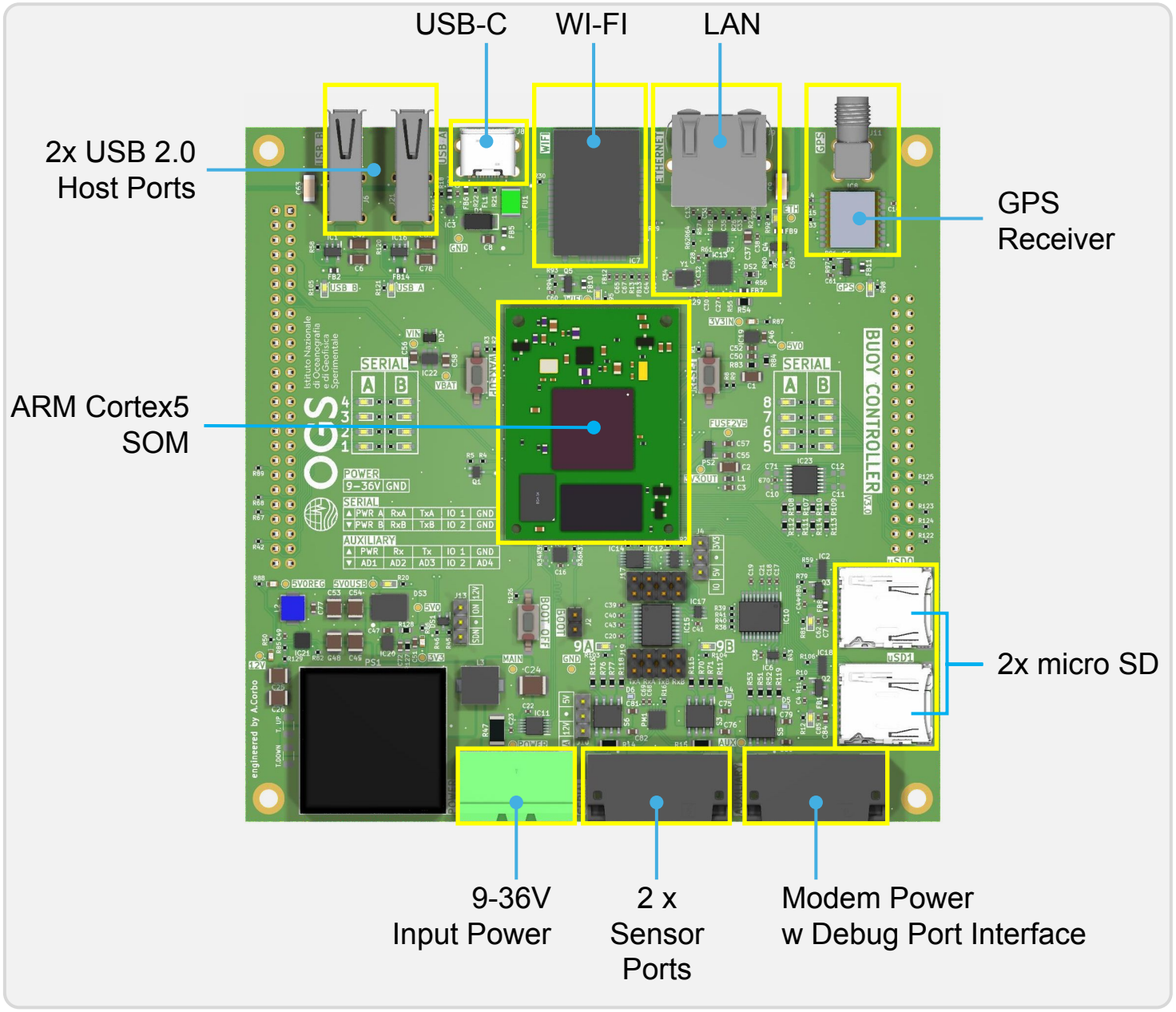
REGIONE AUTONOMA FVG
PROTEZIONE CIVILE civil defence
aRPa FVG agenzia regionale per la protezione dell'ambiente del Friuli venezia giulia

Buoy Controller v 3.0

- Low energy consumption
- Highly robust
- Custom hardware
- In-house Firmware and Software

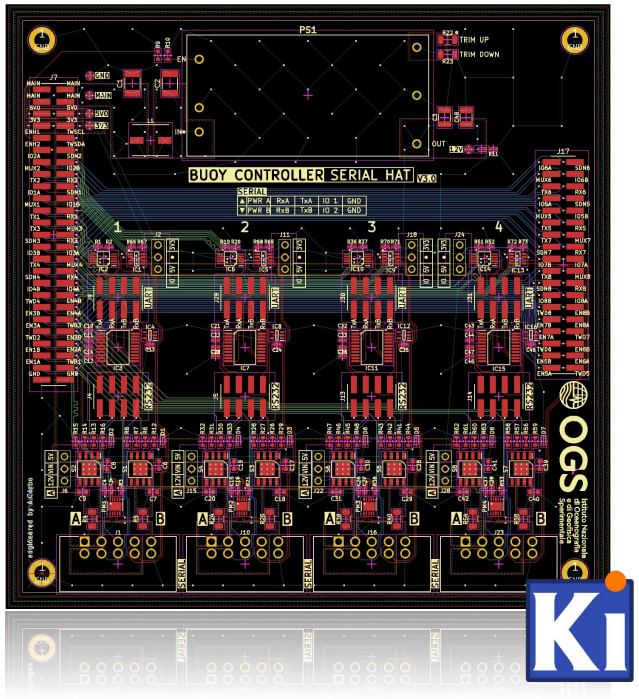


KiCad

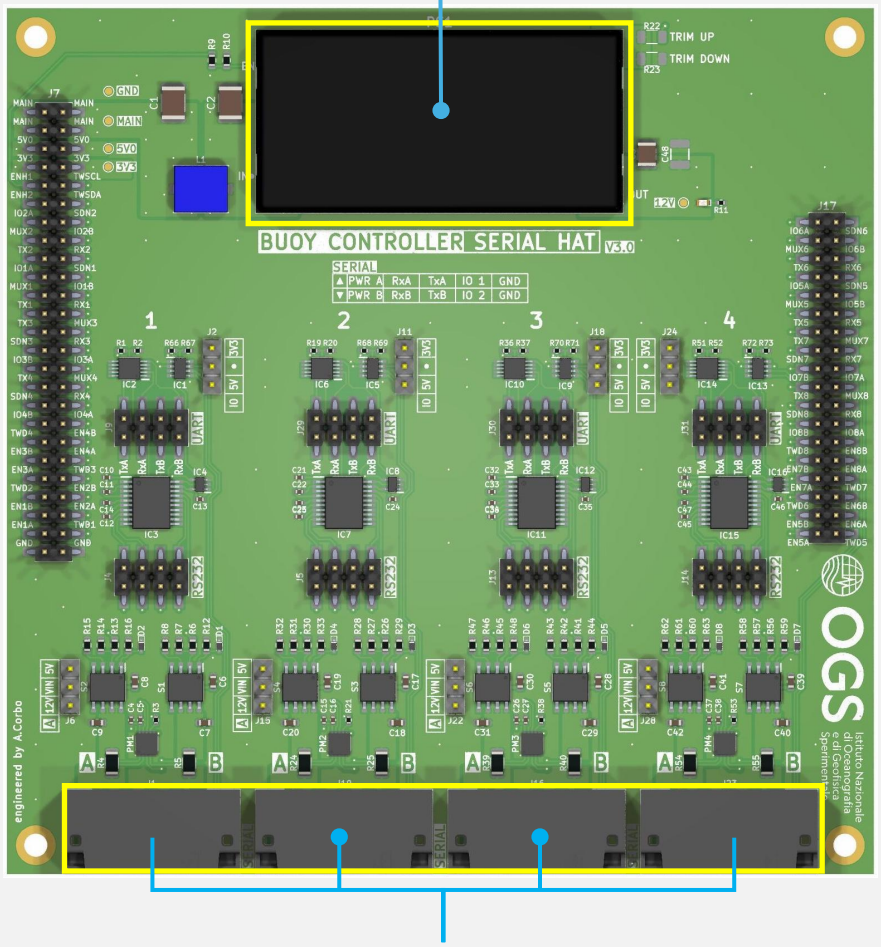


Buoy Controller v 3.0 – **Serial Hat v 3.0**

- Modular
- Scalable
- Application-Specific requirements
- Up to 2 stackable expansion boards
- Up to 18 sensors

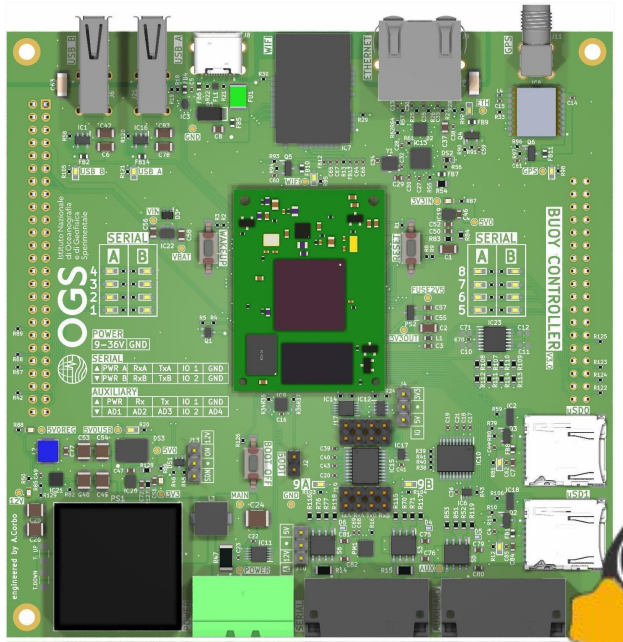


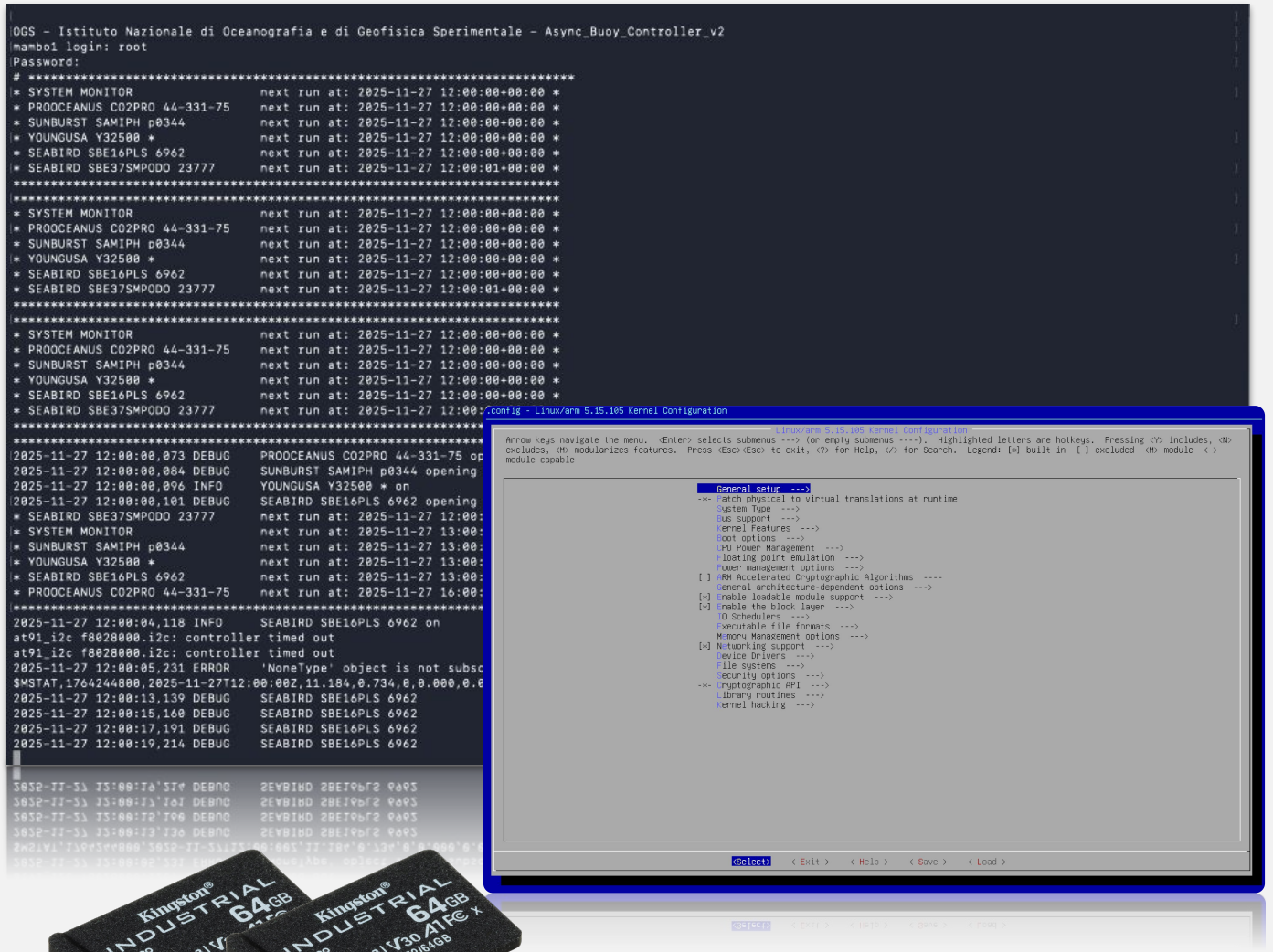
KiCad



Buoy Controller v 3.0 – **Linux Operating System**

- Customized **Linux** OS distribution
- Only necessary **Kernel** modules
- Small size **Root File System**
- Fast boot up and Energy saving
- Industrial grade memory support

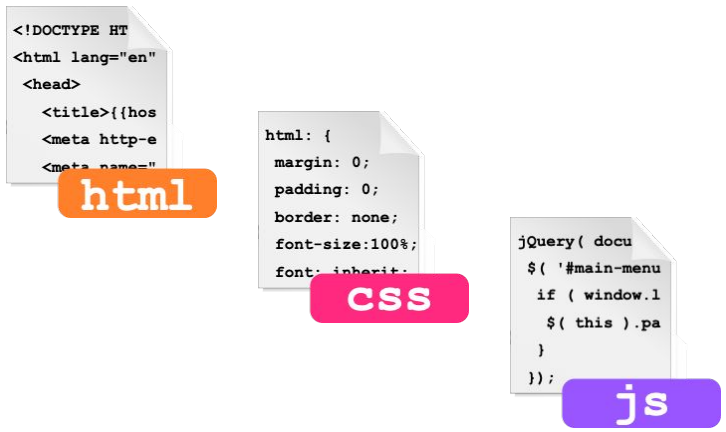


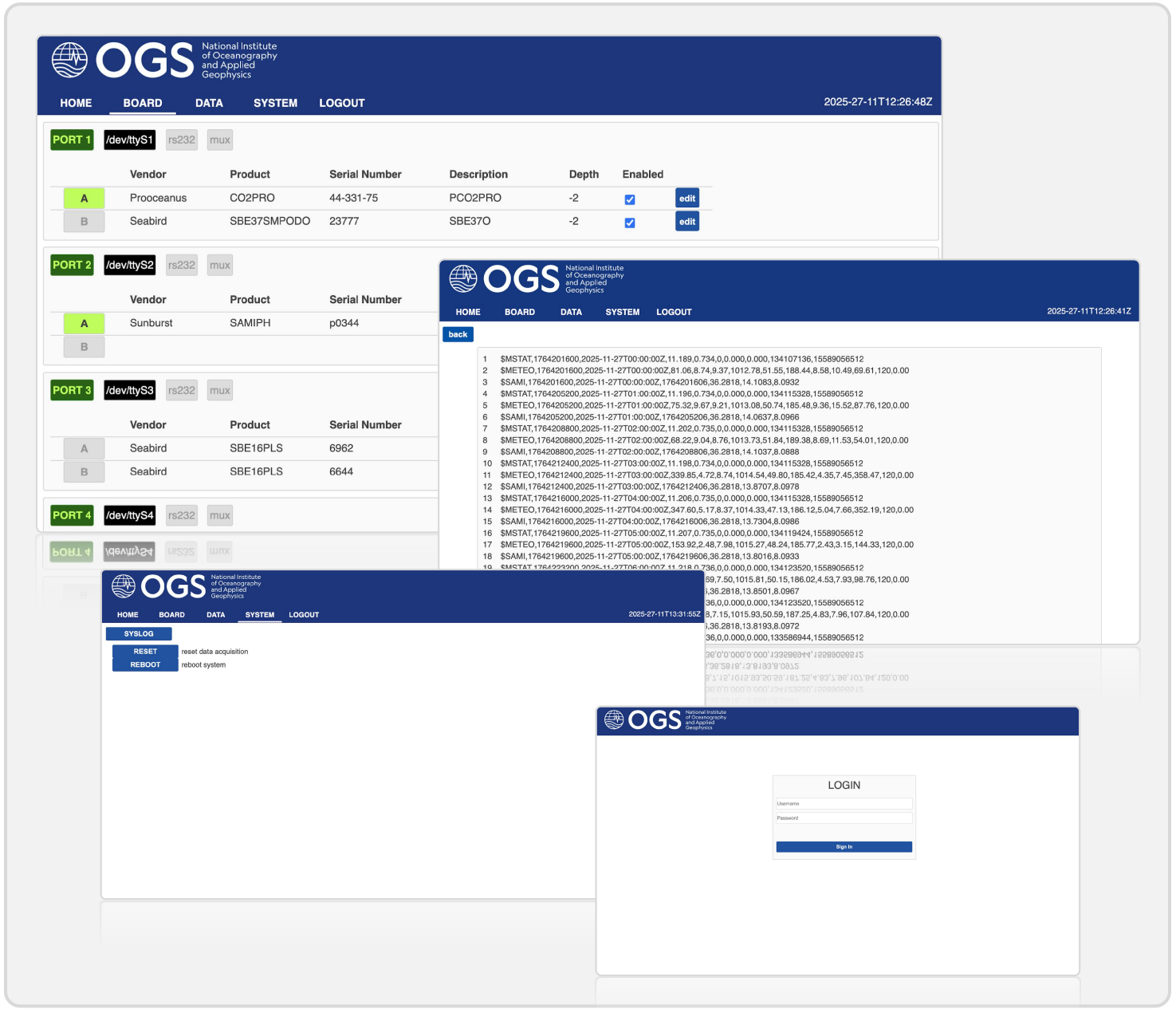


Buoy Controller v 3.0 – **Web Interface**

- Local and remote access
- **G**raphical **U**ser **I**nterface
- System status monitoring
- Real-Time data visualization
- On site configuration

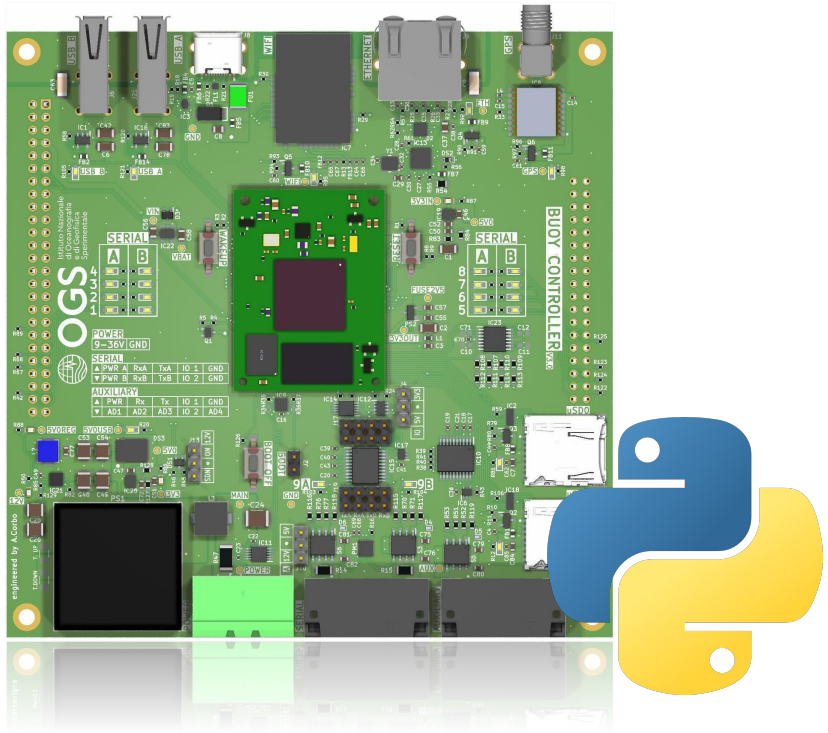
Flask
web development,
one drop at a time





Buoy Controller v 3.0 – **Control Software**

- **A**synchronous **P**ython **Scheduler**
- Modular architecture
- Main common module
- Sensors specific sub-modules
- Json configuration files



Main module

main

instrument

tools

base

board

schedule

Sensors sub-modules

device1.py

device1.json

device4.py

device4.json

device2.py

device2.json

...

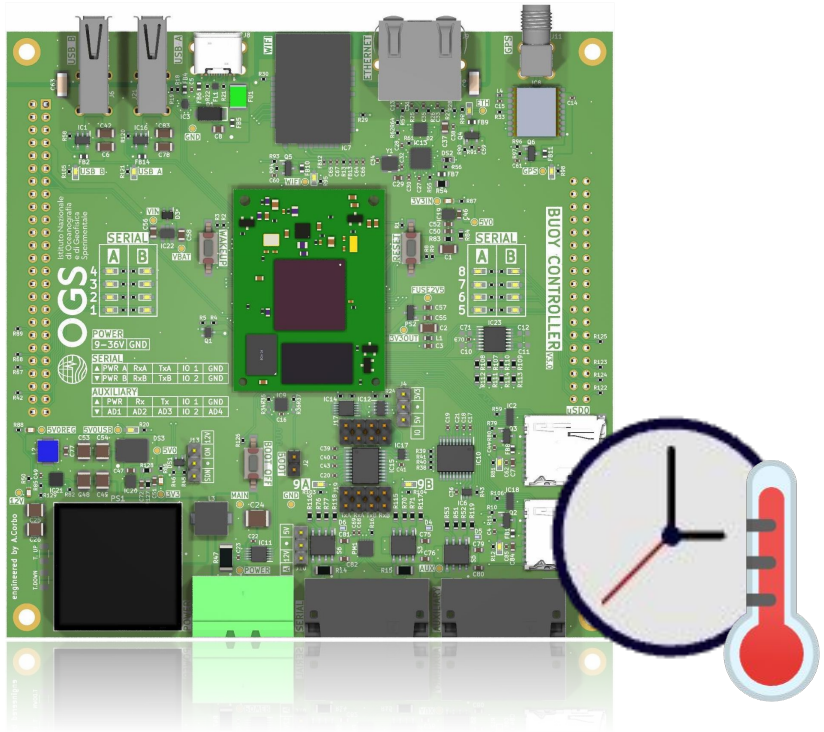
device3.py

device3.json

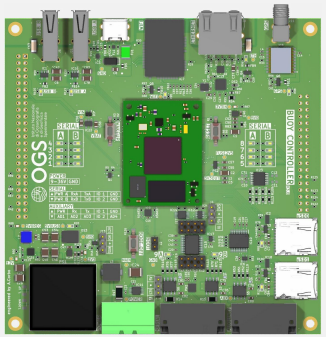
...

Buoy Controller v 3.0 – **Asynchronous Data Acquisition**

- Asynchronous sensors activation
- Sensor initialization
- Sensors power management
- Customized sampling interval
- Streamed or Polled sampling
- Data conversion and formatting



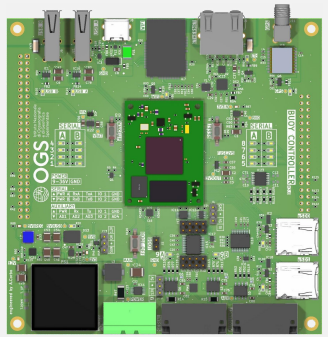
Streamed Sampling



Power

Samples Stream

Polled Sampling



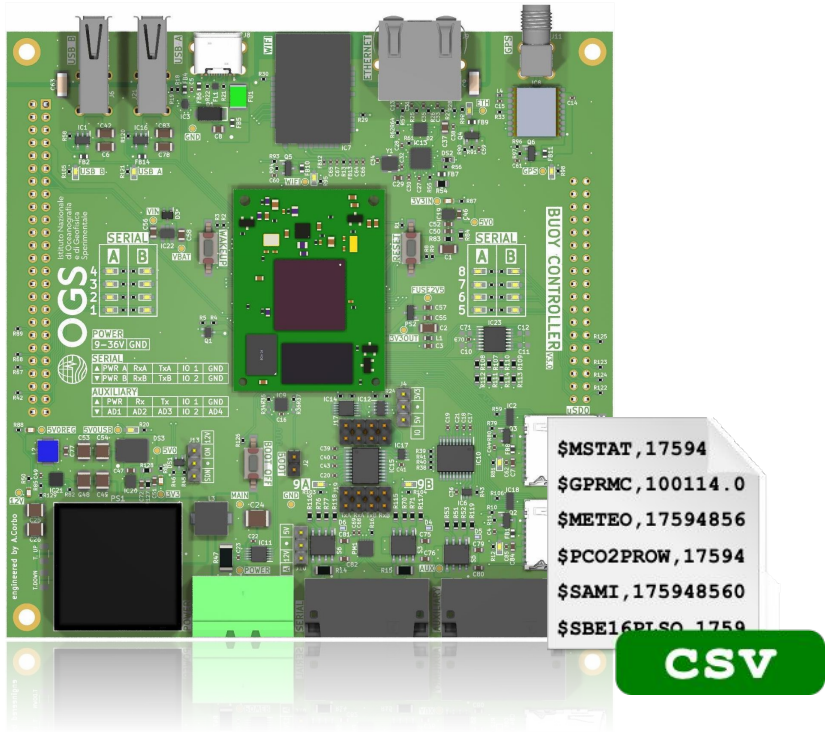
Power

Command

Sample

Buoy Controller v 3.0 – **Data Formatting**

- Plain ASCII strings
- Comma Separated Values
- NMEA sentence syntax
- Controller UTC timestamp
- Daily file / Periodic transmission



Sensors data

label *unix epoch* *iso 8601 timestamp* *instrument payload*

```
$SBE370,1764505800,2025-11-30T12:30:00Z,21.2064,5.39146,5.245…
$SBE16PLS01,1764505800,2025-11-30T12:30:00Z,21.2064,5.391467…
$SBE16PLS02,1764505800,2025-11-30T12:30:00Z,21.2064,5.391462…
$PCO2PRO,1764505800,2025-11-30T12:30:00Z,21.2064,5.39146,5sd.…
```

System Status

label *battery V* *RH%* *uSD used bytes*

```
$MSTAT,..,.12.710,0.233,1,66.445,18.478,115699712,15589056512
```

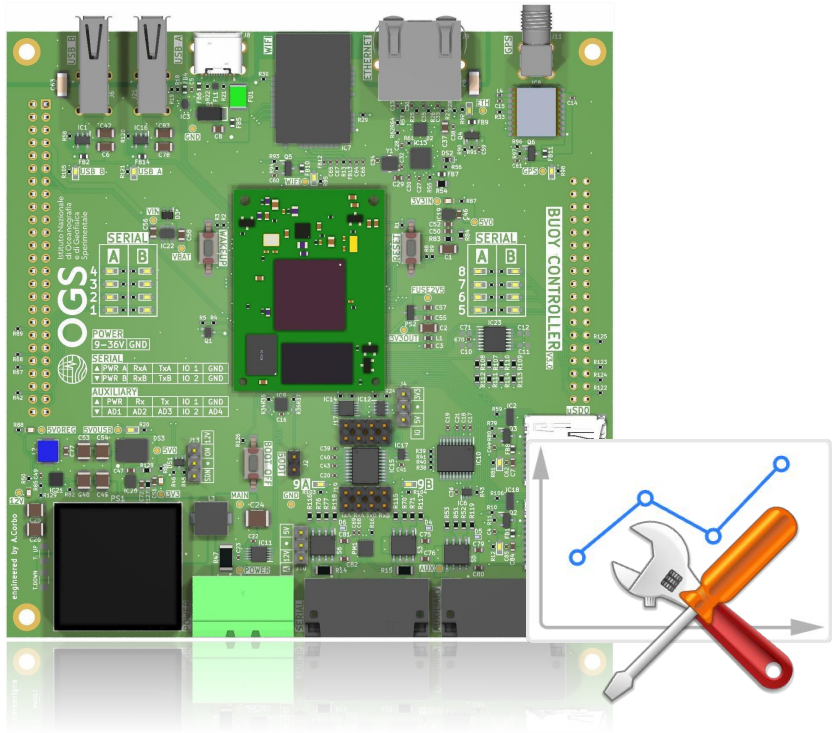
current A *temp °C* *uSD free bytes*

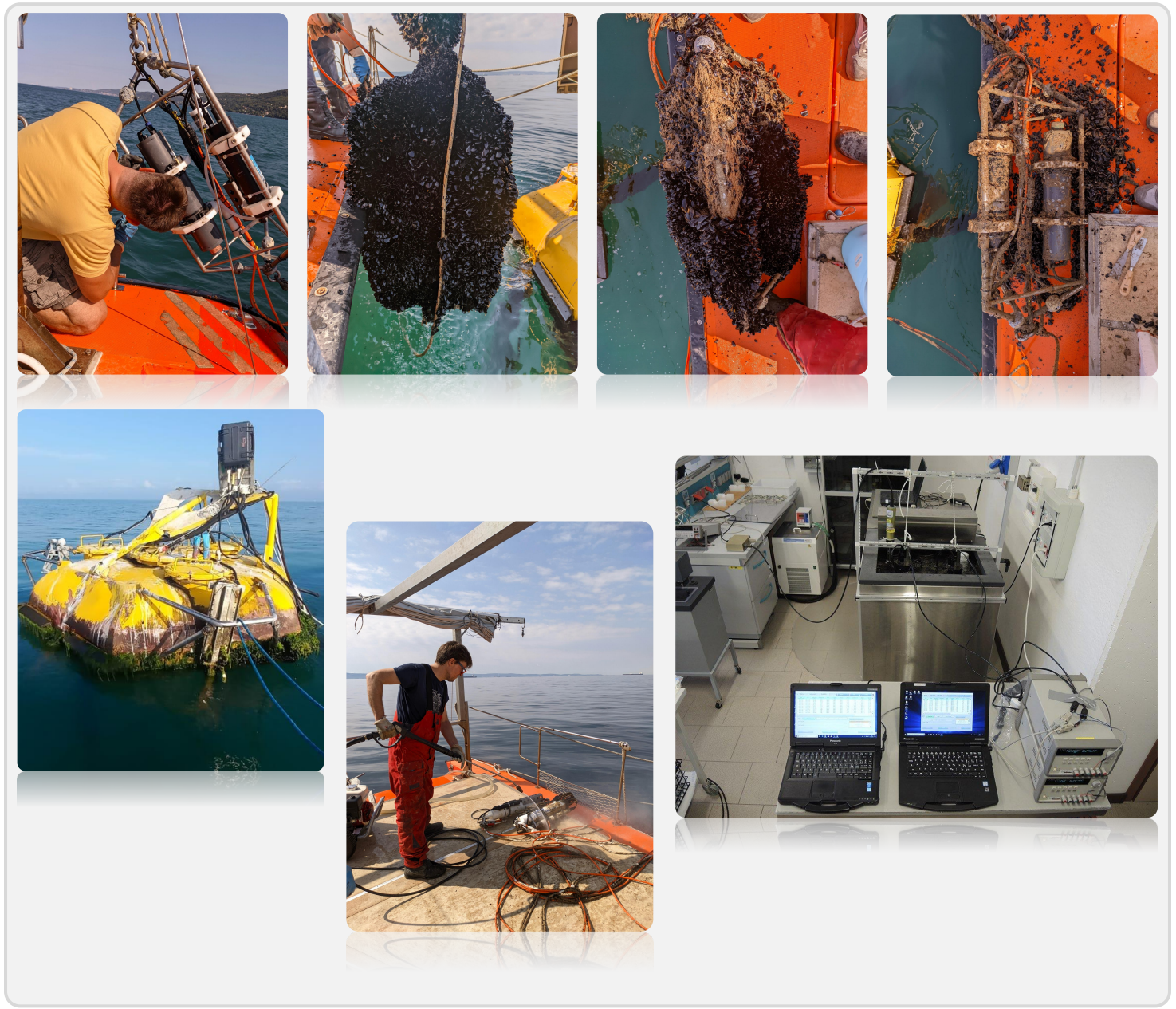
System Log

```
$SYSLOG,2025-11-30 00:15:31,091,CRITICAL,Starting - As…
$SYSLOG,2025-11-30 01:00:05,164,ERROR,'NoneType' objec…
```

Buoy Controller v 3.0 – **Instrument Maintenance**

- Biofouling
- Corrosion
- Vessels collisions
- Extreme weather conditions
- Periodic cleaning and maintenance
- Laboratory calibration - CTMO

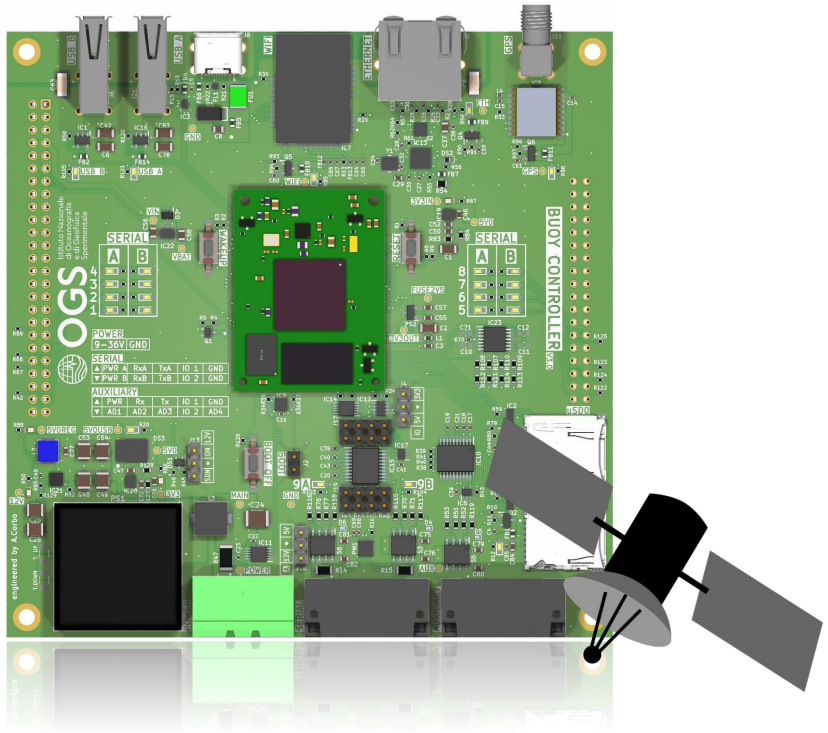




Buoy Controller v 3.0 – **Geofencing**

- Location-based technology
- Virtual boundary around mooring
- SMS alert messages (coastal)
- Server side email with maps

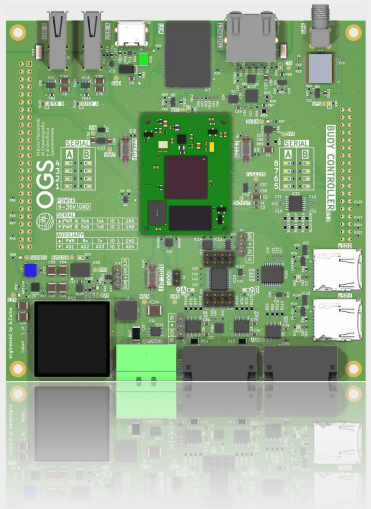
2023 march, E2M3A mooring **S.A.R.** cruise (52nm off the Italian coast at a depth of 1200m). All the instruments that sank due to the mooring break have been recovered.



Buoy Controller v 3.0 – **Near Real-Time Data Transmission**

OGS

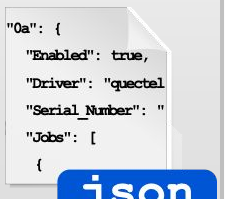
LTE or Satellite Network



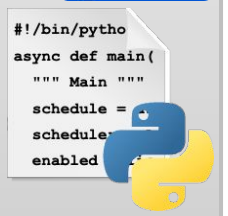
Samples,
System Status and
Logs

Google Cloud

OAuth v2



System Updates



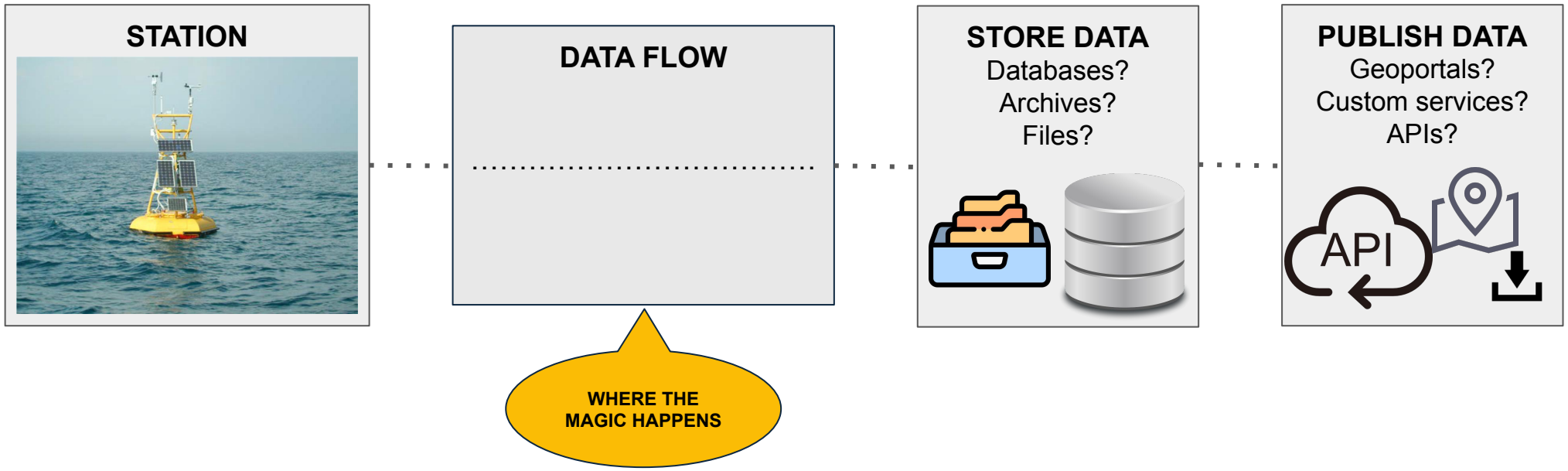
OGS

Data flow and publication

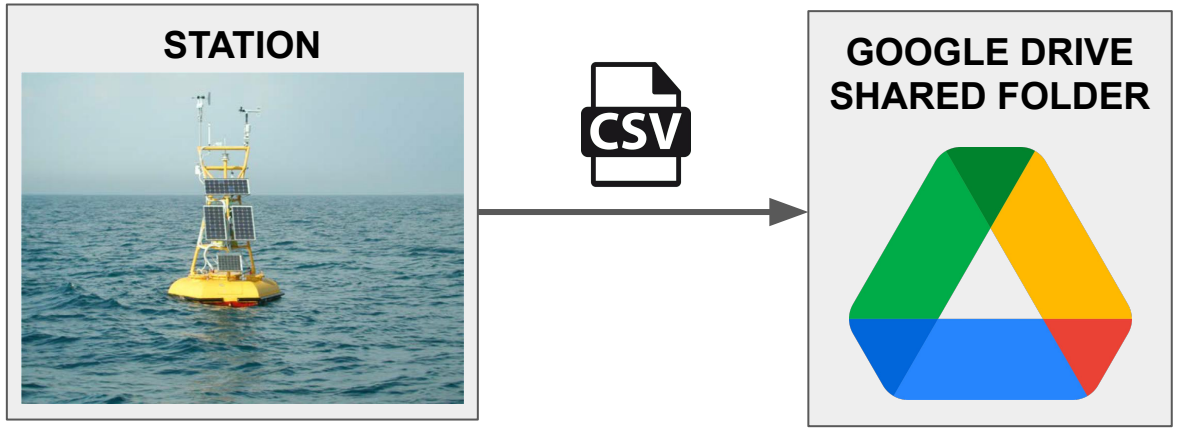
As data managers we want:

1) Store data in a standardized way
   - Databases are a great tool if used correctly
   - Metadata are essential

2) Make data available to the public
   - Services and API that let users download / interpolate / graph data



Data flow - format conversion



Stations collect data and save it in CSV ASCII files

Let's see an Example:

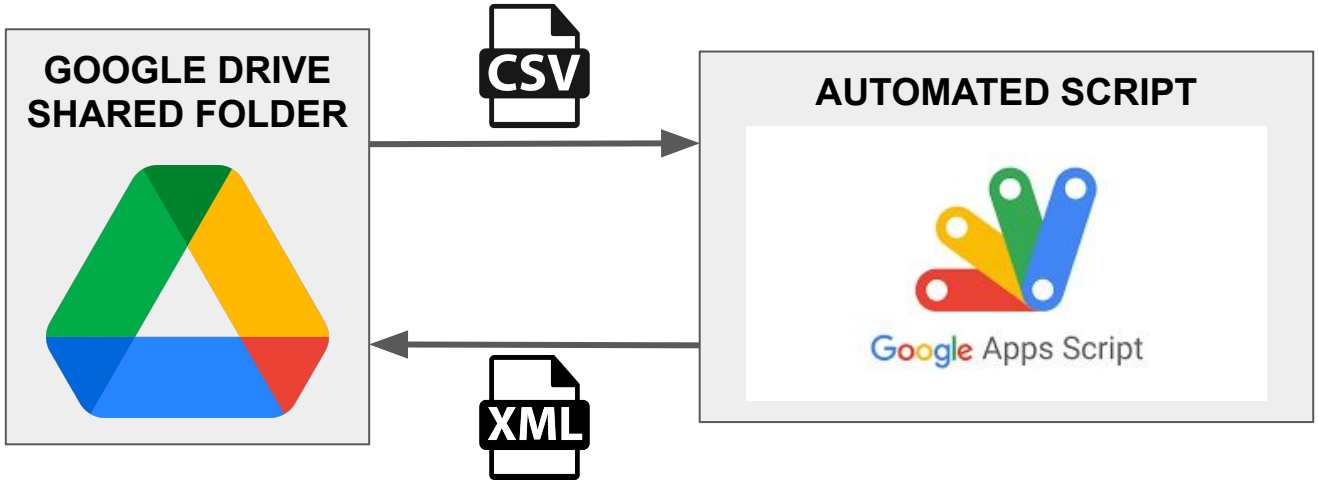
```
$SEAFET,1711252810,2024-03-24T04:00:10.827714Z,SEAFET02003,2024-03-24T
$PCO2PROW,1711252823,2024-03-24T04:00:23.783219Z,W M,2024,03,24,04,00,
$SBE37O,1711252849,2024-03-24T04:00:49.323829Z,13.1879,4.28663,1.935,6
$SBE16PLS01,1711252870,2024-03-24T04:01:10.546259Z,12.5855,4.30839,10.
$METEO,1711252919,2024-03-24T04:01:59.068983Z,68.04,15.32,7.58,859.46,
```

What do this characters and numbers mean?
How can i distinguish different data?

WE NEED METADATA!

Data flow - format conversion

To solve the previous problem we have made an automated script that converts data from the CSV ASCII format to a standardized XML



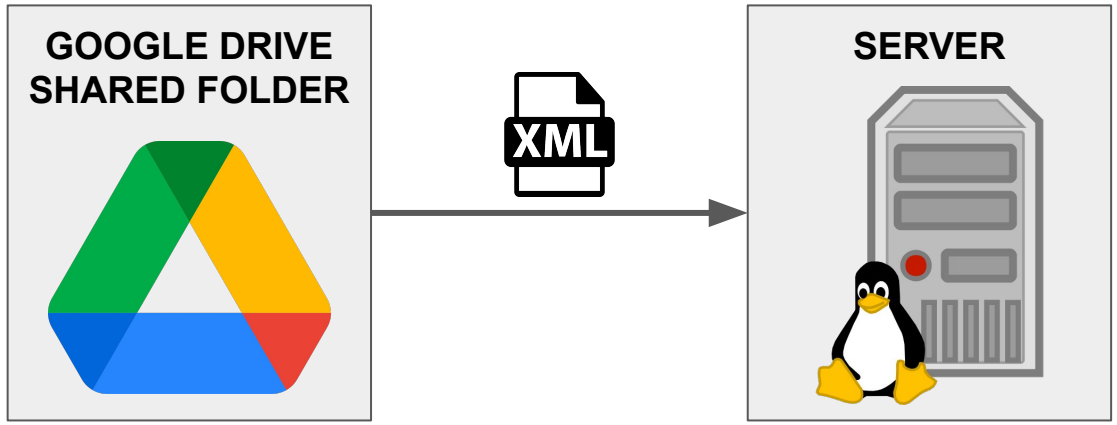
Let's see an Example of the XML:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<capture siteId="MAMBO1">
  <item line-number="1" field-name="temperature">
    <date>2024-03-11T23:01:10.000Z</date>
    <parameter>TEMPPR01</parameter>
    <parameter-unit>UPAA</parameter-unit>
    <instrument-category>130</instrument-category>
    <instrument-identifier>SBE16PLS01</instrument-identifier>
    <results-mean>11.9525</results-mean>
  </item>
  <item line-number="2" field-name="conductivity">
    <date>2024-03-11T23:01:10.000Z</date>
    <parameter>CNDCZZ01</parameter>
    <parameter-unit>UECA</parameter-unit>
    <instrument-category>130</instrument-category>
    <instrument-identifier>SBE16PLS01</instrument-identifier>
    <results-mean>4.28682</results-mean>
  </item>
  <item line-number="3" field-name="pressuredbar">
    <date>2024-03-11T23:01:10.000Z</date>
    <parameter>PRESPR01</parameter>
    <parameter-unit>UPDB</parameter-unit>
    <instrument-category>130</instrument-category>
    <instrument-identifier>SBE16PLS01</instrument-identifier>
    <results-mean>10.626</results-mean>
  </item>
  <item line-number="4" field-name="salinity">
    <date>2024-03-11T23:01:10.000Z</date>
    <parameter>SSALST01</parameter>
    <parameter-unit>UUUU</parameter-unit>
    <instrument-category>130</instrument-category>
    <instrument-identifier>SBE16PLS01</instrument-identifier>
    <results-mean>0.0000</results-mean>
  </item>
</capture>
```

IMPORTANT:

To make the script works correctly a configuration must be kept updated with the current station metadata

Data flow - database insertion



Data in XML format gets transferred into the main server at NODC.

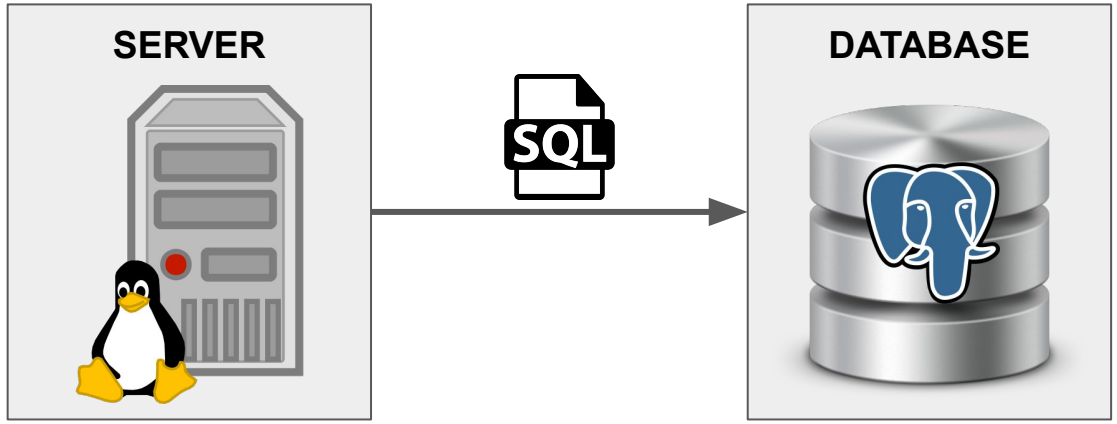
Note:

This step is done automatically by the server, every 5 minutes (**Near** Real-Time)

Keep in mind:

there is another configuration that specifies where the data is located and where it should be stored

Data flow - database insertion

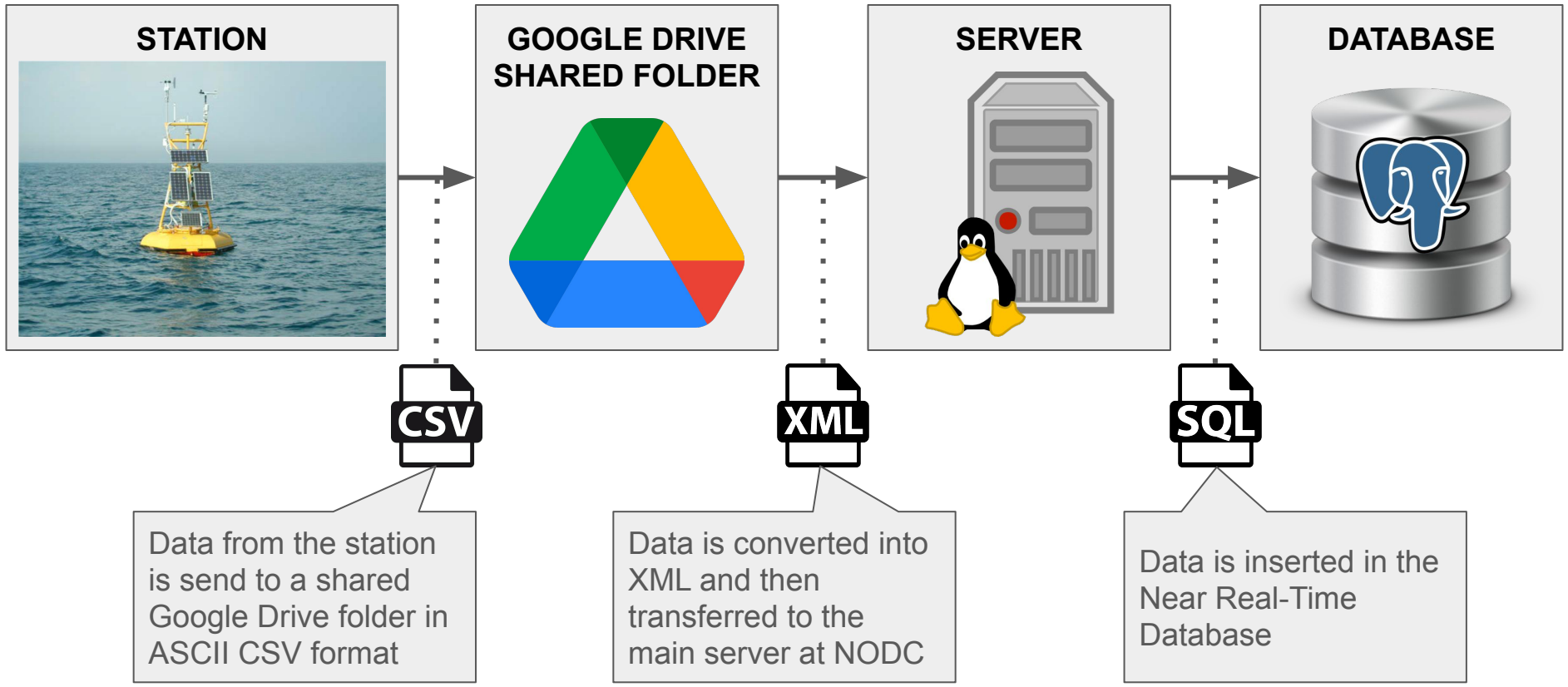


Data is taken from a custom software and inserted into the database

This custom software written in Java has been developed and now maintained internally in NODC

Data flow - format conversion and database insertion

RECAP of the data flow



Data publication - ERDDAP

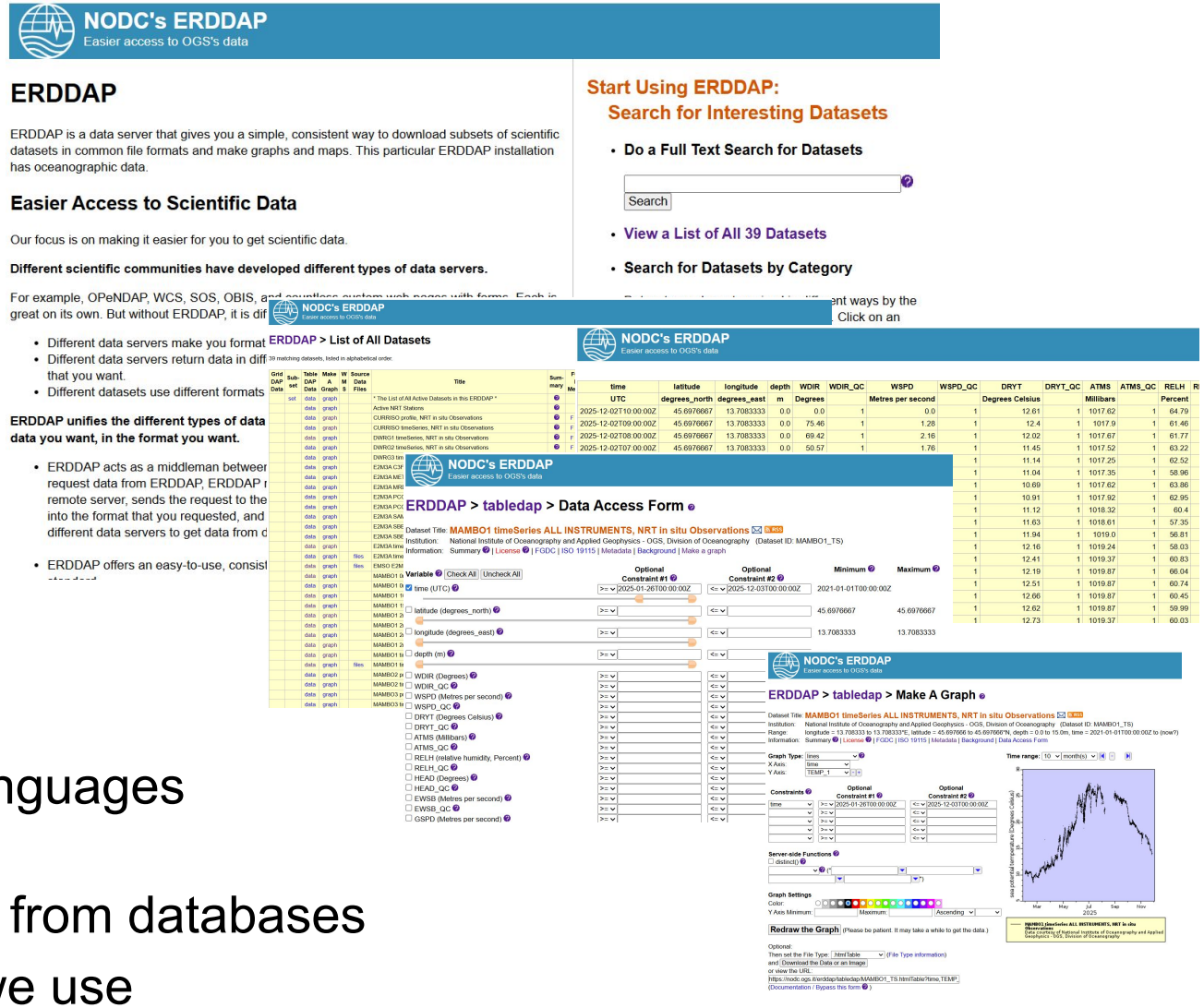
(As said previously)

We want to Make data available to the public

We choose ERDDAP™ data server

Why ERDDAP?

- Open source
  (developed by NOAA)
- Enables users to:
  - filter / interpolate
  - graph
  - download data (lot of standard formats)
- Built in API service
  (based on DAP, Data Access Protocol)
- Custom library for many programming languages
  (python, R, matlab, javascript)
- Easy configuration for publishing dataset from databases
- Customizable with metadata standards we use



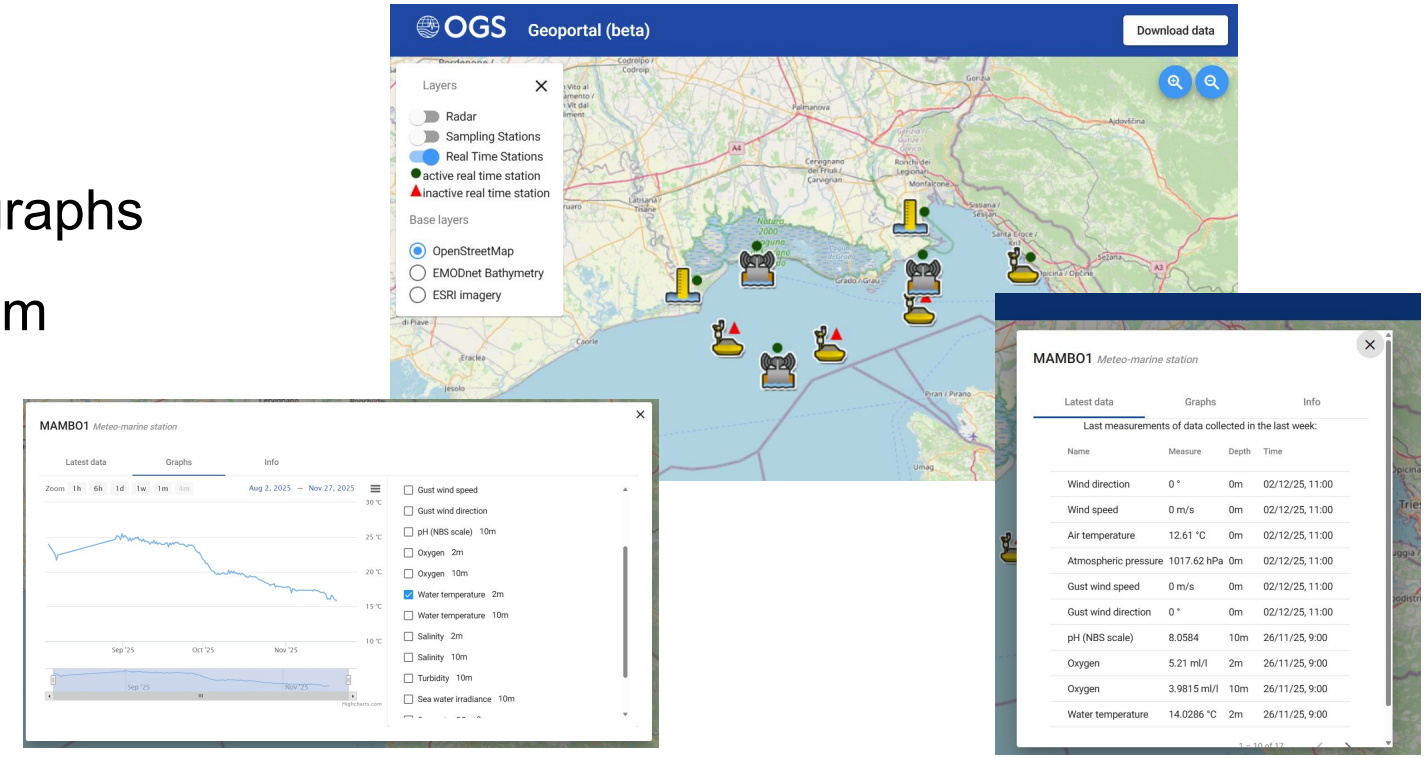
Data publication - NODC Geoportal

Data accessibility is an important theme with data publication!

= Not all users needs or know how to access data servers or APIs

That’s why NODC developed a new Geoportal with the following properties:

- Open source
- Web application
- Easy to use / visualize data in graphs
- Constantly upgraded by our team
- Uses ERDDAP as it’s API

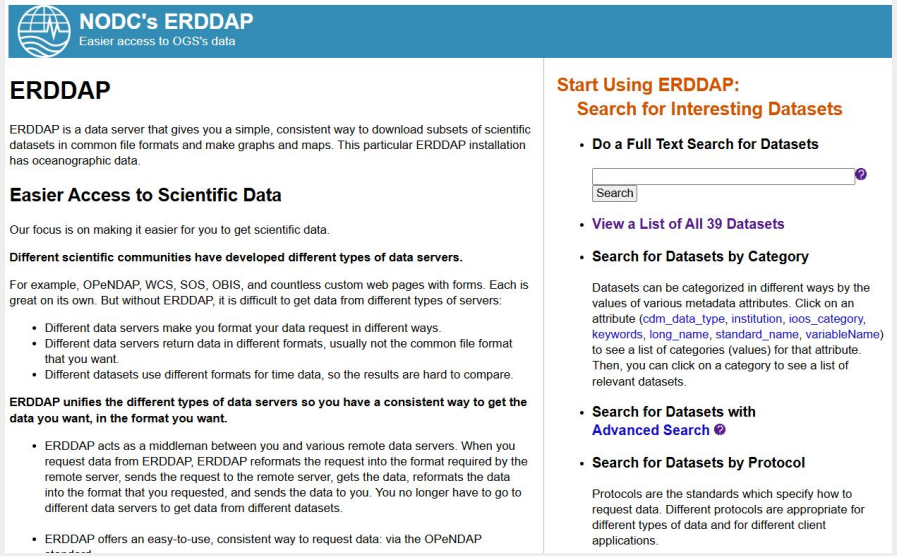


Data publication - ERDDAP and NODC Geoportal

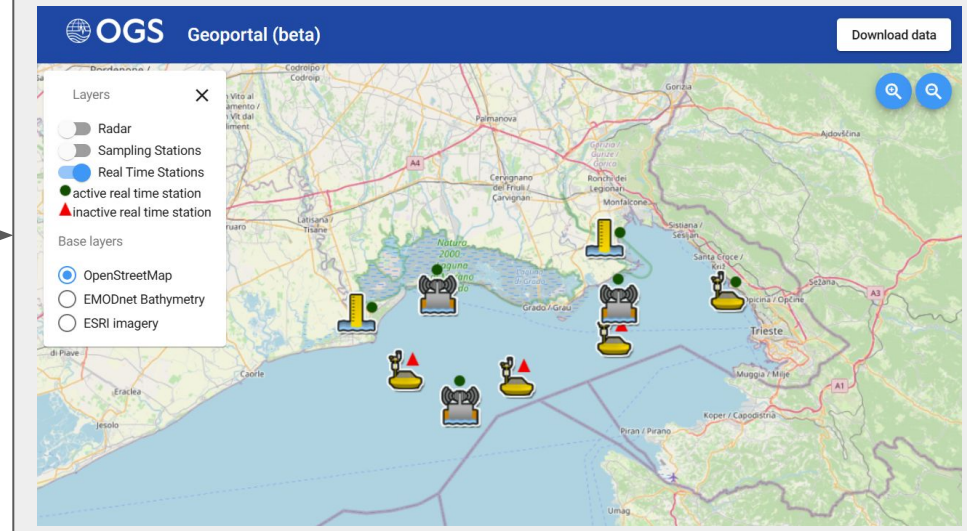
RECAP of the data publication

DATABASE

ERDDAP



GEOPORTAL



An open source data server that gives users the ability to filter, download and graph published data
(Developed by NOAA)

An open source web app that enables users to view data from stations managed by NODC
(Developed internally by NODC)

Why?

To make research data more open (and easier to reuse) and thus promote Open Science more generally

FAIR Principles

Box 2 | The FAIR Guiding Principles

https://www.nature.com/articles/sdata201618

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

ERDDAP is a data server designed to help organizations like NOAA comply with the FAIR (Findable, Accessible, Interoperable, Reusable) principles for scientific data management.
It provides a standardized framework that improves how data is shared and managed, primarily for oceanographic and environmental data.
ERDDAP supports each of the FAIR principles:

ERDDAP helps make data and metadata easy for both humans and machines to find.
Metadata Appending: It allows administrators to easily add missing or required metadata attributes to datasets in the configuration files, ensuring rich documentation.
Standardized Metadata: It can generate standardized metadata documents (such as ISO metadata or Schema.org) on demand, which are harvestable by external systems and search engines.
Persistent Identifiers: ERDDAP supports the creation of packages that can be assigned persistent identifiers, such as Digital Object Identifiers (DOIs), which are crucial for long-term discovery.

Once data is found, ERDDAP ensures it can be retrieved using standardized, open protocols, often through RESTful services.
Multiple Access Formats: It provides access to data in numerous file formats (e.g., .csv, .netCDF), allowing users to select the format that best suits their needs.
Machine-to-Machine Access: It supports machine-readable requests (e.g., OPeNDAP, WMS), allowing computational systems to access data with minimal human intervention.
Open Protocols: Access relies on standard web protocols like HTTPS.

Interoperability ensures data can be integrated and used with various applications and workflows.
ERDDAP acts as a bridge between different systems.
Standardized Units: It enforces the use of standardized, making data from different sources comparable and mergeable.
Middleware Approach: ERDDAP can pull data from different types of source servers (OPeNDAP, OBIS, SOS, WMS) and convert them into a consistent internal format, effectively making data sources interoperable.
Standard Vocabularies: It supports the use of community-accepted languages and controlled vocabularies for metadata, which facilitates data integration.

Reusable

The ultimate goal of FAIR is to maximize the utility of data by making it well-described so it can be reused in future research.

Rich Provenance: The detailed metadata capabilities allow users to include information on how the data was collected or generated (provenance).

Clear Licensing: ERDDAP facilitates the inclusion of clear, human- and machine-readable usage licenses, which defines the terms for reuse.

Through these features, ERDDAP simplifies the process of managing and sharing scientific data in a way that aligns with global best practices for open science.

FOCUS ON -> Interoperability

When are data interoperable? Only if the following hold:

- (Meta)data formats utilize shared **vocabularies** and **ontologies**
- (Meta)data are machine-readable and machine-actionable

FOCUS ON -> Interoperability -> metadata standard

European Directory of Marine Organisations (EDMO)

EDMO contains up-to-date addresses and activity profiles of research institutes, data holding centres, monitoring agencies, governmental and private organisations, that are in one way or another engaged in oceanographic and marine research activities, data & information management and/or data acquisition activities.
Currently, EDMO lists and describes more than 4.000 organisations.

institution_edmo_code	120
institution_edmo_uri	https://edmo.seadatanet.org/report/120

FOCUS ON -> Interoperability -> Standard Vocabularies

Common vocabularies consist of lists of standardised terms that cover a broad spectrum of disciplines of relevance to the oceanographic and wider community. Using standardised sets of terms solves the problem of ambiguities associated with data markup and also enables records to be interpreted by computers.

The BODC NVS2.0 URL follows the following structure:

Controlled Vocabulary: http://vocab. nerc.ac.uk/collection/{listid}/{version}/

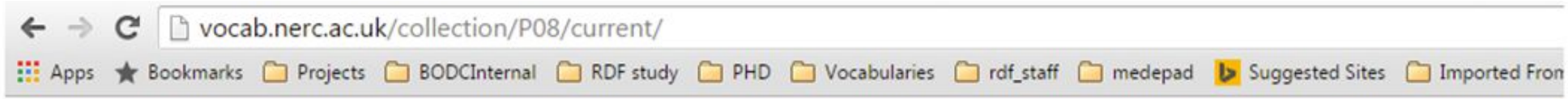
Concept: http://vocab. nerc.ac.uk/collection/{listid}/{version}/{termid}/

FOCUS ON -> Interoperability -> Standard Vocabularies

Extended Checklist

- Vocabularies MUST be documented
- Vocabularies SHOULD be self-descriptive
- Vocabularies SHOULD be described in more than one language
- Vocabularies SHOULD be used by other datasets
- Vocabularies SHOULD be accessible for a long period
- Vocabularies SHOULD be published by a trusted group or organization
- Vocabularies SHOULD have persistent URLs
- Vocabularies SHOULD provide a versioning policy

Reference: [https://www.w3.org/TR/ld-bp/#VOCABULARIES]



title-:	**SeaDataNet Parameter Disciplines**
alternative-:	**SeaDataNet Disciplines**
description-:	**Terms used to classify SeaDataNet Agreed Parameter Groups to provide topic/theme discovery interface.**
date-:	**2008-05-01 02:00:03.0**
publisher-:	**Natural Environment Research Council**
creator-:	**SeaDataNet**
versionInfo-:	**3**
RE_RegisterManager:	**British Oceanographic Data Centre**
RE_RegisterOwner:	**SeaDataNet**
comment-:	**Governance for vocabularies used in the EU SeaDataNet project implemented as cons members of the SeaDataNet Technical Task Team**

⇑ -- Administration and dimensions --

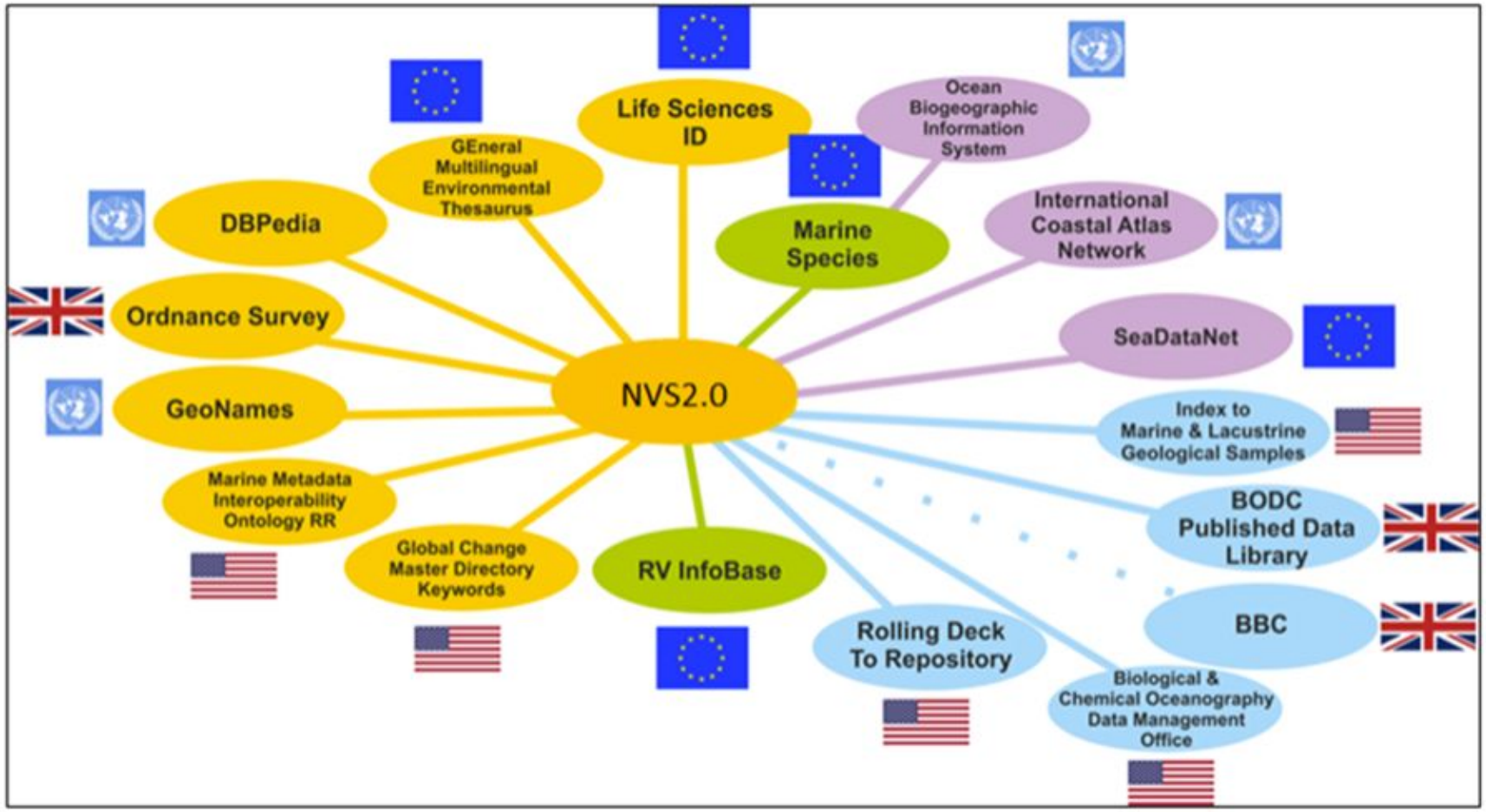
URI	http://vocab.nerc.ac.uk/collection/P08/current/DS07/
Identifier ()	SDN:P08::DS07
Preferred label (en)	**Administration and dimensions**
Alternative label ()	Administration and dimensions
Version Info ()	1
Definition (en)	Parameters related to spatial and temporal co-ordinates, entity referencing (eg record numbering
Deprecated ()	false
Same as ()	http://vocab.nerc.ac.uk/collection/P03/current/Z005/
Broader	http://vocab.nerc.ac.uk/collection/L19/current/001/
Narrower	http://vocab.nerc.ac.uk/collection/W01/current/035/
Date ()	2005-03-10 14:31:52.0

Semantic modeling of vocabularies

There are a few BODC vocabularies that are built from an underlying semantic model, which use other BODC vocabularies as the building blocks or components.
http://vocab.nerc.ac.uk/collection/P01/current/WB000284/

Concentration of lipids per unit wet weight of biota {Mytilus galloprovincialis (ITIS: 79456: WoRMS 140481) [Sex: male Subcomponent: gill]}

- ✔ **Measurement property** Concentration
- ✔ **Measurement statistical qualifier**
- ✔ **Chemical substance** lipids
- ✔ **Measurement-matrix relationship** per unit wet weight of
- ✔ **Matrix** biota
- ✔ **Biological entity** {Mytilus galloprovincialis (ITIS: 79456: WoRMS 140481) [Sex: male Subcomponent: gill]}



RDF Vocabulary resources

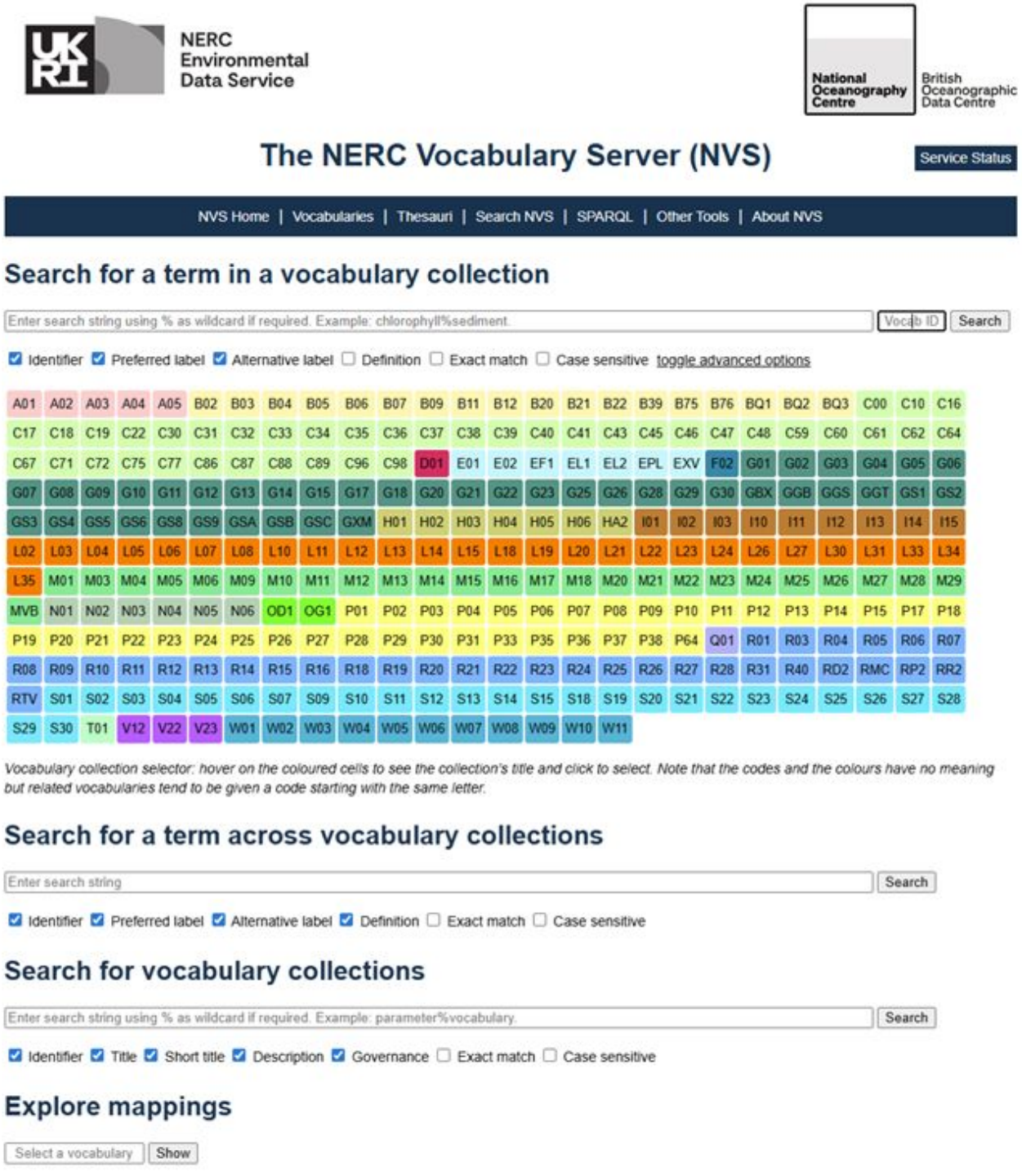
Non RDF Vocabulary resources

Non RDF datasets

RDF Datasets

Vocab Tools:

NVS2 search



Vocab Tools:

NVS2 search

SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

BODC VOCAB LIBRARY

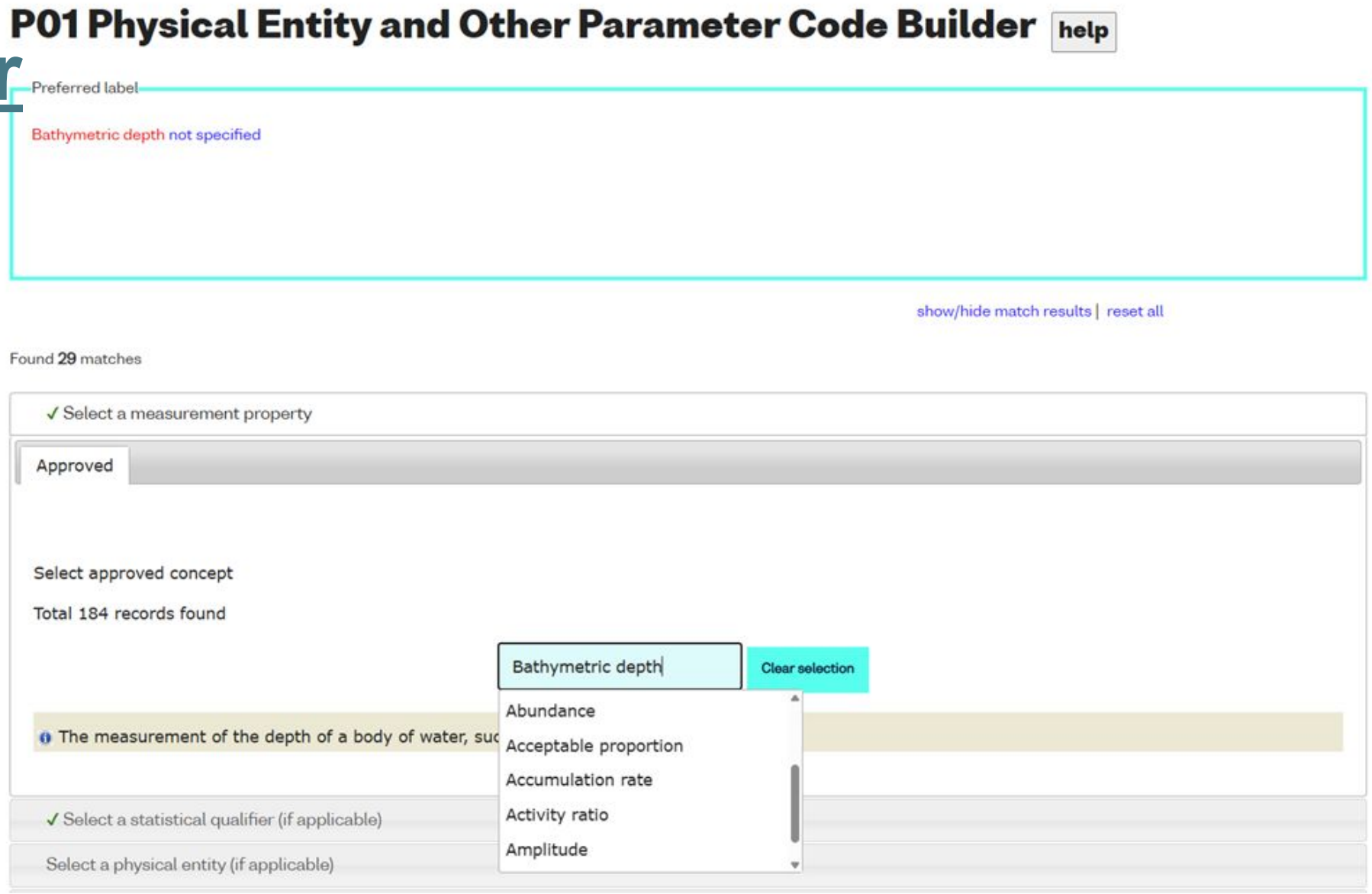
OGS

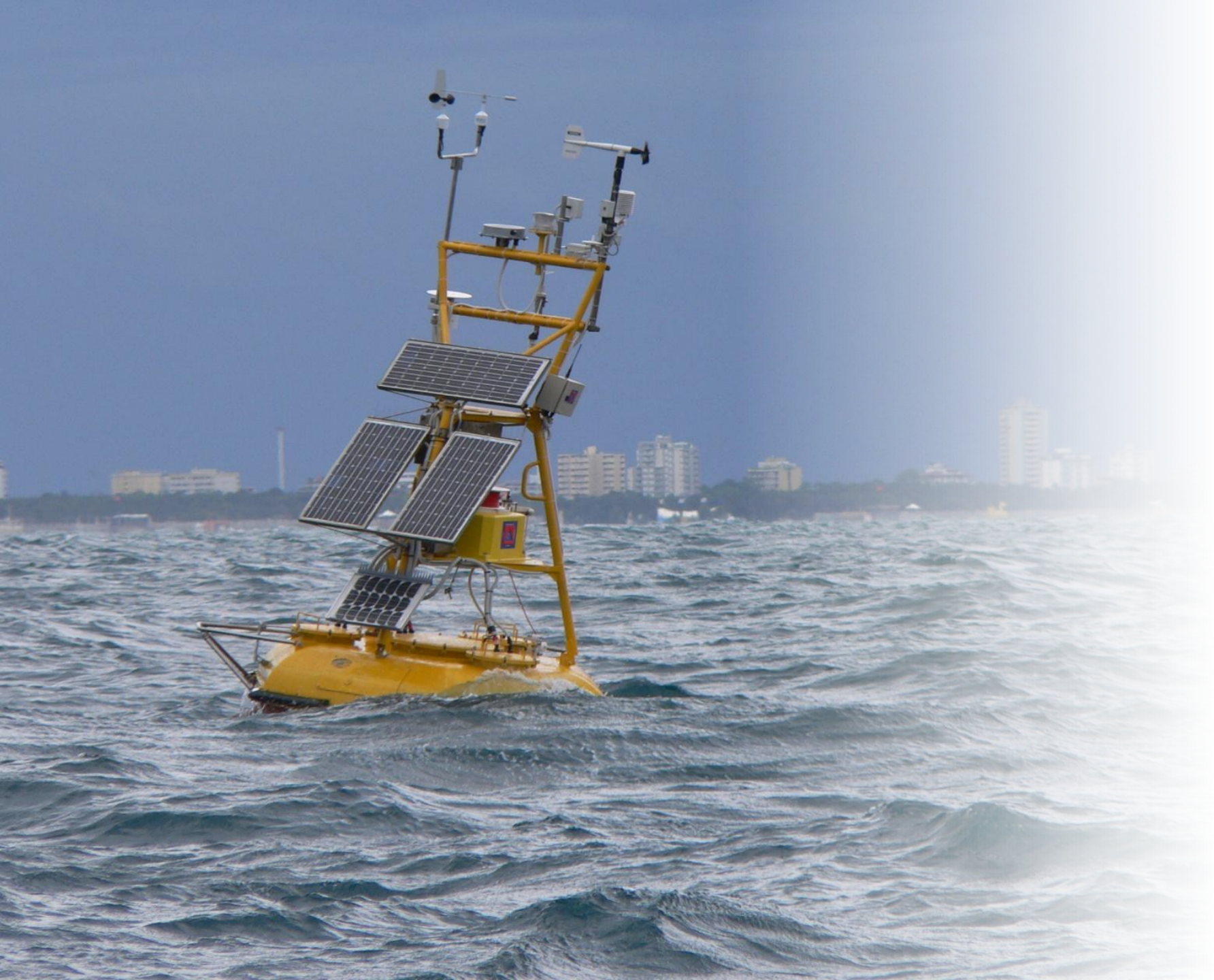
BODC WEBSERVICES V2 (LIBRARIES) CL12

Library	Thesaurus	Title	Alt Title	Version	Members	Modified
A05		AtlantOS Essential Variables	AtlantOS EVs	4	24	3/29/2019 2:01:22 AM
BQ1		EMODnet Bathymetry Quality Indicators of horizontal accuracy (QI_Horizontal)	EMODnet Bathy QI_Horizontal	1	4	7/15/2023 4:00:03 AM
BQ2		EMODnet Bathymetry Quality Indicators of vertical accuracy (QI_Vertical)	EMODnet Bathy QI_Vertical	1	5	7/15/2023 4:00:03 AM
BQ3		EMODnet Bathymetry Quality Indicators of survey purpose (QI_Purpose)	EMODnet Bathy QI_Purpose	1	4	7/15/2023 4:00:03 AM
C16		SeaDataNet sea areas	SDN sea areas	9	127	11/7/2012 2:00:06 AM
C17		ICES Platform Codes	ICES Platforms	1182	13423	11/12/2025 2:00:01 AM
C19		SeaVoX salt and fresh water body gazetteer	SeaVoX water bodies	26	271	8/7/2025 4:00:01 AM
C32		International Standards Organisation countries	ISO countries	10	282	11/18/2020 2:00:03 AM
C34		Activity purpose categories	Purpose categories	4	22	8/27/2011 3:00:05 AM
C35		European Nature Information System Level 3 Habitats	EUNIS3 Habitats	1	56	2/19/2010 2:01:37 AM
C36		Monitoring activity legislative drivers	Monitoring drivers	10	92	4/28/2022 4:00:03 AM
C37		Ten-degree Marsden Squares	Marsden-10	3	612	1/9/2009 2:00:05 AM
C38		SeaDataNet Ports Gazetteer	SeaDataNet Ports	92	4993	8/20/2025 4:00:00 AM
C39		World Meteorological Organisation sea states	WMO sea states	1	10	9/30/2009 3:01:08 AM
C45		Marine Strategy Framework Directive descriptors 2010/477/EU	MSFD descriptors 2010	3	11	2/25/2017 2:00:02 AM
C46		Marine Strategy Framework Directive criteria 2010/477/EU	MSFD criteria 2010	1	29	11/11/2010 2:00:07 AM
C47		Marine Strategy Framework Directive indicators 2010/477/EU	MSFD indicators 2010	1	56	11/11/2010 2:00:07 AM

Vocab Tools:

Vocab Builder





Thank you!

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

nffa-di

E-ARGO

beanTech
IT moves your business

CERIC

This Pilot training activity has been funded by the European Union – NextGenerationEU within the PNRR projects funded pursuant to Article 11, paragraph 1, of Notice 594/2024:

- “NFFA-DI cod. IR0000015, Missione 4, “Istruzione e Ricerca” – Componente 2, “Dalla ricerca all'impresa” – Linea di investimento 3.1, “Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione” – Azione 3.1.1, “Creazione di nuove IR o potenziamento di quelle esistenti che concorrono agli obiettivi di Eccellenza Scientifica di Horizon Europe e costituzione di reti” (CUP B53C22004310006).
- “EFC cod. SSU2024-00002, Missione 4 "Istruzione e ricerca" - Componente 1, “Potenziamento dell'offerta dei servizi all'istruzione: dagli asili nido all'universita'” - Investimento 3.4 “Didattica e competenze universitarie avanzate” - Sub-Investimento “Rafforzamento delle scuole universitarie superiori” (CUP: G97G24000100007).