Supporting Multi-Cloud Model Execution with VLab

M. Santoro (1), P. Mazzetti (1), N. Spadaro (2), S. Nativi (2)

(1) CNR-IIA

(2) EC-JRC

The VLab

- Developed in the context of the European projects ECOPOTENTIAL and ERA-PLANET
- VLab is a cloud-based platform to support the activity of environmental scientists in sharing their models
- The main challenges addressed by VLab are:
  - Minimization of interoperability requirements in the process of model porting (i.e. to simplify as much as possible the process of publishing and sharing a model for model developers)
  - Support multiple programming languages and environments (it must be possible porting models developed in different programming languages and which use an arbitrary set of libraries)

VLab Concept

Ecosystem Expert

Knowledge

Models, ontologies

ECOPOTENTIAL VLAB

Information

Earth Observation data and products

VLab Portal and apps

Knowledge For Informed Decision-Making

Indicators

Environmental/Ecological Decision-maker, Policy-maker

The ECOPOTENTIAL VLAB Concept

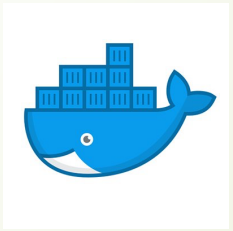
Code Sharing and Containerization Technologies

Git

Git is a distributed version control system designed to handle everything from small to very large projects with speed and efficiency.

Docker

Docker is a tool designed to make to run applications by using containers. Containers allow a developer to package up an application with all of the parts it needs.



Usage of Git

Source **code** of the model

- Python files
- Java files
- Etc.

Scripts to **launch** the model execution

- sh files
- exe files

Machine-readable information about:

- Data input ingestion
- Model execution

Usage of Docker

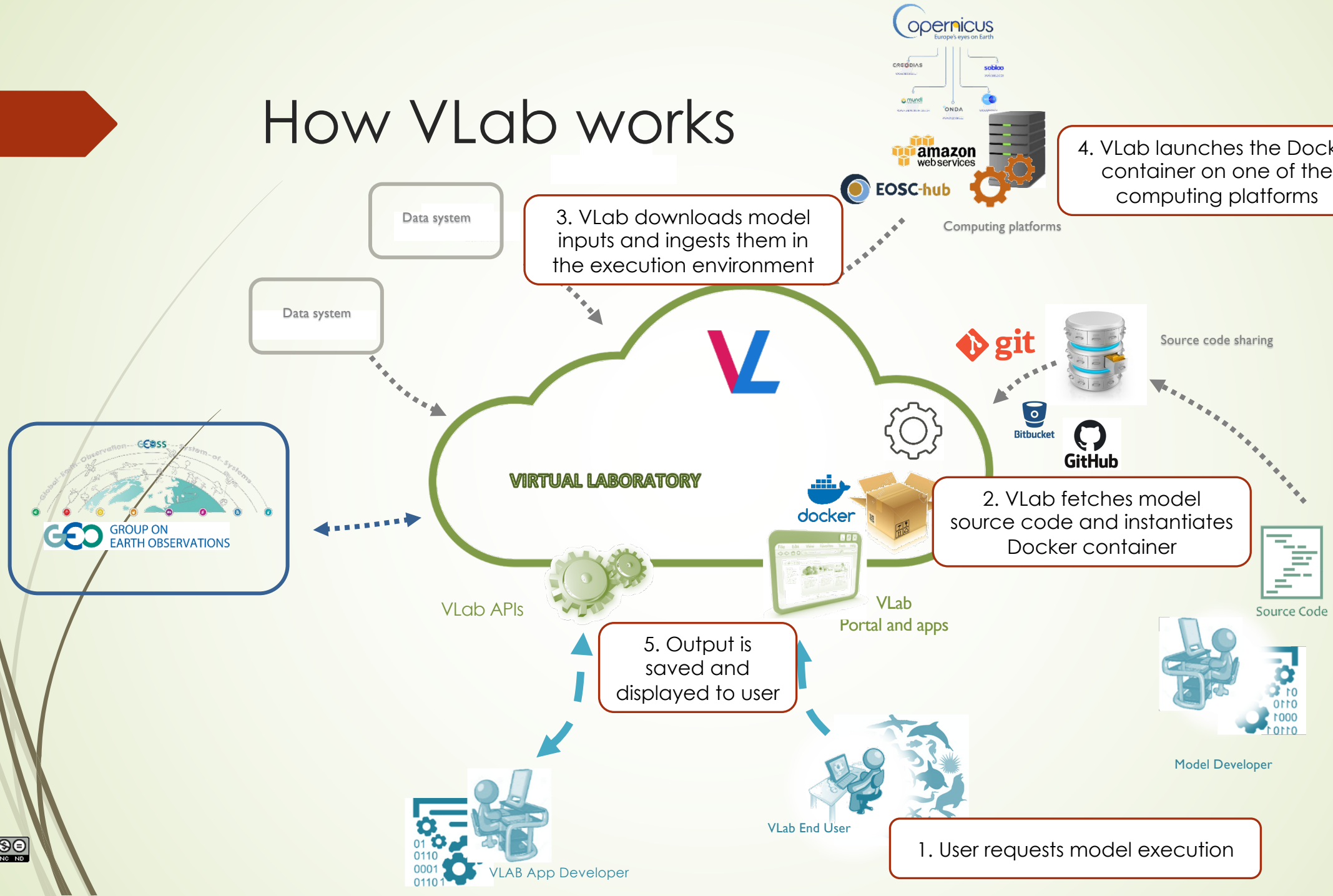
Model Environment

- Vlab uses the provided Docker image to instantiate and run **Docker container**, which is where the actual model execution takes place
  - If the model is coded in python, a Docker image with the required python installation libraries must be defined by model provider

Resources

- When configuring docker image, required resources such as memory (RAM) can be defined

How VLab works



Cloud Services for Model Execution

The design of VLab allows the execution of models on different cloud environments. To this aim, VLab utilizes a set cloud services. Following table lists which cloud specific technology was utilized in the different environments.

Cloud Environment / Cloud Service	Amazon Web Services	OpenStack/Kubernetes
Cloud Hosting	Elastic Compute Cloud	OpenStack Compute (nova)
Auto Scaling	Auto Scaling Group	OpenStack Heat
Container Management	Elastic Container Service	Kubernetes
Shared File System	Elastic File System	Network file System (Shepler, et al., 2010) + Kubernetes Persistent Volume
Web Storage	Simple Storage Service	MinIO (under experimentation)
Hosted Queue	Simple Queue Service	KubeMQ (under experimentation)

Advantages of Supporting Multi-Cloud Execution

- Allows user (or client application) to choose which platform to use for a specific model run, e.g. depending on the required data which might be available on a specific platform only (move code to data).
- Another possible scenario empowered by this multi-platform support feature is the possibility to let the user choose the computational platform and utilize her/his credentials to request the needed computational resources.
- Finally, it is also possible to exploit this feature for benchmarking different cloud platforms with respect to their performances.

Processing Platforms

VLab was experimented on several processing platforms.

CREODIAS

ONDA

aws

EUROPEAN OPEN SCIENCE CLOUD

sobloo

Demonstration

- In the context of the EuroGEOSS Sprint to Ministerial activity , promoted by the European Commission, a web application was developed by the Joint Research Centre of the European Commission (EC JRC) to show the use of Copernicus Sentinel data to calculate Land Cover and Land Cover change in a set of Protected Areas belonging to different ecosystems.
- To this aim, the Earth Observation Data for Ecosystem Monitoring (EODESM) model (Lucas & Mitchell, 2017) was utilized to classify land covers and change.
- The web application utilizes VLab APIs to select the area and dates of interest and request the execution of EODESM model on one of the supported cloud environments.

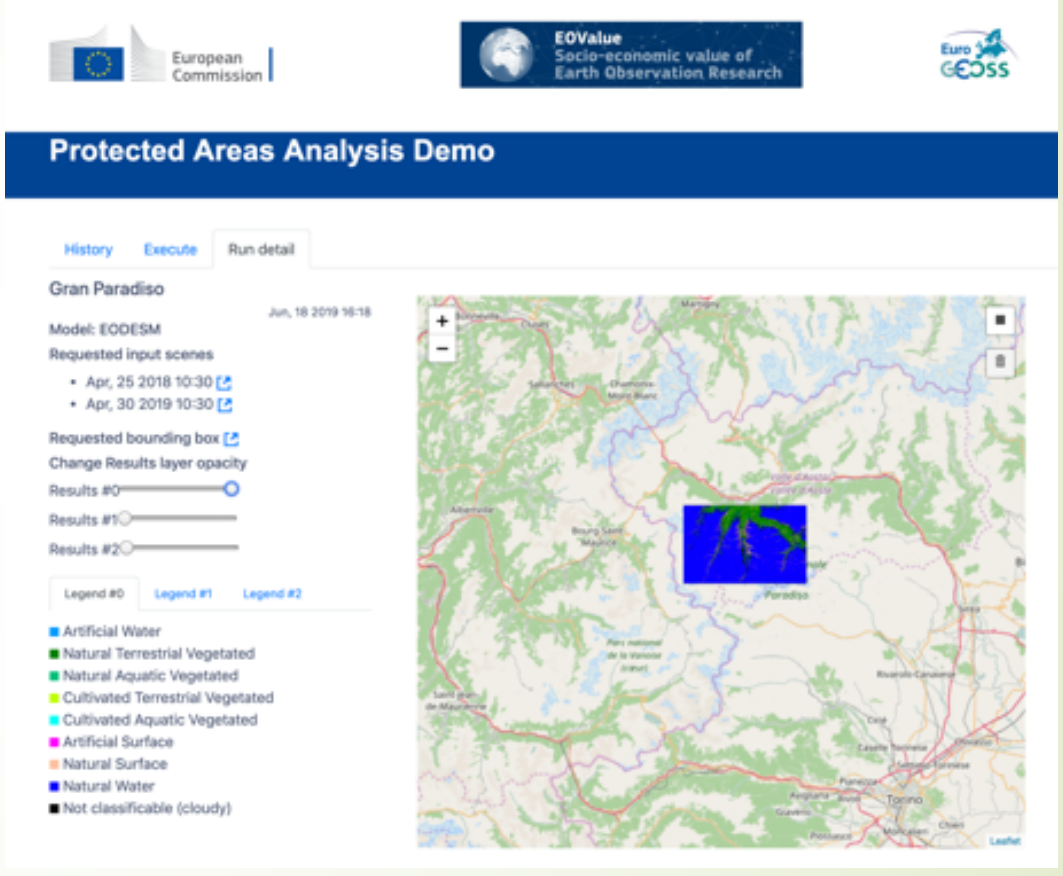


Figure 10 - Protected Areas Analysis Web Application (https://pademo.geodab.org/)

Future Work

- Enhance multi cloud support in Vlab implementing:
  - automatic selection of cloud according to data availability
  - use of users' credentials to request the needed computational resources
- Enhance VLab to support model chaining
- Integrate VLab with Knowledge repositories to increase the interoperability level to semantic and pragmatic

Acknowledgement

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements N. 641762 (ECOPotential), 689443 (ERA-PLANET), 777536 (EOSC-hub), 776136 (EDGE)