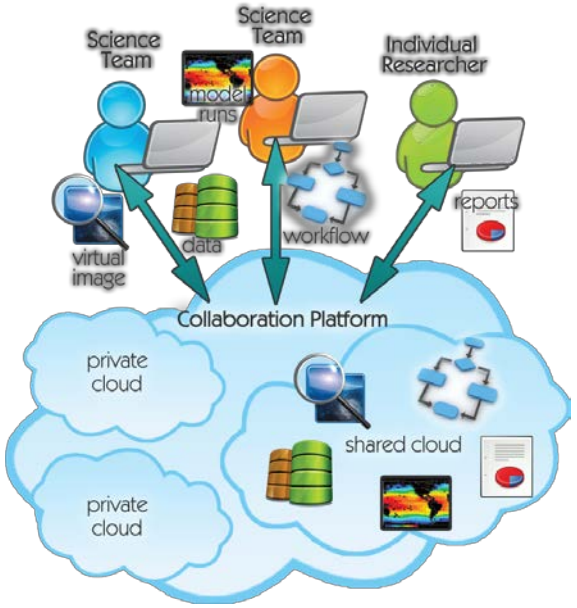
EVER-EST: the platform allowing scientists to cross-fertilize and cross-validate data

Mirko Albani, Cristiano Silvagni, Rosemarie Leone, Fulvio Marelli, Sergio Albani, Michele Lazzarini, Anca Popescu, **Federica Foglini,** Francesco De Leo, Valentina Grande, Stefano Salvi, Elisa Trasatti, Hazel Napier, Tim Aldridge, Steven Cole, Robert Moore, Iolanda Maggio

How can EVER-EST help you TO CROSS-FERTILIZE AND CROSS-VALIDATE DATA

- Remotely access data, software, research results, and documentation
- Organize a scientific workflow in a single digital object, findable and reusable, maintaining attribution through DOI placement
- Collaborate with colleagues located in different parts of the world
- Document scientific work, e.g., encapsulate in a single digital object data and/or results related to a single Supersite event (an eruption)
- Publish grey literature (e.g., project reports, bulletins, etc.) maintaining attribution
- Ensure long term preservation of research work (data, software, results, interpretations)

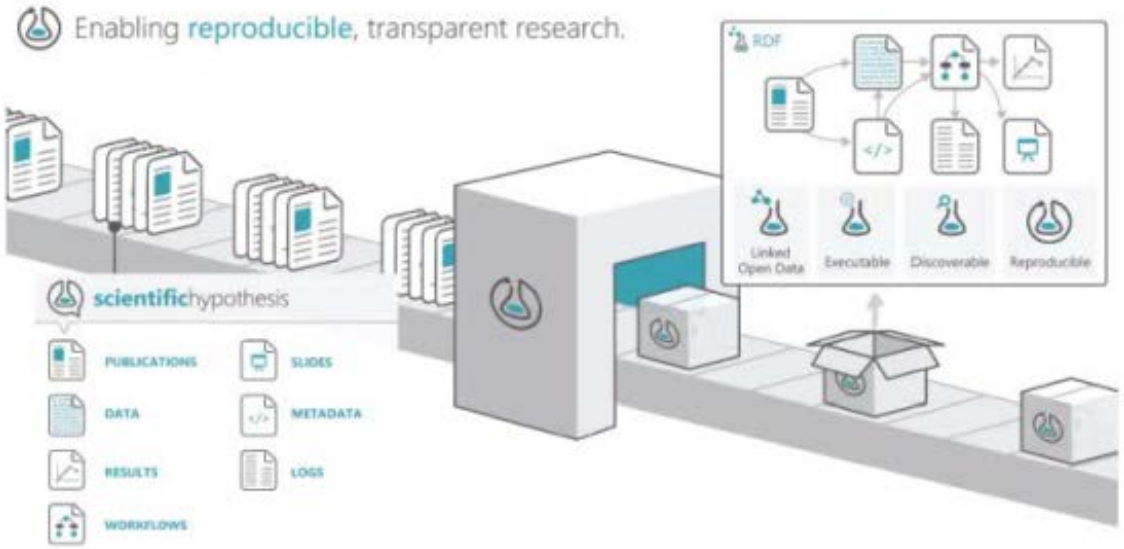


The Research Life cycle management - Research Objects

Aggregation of resources that bundles the content of a research work to facilitate the reusability, reproducibility and better understanding.

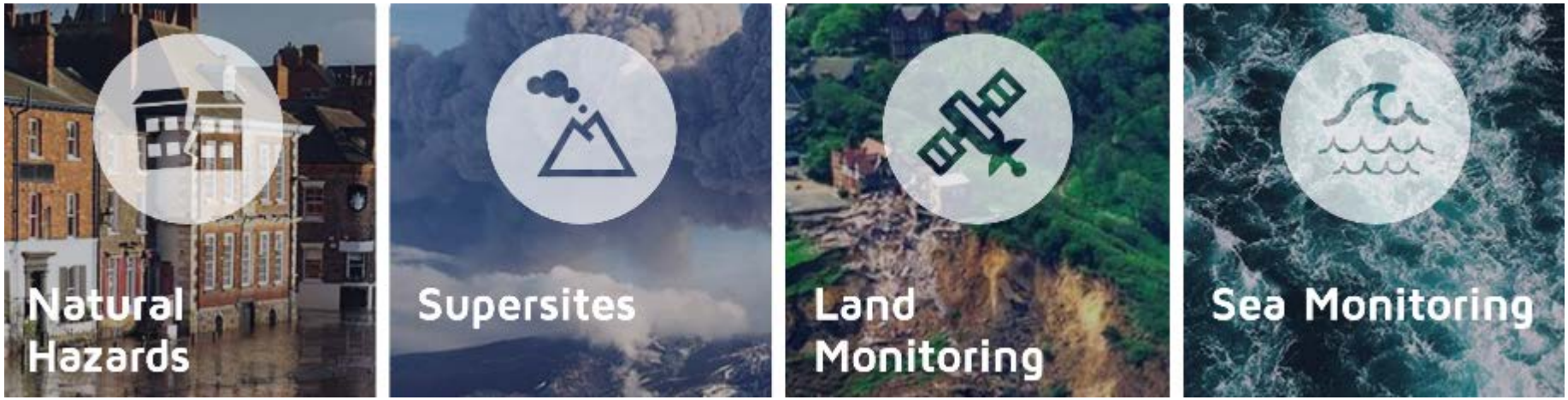
The resources are:

- Data
- Experiments
- Workflows
- Metadata
- Annotations
- Bibliography
- Results
- Provenance
- ...



The Virtual Research Communities

Each of these four Virtual Research Communities has its own specific requirements for data, software, best practices and the community engagement.



EVER-EST project will seek to establish synergies and facilitate dialogue and sharing of information and best practices between the different communities.

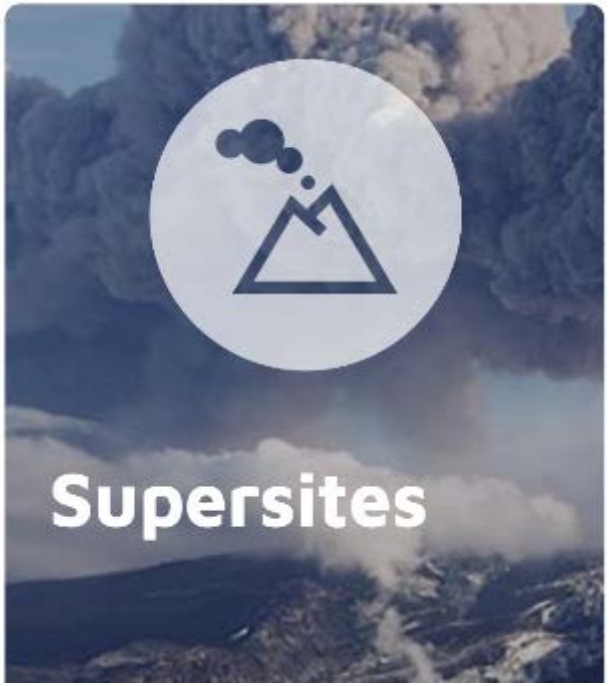
Natural hazards

- **Natural Hazards Partnership**: is a group of 17 collaborating public sector organisations comprising government departments, agencies and research organisations. The NHP provides a mechanism for providing co-ordinated advice to government and those agencies responsible for civil contingency and emergency response during natural hazard events.

Study cases:

- SURFACE WATER FLOODING
- DAILY HAZARD ASSESSMENT (DHA)

Geohazard Supersites

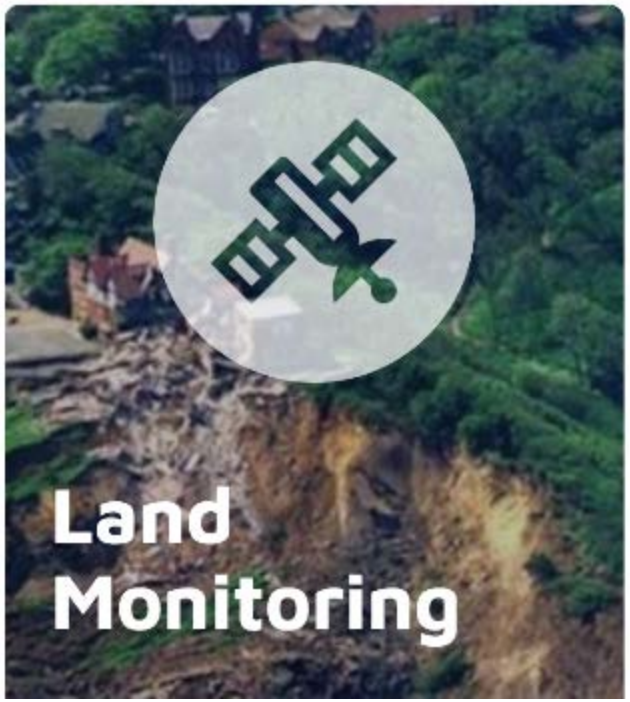


- **Geohazard Supersites and Natural Laboratories:** is a collaborative initiative supported by GEO (Group on Earth Observations) within the Disasters Resilience Benefit Area. The goal of GSNL is to facilitate a global collaboration between Geohazard monitoring agencies, satellite data providers and the Geohazard scientific community to improve scientific understanding of the processes causing geological disasters and better estimate geological hazards.

Study cases:

- VOLCANIC RETRIEVALS PLUME PROCEDURES
- VOLCANIC GEODETIC DATA INVERSION
- INSAR PROCESSING WITH SARSCAPETM ON A WINDOWS VIRTUAL MACHINE

Land Monitoring



- **Land Monitoring**: Monitoring of urban, built-up and natural environments to identify certain features or changes over areas of interest.

Study cases:

- CHANGE DETECTION:

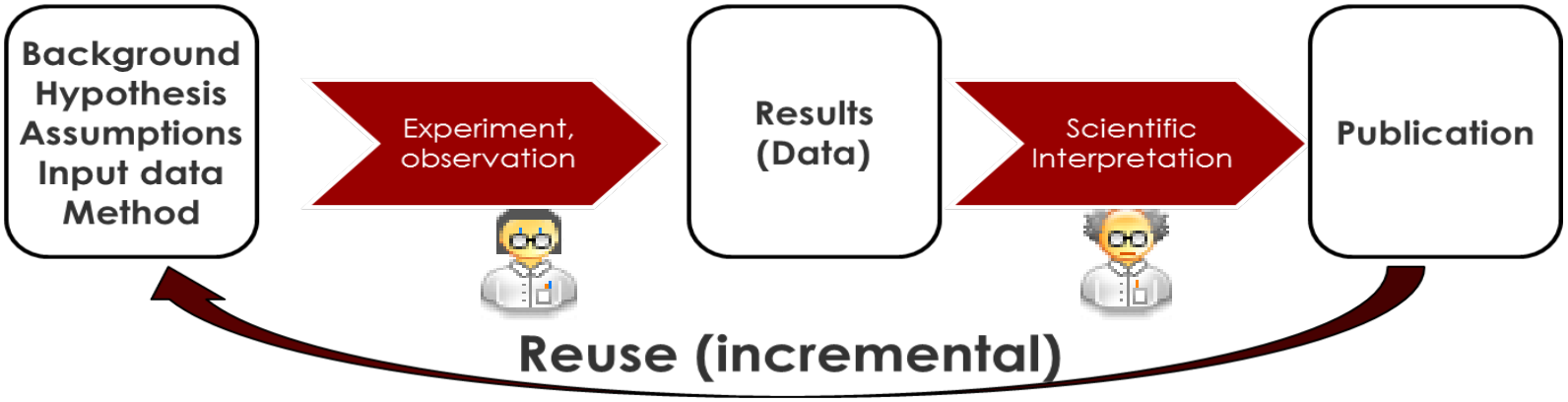
Sea Monitoring

- **Sea Monitoring**: The Sea Monitoring VRC focuses on finding new ways to measure the quality of the maritime environment and it is quite wide and heterogeneous, consisting of multi-disciplinary scientists such as biologists, geologists, oceanographers and GIS experts, as well as agencies and authorities.

Study cases:

- The scientific community has the main role of assessing the best criteria and indicators for defining the Good Environmental Status descriptors defined by the Marine Strategy Framework Directive (MSFD).

The Research Life cycle management – publication



Scientific publication in ISI (international Scientific indexing), peer reviewed **Journals with Impact factor and citation** (DOI and index citation)

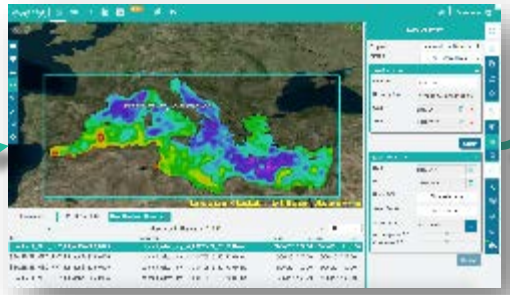
Increase Scientist citation index and scientific credits

Possibility of supplementary materials and data if requested by the journal
And choosing paper licence (open access at different level to protect your work and ensure citation)

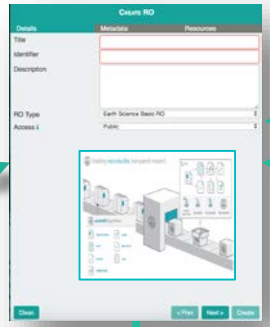
The Research Life cycle management

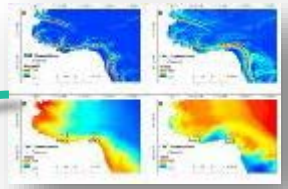
EVER-EST Collaboration sphere

EVER-EST data search interface



EVER-EST Ro creation interface





input

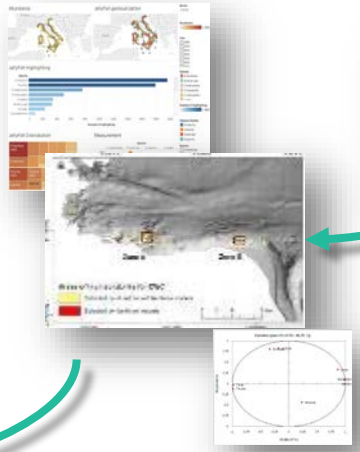
Seafile

Taverna Workflow

EVER-EST Virtual Machine

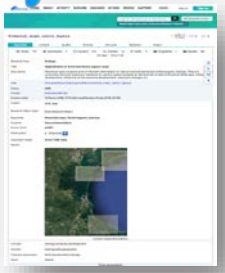
EVER-EST Workflow runner

Results



EVER-EST RO management interface

ROhub



ROhub RO management interface

The Research Life cycle management

Main daily challenges for a CNR scientist:

1. searching of existing data and products;
2. sharing methodologies;
3. working on the same workflows and data;
4. adopting shared powerful tools for data processing

Today solution:

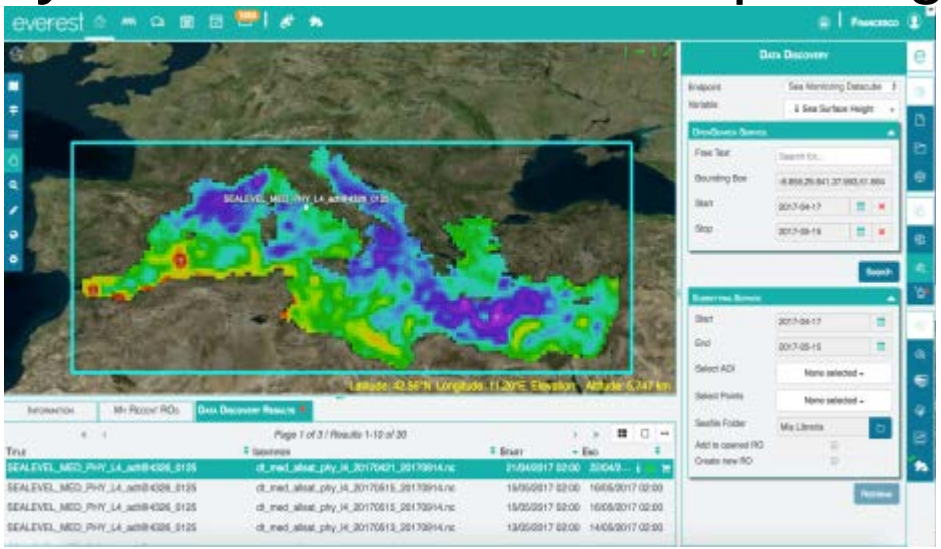
1. searching of existing data and products among many different web site, colleagues and institutional partners
2. sharing methodologies through description in scientific papers publication
3. working on the same workflows and data almost with colleagues in the same place and time (laboratory, workshops and meetings)
4. adopting shared powerful tools for data processing only if are available in the laboratory

Ever-Est Solution

1. Data sharing and Harmonization- reduction of data and knowledge fragmentation.

EVER-EST ROHUB and Collaboration sphere

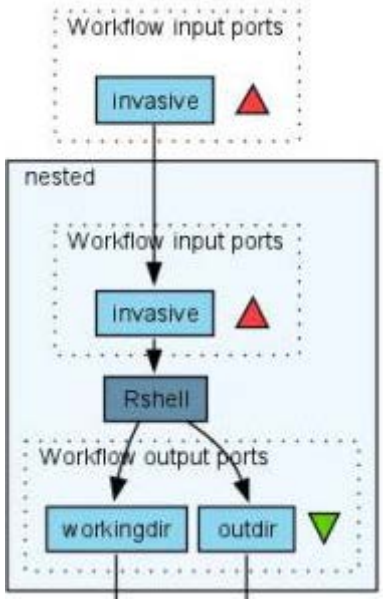
2. Easy data Discovery Re-Use and Re- Purposing of open data



Ever-Est solution

3. On line Data processing – resources and collaboration using a virtual lab

Taverna Workflow



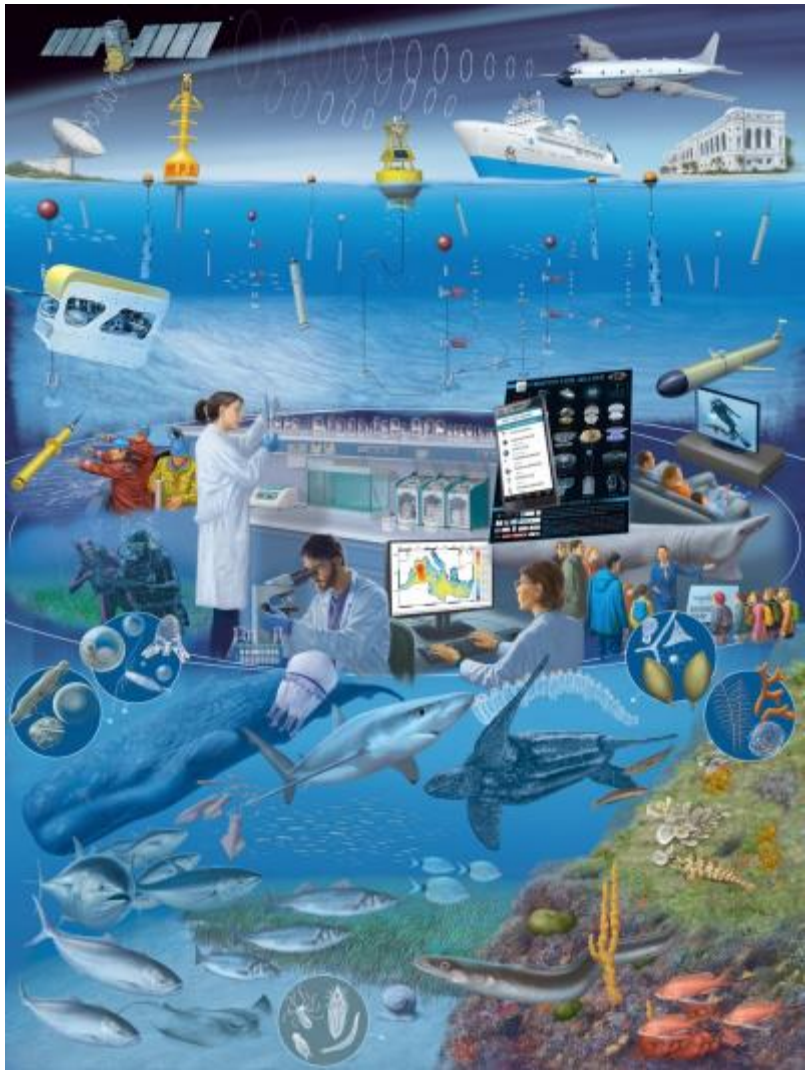
EVER-EST Workflow runner

EVER-EST Virtual Machine

4. Implementing fair data principles through the adoption of Research Object able to encapsulate, share and reproduce the entire research cycle

The Sea monitoring Community

The sea monitoring community is wide and heterogeneous including both multi-disciplinary scientists, national/international agencies and authorities dealing with the adoption of a better way of measuring the quality of the environment.

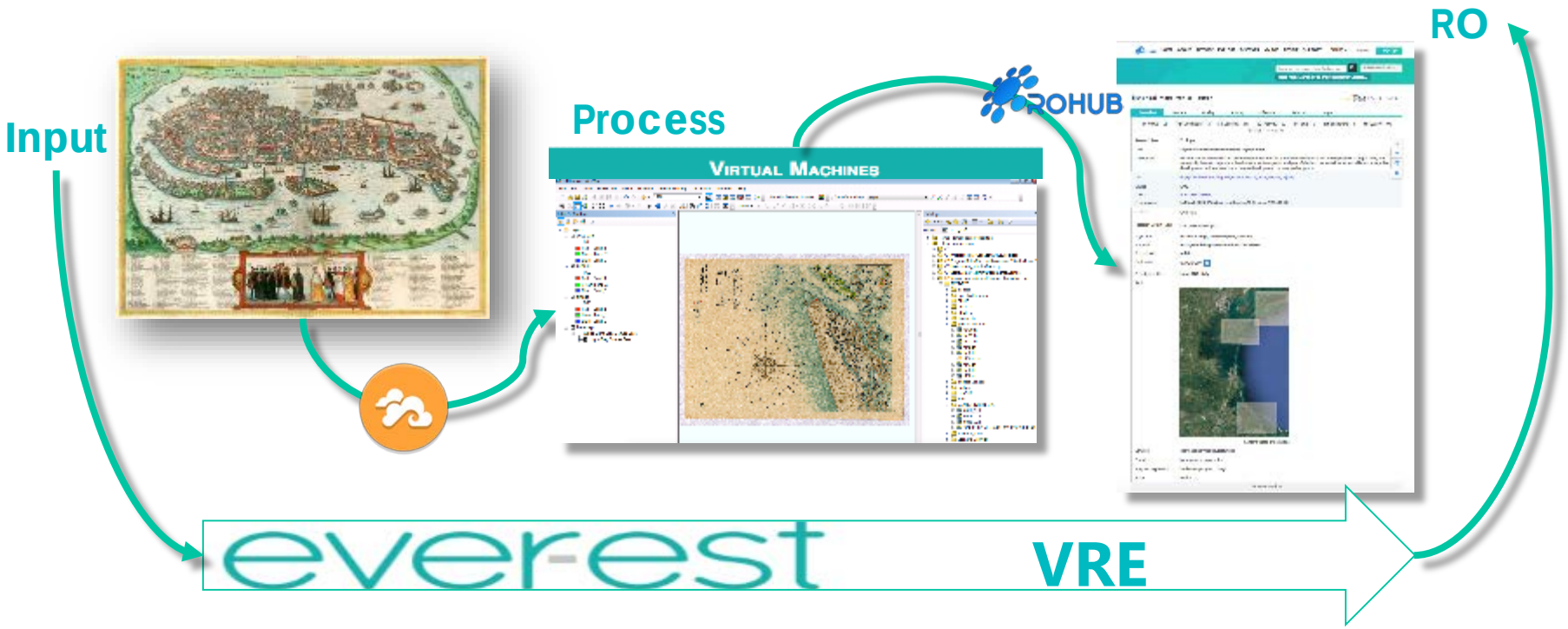


EVER-EST Sea Monitoring CASE STUDIES

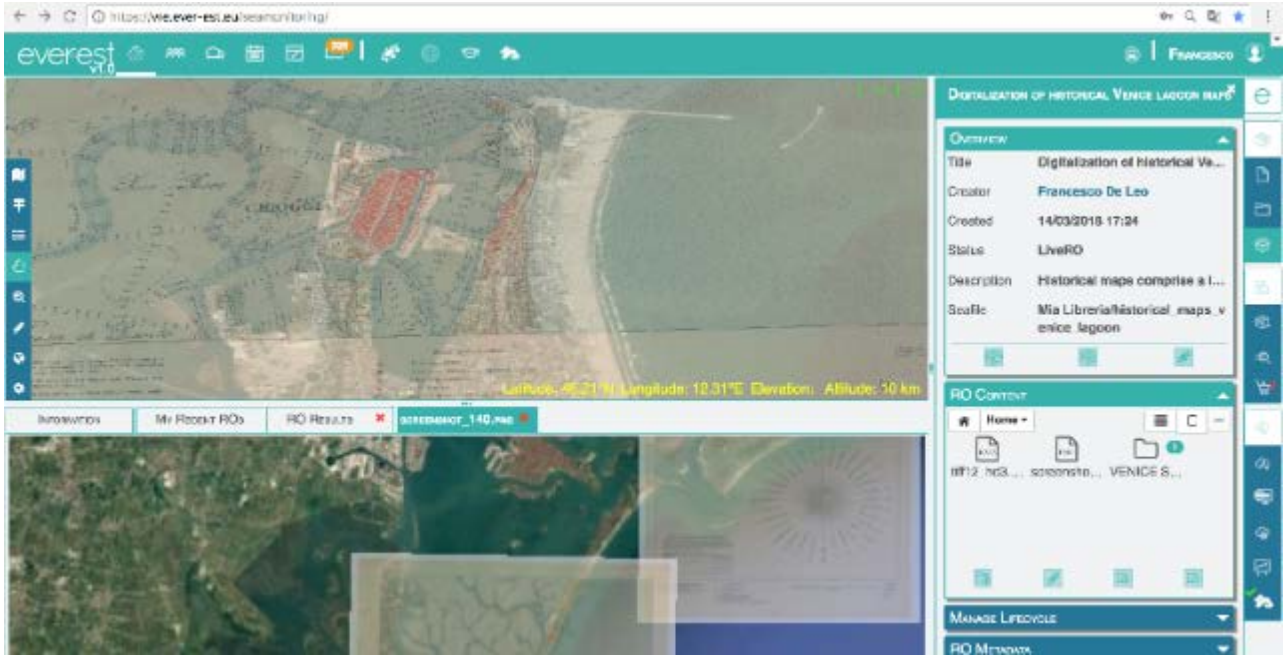
- Habitat extent Cold Water CoralsHabitat suitabilty model
- Jellyfish role to asses indicators in Marine strategy: Trending Species distribution and citizen science, evolution of invasive species.
- Mapping Human impact within lagoons from literature review
- Preserving ancient map of the lagoon of Venice for assessing changes of human footprint
- Posidonia regression along Apulian coast

Long-term “active” data preservation for ancient map

Historical maps comprise a lot of inherent information on natural environmental and anthropogenic changes. They are commonly the most important database for various spatial analyses of the land use



Study case: Preserving ancient map of the lagoon of Venice for assessing changes of human footprint



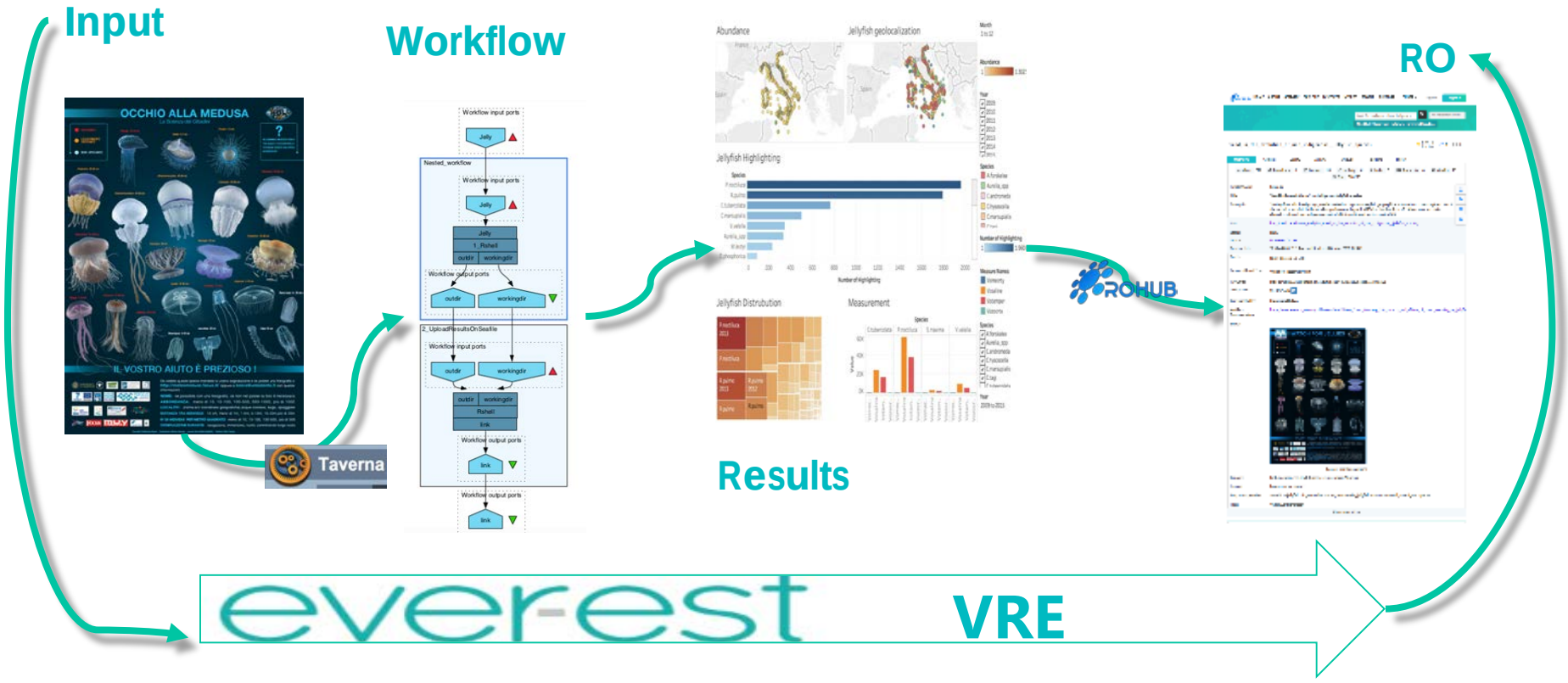
- RO Type: *DATA Ro*
- *Required tool : ArcGis on Platform Virtual Machine and SeaFile*
- Content: *High Resolution Tiff, GeoTiff*

http://www.rohub.org/rodetails/historical_maps_venice_lagoon/overview

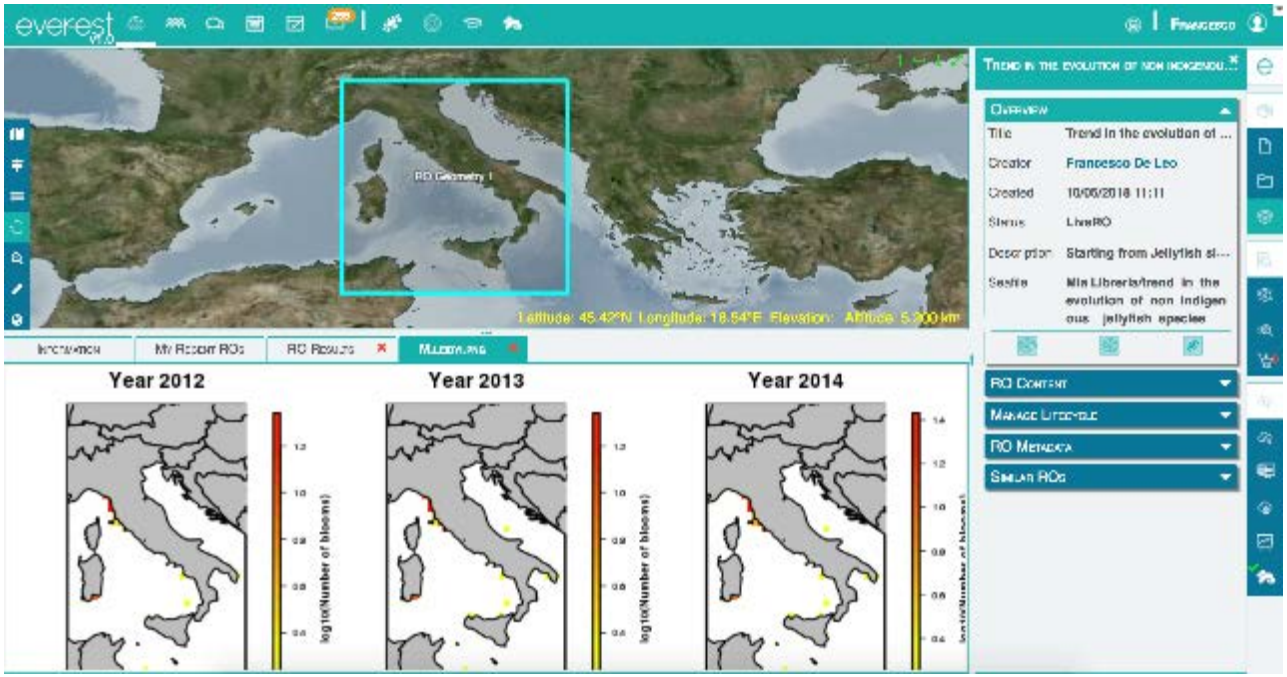
DATA ANALYSIS & CORRELATIONS: in situ observation and satellite data

Species distribution & Non-indigenous species: Jellyfish distribution to asses indicators in MSFDs

Starting from Jellyfish sightings, we elaborate data to produce explicit geographical information concerning trend about the evolution and distribution of alien species according with MSF directive descriptors 2.1: Abundance and state characterisation of non-indigenous species (NIS), in particular invasive species (IAS)



Study case: Trend in the evolution of non indigenous jellyfish species

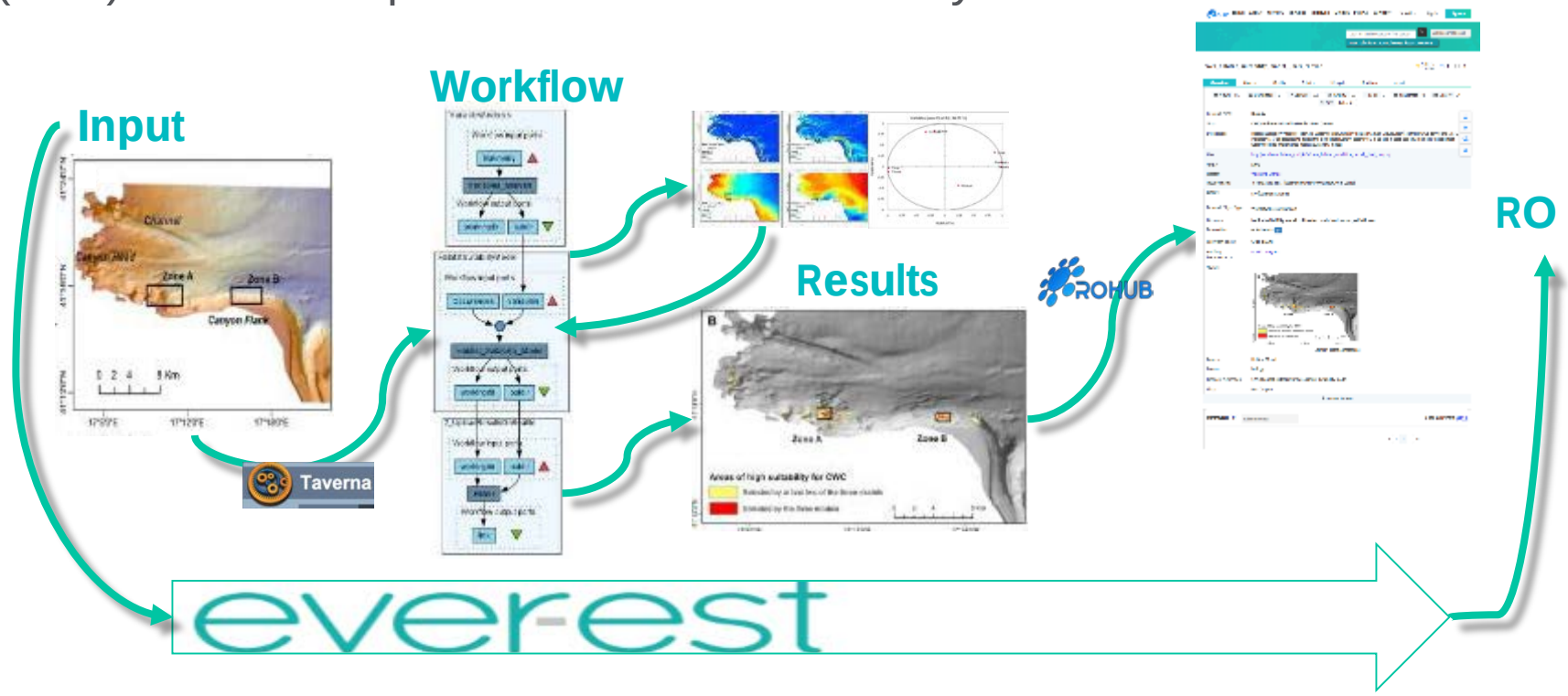


- RO Type: *Workflow Ro*
- *Required tool : R, SeaFile, Taverna Workbench on VM, Workflow runner.*
- *Input: Workflow, Jellyfish sightings*
- Output: *density annual map of the NIS jellyfish blooms*

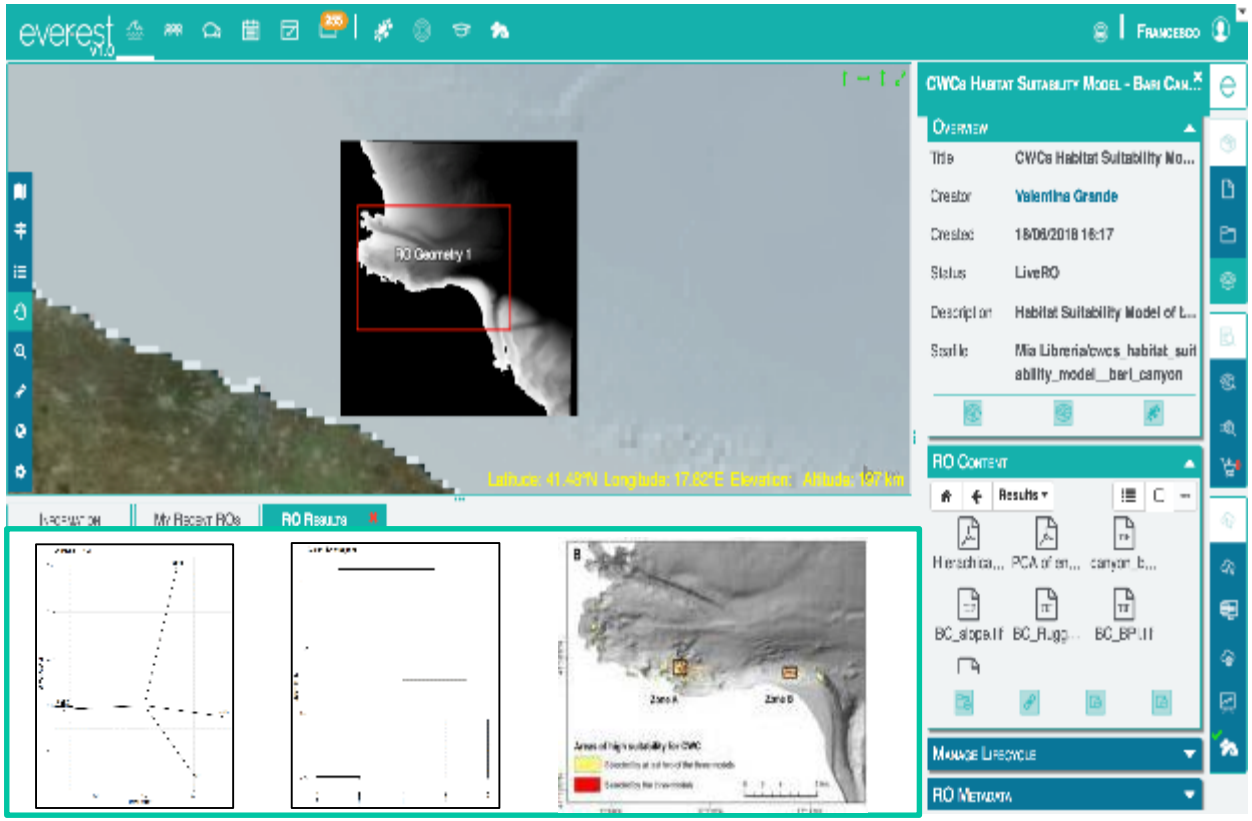
http://www.rohub.org/rodetails/trend_in_the_evolution_of_non_indigenous__jellyfish_species/overview

Habitat extent
Cold Water Corals Habitat suitabilty model

Habitat Suitability Model of the Cold Water Corals (CWCs) in the Bari Canyon (Apulia, Italy). In this RO we derive the MSFD indicator 1.5 (Habitat area) to assess the biological diversity descriptor. To do this in deep sea environment, the scientist (user) needs to implement a habitat suitability model.



Study case: CWCs Habitat Suitability Model - Bari Canyon

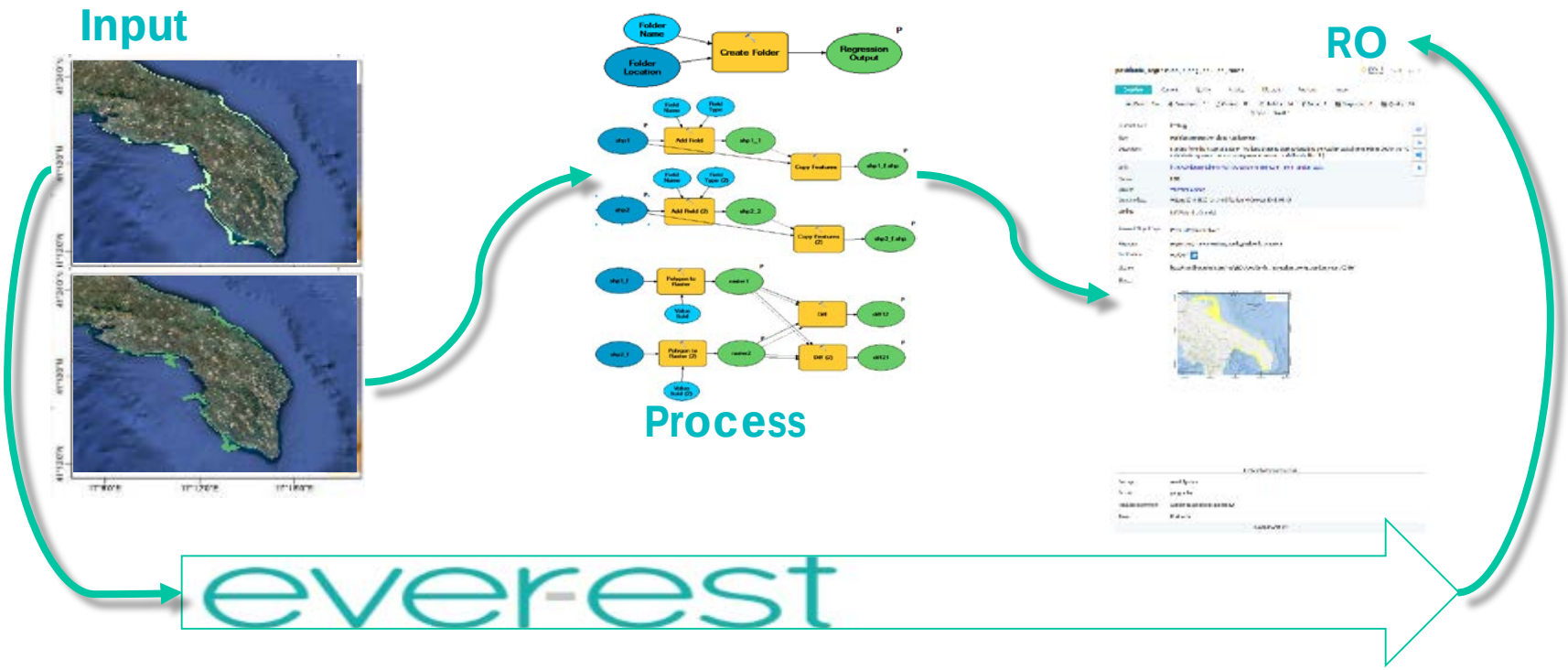


- RO Type: *Workflow Ro*
- *Required tool : R, SeaFile, Taverna Workbench on VM, Workflow runner.*
- *Input: Workflow, high resolution bathymetry, Cwc occurrence*
- Output: CWCs Habitat Suitability Model

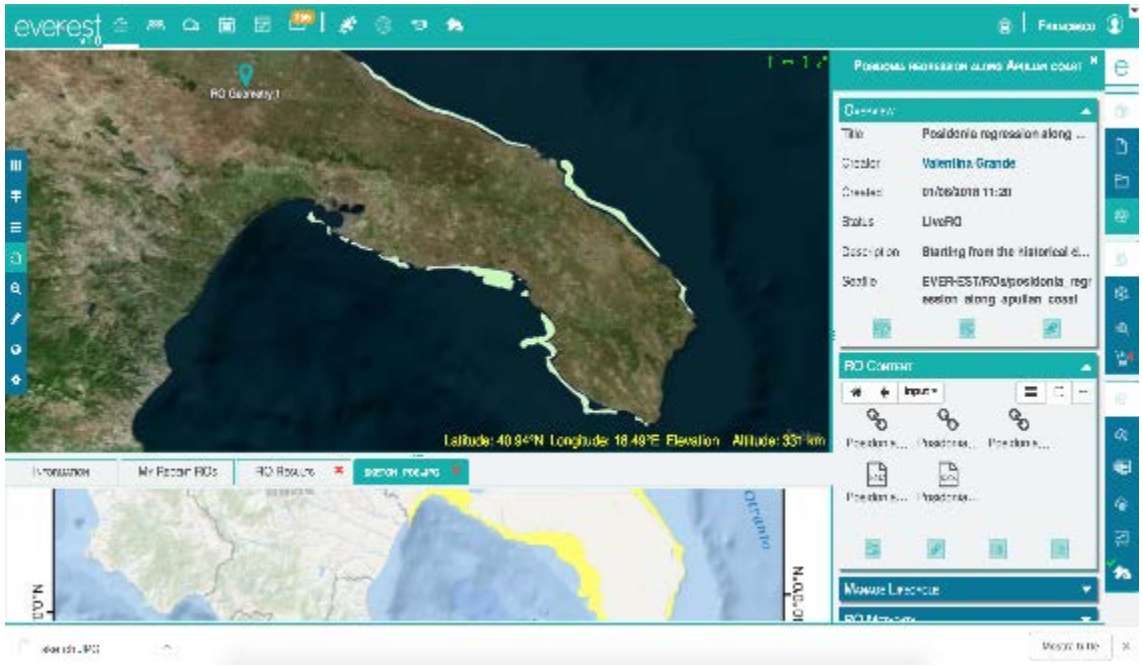
http://www.rohub.org/rodetails/cwcs_habitat_suitability_model__bari_canyon/overview

Habitat distribution and regression patterns

Starting from the historical data on Posidonia oceanica distribution along the Apulian coast (from 1986 to 2006), the RO individuate regression hotspots using a model made in model builder (ArcGIS)



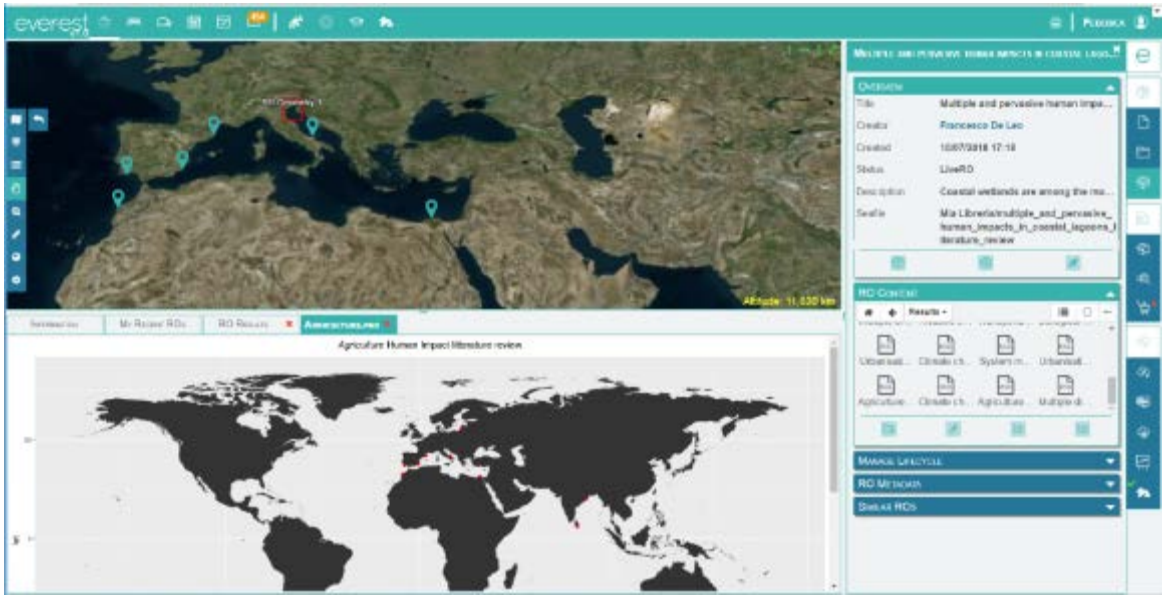
Study case: Posidonia regression along Apulian coast



- RO Type: *Process Ro*
- *Required tool : ArcGis on Platform Virtual Machine and SeaFile*
- Content: *High Resolution Tiff, GeoTiff*

Mapping Human impact within lagoons from literature review

Retrieving info from 125 papers



- RO Type: *bibliographic RO*
- *Required tool : SeaFile and R*
- Content: pdf file shape file

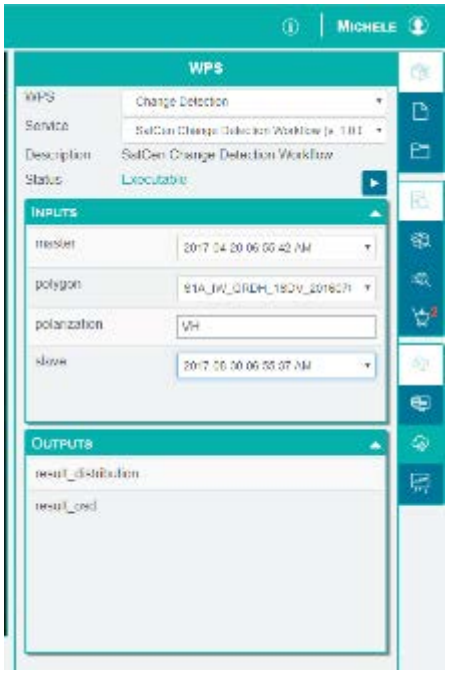
EVER-EST cross fertilisation case study

Evaluate how human activities can cause *Posidonia* meadows regression

In this study case, starting from historical data of Posidonia *meadows distribution, Sea Monitoring (SM) VRC detects regression areas* Descriptor 6, "**Sea-floor integrity"***and compares their distribution with the different human activities dectected by the WPS developed by Land Monitoring (LM) VRC (Change Detection)*

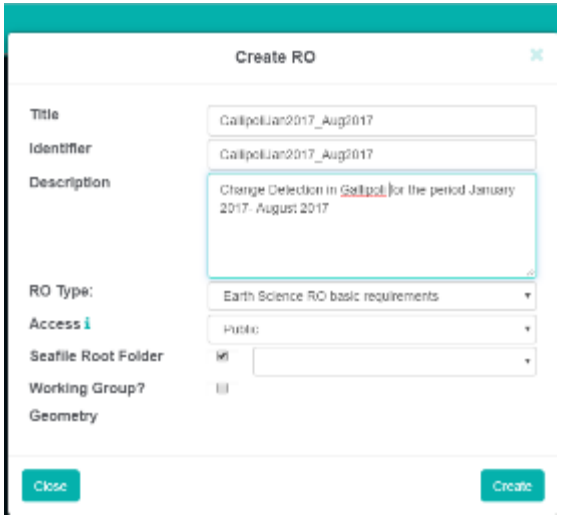
Level 1: Land Monitoring runs the WPS in the VRE in the Apulian region and create a RO with the results

WPS running



Results

RO creation



EVER-EST cross fertilisation case study

Level 2. Sea Monitoring runs a workflow developed to detect *Posidonia* regression using the Virtual Machine and create a RO with data, results, and methodologies.

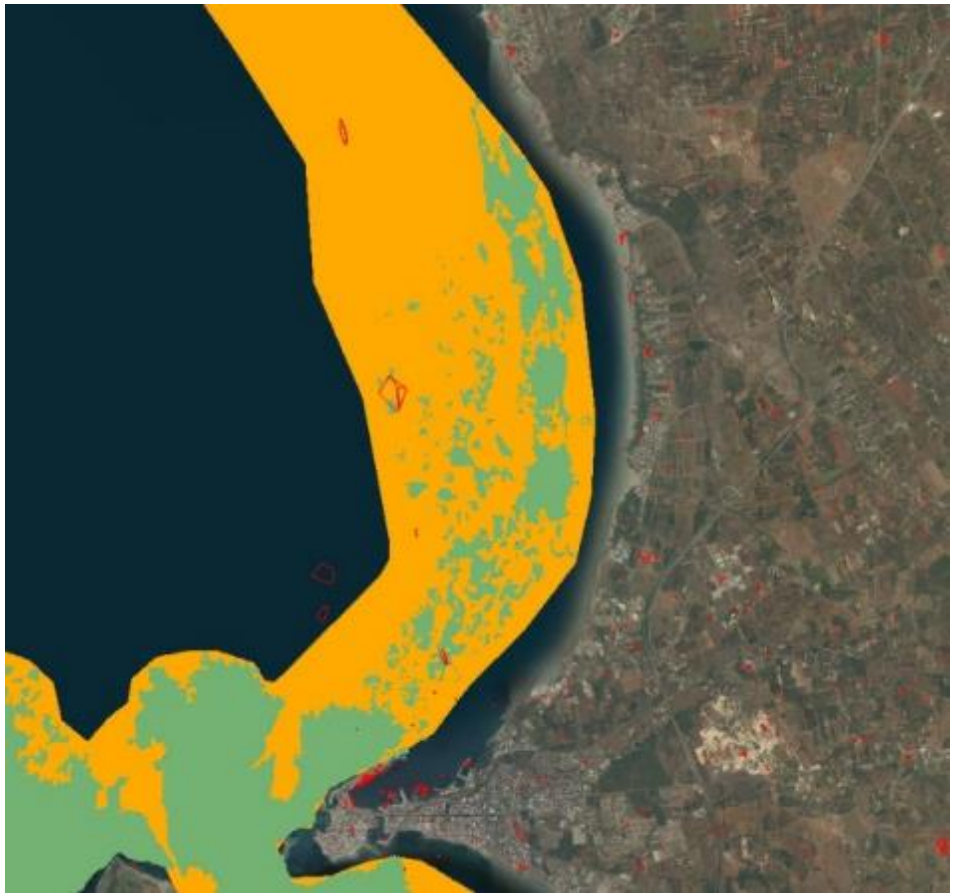
Posidonia distribution in 1986

Posidonia distribution in 2006

Diff analyses result

EVER-EST cross fertilisation case study

Level 3. Visual comparison between the results from LM and SM analysis.



Conclusions: from this first analysis appears to be a correlation between the human activities (anchorage) detected by LM and the *Posidonia* regression off shore Gallipoli detected by SM

Conclusions

- The EVER-EST project has demonstrated the relevance of Research results (Research Object) standardisation and interoperability to boost innovation and open science (FAIR principle)
- ROS (data Ros, Workflow Ros, Bibliographic Ros, Golden ROs) complemented by Data and Publication DOIs enable the bi-directional link between the data and the research output results and assure the automatic recording and tracking of the quality of the research results and ROs
- The functionality of GeoReferencing ROs proves invaluable for Data Provider to assess data set valorisation requirements including historical maps ingestion to built long term data series from satellite images back to historical ground measurement (e.g. sea-monitoring data cubes, ARDs),

Thanks for your attention

Questions?