

DEEP-Hybrid-DataCloud

Digital Infrastructures for Research (DI4R)
Lisbon (Portugal)
October 10, 2018

Álvaro López García
aloga@ifca.unican.es
Spanish National Research Council



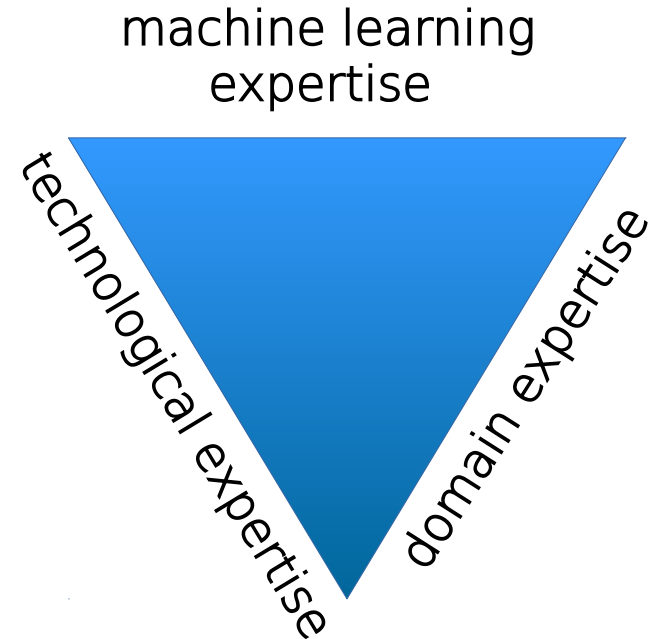
DEEP-HybridDataCloud has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777435.



- We need to build added value and advanced services on top of bare IaaS and PaaS infrastructures
- Ease and lower the entry barrier for **non-skilled** scientists
 - Transparent execution on e-Infrastructures
 - Build ready to use modules and offer them through a catalog or marketplace
 - Implement common software development techniques also for scientist's applications (DevOps)
- Build and promote the use of **intensive computing services** by different research communities and areas, and the support by the corresponding e-Infrastructure providers and open source projects

Who is the user?

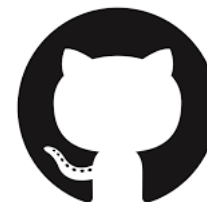
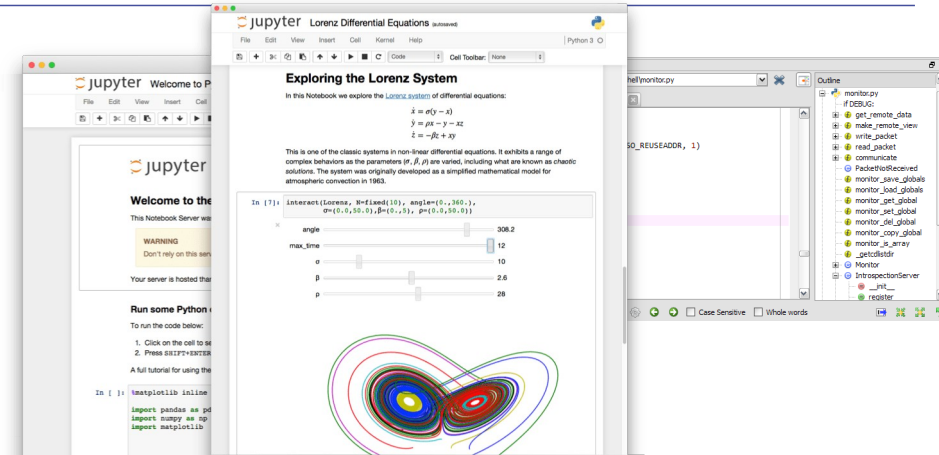
level of knowledge being required depends on the specific use case and the user profile



- Category 1: Deploy a readily trained network for somebody else to use it on his/her data set
 - Domain knowledge
- Category 2: Retrain (parts of) a trained network to make use of its inherent knowledge and to solve a new learning task
 - Domain + machine learning knowledge
- Category 3: Completely work through the deep learning cycle with data selection, model architecture, training and testing
 - Domain + machine + technological knowledge

Previously...

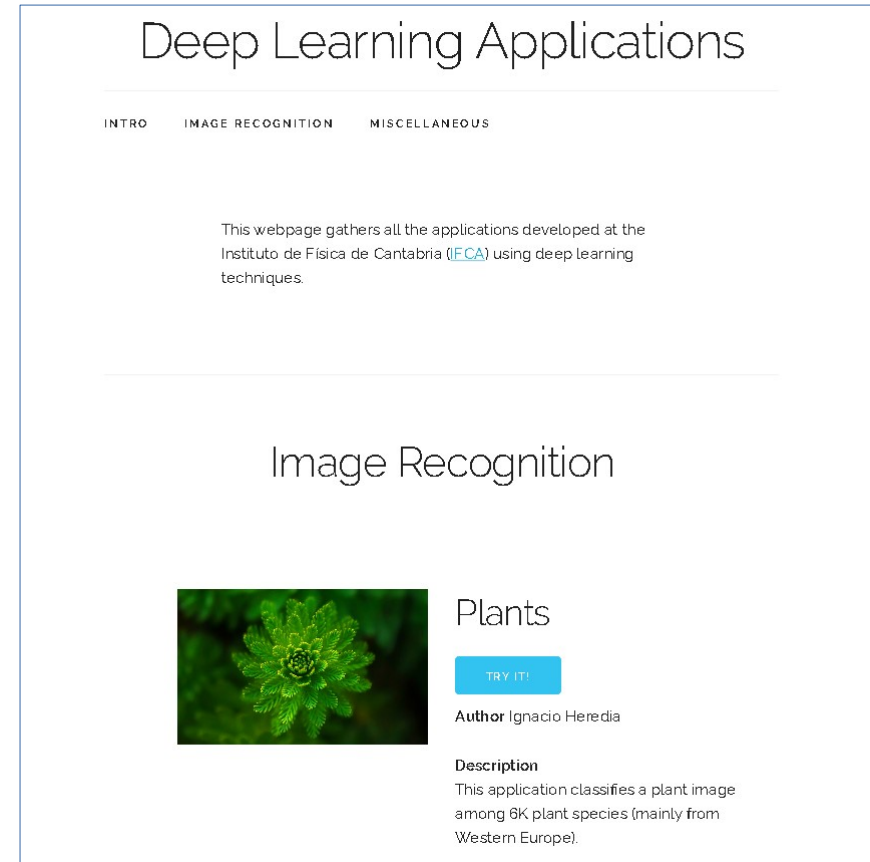
- Scientists create a deep learning application on their personal computers
- The deep learning model is trained in a GPU node (maybe also locally)
 - What happens if they do not have access to one?
- The work is published (or not)
 - Model architecture, configuration, scientific publication, etc.
- But:
 - How can a scientist easily offer it to a broader audience?
 - What about dependencies?



GitHub



- Development of APIs and web applications
- Scientists need to know what an API is
 - REST, GET, POST, PUT...
- Lack of API consistency → hard for external developers to consume them
- Provide users with a generic API (OpenAPI) component where they application can be plugged




Deep Learning Applications

INTRO IMAGE RECOGNITION MISCELLANEOUS

This webpage gathers all the applications developed at the Instituto de Física de Cantabria (IFCA) using deep learning techniques.

Image Recognition



Plants

[TRY IT!](#)

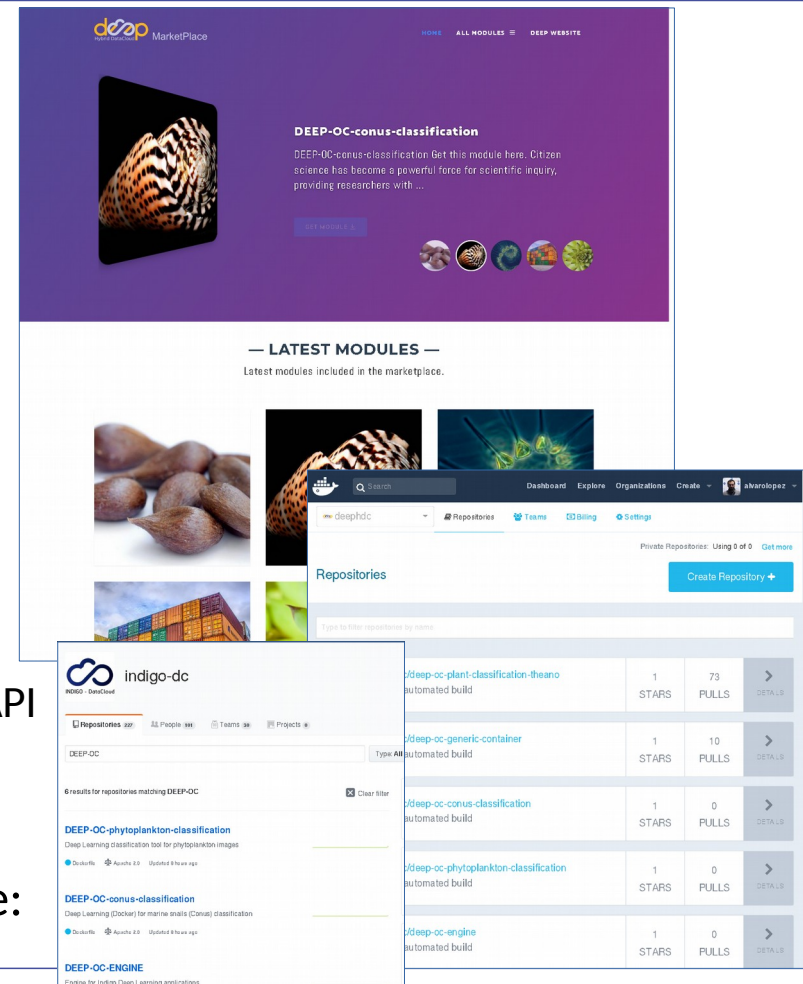
Author Ignacio Heredia

Description
This application classifies a plant image among 6K plant species (mainly from Western Europe).

- An application may consist on several components that need to be deployed, configured, etc → **service composition**
- Service composition, if done properly, provide a way to re-deploy the same topology over different infrastructures → **catalog of components**
- Scientists should not need to deal with technologies and infrastructures they do not care at all
- We need therefore different roles, to perform different tasks
 - Comparison with laboratory technician
- In INDIGO-DataCloud we started with this approach, but this needs to be generalized (and the roles recognized)

DEEP Open Catalog

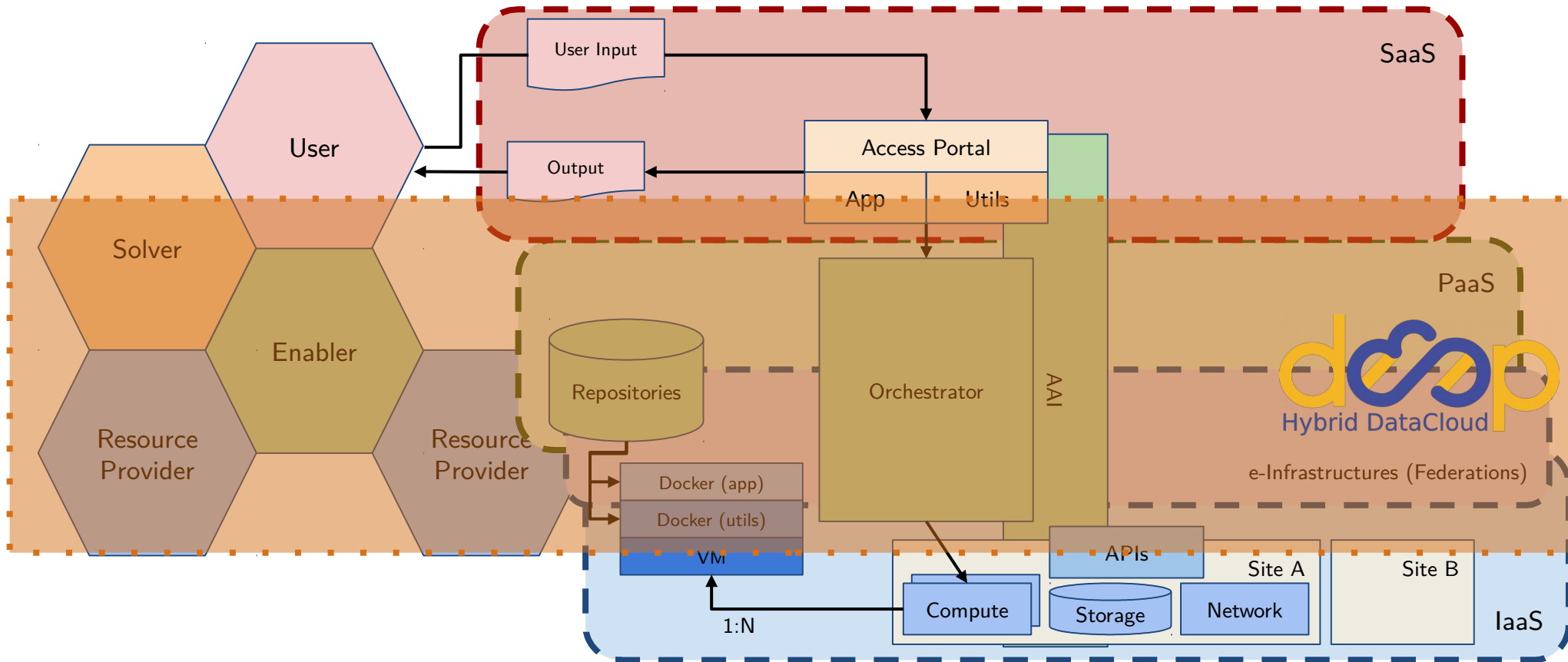
- Collection of ready-to-use modules
 - Comprising machine learning, deep learning, big data analytics tools
 - ML Marketplace
<https://marketplace.deep-hybrid-datacloud.eu>
 - GitHub
<https://github.com/indigo-dc?utf8=✓&q=DEEP-OC>
 - DockerHub
<https://hub.docker.com/u/deephdc/>
- Based on DEEPaaS API component
 - Expose underlying model functionality with a common API
 - Based on OpenAPI specifications
 - Minimal modifications to user applications.
- Goal: execute the same module on **any** infrastructure:



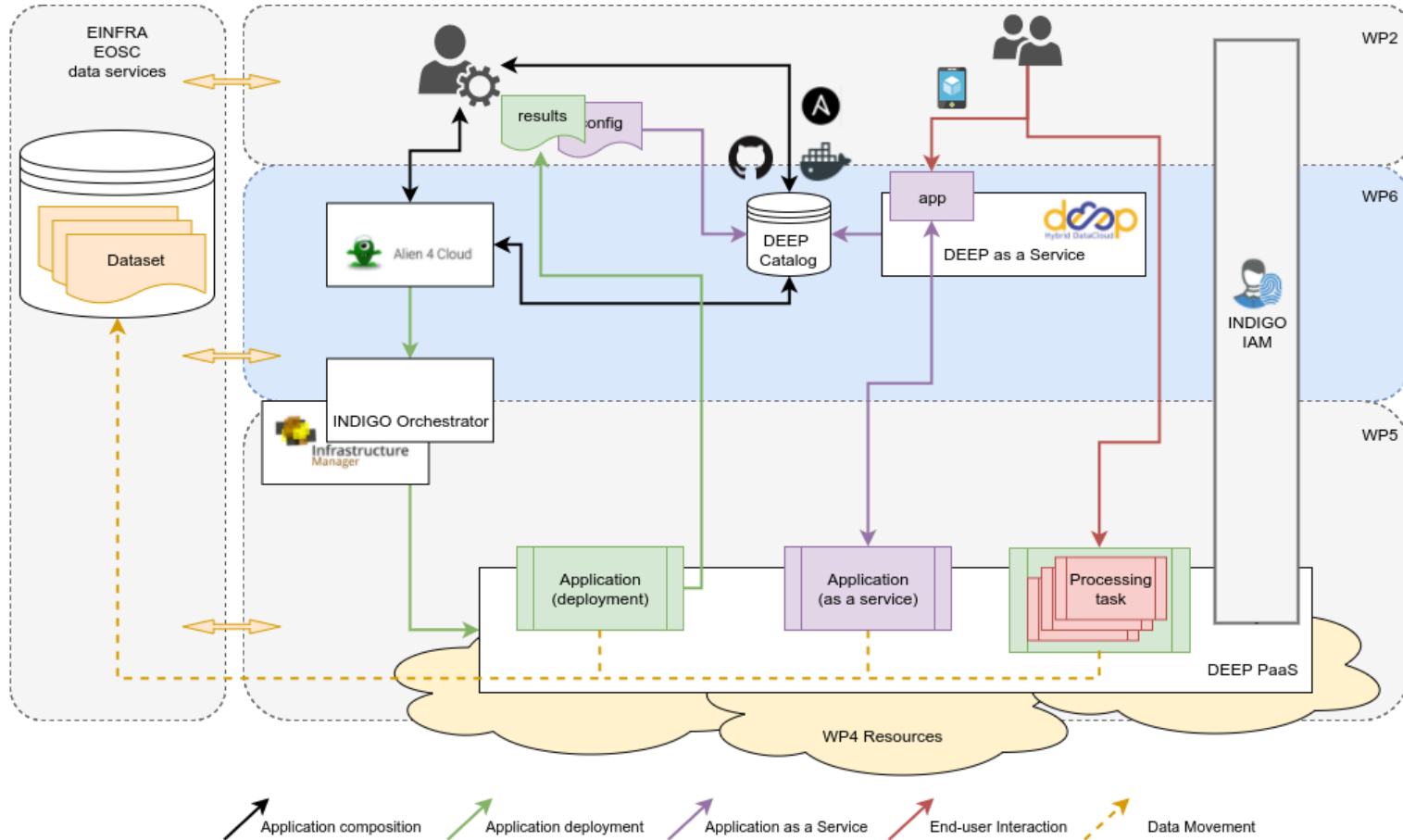
The image displays two screenshots. The top screenshot shows the DEEP Marketplace interface, featuring a purple header with the DEEP logo and navigation links. A featured module, 'DEEP-OC-conus-classification', is highlighted with a large image of a nautilus shell. Below this, a section titled 'LATEST MODULES' shows a grid of smaller module thumbnails. The bottom screenshot shows a GitHub repository page for 'deephdc'. The repository name is 'deephdc' and it is under the organization 'indigo-dc'. The page shows a list of repositories with columns for repository name, stars, and pulls. The following table summarizes the visible repository data:

Repository Name	Stars	Pulls
deephdc/deep-oc-plant-classification-theano	1	73
deephdc/deep-oc-generic-container	1	10
deephdc/deep-oc-conus-classification	1	0
deephdc/deep-oc-phytoplankton-classification	1	0
deephdc/deep-oc-engine	1	0

Different roles for different tasks



DEEP high level Architecture



DEEP 1st release: components

Software component	Functionalities
PaaS Orchestrator	<ul style="list-style-type: none">• Hybrid deployments on multiple sites• Support to specifying specialized computing hardware• Improved support for deployment failures
Infrastructure Manager (IM)	<ul style="list-style-type: none">• Improved support for hybrid deployments• Support for additional TOSCA types
DEEPaaS API	<ul style="list-style-type: none">• Support for training a machine learning application• Support for performing inferences/analysis/predictions• Support only for synchronous requests• OpenID Connect support• Support for standalone service & OpenWhisk action
Alien4Cloud	<ul style="list-style-type: none">• Support for visual composition of TOSCA templates• PaaS orchestrator support
Virtual Router	<ul style="list-style-type: none">• Improvements to reach production level• Virtualized routing over distributed infrastructures
cloud-info-provider	<ul style="list-style-type: none">• Support for GPU and Infiniband resources
uDocker	<ul style="list-style-type: none">• Improved support for GPUs and Infiniband

DEEP 1st release: services

Service	Functionalities	Preview endpoint
Visual application topology composition and deployment	<ul style="list-style-type: none">• Graphical composition of complex application topologies• Deployment through PaaS orchestrator	https://a4c.ncg.ingrid.pt
DEEP as a Service	<ul style="list-style-type: none">• Deployment of DEEP Open Catalog components as server-less functions	https://vm028.pub.cloud.ifca.es/
DEEP Open Catalog	<ul style="list-style-type: none">• Ready-to-use machine learning and deep learning applications, including:<ul style="list-style-type: none">➤ Machine learning frameworks + JupyterLab➤ Machine learning ready to use models➤ Deep learning ready to use models➤ BigData analytic tools	https://marketplace.deep-hybrid-datacloud.eu

- All services are OIDC-ready, following AARC blueprint recommendations
- Also work on:
 - TOSCA templates and TOSCA types
 - Documentation and configuration recipes for GPU support
 - Patches to upstream projects (Apache Libcloud, Apache OpenWhisk, OpenStack)

Web page:

<https://deep-hybrid-datacloud.eu>

Email:

deep-info@listas.csic.es



https://twitter.com/DEEP_eu

Thank you
Any Questions?



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