living planet symposium | BONN 2022

The DYDAS – "DYnamic Data Analytics Services" platform for HPC big data analytics of Earth Observation and Geospatial data

Contacts: matteo.picchiani@gmatics.eu

M. Picchiani[(1)], M. Maranesi[(1)], I. G. Coltea[(2)], G. Pompei[(2)], L. di Giacomo[(2)], V. Pulieri[(2)], M. Mastrucci[(1)], P. Pasquali[(3)], G. Ponti[(4)], G. Sannino[(4)], R. Iacono[(4)], E. Napolitano(4), G. Scavino(5)

[(1)]GMATICS – Via F. 6, 00173 Rome, Italy
[(2)]K2 - Viale dell'Esperanto, 71, 00144, Rome, Italy
[(3)]LINKS Foundation - Via Pier Carlo Boggio, 61, 10138, Turin, Italy
[(4)]ENEA - Lungotevere Thaon di Revel, 76, 00196, Rome, Italy
[(5)]ANCI Lazio - Via dei Prefetti, 41, 00186 Rome, Italy

DYDAS
https://www.dydas.eu/

Nowadays, the ability to handle large amounts of data is related to the need for adequate infrastructure HPC (High Performance Computing) and related implementation techniques. It is in this context that the DYDAS project was born, funded by the Connecting Europe Facility (CEF) programme.

The CEF programme in the telecommunications sector is a key EU instrument. It facilitates international interaction between public administrations, businesses and citizens through the deployment of digital service infrastructures and broadband networks.

In line with the objective of the CEF 2018 work programme and the CEF-T-5 call, the DYDAS project will contribute to the European data infrastructure by improving the sharing and re-use of public and private data and by enabling the use of dynamic data sets such as Earth observation satellite and vehicle data into a HPC-based R&D platform.

The DYDAS project offers a platform capable of handling large volumes of dynamic data, enabling the public sector and industry to benefit from large-scale data analysis.

DYDAS promotes the **sharing and re-use of public and private data in a secure environment** and through innovative monetization mechanisms.

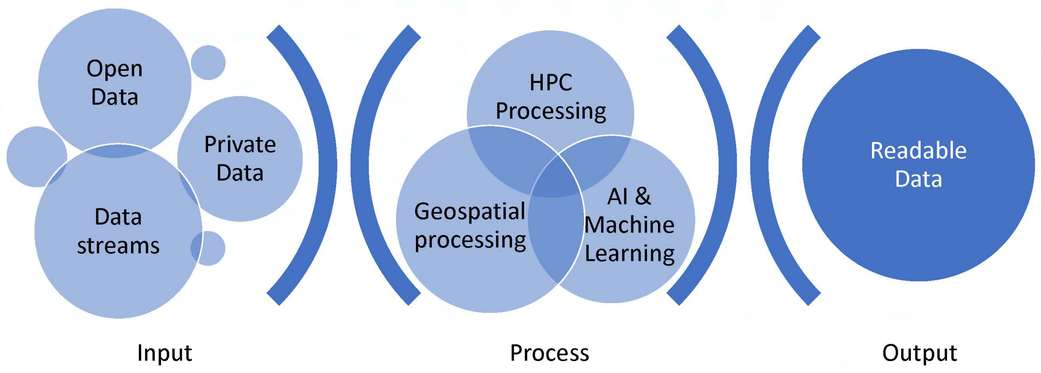
The platform acts as an e-marketplace for data access, and is equipped with HPC-enabled services based on Big Data technologies, machine learning, AI and advanced services.

Data flow management

DYDAS gathers up data from various sources into the Data Lake, ready for processing. Data can be analyzed by end users through custom data modelling and then view results on advanced data analytics dashboards configurable on the platform or through standard OGC APIs with any compatible tool.

Data processing – HPC Infrastructure

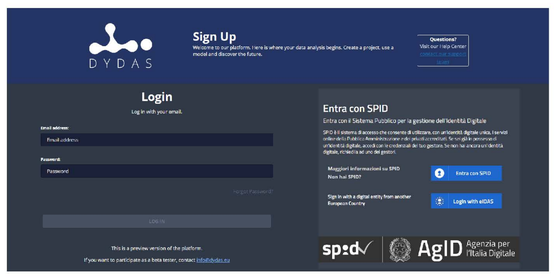
A key factor in data analysis is the expertise to interpret and scattered data into useful information and knowledge that can benefit community and business.
The platform allows to interact with the data gathered into the Data Lake, both open data and private datasets, and develop custom models to obtain readable and ready to use data. Resulting data can be shared or monetized, and presented through customizable dashboards as a service to clients or community.



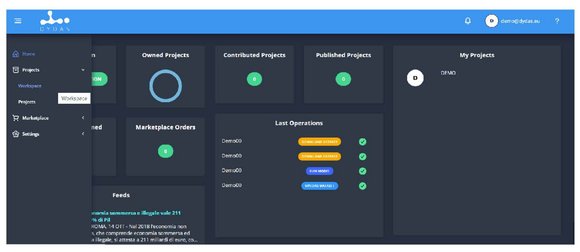
DYDAS Portal

The DYDAS portal is in the pre-operational phase and can be already accessed and used at the link: *https://preview.dydas.eu/*

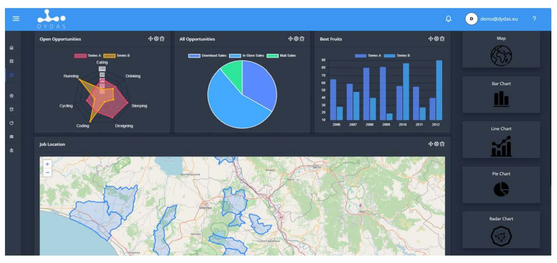
Login page



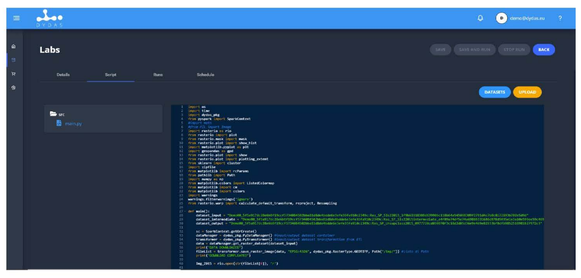
User page



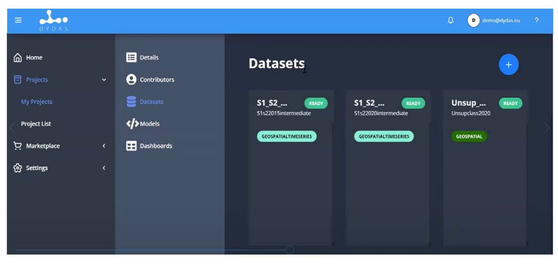
Dashboard for data analysis



Docker for model runs



Data Management

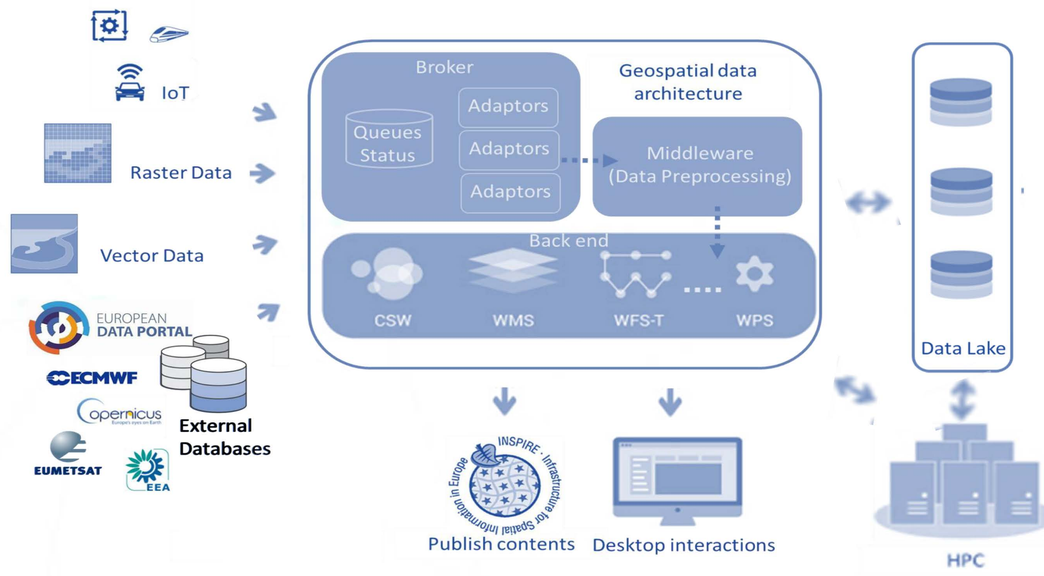


Supported data formats

Vector	Raster
WKT	GeoTIFF
GeoJSON	Georeferenced Jpeg 2000
Shapefile	NetCDF
KML, KMZ	GeoPackage
GeoPackage	HDF4, HDF5
	Grib
	Bufr
	ESA Copernicus SAFE format

DYDAS Geospatial Data Architecture

A key and differentiating element of the project is the implementation of a Geospatial Data architecture. The DYDAS GDA allow the seamless integration and processing of large multidimensional data sets for innovative use modes, through the adoption of a geospatial data model and interoperability rules.



Land Cover use case

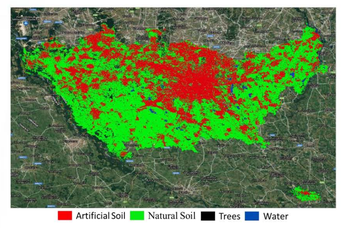
A supervised Neural Network model has been trained using the **DYDAS HPC** with **Sentinel-1 (GRD)** and **Sentinel-2 (L2a)** data. The access to Sentinel's data is directly provided by the DYDAS platform.

A shallow NN topology: 11 – 10 – 5 – 4.

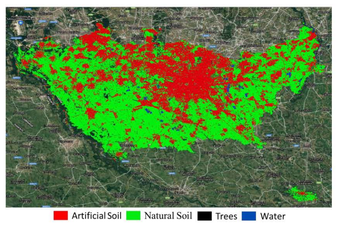
NN input layer:
S1 - sigma$_0$ VH+VV pol.
S2 - B2, B3, B4, B5, B6, B7, B8, B11 and B12

We considered here four macro classes as: **Natural Soil** (including crops, grassland and bare soil), **Trees**, **Water** and **Artificial Soil**.

As training information the CORINE Land Cover of 2018 has been considered.



Reference map derived from the CORINE Land Cover data 2018



Land Cover map achieved by the S1-S2 input data

The **pre-processing** operation runs in few **milli-seconds**.
The NN training has been performed on the HPC infrastructure by using the DYDAS docker with a Python –Tensorflow model (runs **for 1000 epochs < 1 minuts**).
The NN output has been compared with the CORINE reference map, registering an **overall accuracy** of **98%** and **K-coefficient** equal to about **0.89**.

Maritime use case

A primary DYDAS use case has been performed by ENEA by running two **forecast model** of **Sea State** onto the DYDAS HPC platform:

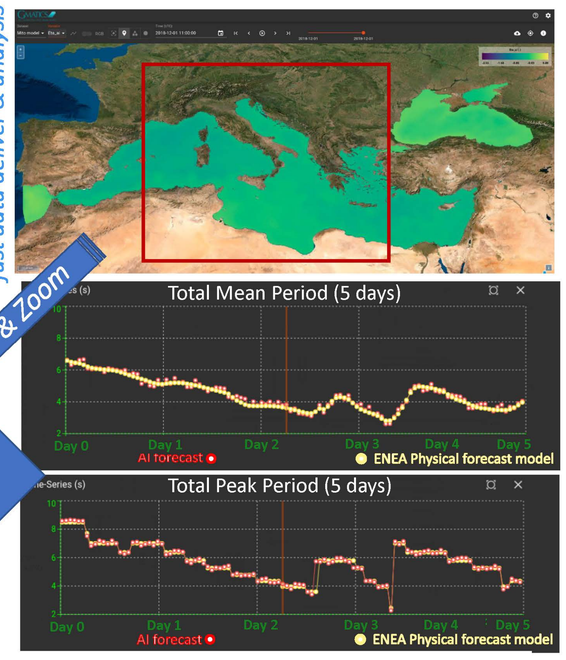
- **MITO Model** - **Velocity**, **temperature** and **salinity** forecast are computed for the **Mediterranean Sea** and **Black Sea** system (5 days/hourly forecast, 2km)
- WAVES - **Significant wave height**, **mean wave period** and **wave energy forecast** are computed over the whole **Mediterranean Sea** (5 days/hourly forecast, 2km)

A LSTM models that replicate the ENEA MITO and WAVE physical forecast models have been also trained and tested over the DYDAS platform producing forecasts over the next 5 days with hourly temporal resolution and 1km of spatial resolution

122.586 core hours have been consumed and 44 TB of data have been processed in order to reduce the AI model runtime for the whole Mediterranean Sea to 15 minutes.

RMSE between AI and ENEA model **< 5%**
R^2 > 90%

External User Interface for fast data deliver & analysis



Co-financed by the European Union
Connecting Europe Facility