The ECOPOTENTIAL VLAB and the modelling tools

Paolo Mazzetti

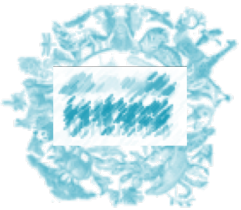
on behalf of the ECOPOTENTIAL Consortium

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

WHY

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

The ECOPOTENTIAL Virtual Laboratory



- The ECOPOTENTIAL Virtual Laboratory is a *knowledge generation platform supporting the activities of the ecosystem science Community of Practice*

From Science to Policy

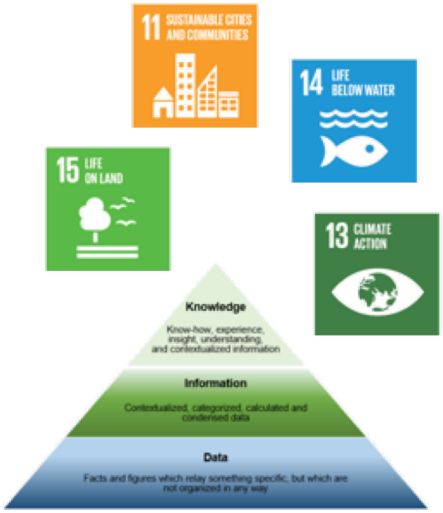
Generated knowledge targeted to policy-makers (e.g. PA managers)
Integration with global efforts for Science-Based Environmental Policy

From Data to Knowledge

Generation of Essential Variables, Indicators and Indices from EO and in-situ data

From Open Data to Open Science

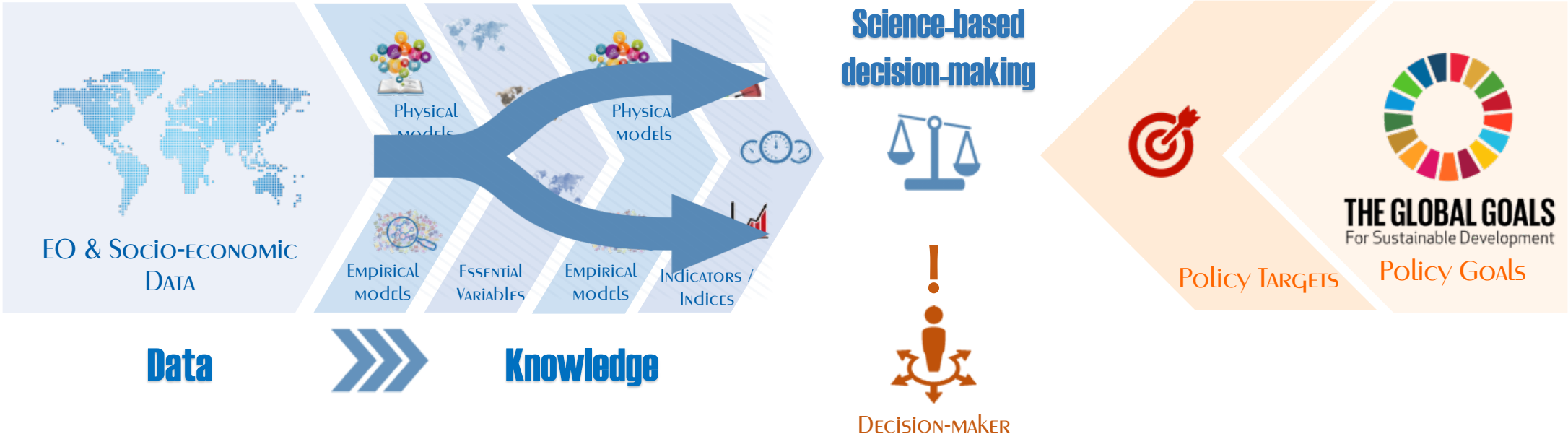
Sharing of knowledge (ontologies), procedures (scientific business process), algorithms (source code) for reusability, reproducibility, etc.



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

Rationale: From (EO) Data to Knowledge for Policy

"Knowledge-based conservation, management and restoration policies are needed to improve ecosystem benefits in face of increasing pressures." [ECOPOTENTIAL DoW]

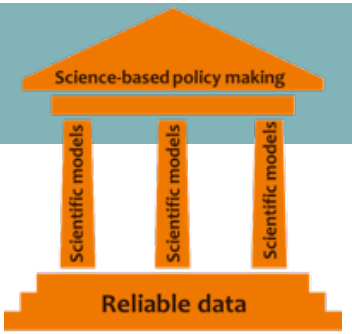


This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

marco.folegani@cnr.it

WHAT

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762



Challenges vs. Opportunities

STRENGTHS	WEAKNESSES	THREATS	OPPORTUNITIES
• LARGE AMOUNT OF ACQUIRED EO/IN-SITU DATASETS	• DARK DATA • UNCLEAR POLICY • LACK OF DOCUMENTATION • QUALITY • GEOGRAPHICAL/TEMPORAL/ TYPOLOGY GAPS	• BIG VOLUME • LARGE VARIETY (FORMATS, RESOLUTION, CRS,…) • SEMANTIC MISMATCH	• IAAS CLOUD TECHNOLOGIES • STANDARDIZATION/MEDIATION SOLUTIONS • SEMANTIC TECHNOLOGIES
• MANY SCIENTIFIC MODELS AVAILABLE	• MODELS DEVELOPED FOR SCIENCE NOT FOR POLICY (DIFFERENT REQS) • EXISTING MODELS UNDERUSING EO DATA	• HETEROGENEITY OF PROGRAMMING/SIMULATION FRAMEWORKS • LACK OF MODEL INTEROPERABILITY	• VIRTUALIZATION/CONTAINERIZATIO N TECHNOLOGIES • DATA SCIENCE/AI ADVANCEMENTS

Data and Model Interoperability

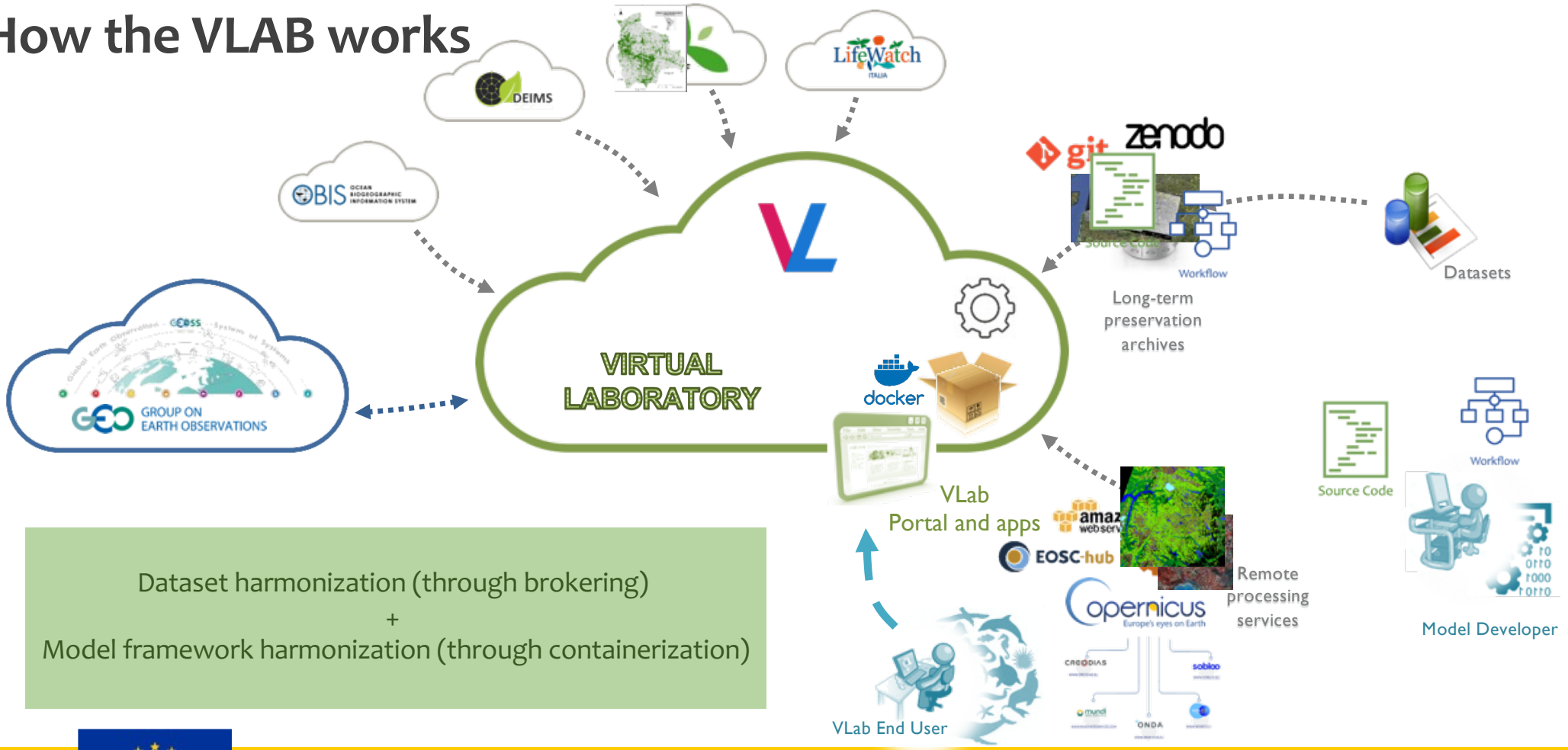
This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

Main Requirements

Requirements	Notes
BIG DATA CHALLENGES	SUPPORT FOR HANDLING BIG DATASETS, LARGE VARIETY, HIGH VELOCITY, ETC. LEVERAGING EXISTING CLOUD PLATFORMS
HETEROGENEITY	NECESSARILY CODE MAY RUN IN DIFFERENT SOFTWARE ENVIRONMENTS AND SIMULATION FRAMEWORKS
CHAINING	NEW WORKFLOWS FOR MULTIDISCIPLINARY APPLICATIONS (MOST INDICATORS ARE MULTIDISCIPLINARY)
DOCUMENTATION	DATA AND CODE DOCUMENTATION FOR SCIENTIFIC EVIDENCE (OPEN SCIENCE)
INSPECTABILITY	DATA AND CODE INSPECTION FOR SCIENTIFIC EVIDENCE (OPEN SCIENCE)
REPRODUCIBILITY	SCIENTIFIC WORKFLOW EXECUTION TO RECREATE RESULTS FOR SCIENTIFIC EVIDENCE (OPEN SCIENCE)
REPLICABILITY	SCIENTIFIC WORKFLOW EXECUTION WITH DIFFERENT DATA, MODELS, PARAMETERS FOR «WHAT-IF» SCENARIOS OR MULTIDISCIPLINARY APPLICATIONS
MULTIPLE UIS	SUPPORT OF DIFFERENT USERS (DECISION-MAKERS, POLICY-MAKERS OR THEIR TECHNICAL PERSONNEL)

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

How the VLAB works



Dataset harmonization (through brokering)
+
Model framework harmonization (through containerization)

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

WHO

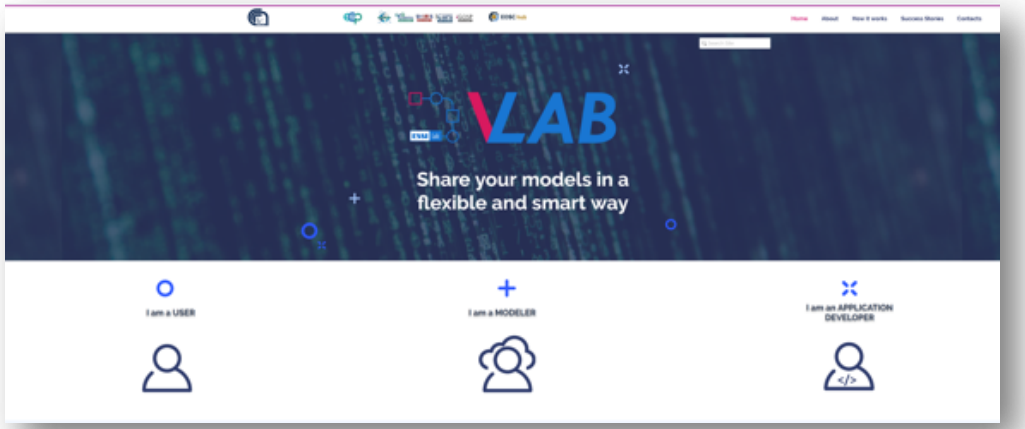
This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

The VLab users

- **Modelers**
  - they can share models developed on «any» programming environment or simulation framework (e.g. Python, R, Java, NetLogo,…)
  - They can build scientific workflows using available data and models

- **Application developers**
  - they can build desktop and mobile applications based on VLab scientific workflows, through the VLab Application Programming Interface (API)
  - Models can run on «any» cloud platform (EOSC, Copernicus DIAS, Amazon currently tested)

- **End users**
  - they can run available VLab-enabled applications

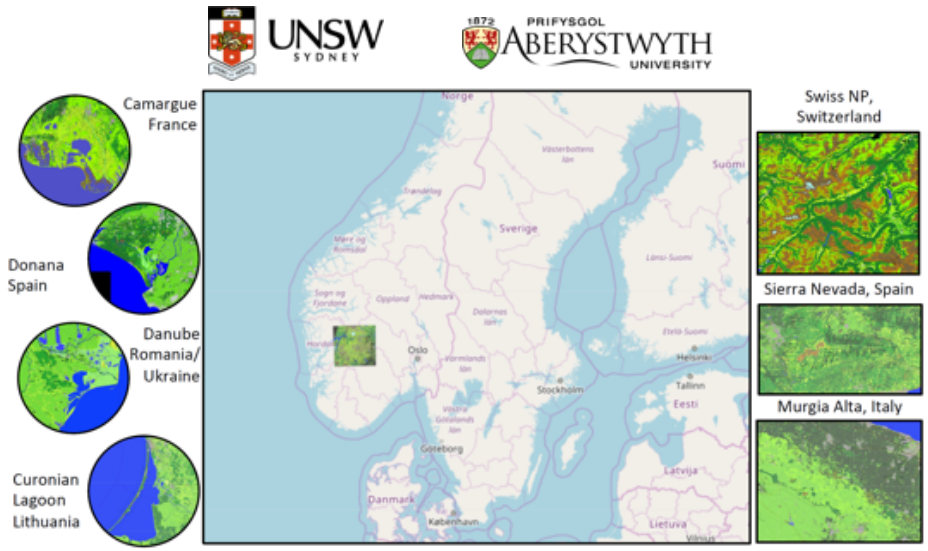


https://vlab.geodab.org/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

An example: EODESM

- The **Earth Observation Data for Ecosystem Monitoring** (EODESM) system classifies land covers according to the Food and Agricultural Organisation's **(FAO's) Land Cover Classification System (LCCS2) taxonomy**.
- The EODESM system can use, as input, any remote sensing or other spatial datasets and at any scale
- Highly detailed and relevant classifications are generated for protected areas and surrounds
- The system is designed for use by a wide range of users and is entirely open source and freely available.

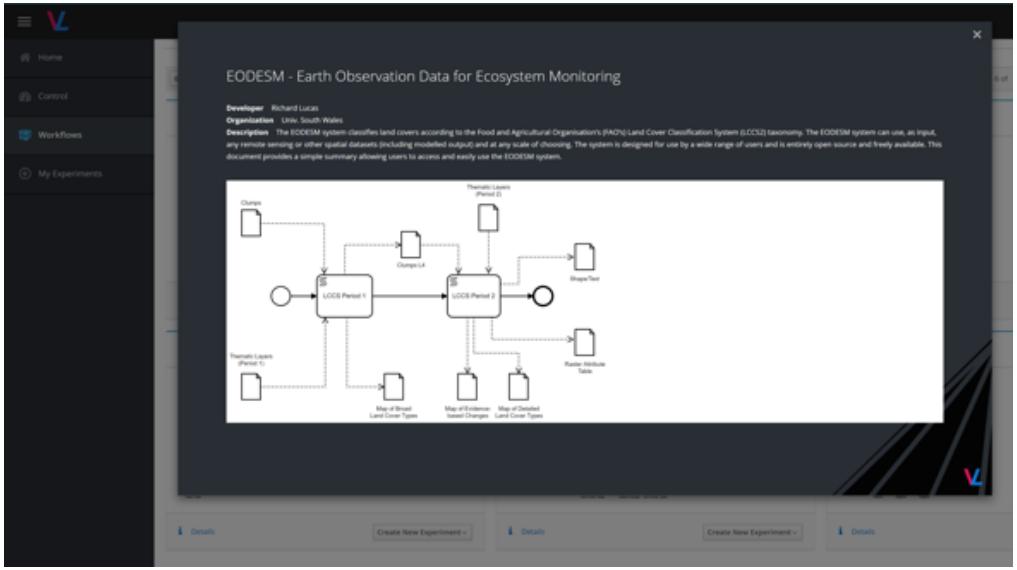


See Richard's presentation later

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

EODESM: 1) porting the model

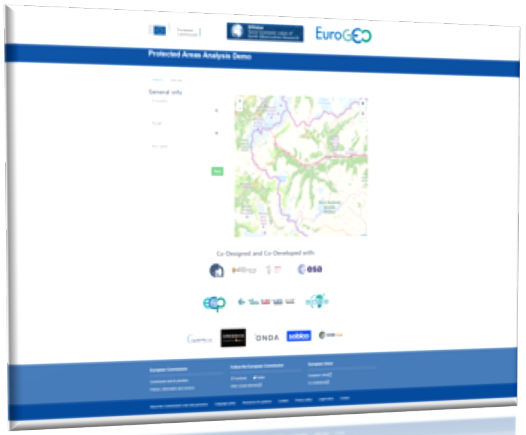
- The source code of the model was published on GitHub.
- The code was tested locally to run in a Docker container
- Information on the model were uploaded in the VLab
  - Source code endpoint
  - Information on the needed container
  - Information on input/output
- The VLab generated a simple workflow
- The model could run from the VLab test environment



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

Land EODESM: 2) Building the app

- ESA in the context of the H2020 EDGE project developed the ECOPOTENTIAL Community Portal using GEOSS Mirror technology and VLab APIs
- EC JRC in the context of the EOVALUE project developed a Protected Areas Analysis Demo application

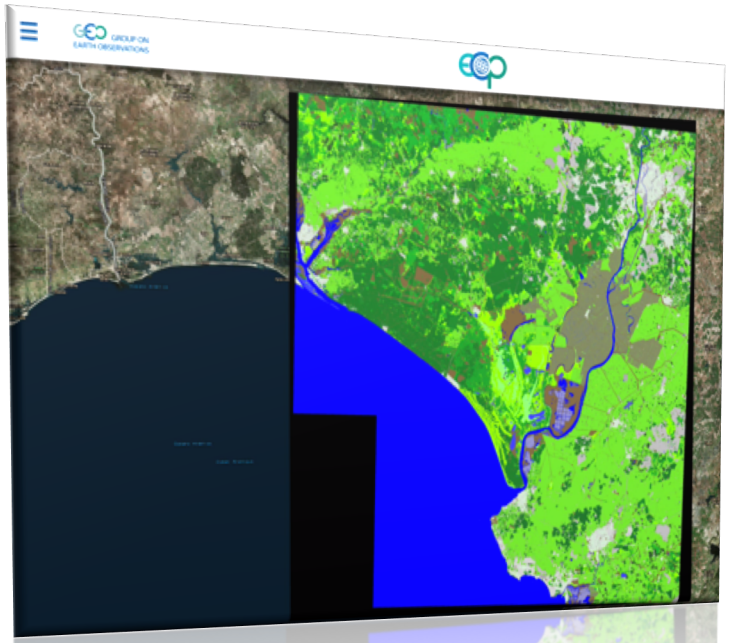


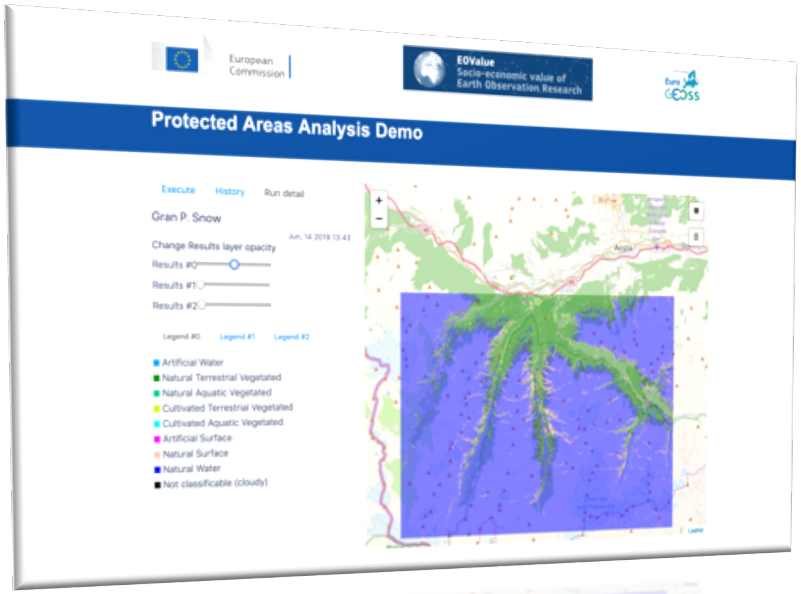
https://pademo.geodab.org/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

EODESM: 3) Using the app

- An end-user can run the model and visualize the output.



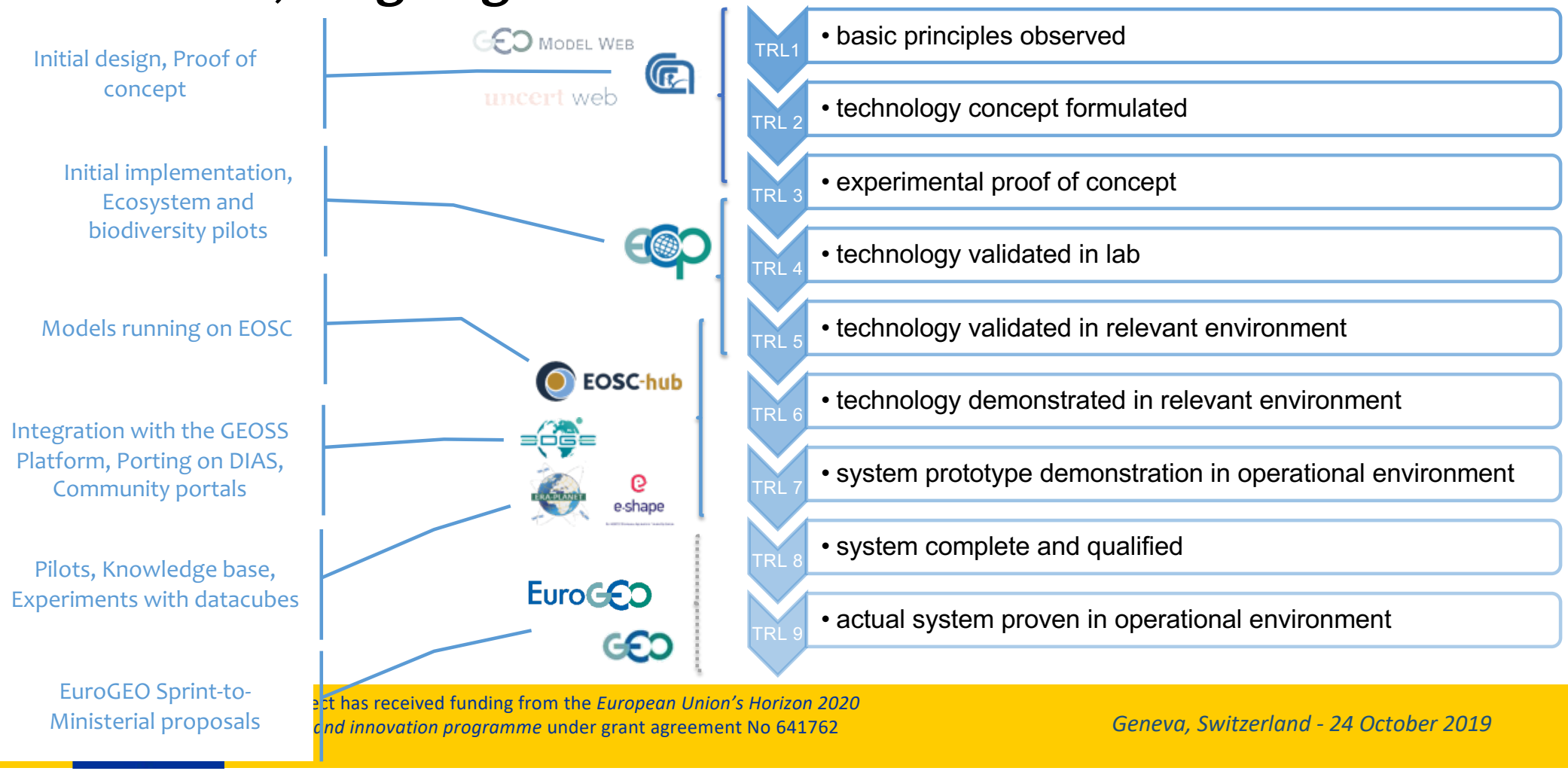


This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

WHEN

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

VLAB Status, on-going and Future Work



ect has received funding from the *European Union's Horizon 2020 and innovation programme* under grant agreement No 641762

VLab on-going and future work

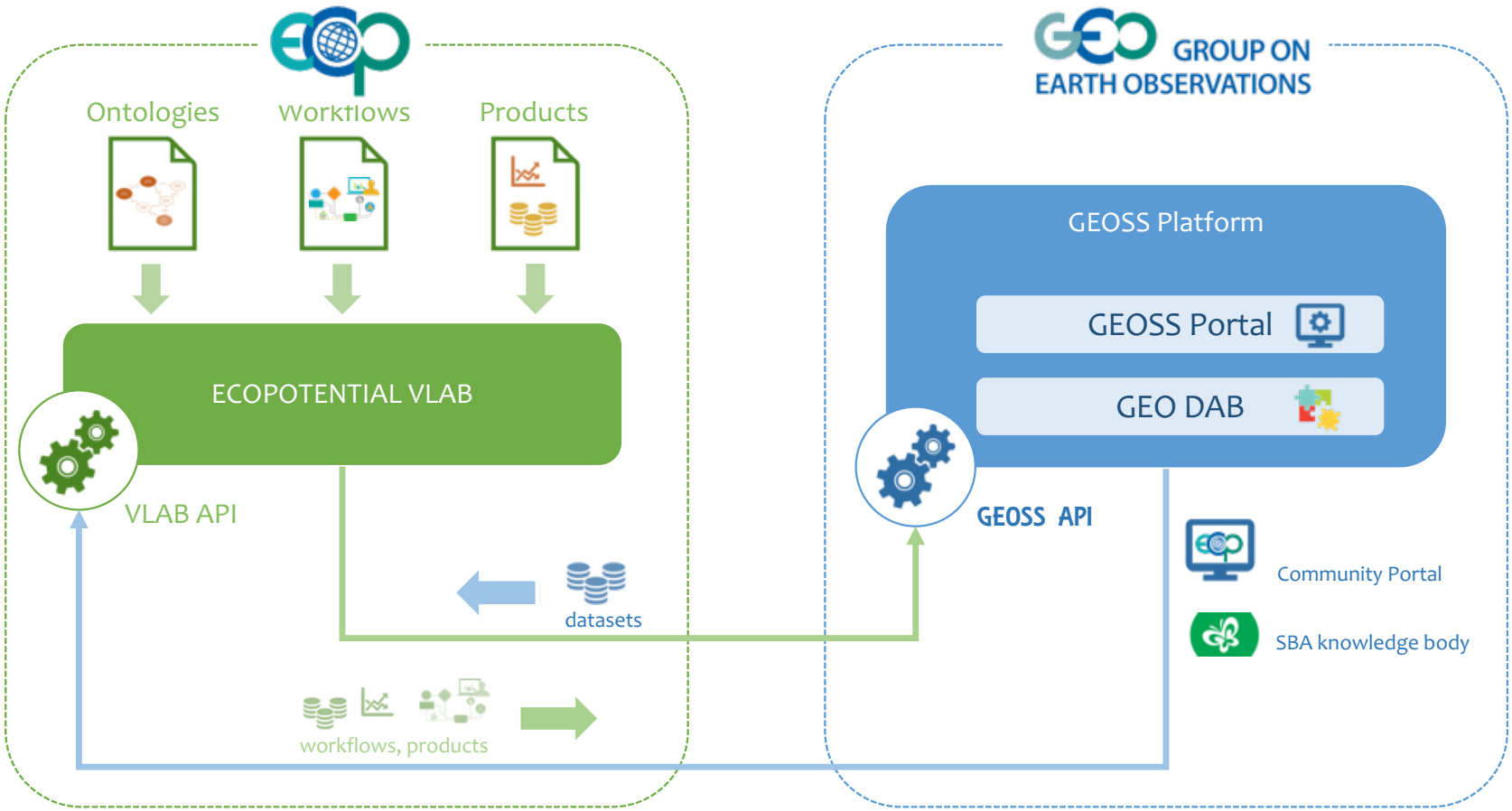
- Lowering remaining entry barriers for modelers (e.g. providing pre-configured Docker images)
- Workflow configuration
- User profiling (e.g. user-based access to cloud platforms)
- Cloud platforms characterization (available datasets, provided services and tools)
- Knowledge formalization for semantic/pragmatic interoperability of models and data
- Integration with emerging technologies: datacubes

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

The VLab and GEO/GEOSS

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

Interoperability with the current GEOSS Platform



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

The VLab and the future evolution of GEOSS

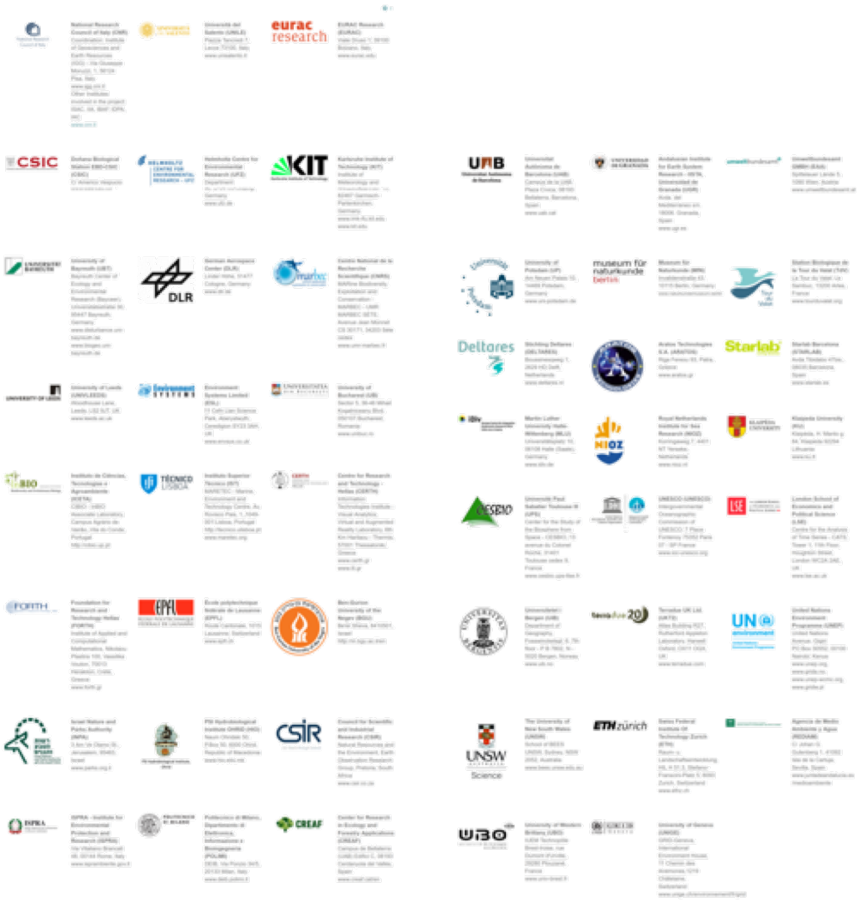
- Results-oriented GEOSS: "*transform the current data focused GEOSS to a knowledge-based GEOSS delivering decision-ready products and services*"
- **The VLab facilitates model sharing and running,** supporting:
  - Public access to code: The VLab may access the source code stored in archives like Git
  - Reproducibility: The VLab can (re-) run workflows to reproduce results
  - Replicability: The VLab can run existing workflows on different datasets
- **The VLab allows developing dedicated apps** using shared workflows through open APIs

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

How could the VLab contribute to GEOSS?

- Strongly depending on the future evolution of GEOSS and Regional GEO services. The VLab could be:
  - An external service linked and called by Regional GEOs or GEOSS
  - A Regional GEO service shared with GEOSS
  - A GEOSS Platform service
- The VLab is not the best solution for everything:
  - Some models have very strict requirements and cannot be «dockerized» (yet?)
  - Some code (e.g. for indicator generation) is very simple and does not need to run on cloud platforms (some code can run directly in the browser)
  - The VLab is not for Rapid Development but for «mature» models sharing (compare with Google Earth Engine, Python Notebooks, etc.)
- The VLab fits to make existing model code run on cloud platforms

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762



Thank you for your attention!

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762