

deep

Hybrid DataCloud

DESIGNING AND ENABLING E-INFRASTRUCTURES FOR INTENSIVE PROCESSING IN A HYBRID DATA CLOUD

Editorial

Use Cases

DEEP ROSSETA:

Second Software Release

Team Members (Part 2)

DEEP Testbed

Collaboration partnership with EOSC DIH

DEEP support to fight COVID-19

Past events



This is the second and the final issue of the Newsletter of the DEEP Hybrid DataCloud project, a 30 months research and innovation action financed by the European Commission under the H2020 Research and Innovation Programme in the topic of Platform driven e-infrastructure innovation. In this issue, we will present the project use cases, also provide an update of the project activities, final results, collaboration initiatives, and assisted events so far.

All Newsletters and other related material are available in the project website:



www.deep-hybrid-datacloud.eu

Don't forget to follow us on social media channels and stay always tuned for new updates and news!



[@DEEP_eu](https://twitter.com/DEEP_eu)



[DEEP Hybrid Data Cloud](#)

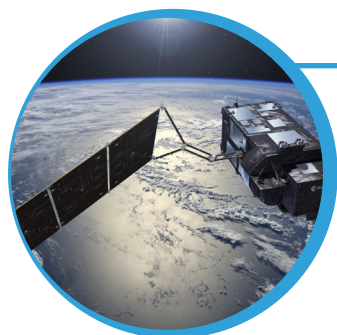


[YouTube channel](#)

USE CASES

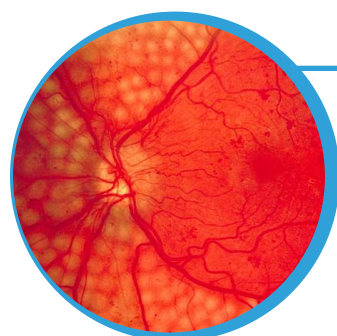
DEEP use cases as deep learning modules are publicly available as Open Source in the DEEP Open Catalogue and publicly accessible in the DEEP Open

Market. They are categorized in the following main extendable groups:



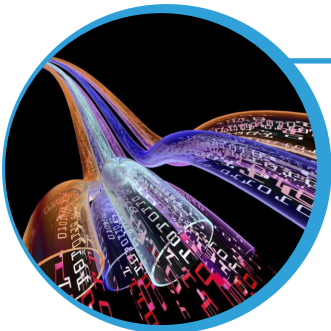
Earth observations

deep neural network applications to perform pattern recognition on satellite images. They can be combined with other in-situ measurements for ecosystems and biodiversity to perform tasks such as remote object detection, terrain segmentation or meteorological prediction. Currently we offer a super-resolution module to upscale low-resolution bands to high resolution for the most popular multispectral satellites from all around the world.



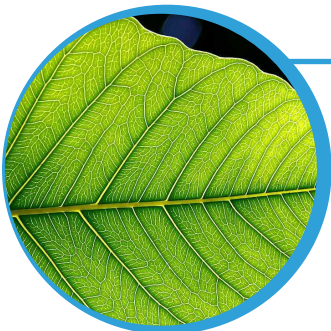
Biological and medical science

deep learning modules for biomedical image analysis have opened new opportunities in how diseases are diagnosed and treated. We currently provide a retinopathy automated classification based on color fundus retinal photography images.



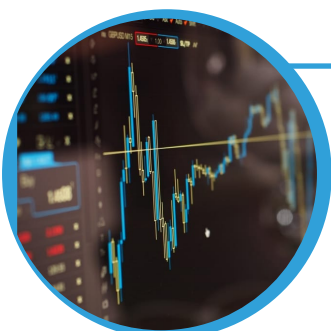
Cyber-security and network monitoring

we provide modules with co-functions enhancements for surveillance Intrusion Detection Systems supervising traffic network flows of computing infrastructure.



Citizen science

deep learning modules for leveraging citizen science in large-scale biodiversity monitoring. The available modules include automatic identification of species from images for a wide range of categories (plants, seeds, conus and phytoplankton). Most of these modules are also available as mobile applications.



General purpose

this includes modules that can be used across a wide range of domains. We currently provide modules for image classification, audio classification, speech-to-text synthesis and pose detection.



In February 2020, our project announces the availability of the second software release and platform. DEEP Rosetta expands on the first version of software generated by the project, called DEEP Genesis, enlarging its functionalities to cover the whole machine learning cycle, enhancing the stability of the different components and adding new features.

All the changes in this new release are oriented towards the common project goal of easing the path for the scientific communities to develop, build and deploy complex models as a service at their local laptop, on a production server or on top of e-Infrastructures supporting the DEEP-Hybrid-DataCloud stack.

As in the previous release, the DEEP components are integrated into a comprehensive and flexible architecture that can be deployed and exploited following the user requirements.

A comprehensive platform for machine learning, deep learning, and artificial intelligence in the European Open Science Cloud. Developing, training, sharing, and deploying your model has never been easier, [click here!](#)

The **DEEP training facility**, accessible through the DEEP training dashboard allows data scientists to develop and train their models, with access to latest generation EU computing e-Infrastructures.

DEEP as a Service is a fully managed service that allows to easily and automatically deploy the developed applications as services, with horizontal scalability thanks to a serverless approach. The pre-trained applications that are published in the catalog are automatically deployed as services to make them available for general use.

DEEP Open Catalog and marketplace comprises a curated set of applications ready to use or extend, fostering knowledge exchange and re-usability of applications. This open exchange aims to serve as a central knowledge hub

for machine learning applications that leverage the DEEP-Hybrid-DataCloud stack, breaking knowledge barriers across distributed teams. Moreover, pre-configured Docker containers, repository templates and other related components and tools are also part of this catalog.

[DEEPaaS API](#) enables data scientists to expose their applications through an HTTP endpoint, delivering a common interface for machine learning, deep learning and artificial intelligence applications.

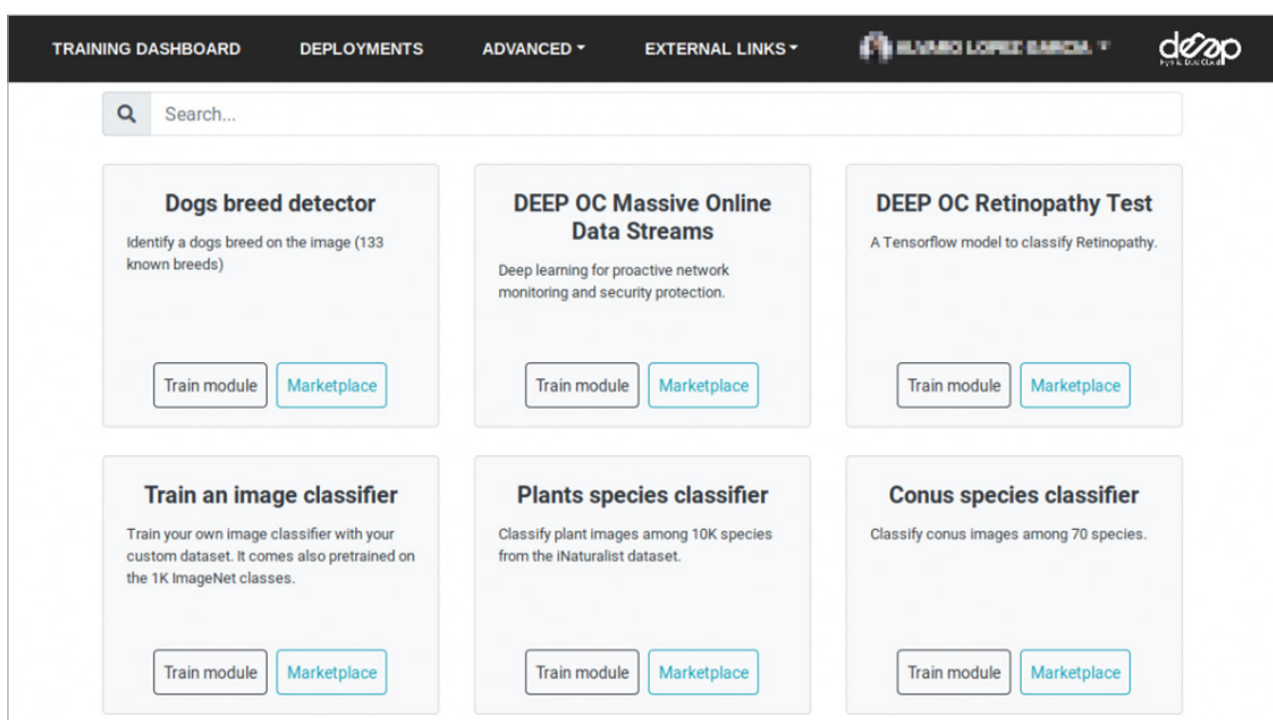
The platform has been extended to support asynchronous training, allowing to launch, monitor, stop and delete the training directly from your web browser. The trained models to perform inference can now be chosen from the models available in the training history. All the documentation on these new

features has been accordingly updated [here](#).

In addition, the user-friendly training dashboard allows now to easily and transparently deploy the modules in a cloud environment. From the dashboard, the user can choose the resources needed for the deployment in terms of memory, type of processing unit (CPU or GPU), the storage client to be used or even to manually configure the scheduling.

[The DEEP training dashboard](#) allows to easily train any existing modules, or your own one.

[The DEEP Open Catalogue](#) has been renewed with a more appealing design, improving the organisation of the modules and the general site interactivity.



The DEEP Rosetta release consists of



10 products

distributed via 22 software packages and tarballs supporting the CentOS 7, Ubuntu 16.04 and 18.04 operating systems.



15 fully containerised ready-to-use models

from a variety of domains available at the DEEP Open Catalogue.



The release notes can be found [here](#). The full list of products together with the installation, configuration guides and documentation can be consulted [here](#).



TEAM MEMEBERS (PART 2)



Atos

Spain

Atos is a global leader in digital transformation with approximately 120,000 employees in 73 countries and annual revenue of around € 12 billion. The European number one in Big Data, Cybersecurity, High-Performance Computing, and Digital Workplace, The Group provides Cloud services, Infrastructure & Data Management, Business & Platform solutions, as well as transactional services through

Worldline, the European leader in the payment industry. With its cutting-edge technologies, digital expertise and industry knowledge, Atos support the digital transformation of its clients across various business sectors: Defense, Financial Services, Health, Manufacturing, Media, Energy & Utilities, Public sector, Retail, Telecommunications, and Transportation.



CESNET

CESNET ZAJMOVE SDRUZENI PRAVNICKYCH OSOB
Czech Republic

CESNET, Association of Legal Entities, is a public non-profit organization established in 1996 by all the public universities and the Academy of Science of the Czech Republic to do research, development and deployment of advanced network technologies and applications. Serving as a recognized Czech NREN and NGI, CESNET's funding is guaranteed by a combination of membership fees, payment for services by non-members, and a governmental contribution. CESNET is a member of several international organizations, including DANTE, TERENA, CEENet (Central and Eastern European

Networking Association), GLIF (Global Lambda Integrated Facility), and EGI.eu (European Grid Infrastructure). CESNET is also involved in many national and international projects, including GEANT and EGEE series of projects, Phosphorus, Lobster, EUAsiaGrid. CESNET was a coordinator of the EGI Design Study project and is currently involved in EGI InSPIRE, EMI and CHAIN projects. In its role as a National Grid Initiative, CESNET not only contributes grid and cloud resources to the European Grid Infrastructure, but also operates a separate Czech national grid and cloud environment.



IISAS

USTAV INFORMATIKY, SLOVENSKA AKADEMIA VIED
Slovenia

The Institute of Informatics carries out excellent research in accordance with the current trends in Applied Informatics, Cybernetics and Microelectronics. The research is conducted by national and international cooperation, especially within the European research area for the long

period of FP5, FP6 and FP7 programs. The focus of our Institute is to apply the research results into practice. The Institute has been for a long term involved in building a strategic pan-European infrastructure for High Performance Computing.



HMGU

HELMHOLTZ ZENTRUM MUENCHEN DEUTSCHES FORSCHUNGSZENTRUM
FUER GESUNDHEIT UND UMWELT GMBH *Germany*

Helmholtz Zentrum München is the German Research Center for Environmental Health. It investigates important common diseases which develop from the interaction of lifestyle, environmental factors and personal genetic background, focusing particularly on diabetes mellitus, allergies and chronic lung diseases.

The mission of Helmholtz Zentrum München is to find new personalized medical solutions for the prevention, diagnosis and therapy of environmentally triggered diseases to promote a healthier society in a rapidly changing world.



UPV

UNIVERSITAT POLITECNICA DE VALENCIA
Spain

The Grid and High-Performance Computing Group (GRyCAP) from the UPV, has a very strong background in Distributed, Cloud Computing and Data Management, especially in the area of Biomedicine. GRyCAP has shared and extended its expertise through active participation in over 30 national and European R&D projects on HPC, Grid and Cloud technologies. These projects include HPCN-TTN Network, EUTIST-M and TT@MED – technology dissemination; EGEE-I, II, III, EGI- InSPIRE and VENUS-C – e-infrastructures. It is one of the founders of the Institute for Instrumentation for Molecular Imaging (I3M); took part in the International HealthGrid Association; has developed a roadmap on the use of grids in health (SHARE), and boasts a solid research line in the distributed storage of medical imaging data through the TRENCADIS platform (CVIMO project). Relevant GRyCAP expertise includes:

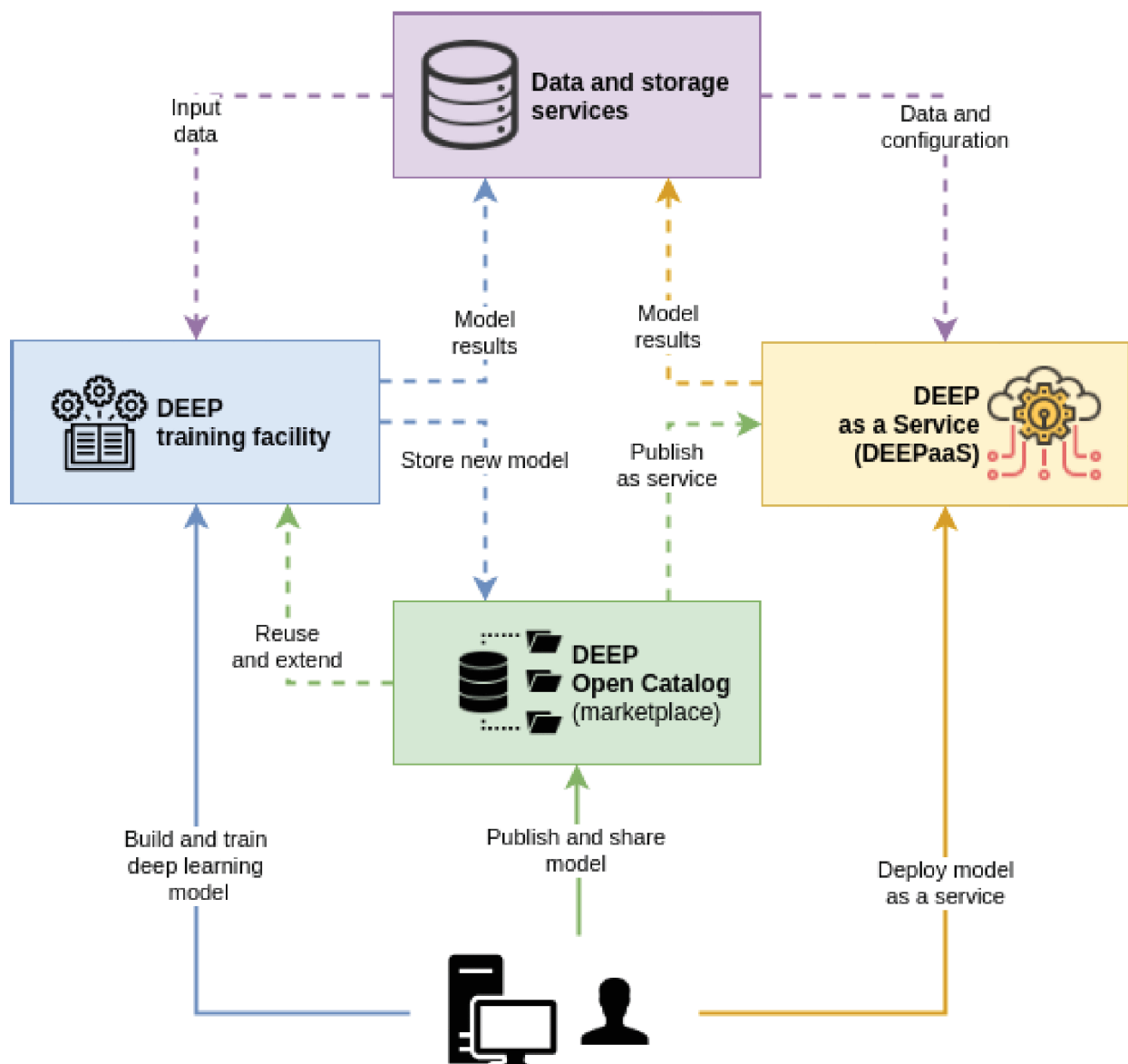
the coordination of the cloud end-user community (27 applications, 20 coming from an Open Call) in the VENUS-C project, one of the first cloud projects to feature in the DAE, the development of distributed and GridComputing platforms with Latin American countries: CyTED-GRiD, EELA and EELA-2, the EU-BrazilOpenBio Project and the EUBrazil Cloud Connect project (the UPV is the coordinator), and development of high-level middleware components for cloud infrastructures in the Spanish Codecloud project (and the recently approved CLUVIEM project) and coordination of the Spanish Network for e-Science to implement the Spanish National Grid Initiative (NGI). The UPV has a close collaboration with the bioinformatics institute of the Research centre Prince Phillip (CIPF) and the Center for Public Health Research (CSISP).

DEEP Preview-Testbed exposing enhanced DEEP-Rosetta services is now available for users to experience the ease of building, developing, training and deploying machine learning models and exploiting the new DEEP training dashboard.

Although called a testbed, in reality it is a small-scale production infrastructure, all software components and services are running the versions released in the DEEP-2 – coded named Rosetta – and are operated and managed as any other production service or platform. A

diagram of the Pilot Preview is shown in the figure below.

Resources made available by project partners are nonetheless significant, one can exploit about 30 high end NVIDIA GPUs distributed across 3 cloud e-infrastructures and a data/storage management system that is federated between 3 providers with about 80TB of total storage. One of the main features is the data locality to the computing resources, allowing a more efficient computation.



Other features worthy to mention are: the cloud resource providers are part of the production EGI Fedcloud infrastructure, the data/storage management system is the result of a tight collaboration between DEEP-HybridDataCloud and eXtreme DataCloud – XDC projects, where storage resources from both projects are federated through the Onedata service (bottom of the diagram); users

are authenticated and authorized through the Federated AAI service called DEEP-IAM (right side of the diagram).

Finally, and the most important highlight, the users can execute ML/AI applications in a production mode with long training of the models and using large datasets (some cases of the order of TBs).

COLLABORATION PARTNERSHIP BETWEEN DEEP-HYBRID-DATACLOUD AND EOSC DIH

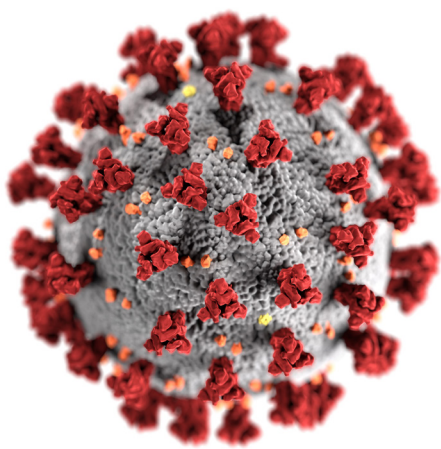
We are pleased to announce that DEEP-Hybrid-DataCloud consortium has signed a collaboration agreement with [EOSC-DIH](#) aiming at boosting the dissemination of DEEP offering and fostering the adoption of project solutions by SMEs. This collaboration has allowed to perform an industrial pilot with [BI Insight](#) which will increase

the acceptance in the market of DEEP solutions.

The EOSC DIH is a mechanism for private companies to collaborate with public sector institutions to access technical services, research data, and human capital.



The last years, artificial intelligence, and more concretely, deep learning, has proved to be a very useful tool for biomedical research, medical related problems and clinical assistance. In the current situation of health emergency a massive amount of data is being produced and need to be understood using the most powerful tools available. The DEEP project is contributing to fight the COVID-19 emergency on different fronts thanks to its capacity to process huge amounts of data, to develop and share deep learning applications in a quick and easy way, and to the resources available at the project testbed. Currently, we are involved in the following initiatives:



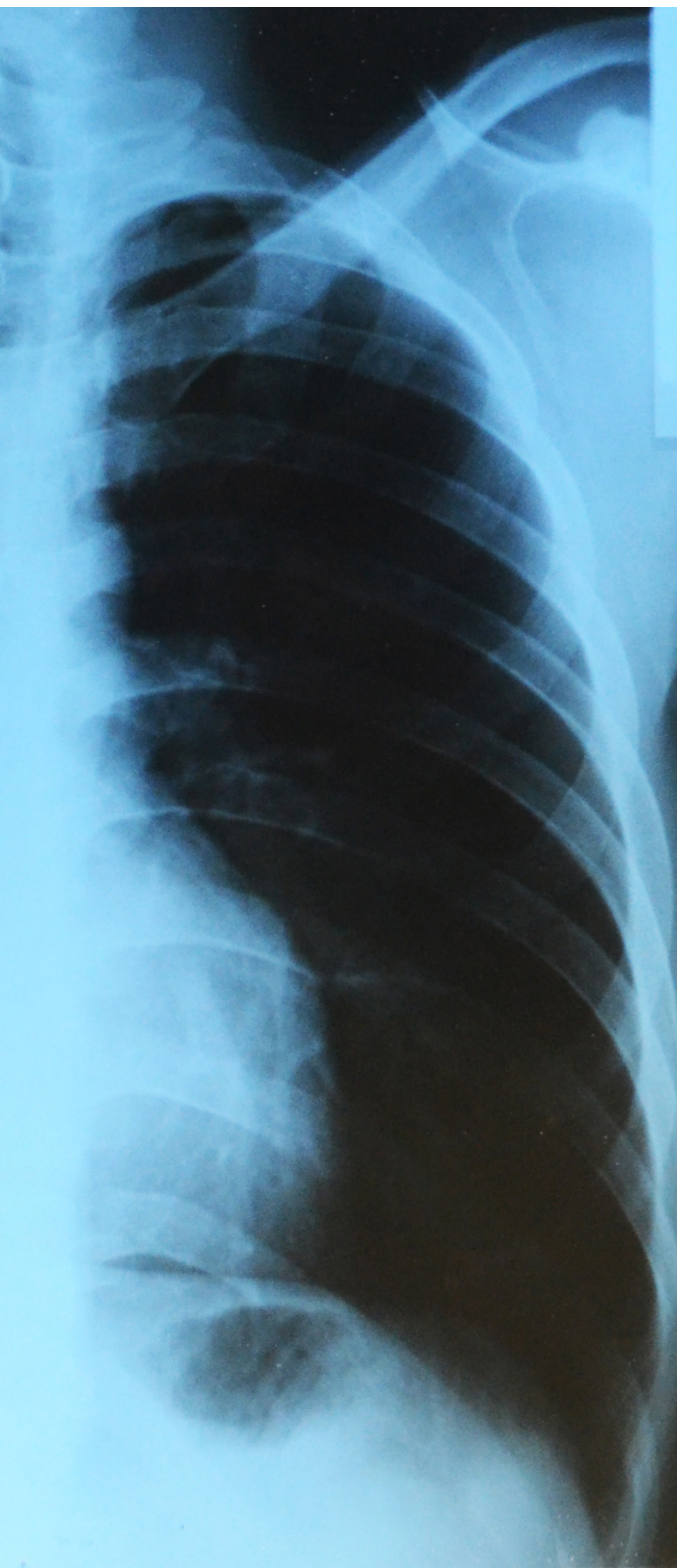
Genetic studies

DEEP has been requested to join a project coordinated by the [Institut d'Investigacions Biomèdiques](#) de Barcelona that aims at discovering any genetic traits explaining why some people without previous pathologies get severe forms of covid-19 leading them to the ICU or even to death. The study will take genetic material

(together with populational and clinical information) of 200 patients who are under 60 years old and who do not have any previous or serious chronic diseases. We want to study the difference between those patients who evolve well and those who get worse and end up in the ICU by discovering whether, at the genetic level, these latter patients have a special susceptibility. In that case this will give us an indicator of which cases are the most vulnerable and should be protected. If this indicator is found, the patients without such genetic condition could get discharged earlier and we could protect those who, besides the elderly, are likely to have serious symptoms of the disease. DEEP will provide extensive data analysis, including the development of a deep learning model that will then be published and available at our [Open Catalog](#), and dedicated testbed resources.

X-ray images classification

Building on our [image clasification module](#), DEEP is collaborating with the [University Hospital Marqués de Valdecilla](#) in order to develop and share a new module trained to classify chest x-ray images that will act as an assistant for the physician and will help with the patients triage. In the current state of health alarm, huge amounts of simple chest x-rays are being produced daily. Due to the saturation of the medical systems, professionals with no x-ray experience are being forced to interpret the chest images, and must systematically resort to the advice of a radiologist who is



overwhelmed with consequent delay in diagnosis. Under these circumstances, a reliable automatic triage system to assist diagnosis using simple chest x-rays would greatly expedite patient management.

Data science to understand confinement effectiveness

Although this project focuses on patients with COVID-19, the developed tools will be equally applicable to other diseases with pneumonia and will be made available at the [Open Catalog](#). European countries have adopted strict confinement measures to fight the COVID-19 spread. The [Spanish National Research Council](#), in cooperation with the Spanish [National Microbiology Center from the Health Institute Carlos III](#), is using data science and computing techniques in order to understand the effectiveness of these measures in Spain. The project is following a multidisciplinary approach involving computing, demography, physics and migration experts; studying high-resolution massive data to gain insights in how mobility and social contacts have changed since the measures were enforced and how these changes are influencing the COVID-19 incidence. These data are then leveraged by computational models (based both on artificial intelligence and mechanistic models), allowing to study different scenarios towards the end of the confinement measures. In this regard, the DEEP-Hybrid-DataCloud stack is being used to develop the AI models, that will be published in the [Open Catalog](#) and served through the [DEEP as a Service](#) component.

PAST EVENTS (2019)



EuSoMII Annual Meeting

18-19 October 2019 Valencia (Spain)



Slovak Matchmaking Fair

24 October 2019 Bratislava (Slovak Republic)



SARIO BUSINESS LINK

Slovak Matchmaking Fair Bratislava 2019

October 24, 2019, Bratislava, Slovak Republic



SC'19: The International Conference for High Performance Computing, Networking, Storage, and Analysis

18-21 November 2019 Colorado (EEUU)



SC19

Denver, CO | hpc is now.



www.deep-hybrid-datacloud.eu



@DEEP_eu



DEEP Hybrid Data Cloud



deep-info@listas.csic.es



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777435