
RESEARCHER PAPER

EVER-EST: The Platform Allowing Scientists to Cross-Fertilize and Cross-Validate Data

Mirko Albani¹, Rosemarie Leone¹, Federica Fogliini², Francesco De Leo², Fulvio Marelli³ and Iolanda Maggio⁴

¹ European Space Agency, IT

² CNR-ISMAR, IT

³ Terradue, IT

⁴ Rhea Group, BE

Corresponding author: Iolanda Maggio (Iolanda.Maggio@esa.int)

Over recent decades large amounts of data about our Planet have become available. If this information could be easily discoverable, accessible and properly exploited, preserved and shared, it would potentially represent a wealth of information for a whole spectrum of stakeholders: from scientists and researchers to the highest level of decision and policy makers. By creating a Virtual Research Environment (VRE) using a service oriented architecture (SOA) tailored to the needs of Earth Science (ES) communities, the EVER-EST (<http://ever-est.eu>) project provides a range of both generic and domain specific data analysis and management services to support a dynamic approach to collaborative research. EVER-EST provides the means to overcome existing barriers to sharing of Earth Science data and information allowing research teams to discover, access, share and process heterogeneous data, algorithms, results and experiences within and across their communities, including those domains beyond Earth Science. The main objective of this paper is to present the EVER-EST platform in all its components describing the most relevant use cases implemented by the Virtual Research Communities (VRCs) involved in the project.

Keywords: Virtual Research Environment; Remote Sensing; Research Object; Cross-fertilization; Data Analysis; Earth Science; Education

1. Introduction

In recent years, Earth Science communities have been facing an important change in the traditional management of data. In particular, Earth Observation (EO) data are constantly growing in terms of variety, volume, velocity, veracity and value, while advances in Information Technology (IT) are boosting emerging approaches for data management, built upon new software architecture styles. Contrary to the traditional analysis in which users process and analyse data locally on their workstations, unnecessary downloads of raw data are avoided in favour of enriched and more digested data.

Earth Scientists need to easily discover, access and exchange reliable (curated) data, and have access to suitable processing power, visualization and analytics tools. They also need to share with peer scientists and communities, for validation and reuse, their methods and approaches, observations, results and – most of all - the lessons they have learnt. Open Science has a growing impact on the entire research cycle, from the inception of research to its publication, and on how this cycle is organised (Bechhofer et al. 2013). New online platforms and VRE aim to improve the capacity to access, process, analyse and visualize this huge amount of heterogeneous data to provide insights with timely, clear and useful information. The EVER-EST H2020 project developed a VRE for Earth Sciences, with the means to manage both the data involved in their disciplines and the scientific methods applied in their observations and modelling. The EVER-EST project followed a **user-centric approach** driven by use cases and scenarios gathered from four preselected user communities, called Virtual Research Communities (VRC). The EVER-EST VRCs both acted as early adopters of the VRE infrastructure and have been committed to promote the scientific paradigm shift towards.

The EVER-EST project successfully demonstrated the concept of a Virtual Research Environment (VRE) for research lifecycle management in Earth Sciences based on a service oriented architecture that enables the

integration of innovative ICT components and state of art services for research data management in Earth Sciences. The E-infrastructure enables scientists to manage the entire research **life cycle** of their scientific investigations, attribute and credit findings, validate claims, preserve and share research materials and results with the scientific community and the general public.

Central to this work is the use of **Research Objects (ROs)** as semantically rich aggregations of data, methods and people in scientific investigations supporting the implementation of FAIR guiding principles and the systemic change of science practices to Open Science, particularly in e-Science. ROs allow encapsulating scientific knowledge and provide a mechanism for preserving, sharing and discovering assets of reusable and reproducible research.

2. Research Object

Modern science requires one to systematically capture the research lifecycle and to provide a unified entry point with accepted (standardized) means to access all sorts of information about the scientific investigation, including e.g. the hypotheses investigated, the data used and produced in a study, the type of analytics and computations used, the derived conclusions, the researchers themselves, and the different versions and licensing of data or software, to name but a few. Research Objects are a key enabler of such vision, with the potential to accelerate the production of scientific knowledge and foster the adoption of good data (and method) management practices, reinforcing the FAIR Data Principles.

A Research Object (RO) is defined as a semantically rich aggregation of resources that bundles together essential information relating to experiments and investigations. The original definition of RO is available in Bechhofer et al. (2013).

Basically, RO is the way to formalize and standardize the all research lifecycle, in order to promote the reuse and or customization of the entire or a part (scrip, dataset etc.) of the process. This information is not limited merely to the data used and the methods employed to produce and analyse that data, but it may also include the people involved in the investigation as well as other important metadata that describe the characteristics, inter-dependencies, context and dynamics of the aggregated resources. As such, a research object can encapsulate scientific knowledge, workflows and provide a mechanism for sharing and discovering assets of reusable research and scientific knowledge within and across relevant communities, and in a way that supports reliability and reproducibility of investigation results (Palma et al. 2014).

The research object contains a workflow, input data and results, along with a paper that presents the results and links to the investigators responsible. Annotations on each of the resources (and on the research object itself) provide additional information and characterize, e.g. the provenance of the results (the results were obtained by executing the workflow on the input data) (ESA WP4-D4.1. et al. 2015). A more intuitive illustration of the research object paradigm is shown in **Figure 1**.

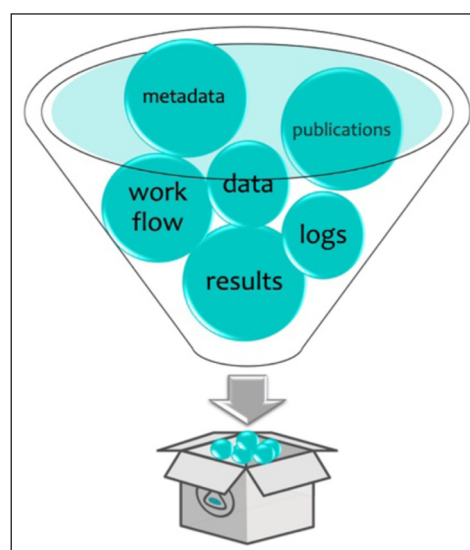


Figure 1: Research Object schematic representation.

Scientific workflows represent a key technology paradigm in the scientific community as they allow scientists to delineate the steps of a complex analysis, record the steps of computational experiments and expose this to peers using workflow design, execution and sharing tools and platforms. A scientific workflow can be defined as a series of structured activities and computations that occur in scientific problem-solving. From a computational perspective, such a workflow could be defined as a directed acyclic graph whose nodes correspond to analysis operations and whose edges specify the flow of data between those operations. In any case, the usefulness of workflows goes beyond the mere description and execution of a set of computations since they play an important role as an executable artefact for sharing, exchanging and reusing scientific in-silico methods, as demonstrated by existing workflow repositories, such as myExperiment and crowdLabs. Their high scholarly value lies in the fact that:

- They allow the assessment of the reproducibility of results;
- They can be reused by the same or by a different scientist;
- They can be repurposed for other goals than those for which they were originally built;
- They can validate the method that led to a new scientific insight;
- They can serve as live-tutorials, exposing how to take advantage of existing data infrastructure.

More specifically, by encapsulating workflows, into research objects and accompanying them with the necessary data and metadata needed for their execution and understanding, one makes the latter more (re-)usable and preservable. This metadata can include, among others, details like authors, versions, citations, etc., and links to other resources, such as the provenance of the results obtained by executing the workflow or datasets used as input. Such additional information enables a comprehensive view of the scientific investigation, encourages inspection of its different elements, and provides the scientist with a clearer picture of the investigation's strengths and weaknesses with respect to decay, adaptability and stability.

Beyond the fascinating concept, the RO needs to be translated in a real object useful in daily environmental research activity. This operation presents a series of constraints. First of all the researcher need to change his way to work formalizing processes and using open source software compatible with the aims of the RO (sharing and reproducibility), in addition the encapsulation and relative metadating of the several components require specific competencies that can be very distant from the operator's background (computer science etc.). In addition, all this activities, especially at the beginning, are time consuming.

As discussed above these critical issues may be an obstacle to the massive utilization of RO.

To solve the problems following a user-centric approach with real use cases driving the implementation of the VRE, EVEREST project assumes as central the concept of the RO. Although several e-laboratories are incorporating the research object concept in their infrastructure, the work done with research objects during EVER-EST, is a novel effort done to adapt the RO model to Earth Science and support automatic generation of research object content-based metadata as presented at the 2017 IEEE 13th International Conference on e-Science (Gomez-Perez et al. 2017). The EVER-EST VRE is the first infrastructure to leverage the concept of Research Objects and their application in observational rather than experimental disciplines.

Research objects aim to account, describe and share everything about your research, including how those things are related (**Figure 2**).

- To provide a logical organization in a single information unit of the materials, methods and outcomes of an investigation
- To uniquely identify and share your research materials and methods with other scientists at discrete milestones of the investigation
- To be recognized and cited
- To provide evidence to findings claimed in scholarly articles
- To enable reproducibility and reuse
- To preserve scientific results, preventing decay

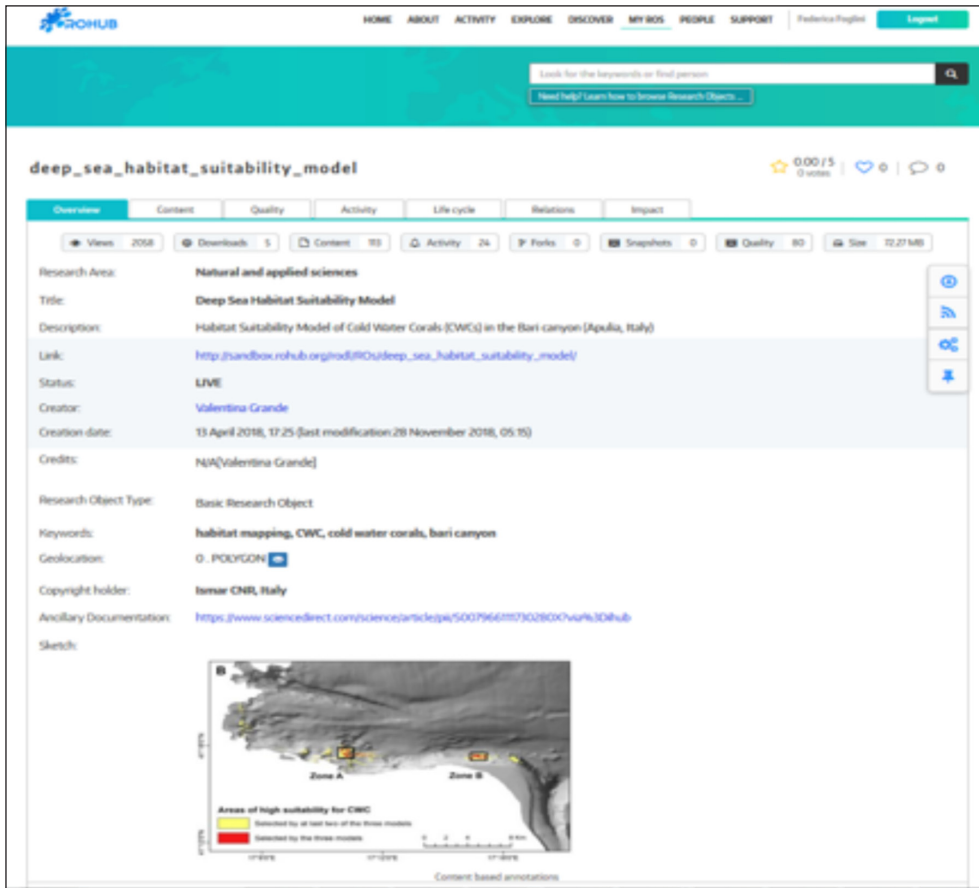


Figure 2: Example of a RO, visible troughs the ROHUB platform, encapsulating the deep Sea Habitat suitability model workflow, data, results and papers.

3. Collaboration Sphere

The Collaboration Spheres provides an alternative user-oriented interface for RO discovery targeted to scientists interested in finding Research Objects based on similarity aspects. Collaboration Spheres is a web user interface for the visualization of correlation between similar objects (e.g., users, Research Objects) based on collaborative filtering and versatile keyword content-based recommendations.

The visual metaphor implemented by the Collaboration Spheres web application is based on a set of concentric spheres centred around a central point that represents the user.



Figure 3: Collaboration Sphere.

These spheres represent different types of similarity metrics between the context of interest and the results obtained by the recommenders. The context is expressed by the user as a collection of research objects as well as other users that the user finds relevant for a particular purpose. The distance between the centre, i.e. the user and the context of interest, and the two external spheres, where recommendation results are displayed, provides a notion of confidence about the recommendations. The closer to the centre, the more specific the recommendation result will be with respect to the user and the current context of interest.

Figure 3 illustrates the Collaboration Spheres user interface.

Around the User there are 3 spheres:

- The Inner Sphere that represents the context of interest. This circle contains the users and research objects that are selected or pre-defined by the active user. In order to create a context of interest, the inner sphere is populated by drag-and-dropping relevant research objects and users from lists ranked by relevance.
- The Intermediate Sphere (model-based recommendation). This circle contains the recommended items obtained by using the context of interest as input for the recommendation techniques. The research objects and members of social network displayed in this sphere correspond to suitable items according to the context of interest defined in the previous sphere. It follows an inside-outside criterion where the inner part flows towards the outside. At this stage the recommendation obtained by this layer uses a content-based recommendation approach based on tags and annotations. This is a social-oriented approach and it is based on the fact that people are many times more reliable than search engines (“I trust my friends more than I trust strangers.”).
- The Outer Sphere (user-based recommendation). It contains recommended research objects and users based on past user actions rather than in the information explicitly defined in the context of interest. In order to populate this sphere, predictive models are used to provide the user with new possible interests.

Concerning the relationship between RO and OAIS, actually part of the research object model is based on OAI-ORE, which is related to the OAIS initiative. In that sense, a RO is an information artefact that can play the role of an AIP but also a DIP or SIP, depending on the use case (archival, preservation or submission).

4. Everest Platform Architecture

EVER-EST is a research and development platform that offers a framework based on advanced services to support each phase of the Earth Science Research and Information Lifecycle. The project follows a user-centric approach which have produced a wealth of innovative and state-of-the-art technologies, systems and tools for e-collaboration, e-learning, e-research and long term data preservation. It provides innovative services to Earth Science user communities for the generation, communication, cross-validation and sharing of knowledge and science outputs.

EVER-EST is currently the only e-Science infrastructure providing Earth Science communities with access to an RO-centric VRE, exploiting the full potential of ROs for the efficient management of their scientific investigations (Belhajjame et al. 2012) (**Figure 4**).

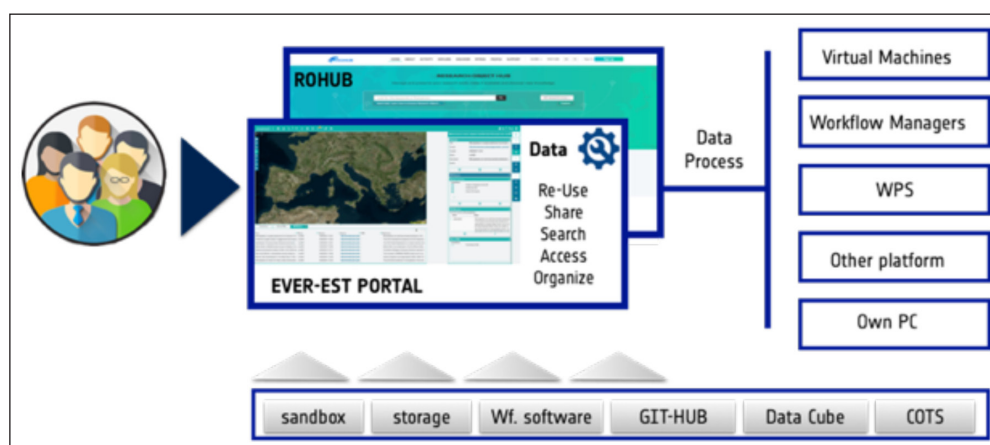


Figure 4: EVER-EST overall logic.

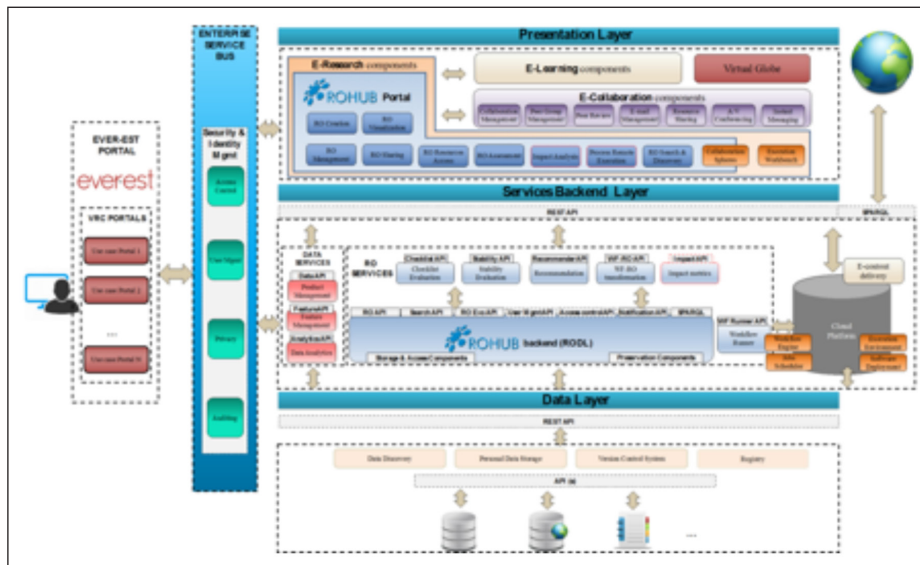


Figure 5: EVER-EST overall architecture.

The EVER-EST architecture (**Figure 5**) reflects the organization of functions across the various layers and with regards to the core element of the design (on the right of the Service Bus vertical component) the following can be identified:

1. Presentation Layer – The EVER-EST VRE offers a web application that includes both the Virtual Research Community portals and the ROHub portal the graphic user interfaces to provide the client-side implementation of e-collaboration, e-learning and e-research services along with the mechanisms for Earth Science RO creation.
2. Service Backend Layer – in the central part of the architectural diagram – provides both generic VRE services and Earth Science specific services. These components represent the reasoning engine of the e-infrastructure and actually orchestrate and manage the services available to the VRE final users. The service layer includes the server-side implementation of e-research and processing services, including cloud resources to instantiate virtual machine properly configured according to VRC specifications.
3. Data Layer – bottom part of the design – references the data holdings made available to the VRCs: data is linked and proper means are provided, where feasible, to access it from the VRE. As a default setting, data will not be copied or duplicated, but will continue to reside on the provider’s local servers unless it is directly retrieved by the user.

The two components on the left part of the core Architecture design include:

1. The VRE Gateway and the VRC GUI’s. This is actually the main Gateway used to access theVRE. Public research objects, a deployment of the Jupiter Notebook, KPI measurements plus additional on-line training modules will be available for guest users. The Gateway will contain the shortcut to each VRC user interface. While some services and functions can be seen as common and transversal to all the VRCs, a set of features and functions will be specific for each of the communities that shall require an ad-hoc design. As anticipated, specific features regarding the user interface are not in the scope of this first version of the document and will be first discussed with the VRCs during future project iterations.
2. The Enterprise Service Bus, with particular focus on the services for Security and Identity management, as the interaction of each user with the VRE will require multiple levels of security and identity controls across the entire infrastructure (e.g. rights to access the VRE, or to execute a process, discover a catalogue, access data, etc.).

EVER-EST platform helps scientists to cross-validate and fertilize data implementing the following functionalities:

- 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)

5. Virtual Research Environment

The EVER-EST e-infrastructure is validated by four virtual research communities (VRC) covering different multidisciplinary Earth Science domains including: ocean monitoring, natural hazards, land monitoring and risk management (volcanoes and seismicity) (ESA WP5-D5.1 et al. 2015) (**Figure 6**).

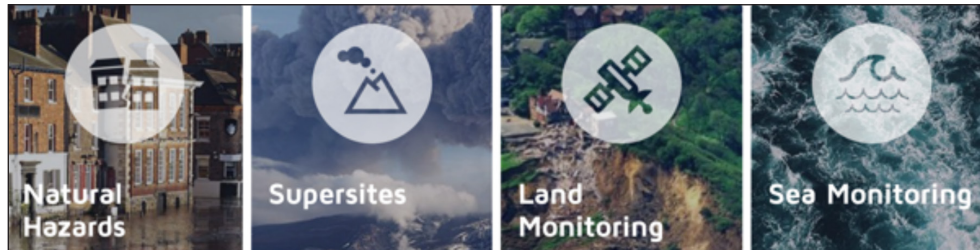


Figure 6: The EVER-EST VRCs.

- **Land Monitoring:** Land Monitoring can refer to the monitoring of urban, built- up and natural environments to identify certain features and anomalies or changes over areas of interest as well as the assessment of natural resources.

The development of a VRE for Land Monitoring scenarios aimed to deal with issues addressing different communities that potentially have different final goals, but which are using the same space assets and similar services/techniques (e.g. security, environment, urban planning). The Land Monitoring pilot is led by the European Union Satellite Centre (SatCen) which represents, in the framework of EVER-EST and in line with the “Secure Societies” Horizon 2020 Societal Challenge, the stakeholders involved in the decision-making process of the EU in the field of the Common Foreign and Security Policy (CFSP).

The design of the conceptual workflow and of the core processing chain was preceded by the collection of user requirements. As such, through the VRC the user is able to: access open data such as Copernicus Sentinel-1, to execute an automatic processing (Change Detection) and to encapsulate the results in a RO (**Figure 7**). The Change Detection service allows to select a pair of Sentinel-1 GRD images from the SciHub, within a timeframe, and to identify changes through suitable algorithms. The service has been deployed on the T2 Sandbox and it can be initiated via the EVER-EST Land Monitoring Portal through a WPS. The service launches a set of chained processing modules based on the Sentinel Application Platform (SNAP): all the chain is transparent to the users, which just display the last output on the GUI and it is meant to represent a pre- operational use of the VRE.

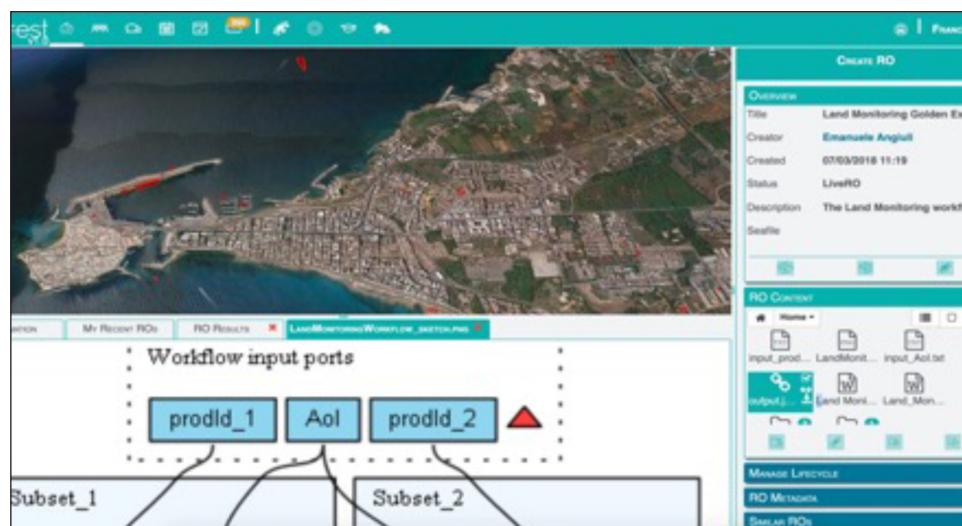


Figure 7: Land monitoring user interface showing a Detection Map.

- **Sea Monitoring:** Sea Monitoring VRC represents a wide and heterogeneous marine community, including both scientists and national/international agencies and authorities (e.g. MPA directors, domain experts from regional agencies like ARPA in Italy, the technician working for the Ministry of the Environment) dealing with the assessment of the Good Environmental Status (GES) of the European seas within the Marine Strategy Framework Directive (MSFD). In this context the VRC developed through the VRE several case studies providing practical methods, procedures and protocols to assess the best GES in their own sub-regions, proving the effectiveness to adopt EVER-EST for managing the whole scientific life cycle. So far the methodologies and the results were focused on 1) benthic habitat mapping such as Cold Water Corals habitat suitability models, 2) mapping the trend in the evolution of non-indigenous jellyfish species; 3) mapping *Posidonia* regression along the Apulian coast (**Figure 8**); 4) preserving ancient maps of the lagoon of Venice for assessing changes in the human footprint.

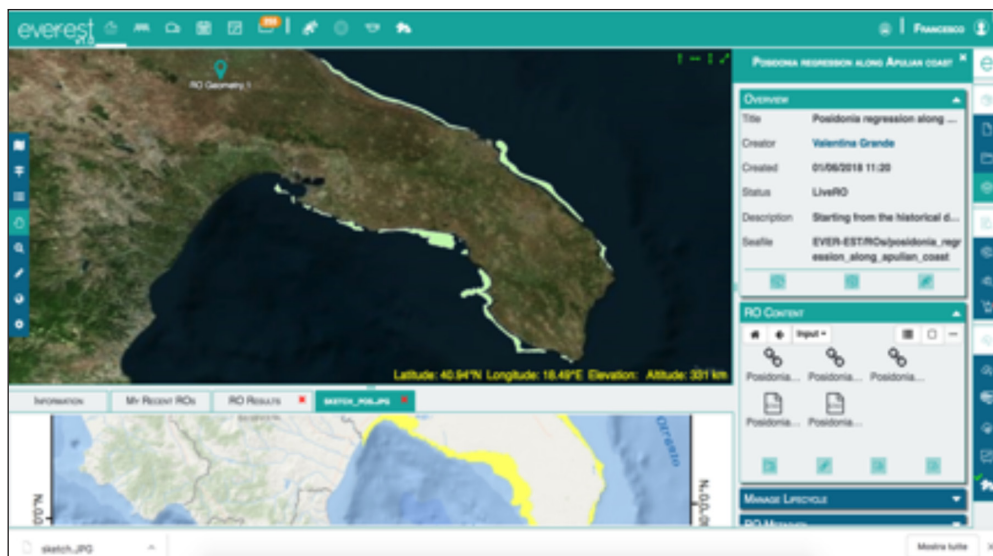


Figure 8: Sea monitoring user interface showing the *Posidonia* regression along Apulian coast.

The operational scenarios are concerned with fulfilling integration and homogenization of available data and protocols from different sources (including literature), the discovery of the data and documents related to the MSFD and the sharing of this information among the MSFD community. The Sea Monitoring portal provides the main user web interface to create and share Earth Science ROs, to discover different kinds of marine data including in situ observations and satellite images, to access, to process and visualize services relying on OGC standards (OpenSearch, Web Coverage Service, Web Processing Service, Web Map Service), to manage ROs, and finally, to execute remote workflow implemented via Taverna. Moreover, the VRE provides different user interfaces for integrated functionalities such as: a) collaboration spheres, for the visualization of correlation between similar objects (e.g., users, ROs) based on collaborative filtering and versatile keyword content-based recommendations; b) RoHub, the reference platform for RO management supporting the preservation and lifecycle management of scientific investigations, research campaigns and operational processes; c) Jupyter Notebook, for capturing the whole computation process: developing, documenting, and executing code, as well as communicating the results; and d) Virtual Machine.

- **Supersites VRC:** The Geohazard Supersites and Natural Laboratory initiative (GSNL)¹ is a voluntary international partnership formalised under GEO – Group on Earth Observations, aiming to improve, through an Open Science approach, geophysical scientific research and geohazard assessment, promoting rapid uptake of scientific results for Disaster Risk Reduction. GSNL is focused on seismic and volcanic hazards, and the Supersite community is composed of globally dispersed scientists studying these high risk regions. To develop and test the VRE user scenarios were defined focusing on the Supersites Campi Flegrei, Mount Etna and Icelandic Volcanoes. The VRE has been designed in order to provide a series of tools and services to make the scientist's work more effective, satisfying the requirements of the community regarding the collaboration, the attribution and recognition, long

term preservation and re-use of scientific resources. The Supersite VRE provides access to SAR data (Sentinel-1, ERS, ENVISAT, COSMO-SkyMed) and optical (Modis) catalogues, used to map short- and long-term surface deformation due to volcanic/seismic activity, and volcanic plumes, respectively. The Supersites community considers the ROs management services provided by the VRE a very important tool to support the full implementation of Open Science. ROs facilitate the findability, reusability and reproducibility of research work, but they also guarantee the proper attribution and preservation of scientific findings. The VRE allows to search, display, edit or execute the content of ROs. The main RO use cases for this community are: to store and refer bibliographic resources, to document datasets and scientific results, to share workflows and executable software. To do this, the VRE provides access to computational resources (on Linux and Windows OS) where various software are available for workflow execution and data analysis. To date, over 30 Supersite users spread in four continents are using the VRE and the EVER-EST computing resources (**Figure 9**).

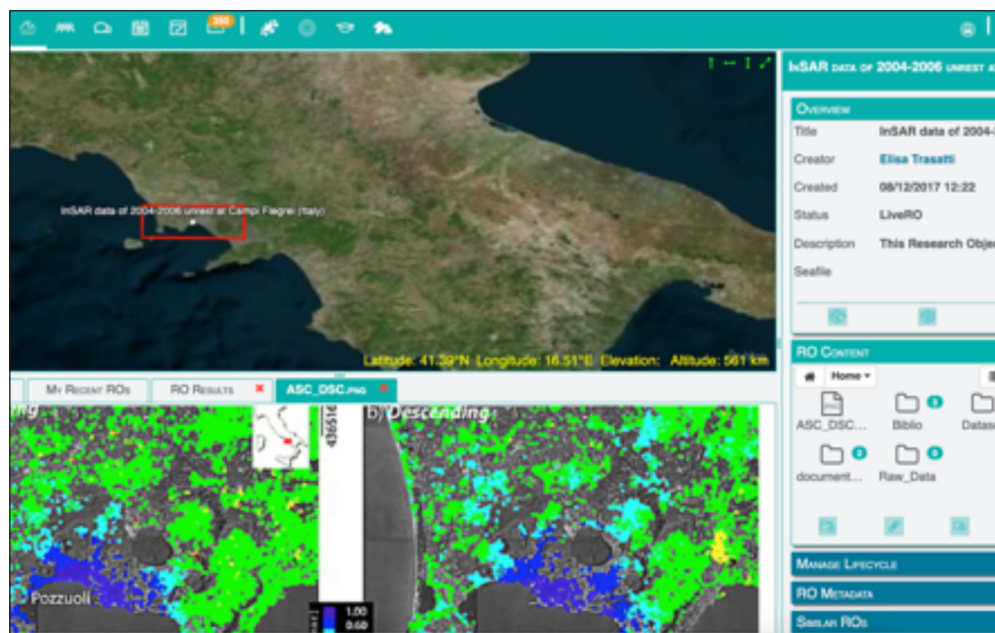


Figure 9: Visualisation in the VRE of a RO containing executable resources.

- Natural Hazards Partnership:** The Natural Hazards Partnership (NHP) is a group of 17 UK 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. The NHP provides daily assessments of hazard status via the Daily Hazard Assessment (DHA) to the UK responder and resilience communities, pre-prepared science notes providing descriptions of all relevant UK hazards and input to the National Risk Assessment. The VRE is allowing the NHP partners to explore novel ways of collaborating in a virtual environment through the provision of tools and services. Collaboration tools are complemented by the adoption of ROs by the NHP's Hazard Impact Modelling group. Early effort was focused on using the platform as a research environment to test development of surface water flooding hazard impact models resulting in reusable science and research; scientific knowledge is encapsulated in a shareable format that can be easily used by partners working on the same model but within their areas of expertise. The Natural Hazards portal allows the user to access the Taverna workflows and all necessary data files held on the Seafile cloud storage that are used to run and test hazard impact models. The Taverna workflow is stored as a workflow RO and is therefore searchable and useable by other NHP scientists to test different modelling scenarios based on historical hazard occurrences. Whilst use of the VRE is still in research phase, and therefore use is limited to a few scientists, it is hoped that with refinement of the hazard impact modelling on the VRE (**Figure 10**) and enhanced collaboration, that the VRE may support operational delivery of forecast impact outputs in the future.

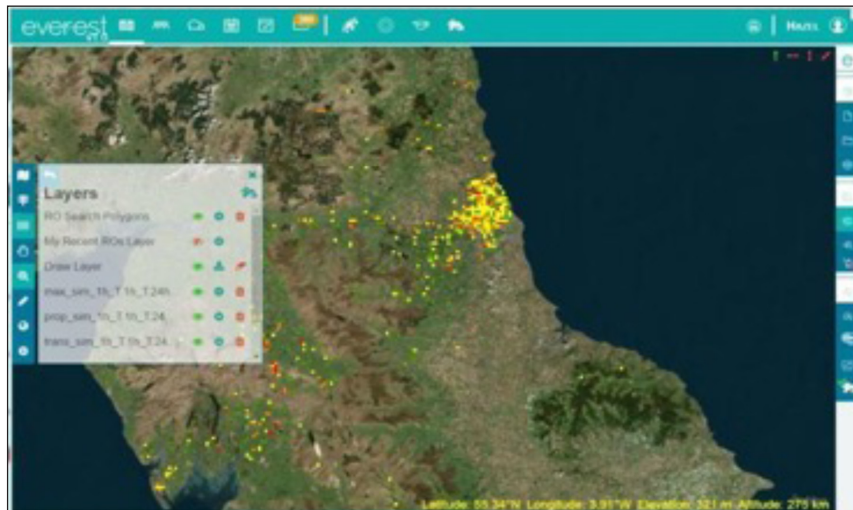


Figure 10: Visualisation of hazard impact modelling output grid generated using Taverna workflow on VRE.

Use of bibliographic ROs has also provided the capability to store and archive the NHP's Daily Hazard Assessment and all contributing evidence from NHP partners in one place providing an audit trail of decision making that is easily searchable and reusable. Each VRC uses the virtual research environment according to its own specific requirements for data, software, best practice and community engagement. This user-centric approach allows an assessment to be made of the capability for the proposed solution to satisfy the heterogeneous needs of a variety of Earth Science communities for more effective collaboration, greater efficiency and innovative research.

6. Case Study: Land Monitoring

- CHANGE DETECTION:** The Change Detection service allows to select a pair of Sentinel-1 GRD images, within a timeframe, and to identify changes through suitable algorithms (ESA WP3-D3.1. et al. 2015). The service has been deployed on the T2 Sandbox and it can be initiated via the EVER-EST Land Monitoring Portal through a Web Processing Service (WPS). It represents a pre-operational use of the EVER-EST infrastructure for the targeted community. The service launches a set of chained processing modules based on the Sentinel Application Platform (SNAP): Thermal Noise Removal, Orbit-Based Correction, Calibration, Terrain Flattening and Terrain Correction. Successively, the images are co-registered and a Change Detection algorithm identifies the areas with changes. The output of the Change Detection is a raster product containing the pixels where changes have been detected. The Land Monitoring use case provided a concrete case of cross-disciplinary interaction between Earth Scientists and Institutional entities to transfer knowledge, best practices and tools. Through the VRC, the user is able to access open data such as Copernicus Sentinel-1, to execute an automatic processing (change detection chain) and to store the results in a Research Object which is then made available for different communities.



Figure 11: Land Monitoring Use Case scheme.

The output of the Land Monitoring Change Detection service can be encapsulated in a RO (**Figure 11**). The user has the flexibility to define which components of the result should be included in the RO, to customize the annotations and metadata according to the needs of the community of practice. In this way, the RO allows to share the results and the metadata related to the workflow between different communities which might have different aims but use the same methodology (e.g. change detection).

The RO contains information on the process (e.g. master and slave image names, polarization used, Area Of Interest definition) and the results are saved as .pngw (to be projected and visualized on a user GIS interface). The RO can be saved as public or private: the RO privacy selection satisfies the requirements of the community, where stakeholders asked for a tool to share processing results but also the possibility to maintain the security of them in some specific cases (e.g. the application of the developed tool on sensitive areas).

7. Cases Studies: Sea Monitoring

- EVALUATE HOW HUMAN ACTIVITIES CAN CAUSE POSIDONIA MEADOWS REGRESSION:** This case study represents a of cross fertilization examples between two VRCs, the Sea monitoring and the Lan monitoring (ESA WP3-D3.1. et al. 2015). Coastal anthropogenic activities increased worldwide in the last half century, amplifying the pressures on marine coastal ecosystems. The management of those multiple and simultaneous threats requires reliable and precise data on the distribution of the pressures and of the most sensitive ecosystems. In this case study, starting from historical remote sensing data of Posidonia meadows distribution, the Sea Monitoring (SM) VRC detected Posidonia regression areas off shore the Apulia region in Italy and compared their distribution with the different human activities identified by the Change Detection WPS developed by Land Monitoring (LM) VRC. LM run the change detection WPS using the EVER-EST VRE service in the Apulia Region and created a RO encapsulating the Taverna workflow and the results as .shp file. In parallel SM runs a workflow implemented to detect Posidonia regression using the EVER-EST VRE Virtual Machine and created a RO with data, results, and workflows. Overlaying through the EVER-EST VRE globe the results from the LM and SM research object it was possible to visually identify a correlation visual between the human activities detected by LM and the Posidonia regression off shore Gallipoli detected by SM (**Figure 12**). Among the various types of human activities, the mechanical damages resulting from boats anchoring in shallow coastal waters appear to be responsible for localized regressions of Posidonia oceanica meadows.

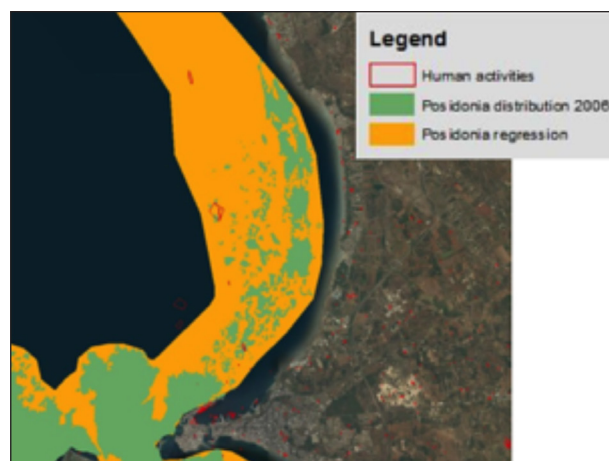


Figure 12: Overlay between SM and LM Research Object results.

- CORRELATION BETWEEN ENVIRONMENT SATELLITE VARIABLES AND JELLYFISH OUTBREAKS:** Cross-Fertilisation study in synergy between University of Tor Vergata in Rome and CNR ISMAR biological researchers group (Benedetti-Cecchi et al. 2015). The group is specialized on the quantification of deterministic and stochastic components of environmental change that lead to outbreaks of maritime species: in this specific case, the jellyfish. The Research Objects created have been cross-fertilized with the RO on “Mediterranean Sea Anomalies detection” developed by University of Tor Vergata research group. This can be considered as a good example of joint work between two communities – Earth Observation researchers and Marine Biologist – which could be not necessar-

ily strictly linked in their everyday activities and that was de facto facilitated by the common use of RO's and the adoption of the EVER-EST infrastructure as working environment. The analysis led to detect identification of correlations between the jellyfish blooms and environmental variables over specific areas of the italian coast (**Figure 13**).

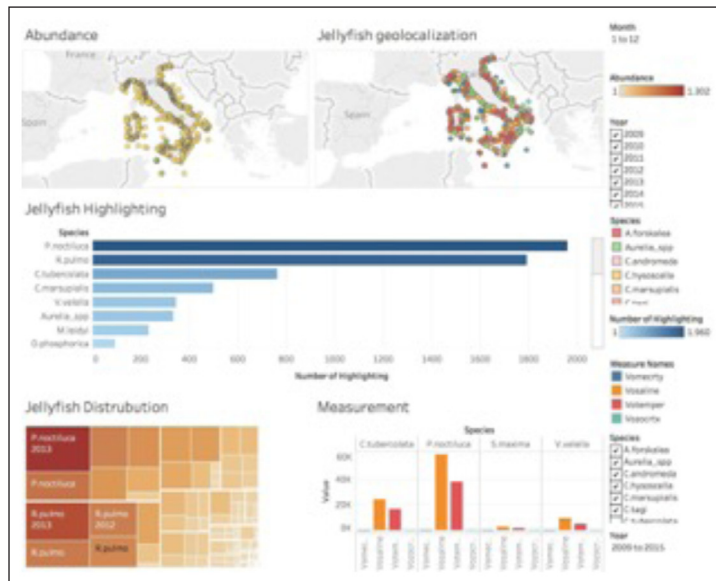


Figure 13: Jellyfish analysis Dashboard.

Some results have been graphically represented using an EVER-EST GIS tool overlapping all information produced by both studies (**Figure 14**).

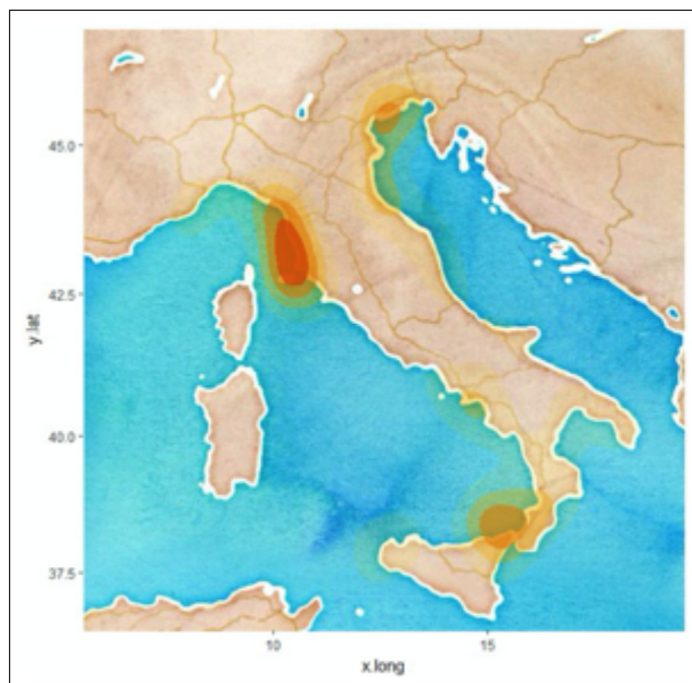


Figure 14: Observations density.

Partial results were collected in terms of light correlations with temperature, chlorophyll and particulate (**Figure 13**). We have identified some years to be better analysed and the need to create a model to solve the seasonality problem. Further analysis needs to be performed applying an ad-hoc non-metric multidimensional scaling (NMDS) analysis for defining principal components. The plots on the density help the further verification in terms of reduction of the pixel polygon for time series definition and reduction in time consuming and performances.

8. Cases Studies: Supersites

- VOLCANIC PLUME RETRIEVALS PROCEDURES:** During eruptions, volcanoes emit large quantities of particles and gases into the atmosphere (ESA WP3-D3.1. et al. 2015). The Volcanic Plume Retrieval procedure has the capability, simultaneously and in real time, to estimate physical parameters of volcanic ash and SO₂ clouds from multispectral MODIS data in the Thermal InfraRed (TIR) spectral range. Plume altitude and temperature are the only two input parameters required to run the procedure. By linearly interpolating the radiances surrounding a detected volcanic plume, the VPR procedure computes the radiances that would have been measured by the sensor in the absence of a plume, and reconstructs a new image without plume. The new image and the original one allows computation of plume transmittance in the TIR-MODIS bands 29, 31, and 32 (8.6, 11.0 and 12.0 μm) by applying a simplified model consisting of a uniform plume at a fixed altitude and temperature (**Figure 15**). The transmittances are then refined using a polynomial relationship obtained by means of MODTRAN simulations adapted for the geographical region, ash type, and atmospheric profiles.

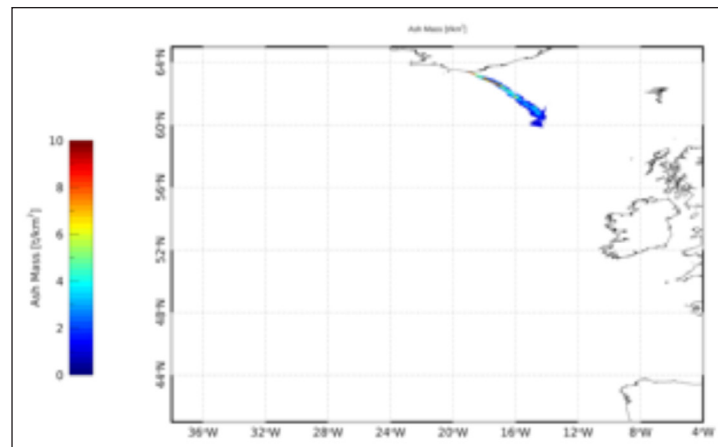


Figure 15: Map of computation of plume transmittance in the TIR-MODIS.

- VOLCANIC GEODETIC DATA INVERSION:** The RO was created to invert 2004–2006 ground deformation data for the Campi Flegrei volcano. The inverted datasets were ascending and descending Line of Sight ground displacements from COSMO-SkyMed InSAR time series. The data were modelled with a spherical magma chamber. At the end of his inversion procedure, researcher created a RO containing the input data, the inversion workflow, and the output results (**Figure 16**), then added some descriptive information and finally archived the RO with a DOI to ensure authorship of the research.

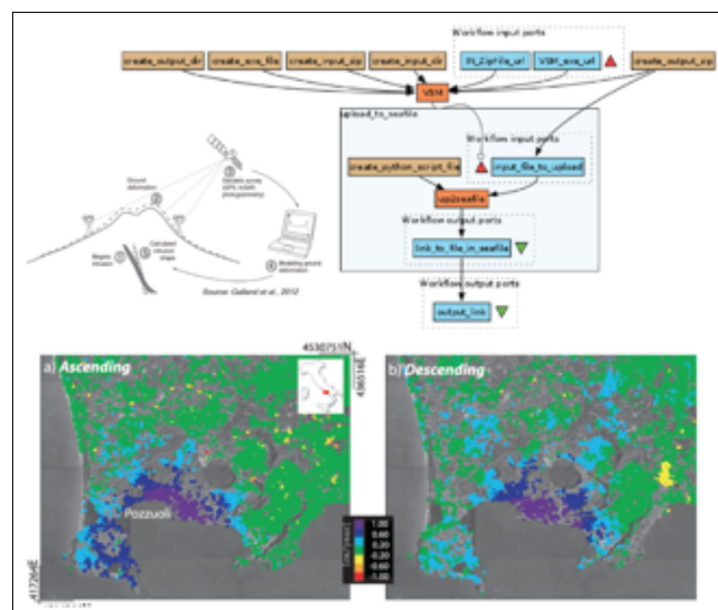


Figure 16: Schematic view of the RO content.

- INSAR PROCESSING WITH SARSCAPETM ON A WINDOWS VIRTUAL MACHINE:** This use case shows how to download Sentinel 1 SAR image data from the EVER-EST VRE interface, and launch the SARscape SAR processing software in a Windows Virtual Machine to carry out Interferometric SAR processing. The VSM RO content opened in the VRE. The following map (**Figure 17**) shows the 2004–2006 InSAR ground deformation used by the RO.

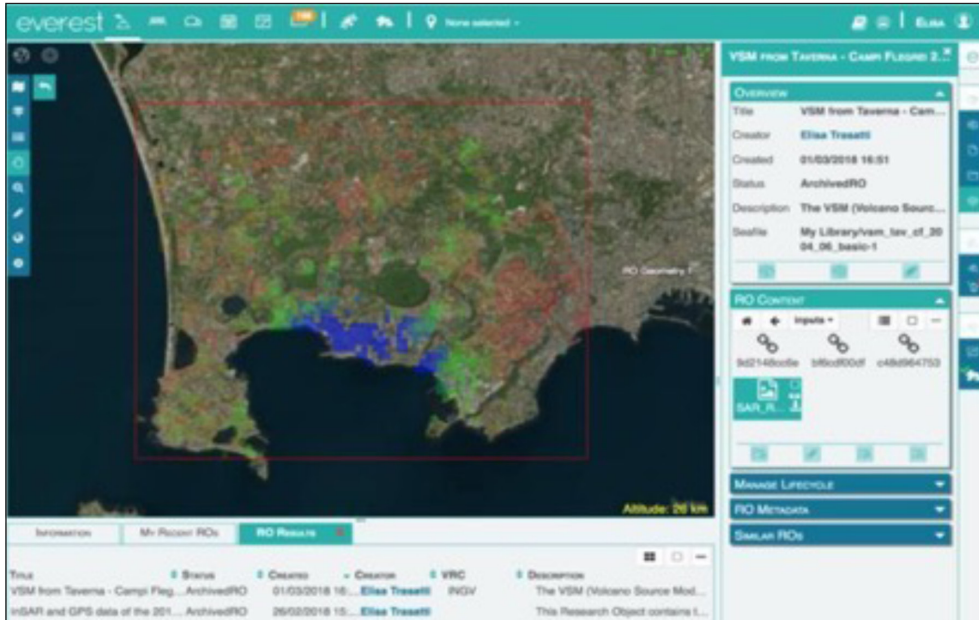


Figure 17: The VSM modelling tool performs geodetic data inversion of magmatic source.

9. Cases Studies: Natural Hazards

- SURFACE WATER FLOODING:** The Surface Water Flooding Hazard Impact Model (SWF HIM) is a well-developed Hazard Impact Model approaching operational deployment with on-going work focused on validation of impacts through chosen case studies (ESA WP3-D3.1. et al. 2015). This involves running a countrywide (1 km grid, 15 min time-step) Grid-to-Grid (G2G) hydrological runoff and routing model (CEH) using rainfall inputs (Met Office), and linking its surface runoffs to potential impacts (HSE) and verifying these against observed impacts (**Figure 18**).

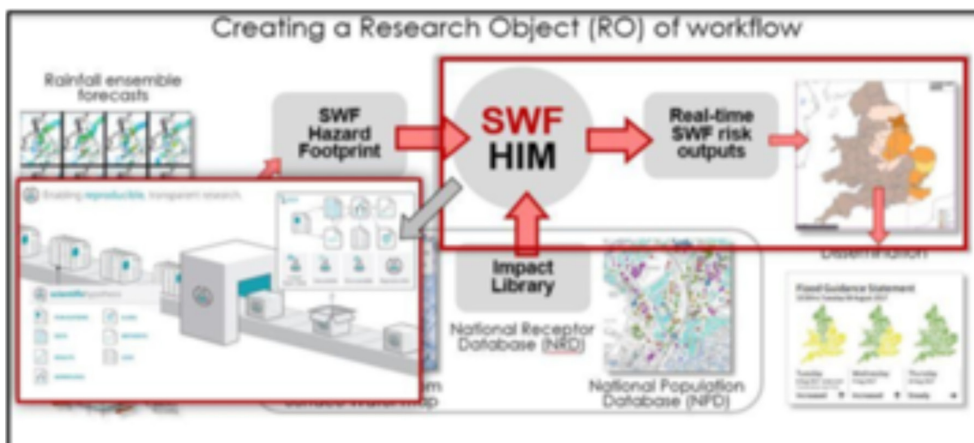


Figure 18: Schematic view of the RO content.

The workflow utilises R and was successfully implemented on the VRE using the workflow management software Taverna. Each part of the workflow was re-written so it could run online in a virtual environment with all input data stored in the cloud (Seafile), and using Taverna Server and Rserve applications hosted on the VRE, instead of desktop software. Further, the workflow was pa-

parameterised so that users working in the VRE could specify key input variables without amending code. This enabled the process to be automated giving the potential for other VRC members to run the workflow. The completed automated workflow is stored as a workflow Research Object (Fig16). This workflow is accessible to all VRC members that have access to the correct Seafire libraries. This workflow can easily be searched using the RO Search option and run multiple times. Workflow results, usually png and shape files, can be inspected within the VRE using the Virtual Globe or the integrated reader. Furthermore, they can be downloaded, stored in the current Research Object or in a new one.

- **DAILY HAZARD ASSESSMENT (DHA):** The DHA is a summary of forecasted hazards released on a daily basis to the responder community, local government and national agencies. It is based on information provided by various partner organisations including the FGS, the NSWWS and the DLHA. Each piece of evidence is linked by date, however if any of the evidence is updated due to a change in the hazard forecast, then an updated piece of evidence is submitted for inclusion in the DHA. The VRC decided to test the storage of each DHA and its contributing evidence in a bibliographic Research Object.

10. Conclusions

During the three-year project, the EVER-EST consortium developed a VRE for Earth Sciences, where the requirements of four communities were addressed. The VRE has been recognized as a successful solution to boost open science and innovation by enabling research life cycle management, long term data preservation, EO data exploitation and capacity building. A sustainability plan has been presented to maintain the findings after the end of the project: in addition, further efforts will be focused on make the platform fully operational (e.g. services improvement, architecture optimisation and user support), to improve the services model and to engage new communities.

The EVER-EST project has demonstrated the relevance of Research results (Research Object) standardisation and interoperability to boost innovation and open science (FAIR principle). The 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.

We documented through the VRCs case studies the extensive results produced during the journey of the Research Object adoption and adaptation to the Earth Science domain in EVER-EST, which include:

- An extended research object model adapted to the needs of earth scientists;
- The provisioning of digital object identifiers (DOI) to enable persistent identification and to give due credit to authors;
- The generation of content-based, semantically rich, research object metadata through natural language processing, enhancing visibility and reuse through recommendation systems and third-party search engines;
- The inclusion of various types of checklists that provide a compact representation of research object quality as a key enabler of scientific reuse.

In doing so, the main lessons learnt are: i) rich and expressive metadata is a key factor for sharing and reuse, ii) scientific results need to be visible and easily discovered, iii) scientists need to receive due credit for their work, and iv) research object management capabilities need to be integrated in existing analytic tools already in use by earth scientists in order to foster adoption.

Competing Interests

The authors have no competing interests to declare.

References

- Bechhofer, S, Buchan, I, De Roure, D, Missier, P, Ainsworth, J, Bhagat, J, Couch, P, Cruickshank, D, Delderfield, M, Dunlop, I, Gamble, M, Michaelides, D, Owen, S, Newman, D, Sufi, S and Goble, C. 2013. Why linked data is not enough for scientists. *Future Generation Computer Systems*, 29(2): 599–611. Special section: Recent advances in e-Science. DOI: <https://doi.org/10.1016/j.future.2011.08.004>
- Belhajjame, K, Corcho, O, Garijo, D, Zhao, J, Missier, P, Newman, DR, Palma, R, Bechhofer, S, Garcia-Cuesta, E, Gomez-Perez, JM, Klyne, G, Page, K, Roos, M, Ruiz, JE, Soiland-Reyes, S,

- Verdes-Montenegro, L, De Roure, D and Goble, C.** 2012. Workflow-centric research objects: A first class citizen in the scholarly discourse. In: *2nd Workshop on Semantic Publishing (SePublica), number 903 in CEUR Workshop Proceedings*, 1–12. Aachen.
- Benedetti-Cecchi, L, Canepa, A, Fuentes, V, Tamburello, L, Purcell, JE, Piraino, S, Roberts, J, Boero, F and Halpin, P.** “Deterministic Factors Overwhelm Stochastic Environmental Fluctuations as Drivers of Jellyfish Outbreaks”. DOI: <https://doi.org/10.1371/journal.pone.0141060>
- ESA, NERC, INGV, ISMAR, SatCen.** “Use Cases Description and User Needs”. *EVER-EST DEL WP3-D3.1*.
- ESA, NERC, INGV, ISMAR, SatCen.** “VRE Architecture and Interfaces Definition”. *EVER-EST DEL WP5-D5.1*.
- ESA, NERC, INGV, ISMAR, SatCen.** “Workflows and Research Objects in Earth Science – Concepts and Definitions”. *EVER-EST DEL WP4-D4.1*.
- Gomez-Perez, JM, Palma, R and Garcia-Silva, A.** Oct 2017. Towards a human-machine scientific partnership based on semantically rich research objects. In: *2017 IEEE 13th International Conference on e-Science (e-Science)*, 266–275. DOI: <https://doi.org/10.1109/eScience.2017.40>
- Palma, R, Hołubowicz, P, Corcho, O, Gomez-Perez, JM and Mazurek, C.** 2014. Rohub—a digital library of research objects supporting scientists towards reproducible science. In: *Semantic Web Evaluation Challenge*, 77–82. Springer. DOI: https://doi.org/10.1007/978-3-319-12024-9_9
<http://www.researchobject.org/>.
<http://www.scidip-es.eu/>.
<http://ceur-ws.org/Vol-679/paper4.pdf>.
<http://www.geowow.eu/>.
<http://www.envriplus.eu/>.


How to cite this article: Albani, M, Leone, R, Foglini, F, De Leo, F, Marelli, F and Maggio, I. 2020. EVER-EST: The Platform Allowing Scientists to Cross-Fertilize and Cross-Validate Data. *Data Science Journal*, 19: 21, pp.1–16. DOI: <https://doi.org/10.5334/dsj-2020-021>

Submitted: 26 January 2019

Accepted: 15 October 2019

Published: 08 May 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 