



## INDIGO-DataCloud solutions for Earth Sciences

Fernando Aguilar Gómez (1), Jesús Marco de Lucas (1), Sandro Fiore (2), Stephen Monna (3), and Yin Chen (4)  
(1) IFCA-CSIC, Santander, Spain (aguilarf@ifca.unican.es), (2) CMCC Foundation, Italy, (3) INGV, Italy, (4) EGI.eu

INDIGO-DataCloud (<https://www.indigo-datacloud.eu/>) is a European Commission funded project aiming to develop a data and computing platform targeting scientific communities, deployable on multiple hardware and provisioned over hybrid (private or public) e-infrastructures.

The development of INDIGO solutions covers the different layers in cloud computing (IaaS, PaaS, SaaS), and provides tools to exploit resources like HPC or GPGPUs.

INDIGO is oriented to support European Scientific research communities, that are well represented in the project. Twelve different Case Studies have been analyzed in detail from different fields: Biological & Medical sciences, Social sciences & Humanities, Environmental and Earth sciences and Physics & Astrophysics.

INDIGO-DataCloud provides solutions to emerging challenges in Earth Science like:

-Enabling an easy deployment of community services at different cloud sites. Many Earth Science research infrastructures often involve distributed observation stations across countries, and also have distributed data centers to support the corresponding data acquisition and curation. There is a need to easily deploy new data center services while the research infrastructure continuous spans.

As an example: LifeWatch (ESFRI, Ecosystems and Biodiversity) uses INDIGO solutions to manage the deployment of services to perform complex hydrodynamics and water quality modelling over a Cloud Computing environment, predicting algae blooms, using the Docker technology: TOSCA requirement description, Docker repository, Orchestrator for deployment, AAI (AuthN, AuthZ) and OneData (Distributed Storage System).

-Supporting Big Data Analysis. Nowadays, many Earth Science research communities produce large amounts of data and are challenged by the difficulties of processing and analysing it.

A climate models intercomparison data analysis case study for the European Network for Earth System Modelling (ENES) community has been setup, based on the Ophidia big data analysis framework and the Kepler workflow management system. Such services normally involve a large and distributed set of data and computing resources. In this regard, this case study exploits the INDIGO PaaS for a flexible and dynamic allocation of the resources at the infrastructural level.

-Providing Distributed Data Storage Solutions. In order to allow scientific communities to perform heavy computation on huge datasets, INDIGO provides global data access solutions allowing researchers to access data in a distributed environment like fashion regardless of its location, and also to publish and share their research results with public or close communities.

INDIGO solutions that support the access to distributed data storage (OneData) are being tested on EMSO infrastructure (Ocean Sciences and Geohazards) data. Another aspect of interest for the EMSO community is in efficient data processing by exploiting INDIGO services like PaaS Orchestrator.

Further, for HPC exploitation, a new solution named Udocker has been implemented, enabling users to execute docker containers in supercomputers, without requiring administration privileges.

This presentation will overview INDIGO solutions that are interesting and useful for Earth science communities and will show how they can be applied to other Case Studies.