This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No [308417]".

New Directions in Seismic Hazard Assessment through Focused Earth Observation in the Marmara Supersite

Grant Agreement Number: 308417
co-funded by the European Commission within the Seventh Framework Programme
THEME [ENV.2012.6.4-2]
[Long-term monitoring experiment in geologically active regions of Europe prone to natural hazards: the Supersite concept]

D10.4

Report on integration of Earth Observation data, products and toolboxes

Project start date	1 November 2012
Project duration	36 Month
Project coordinator / organization	Nurcan Meral Özel / KOERI
Work Package number	10
Deliverable name/ number	Report on integration of Earth Observation data, products and toolboxes / D10.4
Due date of deliverable	30 April 2015
Actual submission date	30 April 2015
Author(s) / organization	Pierre-Philippe Matthieu / ESA
Reviewer(s) / organization	John Douglas / BRGM

Dissemination level		
PU	Public	
PP	Restricted to other programme participants (including the Commission)	
RE	Restricted to a group specified by the consortium (including the Commission)	
CO	Confidential, only for members of the consortium (including the Commission)	

TABLE OF CONTENTS

TABLE OF FIGURES

ABBREVIATIONS

ASAR	Advanced Synthetic Aperture Radar
CEOS	Committee on Earth Observation Satellites
CEOS WGISS	CEOS Working Group on Information Systems and Services
DLR	Deutsches Zentrum für Luft- und Raumfahrt
DOI	Digital Object Identifier
EC	European Commission
ESA	European Space Agency
EO	Earth observations
ERS	Earth Resources Satellite
GEO	Group on Earth Observations
GEOSS	Global Earth Observation System of Systems
GEOWOW	GEOSS interoperability for Weather, Ocean and Water
GEP	Geohazards Exploitation Platform
G-POD	Grid-Processing on Demand
GSNL	Geohazards Supersites and Natural Laboratories
IaaS	Infrastructure-as-a-Service
IETF	Internet Engineering Task Force
InSAR	Interferometry Synthetic Aperture Radar
L1	Level 1 EO products (engineering units, calibrated and geolocated data)
L2	Level 2 EO products (geophysical units, re-tracked data)
OGC	Open Geospatial Consortium
PaaS	Platform-as-a-Service
SaaS	Software-as-a-Service
SAR	Synthetic Aperture Radar
URL	Uniform Resource Locator
VM	Virtual Machine
VNC	Virtual Network Computing
WMS	Web Map Service
WPS	Web Processing Service

1. Introduction

1.1. Purpose and scope

This report presents the ESA contribution to the MARsite project, as part of the Task 3 “Fostering access and integration of Earth observation data products, tailored to the needs of the Geohazards community” of the Work Package 10 “Integration of data management practices and coordination with on-going research infrastructures”.

Within this task, ESA provides to the MARsite partners a range of innovative processing services, tailored for the generation of Earth observation (EO)-based products. The processing services are usable by the MARsite partners, and more generally the geohazards research community, on a geospatial data platform, the Geohazards Exploitation Platform (GEP). Processing results are made available via interoperability standards, so they can be readily integrated in the MARsite portal.

With this approach, ESA contributes to the MARsite project with data provision and integration of satellite data and new products elaborated on GEP, tailored to the needs of the geohazards researchers community working on the Marmara sea region.

1.2. Documents of reference

DR1 MARsite D10.2 Report - Second report on the integration and links to other initiatives, 3rd December 2014

DR2 MARsite D10.3 Report - Report on international standards on architecture principles, metadata, data models and services, and improving links with other projects, 28th April 2014

2. An approach for integration of EO data products and toolboxes

ESA developments through MARsite address the main goal of facilitating ESA data access and exploitation, and of conducting preliminary experiments on cloud processing services. This approach aims at making discoverable and accessible a large number of ESA products that are of interest for the geohazards community in MARsite. These data include both L1 and L2 data and value-added products that are generated in the project.

As far as the value-added products are concerned, ESA focuses effort on hosted processing capabilities offered by the ESA Geohazards Exploitation Platform (GEP), providing users with a range of data processing and dissemination services. In general, the processing services are made available to users as-a-Service, where the users will be enabled to define the set of input data, the processing parameters, and trigger the execution of the algorithms. These services can be invoked from the MARsite portal through a standard OGC Web Processing Service interface. The obtained results are then published on a dedicated archive and will become discoverable and available through the MARsite portal.

In order to support the development of new algorithms, the ESA GEP also provides a Platform-as-a-Service capability (PaaS). In particular, Cloud Sandboxes enable developers and integrators to easily implement new algorithms. This solution makes use of the Virtual Machines technology and includes a middleware providing transparent interfaces to Cloud services, used for scaling-up the processing when increasing the dimensions of the input dataset. In case researchers in MARsite are interested in experimenting with their own algorithms on ESA data, GEP offers them direct access to dedicated Cloud Sandboxes.

The use of GEP removes the need of transferring huge amounts of input product data from the ESA archives to the users' machines, resulting in significant savings for the users.

2.1. The Open Science model

With GEP, the agency is providing partners with tools and infrastructure aimed at supporting geohazards researchers and practitioners with easy and open access to the ESA sensors data, community knowledge and expertise, and collaborative research.

The e-Science principle of open, reproducible, verifiable research experiments is a critical goal of Open science, tackling key scientific challenges such as the transparency of experimental approaches, traceability of the collection of observations and the public availability and reusability of scientific data.

Especially with regard to the exploitation of Earth observation (EO) data, there is a particular need to support researchers in setting up computer-intensive experiments, and for practitioners to better select and apply research outcomes in their exploitation tasks. In this perspective, GEP supports researchers in transitioning from siloed computational science activities addressing the simulation of complex phenomena, to e-Science scenarios expanding current capabilities to data exploration approaches: more integrated data access and processing, more support for documentation and sharing of processing experiments, and enhanced traceability across data and scientific literature.

In the geohazards domain, science users require satellite EO to support mitigation activities designed to reduce risk. These activities are carried out before the earthquake (or other geological peril) occurs, and they are presently the only effective way to reduce the impact of earthquakes on society. Short-term earthquake prediction today offers little promise of concrete results. The assessment of seismic hazard requires gathering geo-information for several aspects: the parameterization of the seismic sources, knowledge of historical and instrumental rates of seismicity, the measurement of present deformation rates, the partitioning of strain among different faults, paleo-seismological data from faults, and the improvement of tectonic models in seismogenic areas. Operational users in charge of seismic risk management have needs for geo-information to support mitigation. Satellite EO can contribute by providing geo-information concerning crustal block boundaries to better map active faults, maps of strain to assess how rapidly faults are deforming, and geo-information concerning soil vulnerability to help estimate how the soil is behaving in reaction to seismic phenomena (read more from https://geohazards-tep.eo.esa.int/#!pages/initiative).

The ESA GEP exploits different types of cloud appliances. With GEP, it is possible to run computer-intensive workflows, enabled by specific cloud appliances that are built on "Hadoop Cloud Sandbox" components. Such Cloud Sandbox components were initiated by ESA funding as part of its G-POD Cloud evolutions (Cloud Interoperability Operational Pilot: An EO Sandbox Service) and then further developed through other EC FP7 projects, e.g. GEOWOW for the interface with the GEO Data Access and Broker (DAB).

As activities in the EC Horizon 2020 calls (e.g. EINFRA-1-2014 and EINFRA-2-2014) have a focus on supporting e-infrastructures for Open Access (Open Access mandate covering all Horizon 2020 publications outputs) and to provide services for managing the life cycle of data that they collect or produce within their projects (e.g. deposition, storing, access and preservation), the data policies of ESA for its operated satellites is more than ever at the front of Open Science applications within the European Union and globally. The approach implemented for MARsite is promoting open science and open access to research results, and ESA is supporting the MARsite partners with such an approach for the integration of EO data products and toolboxes.

For years now, ESA has been supporting the development and adoption of open source toolboxes (cf. https://earth.esa.int/web/guest/pi-community/toolboxes) for reading, processing, analysing and visualising ESA (ERS-1/2, Envisat, Sentinel-1) and other spaceborne sensor data, in SAR, Altimetry or Raster modes. The geohazards community have also been developing SAR data processors. For the first time, geohazards researchers and practitioners can join a shared platform where such toolboxes can be integrated, connected to large volumes of data, and exploited as-a-Service by end-users.

This approach for the integration of earth observation data, products, and toolboxes is central to the newly-launched ESA Thematic Exploitation Platforms (TEPs). The TEPs architecture is developed from a cloud-based e-Infrastructure approach focusing on the virtualization and federation of satellite EO applications. The TEPs are e-infrastructures providing scientists with a "testbed" to do their research without wasting time on ICT. Users are supported in exploiting resources made available as-a-service, in developing and testing their own computer-intensive processing chains, and overall in sharing data, knowledge and processing services across communities.

In this context, the GEP is providing a set of capabilities, made available to the MARsite partners, that support Open Science activities through services for EO Data discovery, EO data access over distributed repositories, EO data consumption accounting, scalable on-demand EO processing, and sharing of value-added products.

In this approach, the GEP is a contribution to the global interoperability of open access data e-infrastructures, and provides a linkage with other global initiatives under GEO and Committee on Earth Observation Satellites (CEOS) umbrellas. This will allow the MARsite partners to deliver physical access to research resources, including EO data access and value-added products dissemination.

Considering drivers for change, open is better for the strengthening of links between science and society.

2.2. Links with GEOWOW

The EC FP7 GEOWOW project (http://www.geowow.eu) developed the following vision in support of discovery, access and exploitation of data at different levels.

- **Data scientists**: they are mainly interested in observations and raw data (i.e. data still to be processed to be turned into meaningful information) and they usually have their own community-specific clients or applications to analyse and further elaborate them into information that is meaningful for different kinds of users;
- **Decision makers**: they can also be interested in observations, but often need to access data products and information extracted from lower level data. This means that they need to access information produced by, e.g., the data scientists; and
- **Citizens**: decision makers in turn may act as information "providers" for citizens (e.g. they may inform the citizens about the outcome of their analyses regarding citizen-related issues) and other generic users who may be interested in different kinds of information according to their needs.

Within GEOWOW, the project partners ESA and Terradue undertook joint activities for innovative Cloud platform development and integration activities, to further develop the infrastructure services necessary to provide access not just to data, but also to selected analytical models and workflows. This was about moving beyond the quest to find data from heterogeneous distributed sources, to the more important question of how to use the data effectively to address complex scientific and policy problems. It translated into a consolidated review of the compliancy of ESA data with the GEOSS Data-CORE and GEOSS Data Sharing Principles on one hand, and on improved PaaS capabilities (Platform-as-a-Service) of the Developer Cloud Sandbox technology (engineered by Terradue), that are now embedded in the 'reproducibility' and 'sharing' functions of the GEP as described in the next section.

An outcome of the GEOWOW project was to improve the management of workflow development tasks, especially for experiment reproducibility, by integrating access to and interaction with GitHub repositories, from within a PaaS environment. Support for cloning and group sharing of Cloud Sandboxes applications was provided to users with administration rights to allow scientific work to be reproducible (as-is, or adapted to a related purpose), for example, by peer-reviewers or by other research laboratories.

This evolution of the vision and improvement in technology were presented as part of the European Commission's contributions to GEO / GEOSS. Integration of data access and Geoprocessing is a long-awaited next step in GEOSS developments, and raising the interest from the communities of practice that are gathered by GEOSS (e.g. the Geohazards Supersites and Natural Laboratories) was a key achievement for GEOWOW.

This approach was presented at the GEO-X plenary meeting in Geneva, January 2014 and is available online (Open Science and GEOSS: the Cloud Sandbox enablers - http://www.slideshare.net/terradue/open-science-and-geoss-the-cloud-sandbox-enablers).

While GEOWOW marked a key evolution for GEP and GEOSS, there are other important and developing linkages between GEP and GEOSS, presented in the following section.

2.3. Links with GEOSS

We discussed how GEOWOW promoted new concepts and related tools as a European contribution to the GEOSS Common Infrastructure (GCI) and the GEOSS Data-CORE.

The PaaS capability of GEP inherited from these efforts, and provides today an environment for scientists to prepare data and scalable processing chains (including development and testing of new algorithms), designed to automate the deployment of the resulting environment to a Cloud computing facility, that allows running compute-intensive tasks.

This approach towards the GEOSS community is being currently evolved and coordinated by ESA, through the following activities around GEP.

- First, by coordinating the GEP developments with the CEOS. GEP follows the SuperSites Exploitation Platform (SSEP), originally developed in the context of the Geohazards Supersites and Natural Laboratories initiative (GSNL). The geohazards platform has been expanded to address broader objectives of the geohazards community. In particular it is a contribution to the CEOS WG Disasters to support its Seismic Hazards Pilot and the terrain deformation applications of its Volcano Pilot.
- Second, by supporting users from the EC Supersites FP7 projects (MED-SUV, MARsite and FutureVOLC). These three projects are the EC contribution to the GEOSS GSNL. ESA in developing the GEP has accommodated a Validation phase starting in Q1 2015 for a set of Early Adopters, with a formalized relationship between a user and the GEP following its acceptance, to activate exploitation scenarios with specific guidance and support.

Support to the GEOSS researchers is also provided through a level of automation enabling users to reference work done on the Platform with the use of Digital Object Identifiers (DOI), where each application gets a DOI to track the service's impact through citations. This functionality is completed in GEP with job processing reproducibility (users can save and share the parameters of a processing job so that partners or reviewers can re-execute a job) and sharing functions (users can advertise their activity and results on major social media platforms).

As planned in May 2012, when ESA and the GEO Secretariat convened the International Forum on Satellite EO for Geohazards (the "Santorini Conference"), the seismic community has set out a vision of EO's contributions to an operational global seismic risk program. In 5 to 10 years' time, EO could provide fundamental new observations of the seismic belts - around 15% of the land surface and improved understanding of seismic events through the work of the GSNL. At that same Santorini Conference, the volcanic community identified priorities for satellite support to geohazards. In the long-term, the community aims to monitor all 1,500 Holocene era volcanoes on a global basis, a dramatic increase from the roughly 10% that are monitored now using both satellites and terrestrial sensors.

One of the core user communities for the GEP is the group of users and practitioners working on the CEOS Seismic Hazards Pilot, a three-year demonstration project of CEOS to showcase how satellite EO can be applied to seismic hazard research.

2.4. Links with the EPOS (ENVRI) infrastructure

EPOS (the European Plate Observing System) is a long-term integration plan of Research Infrastructures for solid Earth science in Europe. EPOS is promoting and making possible innovative approaches for a better understanding of processes controlling earthquakes, volcanic eruptions, tsunamis, surface dynamics and tectonics. EPOS is developing new concepts and tools to better address the grand challenges in solid Earth science. In particular, EPOS is promoting open access to geophysical and geological data as well as modelling/processing tools, enabling a step change in multidisciplinary scientific research for Earth sciences.

From the EPOS Preparatory Phase, Working Group 8 (Satellite Information Data) addressed the goals of involving European data providers, i.e. the space agencies, defining a strategy for the acquisition of satellite data available to the EPOS Community, and exploring solutions for data repositories. One contribution of MARsite to this EPOS endeavour is to link EPOS Core Services to ESA Data Repositories.

In the light of the EPOS integration plan, the EC Supersites projects are part of the EPOS Community Layer - Thematic Services, while the GEP is part of the EPOS Integration Layer - Integrated Services (access to high performance computing, processing tools, access to data products). Having MARsite and GEP working together in the frame of the MARsite project provides a direct contribution to the EPOS vision and integration plan, especially when considering the EPOS principle of organising the Community Layer as a set of distributed data archives and services.

Other key EPOS concepts well supported by the GEP contribution in support of MARsite activities are the concept of "Product incubation chamber" (where scientists develop, test and improve new products before they are established as standard products). As a component of EC infrastructure projects, MARsite will benefit from the GEP added value to facilitate data discovery, enable repeatable workflows, facilitate effective large-scale data management, facilitate multidisciplinary analysis, and enable adequate data source identification. This will empower the MARsite project partners for their community involvement in setting the requirements for the EPOS integrating core services, and ultimately in building it.

3. ESA return of experience in the TEP Geohazards

In 2012, the European Space Agency started the SSEP flagship application within the Helix Nebula, Science Cloud initiative.

SSEP promoted a synthetic aperture radar (SAR) data exploitation platform focused on a few 'natural laboratories' around the world as defined within the GEO GSNL (http://www.earthobservations.org/gsnl.php) effort.

The SSEP brought together existing SAR interferometry EO toolboxes and EO data in a workspace that allowed researchers to run algorithms over data in a Cloud Computing environment. A paradigm shift was implemented that no longer relied on downloading data locally and running desktop software. Users were empowered with a set of tools hosted on the platform, using cloud and grid resources to enable compute-intensive tasks over massive data repositories.

The SSEP effort presented the first view over the assembled knowledge and experience from a large number of ESA assets. First, it was tied to the development and maintenance of the Agency's Grid Processing on Demand (G-POD, https://gpod.eo.esa.int/) system, and the related activities for integration of scientific applications. Second, it was empowered by new assets brought by the developments of Developer Cloud Sandbox services, matured through the Cloud Computing Interoperability Pilots (CIOP) project (http://wiki.services.eoportal.org/tiki-index.php?page=CIOP), and evolved through the EC FP7 GEOWOW project (http://www.geowow.eu/) that concluded in September 2014.

Today, building on SSEP insights, the TEP Geohazards 'QuickWin' platform is a progression and enhancement towards a phase of GEP pre-operations. It brings the SSEP developments, mainly focused on proof-of-concept efforts, to a next level of federated architecture, with more power, additional tools (including pre-negotiated software licences with pay-per-use conditions, software renting,…), strengthened Web and Cloud Application Programming Interfaces (APIs) and a dedicated governance process to on-board users with a phased and coordinated approach (including user support services). At the current stage of these developments, outputs of the ESA's 'QuickWin' project are empowering a first community of 'early adopters' (from leading organizations like BGS, INGV, NOA,…), providing them with a Cloud platform having large data hosting capabilities, applications and services (e.g. InSAR processors and EO toolboxes) running on scalable Cloud resources without vendor lock-in (interoperability with several Cloud providers), and being exploitable through a new geobrowser service to access data, trigger processing tasks, and share the availability of EO-based products with other users on the platform as well as towards the larger community.

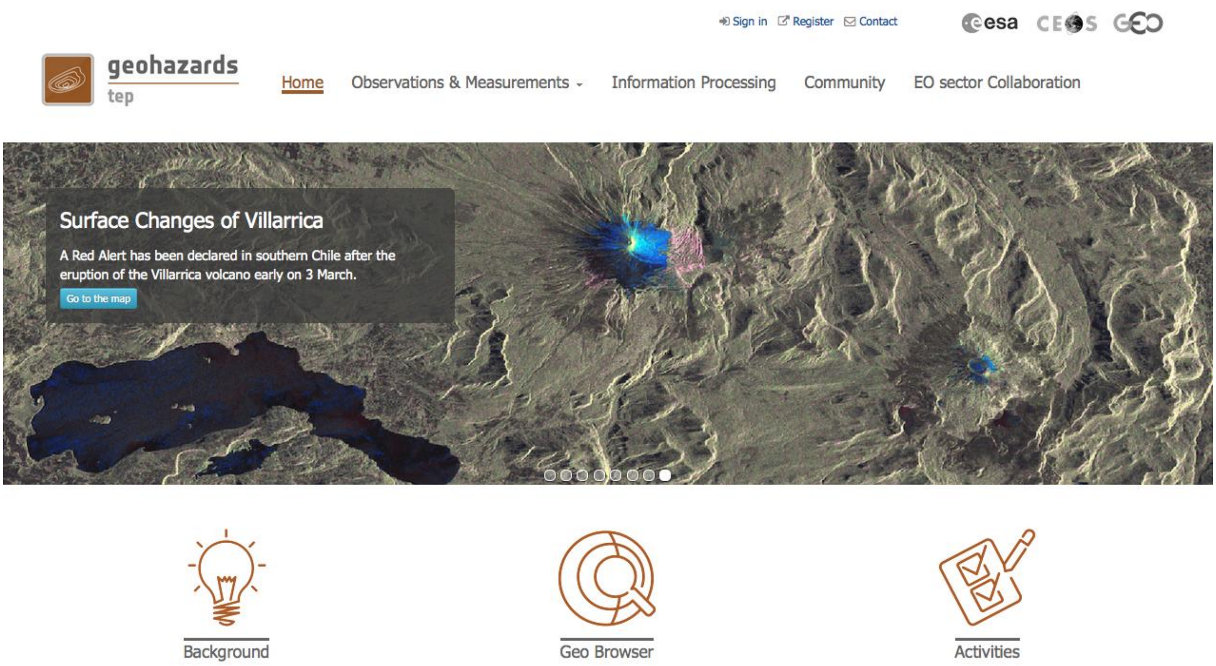
3.1. Leveraging the GEP to integrate Cloud Appliances as Processing Services

An "Exploitation Platform" refers to a virtual ICT environment, often cloud-based, providing users with very fast access to: (i) a large volume of data (EO/non-space data), (ii) computing resources (e.g. hybrid cloud/grid), and (iii) processing software (toolboxes, RTMs, retrieval schemes and visualization routines).

The idea underpinning exploitation platforms is to enable users to perform effectively data-intensive research by providing them with a virtual machine running dedicated processing software close to the data, thereby avoiding moving large volumes of data through the network and spending non-research time on developing ICT tools.

The GEP portal is already accessible online for users (including public access level) at: http://geohazards-tep.eo.esa.int (cf. Figure 1).

Figure 1 - The Geohazards Exploitation Platform portal



It is providing the following set of capabilities, which are made available to the MARsite partners.

- **EO data discovery service** through a single point of access to visualize data collections in terms of acquisition footprints and sensor parameters, with resources available from ESA missions (especially the SAR missions from ENVISAT, ERS and SENTINEL-1) and third party missions (currently DLR TerraSAR-X and, upcoming for, ASI Cosmo-Skymed and CNES).

- **EO data access service over distributed repositories** supporting the dissemination of imagery either stored in the GEP cloud platform environment or accessed through the GEP portal in other remote data repositories from the pool of contributing agencies. Data access is based on the authentication of registered users and the granting of data dissemination according to the user profile. For instance, EO data constrained by license terms and distribution restrictions can be accessed from the platform's geographic interface via active links to the repository of the data provider.

- **Accounting service for EO data consumption** allowing the monitoring of the volumes of data use per EO source and according to the activity associated to the user profile. The accounting service can be used to support reporting concerning the exploitation of EO data, either from the Platform (e.g. hosted processing) or in the framework of application projects (e.g. CEOS pilots).

- **EO processing services** for on-demand processing, exploiting software to transform EO data into measurements; the user may run an EO processor provided on the platform (ready to use software-as-a-service, SaaS), or integrate an application he/she has developed (platform-as-a-service capabilities, or PaaS).

The SaaS Processing can be invoked either interactively though a web browser, or through scripting using the OGC Web Processing Service (WPS) interface; The PaaS provides software development and integration tools, and enables users to perform their data exploitation activities with large flexibility and autonomy, by using one or several virtual hosts directly provisioned on the cloud platform and deployable on demand.

- **Access to Value-Added products** generated on the GEP, or products contributed by third parties. The platform allows cataloguing and dissemination of products relevant to the geohazards community. It can be used to provide access to elaborated products in support of thematic exploitation goals.

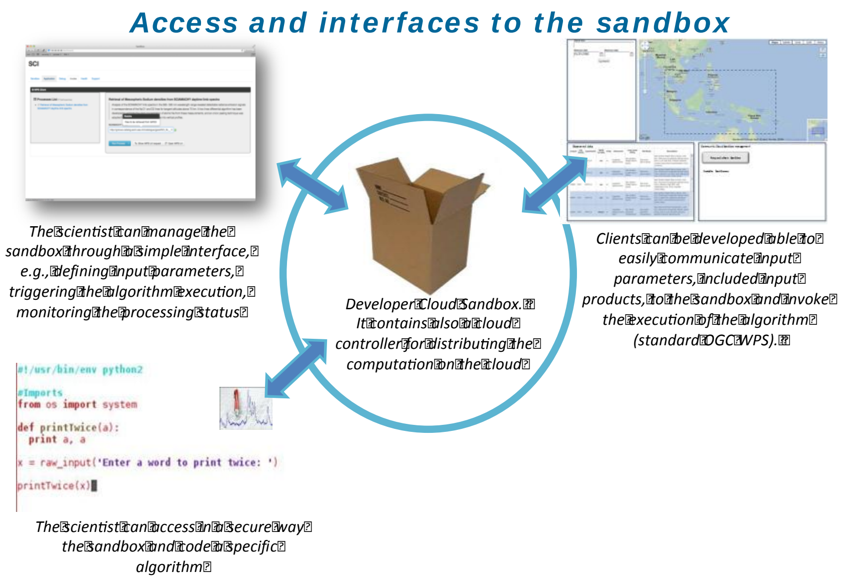
3.2. The Developer Cloud Sandbox computing model

The Developer Cloud Sandbox (Figure 2) is a computing model that allows scientists to develop and test EO services with:

Directed Acyclic workflows;

- Parallel computing;
- OGC WPS submission interface, that makes it easy to:
    - Invoke the service from the MARsite portal;
    - Develop any custom web-client able to interface the WPS;
- Tools to query and access EO data;
- Output data published and if needed registered to GEOSS.

Figure 2 - Access and interfaces to a Cloud Sandbox

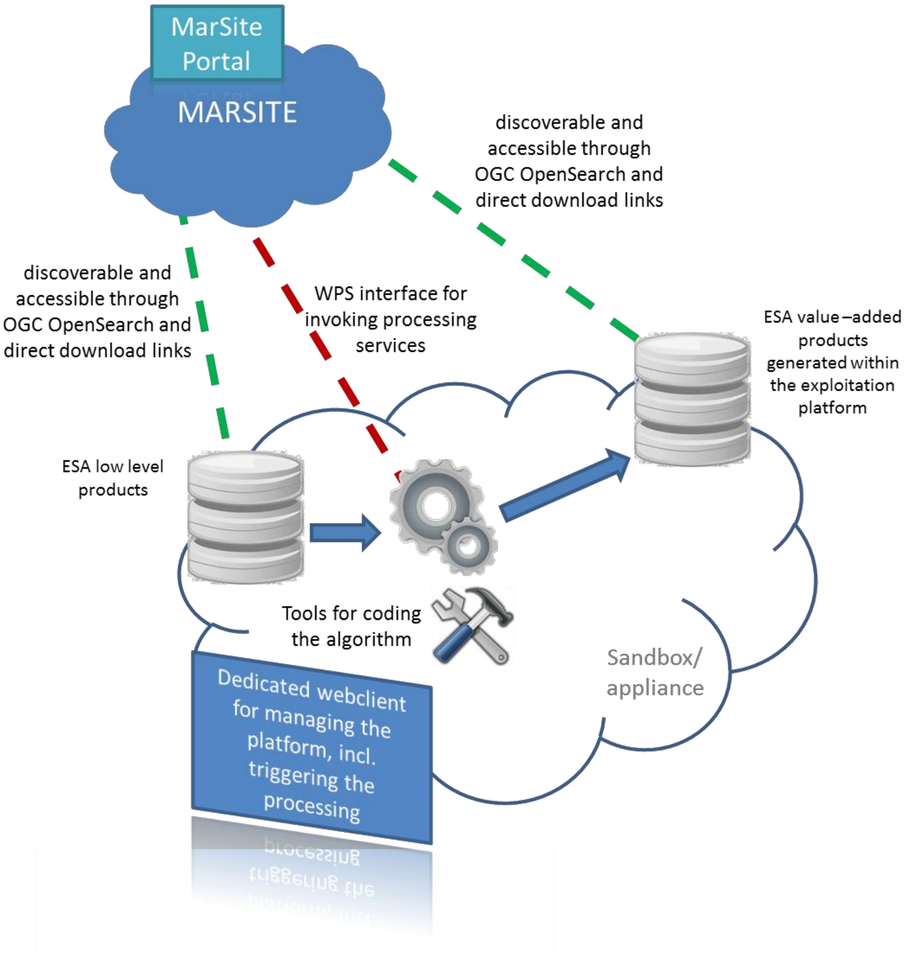


The Developer Cloud Sandbox is a computing model that allow scientists to develop and test EO services with Directed Acyclic Graph workflows, OGC WPS submission interface that makes it easy to develop any custom web-client able to interface the WPS, and tools to query and access EO data.

With this technology embedded, the GEP targets the evaluation of the different opportunities to move beyond EO data discovery bottlenecks and focuses on the computational science for data intensive analysis (Figure 3).

- Use of Infrastructure as a Service (IaaS) enables data and resource sharing, provides optimized costs and allows for a massively scalable ICT infrastructure.
- Adoption of pay-per-use models gives access to resources that users would not be able to afford on their own.
- Evaluation of innovative "EO application stores" and EO Software-as-a-Service (SaaS) concepts, along with options for sourcing of content (data) and applications (processors) from both open stores and commercial providers.

Figure 3 - How GEP Cloud Sandbox fits in the bigger MARsite picture



3.3. Interoperability: OGC OpenSearch and OGC WPS interfaces

GEP is making available best practices for search services, using the OGC OpenSearch interface standard with Geo, Time and EO extensions, as defined by CEOS, that allows standardized and harmonized access to EO catalogues from satellite EO data providers worldwide.

Starting in 2008, Terradue has pushed the OpenSearch approach towards its international standardization by proposing and editing the specifications from OGC. This interface was recognized and adopted in 2014 by the OGC as the preferred baseline for the catalogue services. ESA selected this interface for the implementation of the Agency's Next Generation User Services for EO (ngEO), NASA applied it to the newly deployed system of Earth Observing System (EOS) Clearing House (ECHO) and their interface to the Federation of Earth Science Information Partners (ESIP) that enables the discovery and access to the totality of NASA archives. The CEOS WGISS Integrated Catalog (CWIC) recently established a common interoperability best practice of OpenSearch in order to allow for standardized and harmonized access to metadata and data of CEOS agencies.

This track record of OGC OpenSearch implementation in the EO domain is strengthening interoperability for the whole geohazards community.

GEP is also putting into practice the OGC WPS interface standard, a web interface allowing for the dynamic invocation of user processes in a distributed computing environment. The OGC WPS interface handles the description of processing offerings, the triggering of processors with dynamic input parameters, and retrieval of processing outputs. Combined with encoding standards for processing results like OpenSearch description documents and/or IETF Metalink specification (for products download / retrieval), OGC WPS provides a simple and efficient solution to bind geoprocessing services with web applications.

Other notable interoperability protocols or encoding formats available from GEP are OGC WMS (with upcoming functionality of automatic layer publication after a processing job) and OGC OWS Context (for the exchange of data packages and other contextual information).

4. MARsite new developments

4.1. Inventory of relevant data from GSNL

Today the GEP has a primary focus on mapping hazard-prone land surfaces and monitoring terrain deformation. It allows users to access and the exploit large collections of ENVISAT ASAR and ERS SAR data hosted in the ESA clusters and in ESA's Virtual Archive, plus the SENTINEL-1 SAR data hosted on the S-1 SciHub, and also SAR missions from third-party providers (currently, DLR's TerraSAR-X).

We present hereafter the details of this data availability from the GEP.

EO data collections (i.e. the sensor, the type of EO data product and the time span of the data collection) are regularly added to GEP.

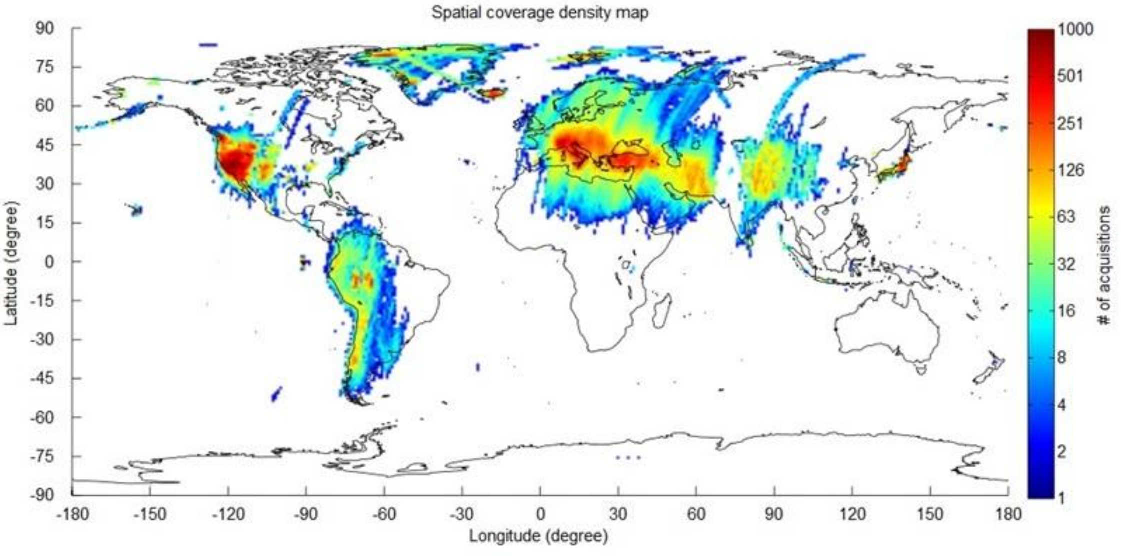
GEP users are invited to search for data directly on GEP, and eventually to also perform queries on the Virtual Archive 4 (http://eo-virtual-archive4.esa.int/) and EOLI SA (https://earth.esa.int/web/guest/eoli) to verify the availability of the data collections that are needed but would not yet be referenced by GEP.

A large collection of ENVISAT ASAR and ERS SAR are available in the platform through the Virtual Archive 4, spanning 1992-2011, and are freely usable by the geohazards community (e.g. for the EPOS community).

In 2014 an additional 40+ Terabytes of ERS and ASAR data was added in response to requirements of the CEOS Pilot on Seismic Hazards.

The following SAR data archives from ERS and ENVISAT Level 0 (in number of acquisitions, cf. Figure 4) are already available on GEP, integrated from the Virtual Archive 4 (VA4).

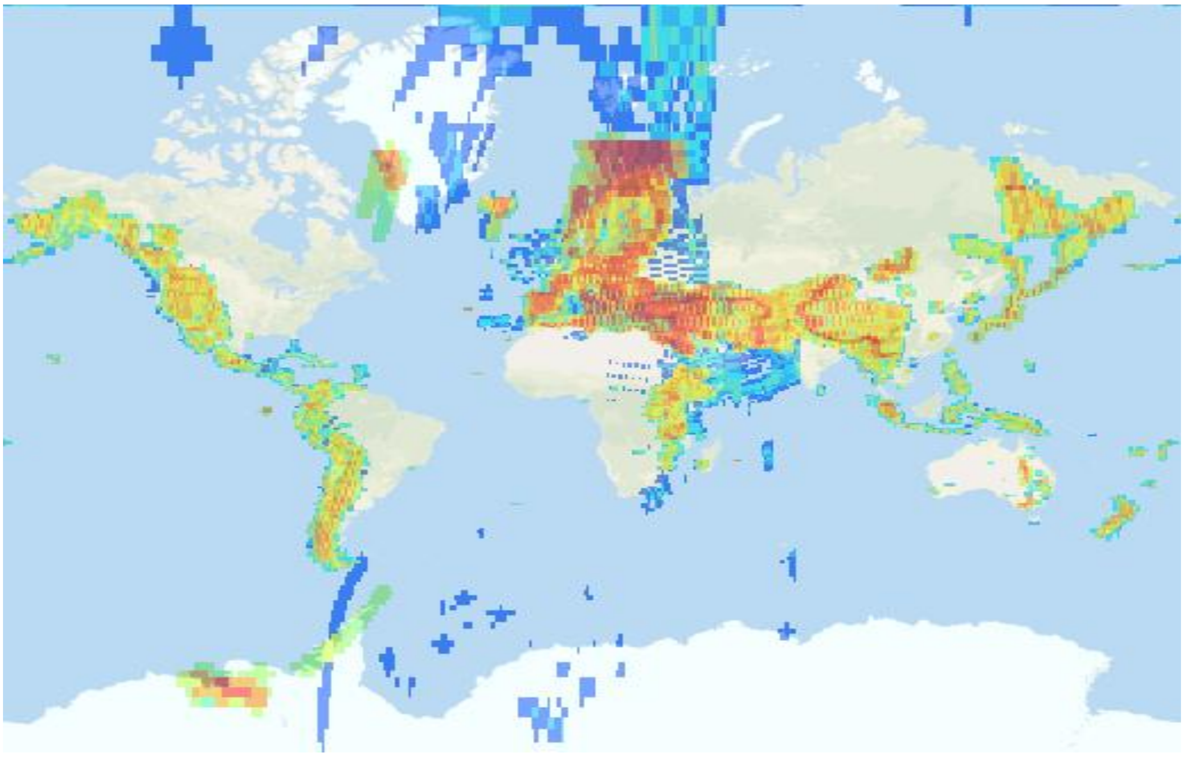
Figure 4 - ERS SAR & ENVISAT ASAR data archive in GEP (as of March 2015)



The GEP is also used to help users gradually access Sentinel-1 data as it flows from the Sentinel-1 SciHub (http://scihub.esa.int).

The following SAR data archives from SENTINEL-1 are already available via the GEP (in terms of number of acquisitions, see Figure 5).

Figure 5 - SAR data archive in GEP from S-1 SciHub



Regarding Sentinel-1 data, the assessment of data collections is done via the 'geobrowser' service of the GEP (http://geohazards-tep.eo.esa.int/geobrowser).

The data provisioning activity in GEP also intends to support access to other EO missions. Currently the DLR TerraSAR-X mission is accessible from the GEP geobrowser (the source catalogue being https://supersites.eoc.dlr.de/); ASI CosmoSkymed mission is upcoming.

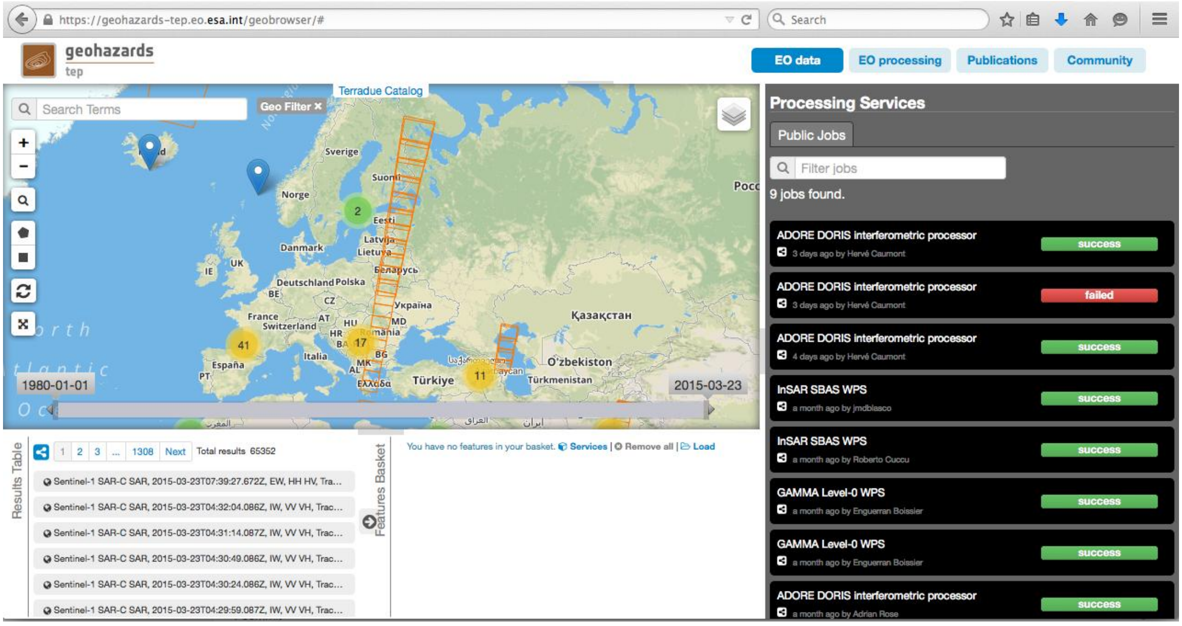
4.2. ESA low level products discovery and access

The first step for the MARsite partners willing to work with GEP and to create new products, is to search, access and select from GEP the ESA low-level products of interest, and consider using the publicly-shared processing services as their primary production tools.

The platform is currently in its Validation phase, which will explore up to October 2015 a growing set of functionalities. As of today, the Platform already allows public data search over a large archive of EO data from ESA sensors and from third-party missions (CEOS partners) like DLR's TerraSAR-X.

Practitioners can access the GEP infrastructue through the Geobrowser service, which is now made available with a set of baseline functionalities to all users (unregistered users / general public) for data search, data selection and data processing at: http://geohazards-tep.eo.esa.int/geobrowser (cf. Figure 6).

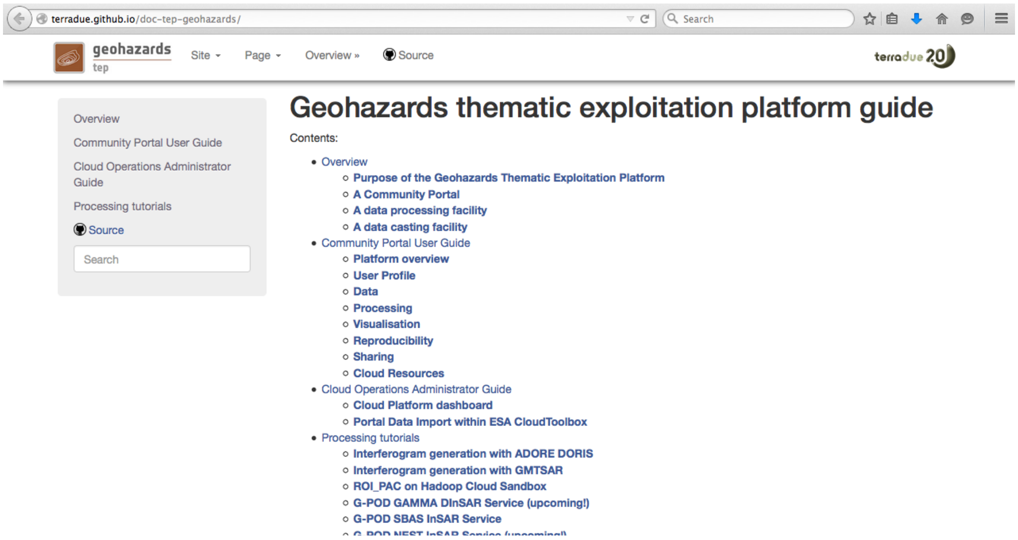
Figure 6 - User access to the Geohazards Exploitation Platform



The user documentation for GEP is also available online. It features an overview of the Platform concepts, a Community Portal User Guide, a Cloud Operations Administrator Guide and a growing set of data processing tutorials (SAR processing with ADORE DORIS, GMTSAR, ROI_PAC, and a set of G-POD services such as GAMMA-L0, SBAS).

This user documentation will continue to evolve in the coming months and it is available at:http://terradue.github.io/doc-tep-geohazards/ (cf. Figure 7).

Figure 7 - User documentation for the Geohazards Exploitation Platform



The next step for the MARsite partners would be to apply to become a partner in the ESA GEP validation phase, filling in the GEP registration form that ESA provides to requesters, and to describe in this form the type of activity foreseen: EO data exploitation, new EO Service development, or new EO Product development.

Registration on GEP opens privileged access to Cloud resources on the GEP, where users can create a new Virtual Machine with dedicated EO data management tools, like with the ESA CloudToolbox VM type (featuring pre-installed EO data processing software and toolboxes) and the Hadoop Cloud Sandbox VM type (featuring software integration tools and APIs to integrate processors in scalable processing chains that can be deployed from GEP on large clusters, provisioned on selected European Cloud Providers).

Some auxiliary services are also made available to the MARsite partners: DEM generation, catalogues with orbital data and catalogues with interferometric search. The overarching goal is to create cloud appliances to ease the generation of interferograms and time-series.

4.3. SAR processing with Hadoop Cloud Sandboxes as Cloud Appliances

As introduced before, registered users (currently, the early adopters that have applied for the GEP Validation phase activities) can access Cloud resources from GEP. One type of GEP cloud resources is the Developer Cloud Sandbox configured with the Hadoop framework enabling a massively parallel computation model.

The Developer Cloud Sandbox is a Virtual Machine (VM) that provides scientific developers with an Exploitation Platform-as-a-Service (PaaS). It consists of a development environment for processor integration and testing, and a framework for Cloud provisioning. The Developer Cloud Sandbox PaaS allows you to plug scientific applications written in a variety of languages (e.g. Java, C++, IDL, Python, R), then deploy, automate, manage and scale them in a very modular way. The algorithm integration is performed from within a dedicated Virtual Machine, running initially as a simulation environment (sandbox mode) that can readily scale to production (cluster mode). Accessed from a harmonized Shell environment, support tools also facilitate the data access and workflow management tasks.

Current outcomes from using the GEP Hadoop Cloud Sandbox (including for MARsite users) are the availability 'as-a-Service' of several SAR processors on the platform: ROI_PAC for co-seismic interforogram generation, ADORE DORIS for interferograms generation, PF-ASAR Level 0 to Level 1 Instrument Processing Facility, GMTSAR for interferogram stack generation. More processors are planned for being integrated in the coming month (DIAPASON, StaMPS, DLR's InSAR-QL and so forth).

Once integrated and validated, the resulting Cloud Appliance is a computing model that allows scientists to execute EO services with the processing power of parallel computing (leveraging the Hadoop Framework), an OGC WPS submission interface that makes it easy to invoke the service from an Exploitation platform or from a custom web-client application, and to manage output data to be published externally and if needed registered into the GEOSS Common Infrastructure.

As indicated previously, the MARsite partners willing to integrate their own processing chain and make them available on GEP as-a-Service can also apply as early adopters, as part of the ongoing GEP validation phase.

4.4. EO software toolboxes integration as Cloud Appliances

The ESA CloudToolbox is a Virtual Machine that offers a flexible amount of CPUs, RAM and dedicated storage, tailored to a user needs and type of machine required. This provisioning is flexible, so that users can request upgrades of the configuration if needed (for example, asking more processing power), in compliance with the Cloud infrastructure constraints of resource types that are made available by Cloud providers.

A pre-built Virtual Machine template offers ready-to-use machines for SAR Interferometric processing or generic EO data processing. Moreover, besides the free and/or licensed software tools (e.g. Sentinel-1 toolbox, NEST, GAMMA, Matlab, etc.) that can be installed on the machines, users may also request installation of additional tools. A data access tool to ingest a ‘data package’ (the result of a search and select process via the GEP geobrowser) is also available from that environment.

To create a CloudToolbox, registered users simply access the GEP cloud dashboard (a user interface for cloud resources management), click to create a new Virtual Machine, set a name (e.g ‘my InSAR toolbox’), select the “ESA CloudToolbox” template, click on create for the VM to be deployed, and then get the <ESA CloudToolbox IP> address. From there, users connected to the GEP Virtual Private Network can access the Virtual Machine through a VNC client.

As indicated previously, the MARsite partners willing to work and experiment with such cloud-based EO processing software and data access, can also apply as early adopters, as part of the ongoing GEP validation phase.

5. Contributed capabilities to the MARsite portal

5.1. Dedicated web client capabilities

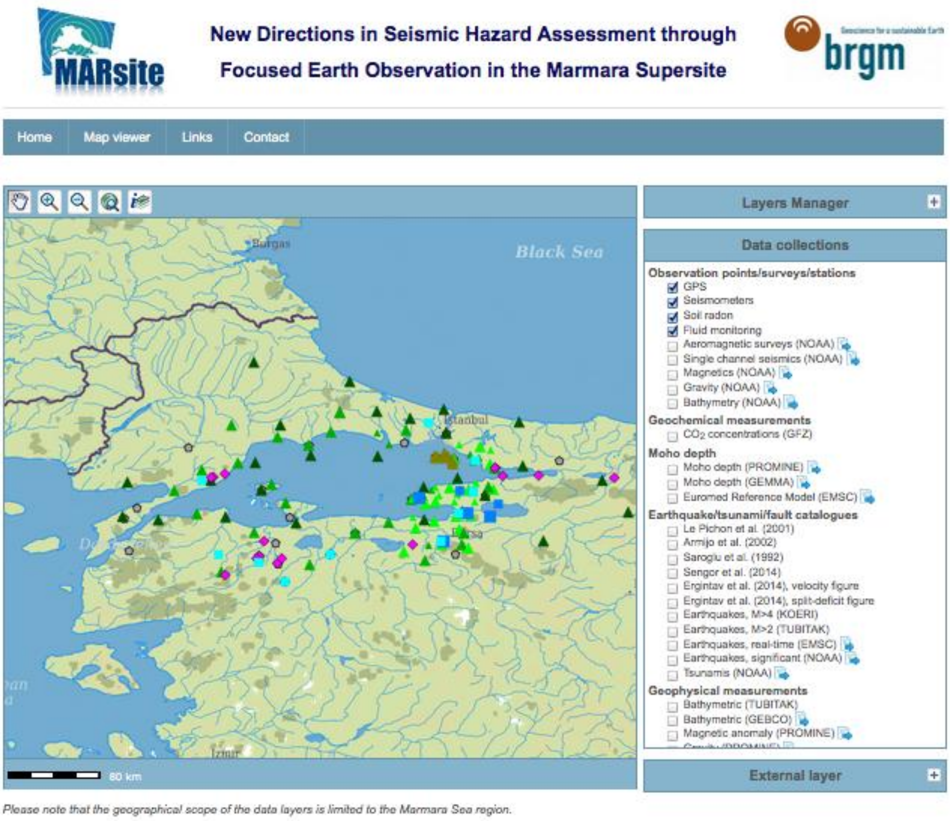
One of our aims with this Work Package 10 activity is to contribute EO-related web services to the pilot implementation of a MARsite interoperable portal (under the responsibility of the MARsite partner BRGM, with development instance here http://marsite.brgm-rec.fr/marsite/), so that BRGM can ensure their proper ingestion according to acknowledged interoperability standards. That MARsite portal will allow access to data relevant for the scientific goals of the MARsite project. It will show the benefits of adopting international standards, in line with GEOSS recommendations and best practices, and other on-going initiatives, such as EPOS and EMSO.

We present hereafter several OGC-compliant web services that are made available by the ESA GEP. Following a due process of interoperability arrangements between ESA and the MARsite project, they are contributed for being ingested by the MARsite portal under the responsibility of BRGM.

5.2. Visualise data layers through OGC WMS interface

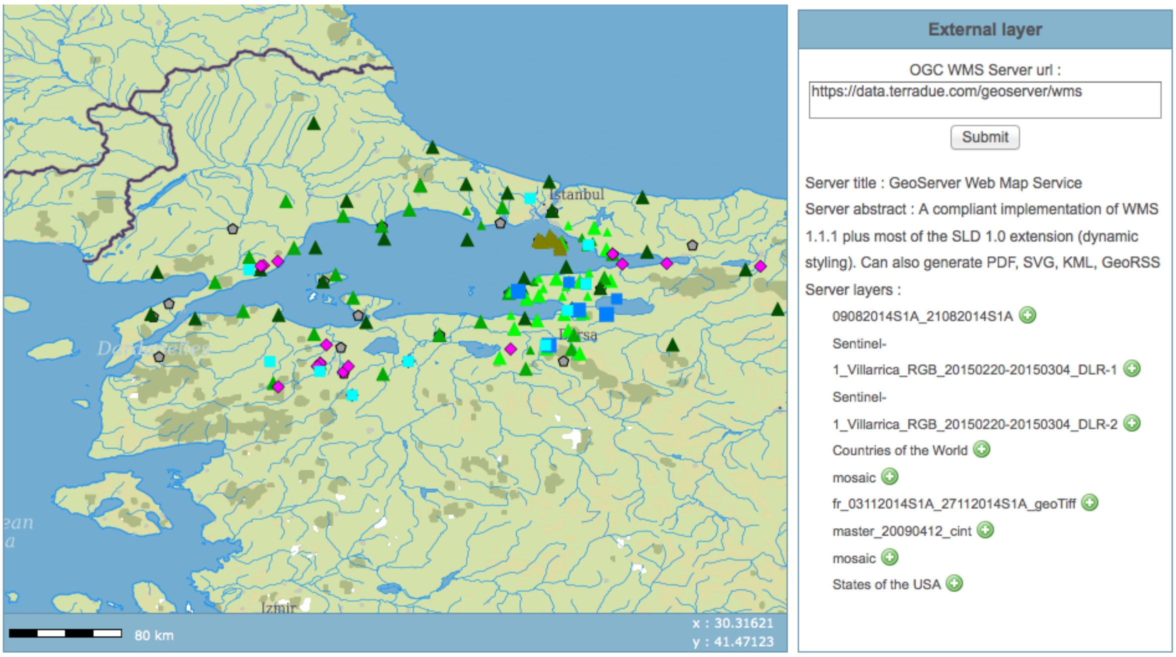
In its current state, the MARsite Map Viewer allows the visualisation of OGC-compliant map layers, leveraging the OGC Web Map Service (WMS) interface. Already, a selection of data from previous projects and also data compiled, collected or simulated within MARsite is available in this viewer (cf. Figure 8).

Figure 8 - MARsite Map Viewer – Data collections WMS layers



The interoperability between the MARsite Map Viewer and the GEP WMS server have been successfully tested, using the “External layer” functionality of the Map Viewer to import the layers served by GEP (see Figure 9).

Figure 9 - MARsite Map Viewer – Adding external WMS layers from GEP



Please note that the geographical scope of the data layers is limited to the Marmara Sea region.

At this stage of the GEP activities, only a few WMS test layers are available, like the colour composite image created by DLR with ESA Sentinel-1 image pair after the eruption of the Villarrica Volcano on 4th March 2015.

As the MARsite Map Viewer features a map view limited to the Marmara sea region, we generated a test product on the GEP platform to validate the layer inclusion into the viewer layers list, as described in the following section.

Partners creating EO-based products from GEP can easily share their products as WMS layers that can be accessed by the users of the MARsite Portal.

5.3. Invoke processing services through OGC WPS interface

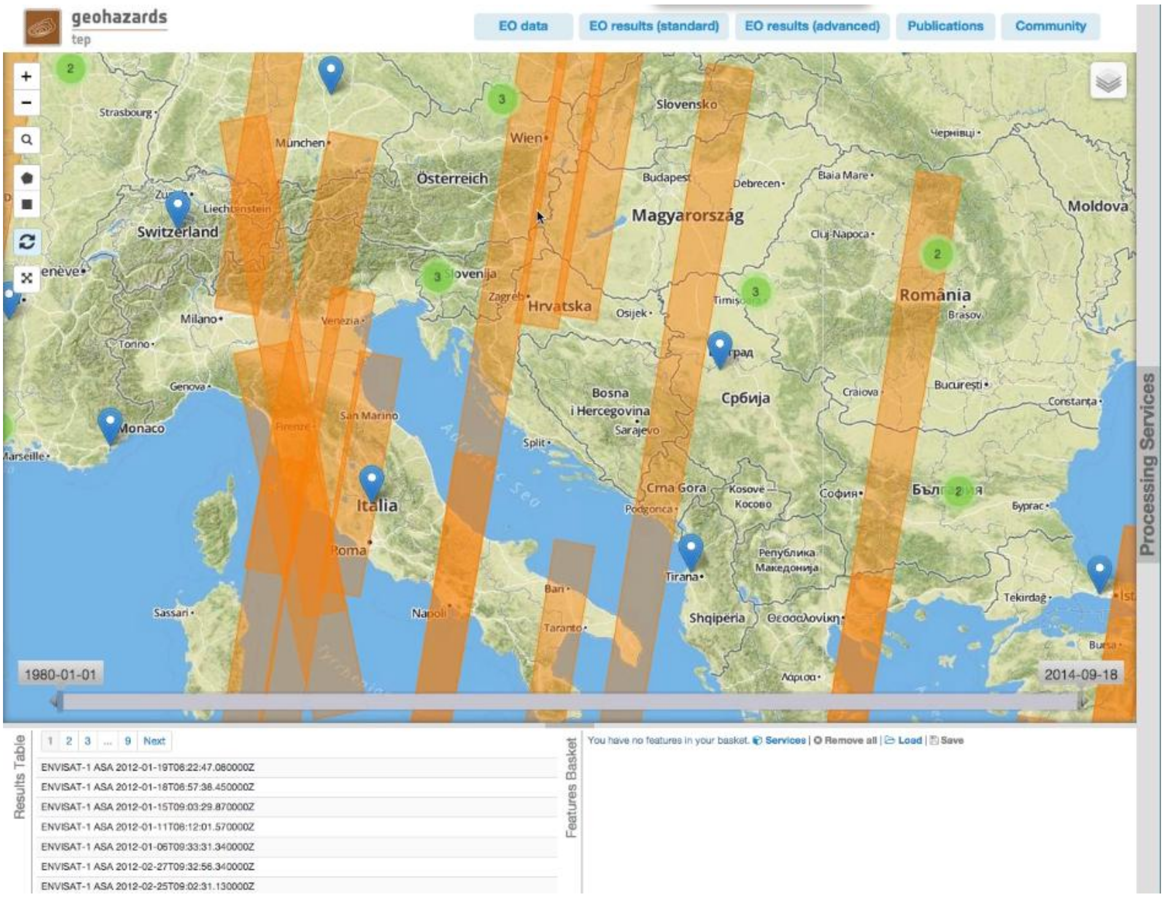
5.3.1. Input data search and selection

The GEP leverages the OGC OpenSearch interface standard for discovery and access. Search is done through Geo and Time filters. Search results are presented over the map and can be further refined (cf. Figure 10).

Currently, search can also be filtered by the intersection with a point of interest, e.g. a Disasters Charter activation event, as currently made available as a layer (others will be added) in the geobrowser (cf. Figure 11).

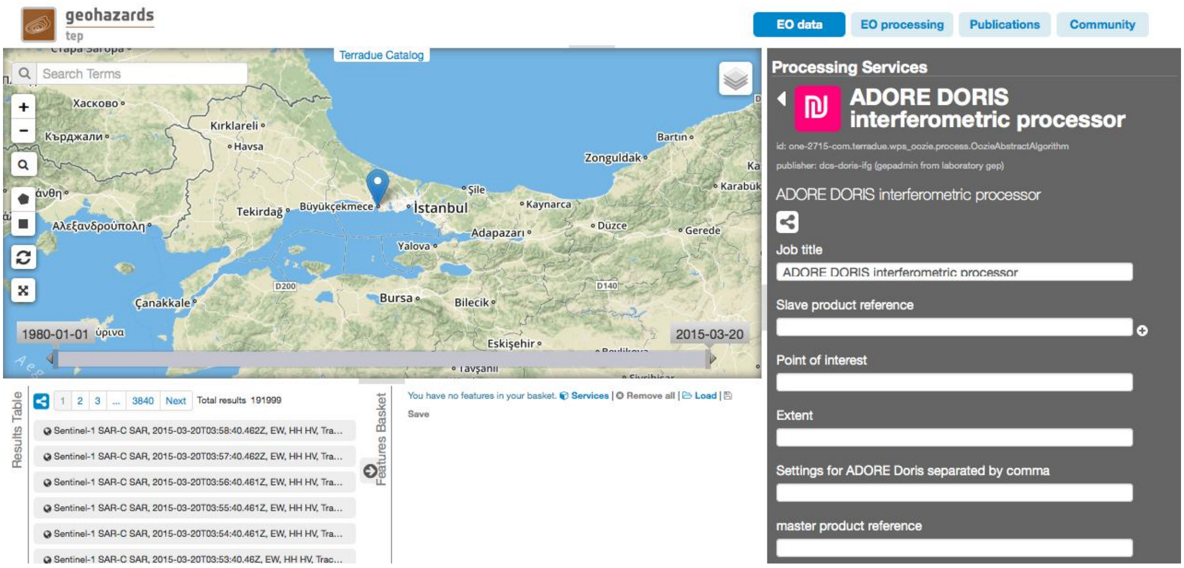
The filtered search results are presented as EO footprints on the map, and as a list on the search results panel from which selections can occur to create a data package (managed in a ‘my basket’ panel).

Figure 10 - GEP input data browsing



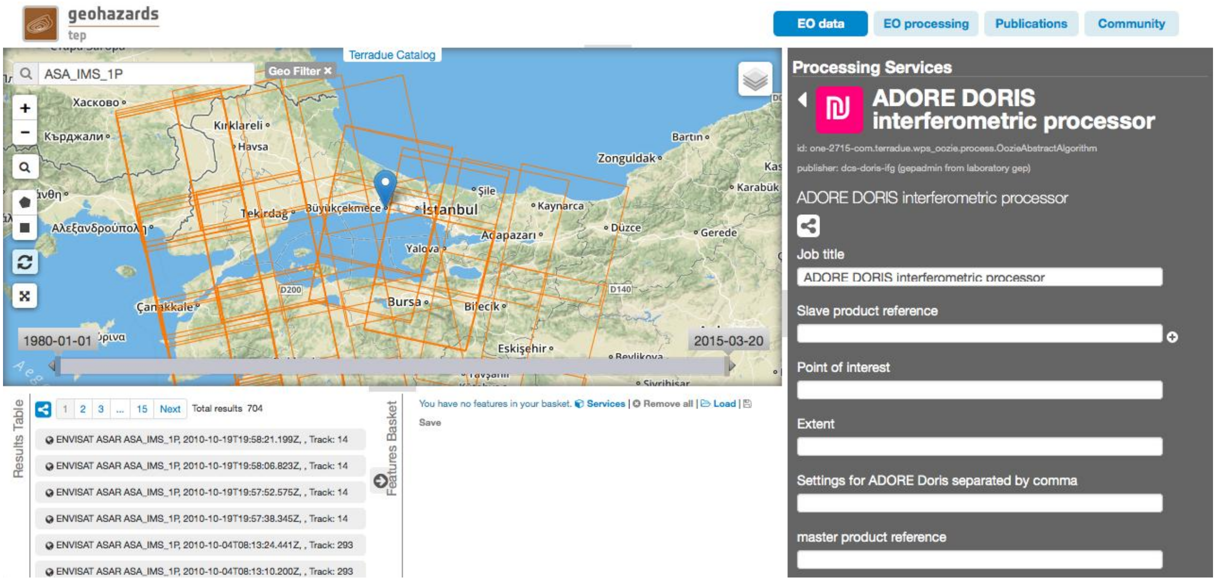
Hereafter (Figure 11) in the results panel, the EO sensor data (e.g. Sentinel-1) that are intersecting the provided point of interest are displayed. In the right-hand panel, the selected processing service with its processing parameters that must be filled in.

Figure 11 - GEP input data area of interest



Search results can be further filtered by sensor type and product level, like here for DORIS ADORE we are looking for ENVISAT ASAR level 1P products (Figure 12).

Figure 12 - GEP input data filtering by sensor type



Then products for processing can be selected and added to the user basket, like here choosing overlapping products on the same acquisition track (Figure 13), to become the master and slave inputs of the InSAR processor (Figure 14).

Figure 13 - GEP select SAR data with same track

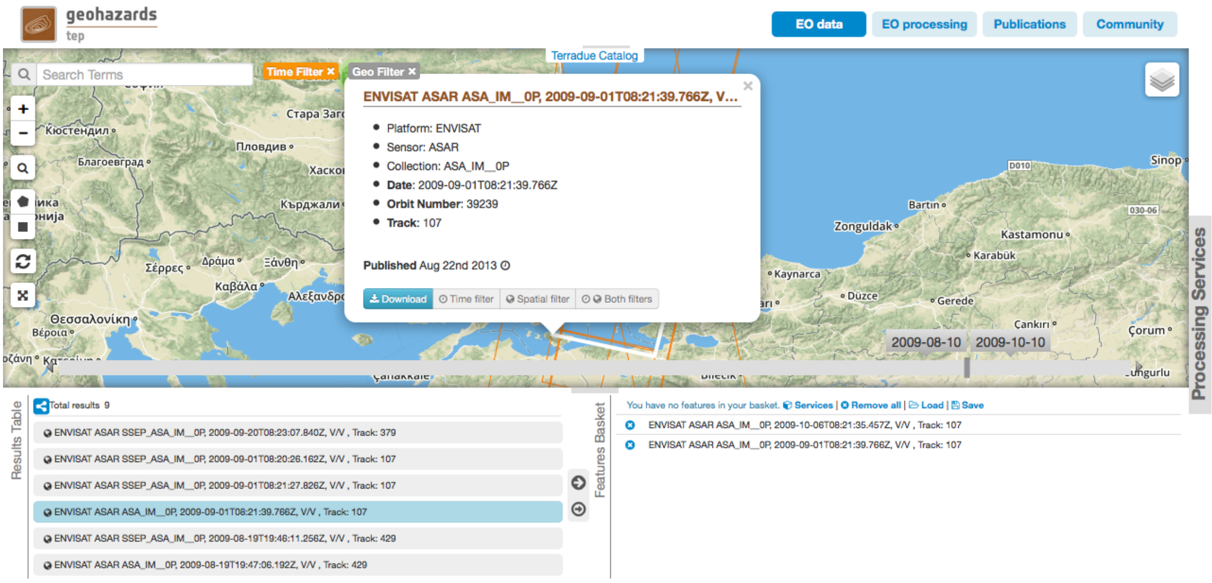
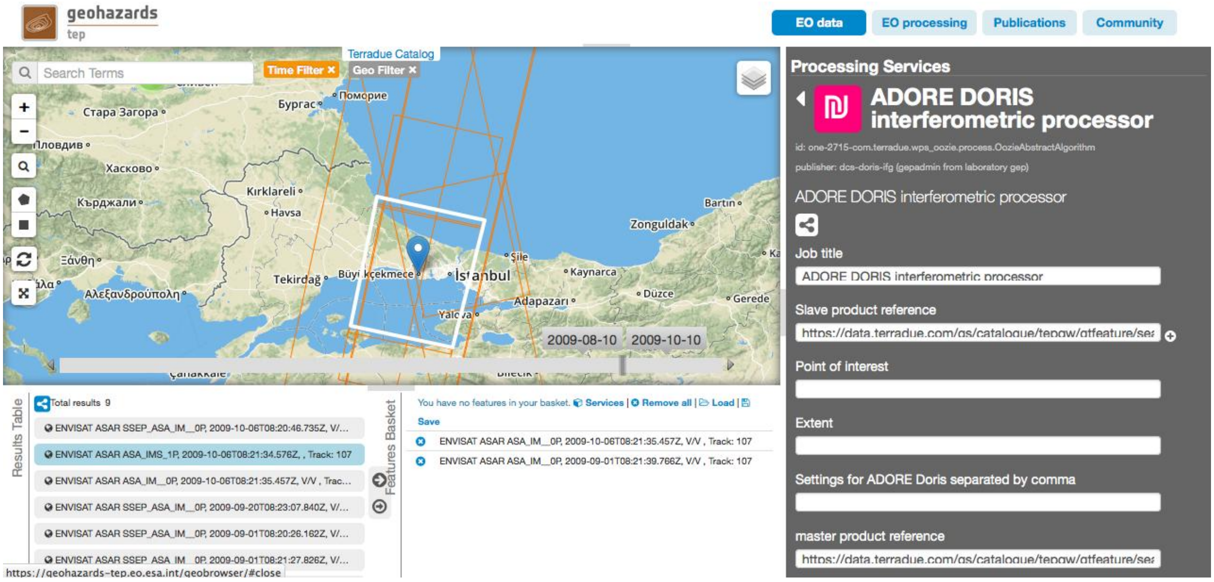


Figure 14 - GEP drag and drop inputs in processor parameters fields

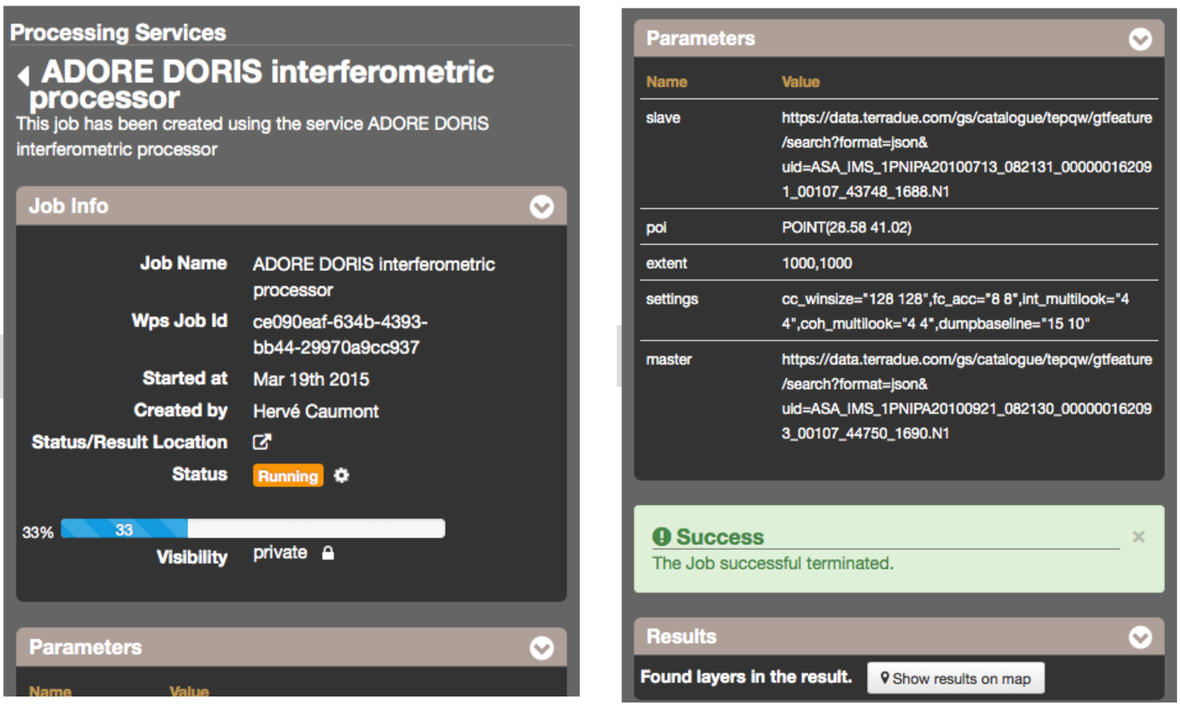


5.3.2. Process monitoring

GEP aims at improved usability in transitioning from EO data resources selection to processing services configuration and triggering.

Once the parameters are filled in, the processor can be triggered and monitored. Once successfully completed, and results analysed and resented on the map (here at this stage of the DORIS ADORE integration, results are only available in SAR geometry).

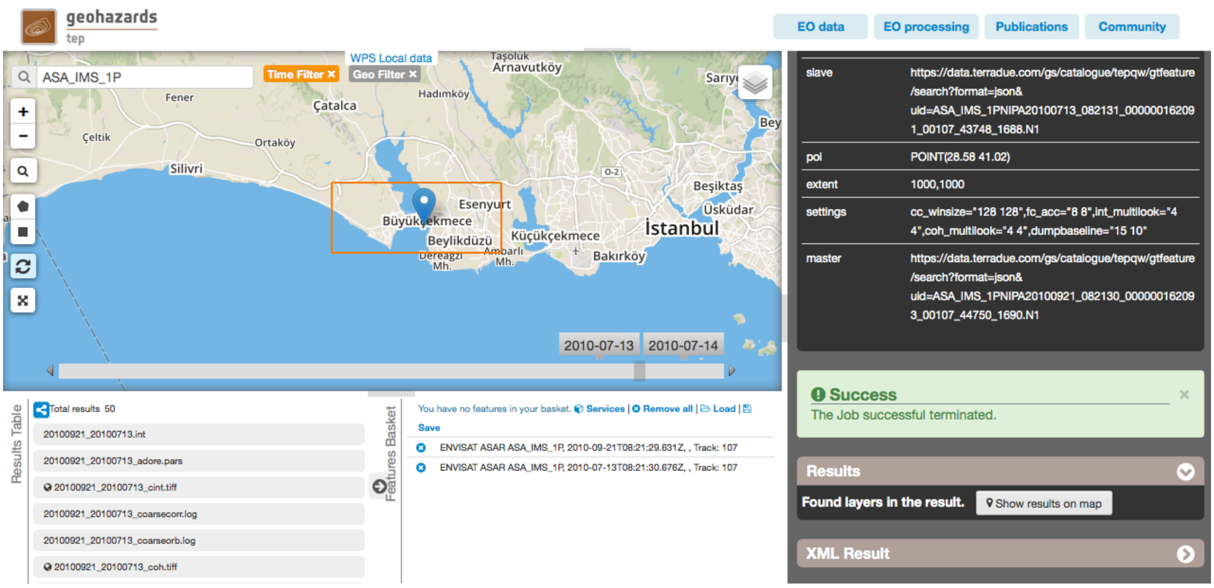
Figure 15 - GEP Process monitoring



5.3.3. Results management

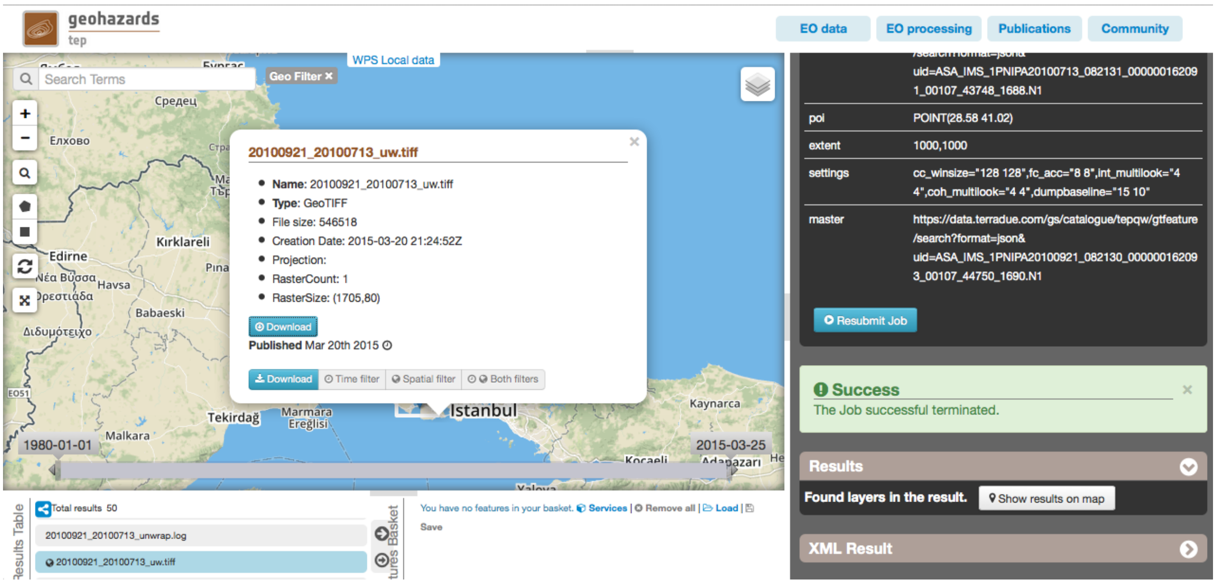
Once a processing job is completed (Figure 16), users can analyse results, and make private jobs public to share the job and results with other users on the platform (allowing them to perform the same tasks) through the user management of processing tasks and of processing results (value-added products).

Figure 16 - GEP retrieve processing result



Results analysis from the geobrowser is supported by processed products download (featured function "Show results on map") and by processing logs and results review and download (featured function "XML Result").

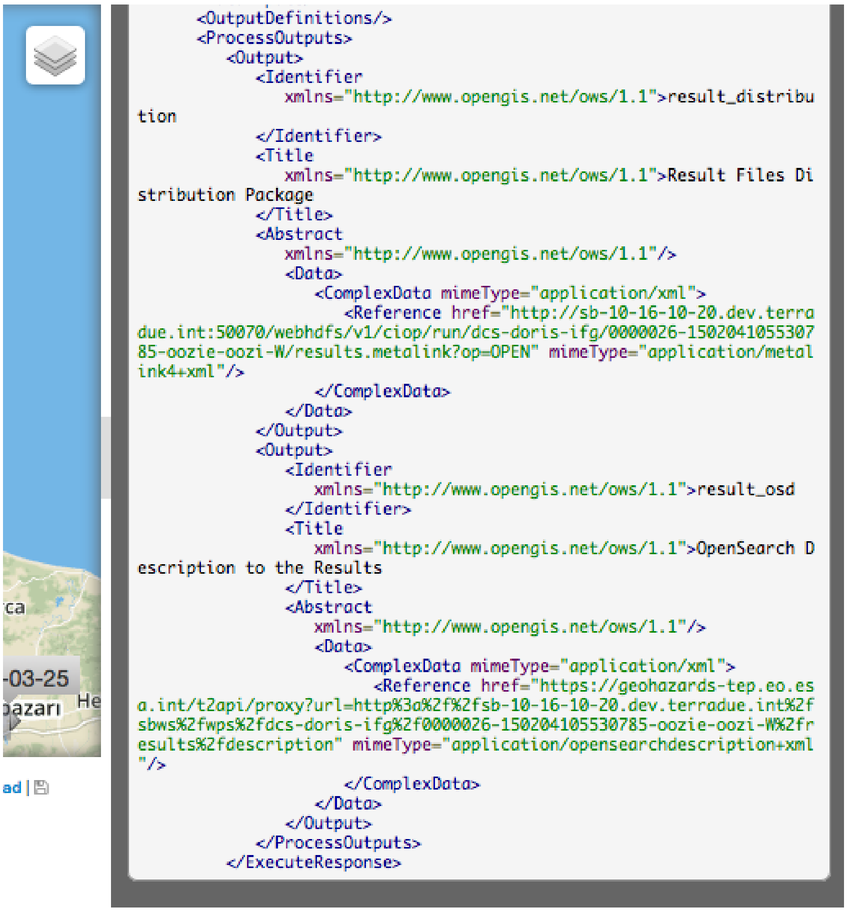
Figure 17 - GEP results analysis



Accessing the XML results provides with the WPS output URL.

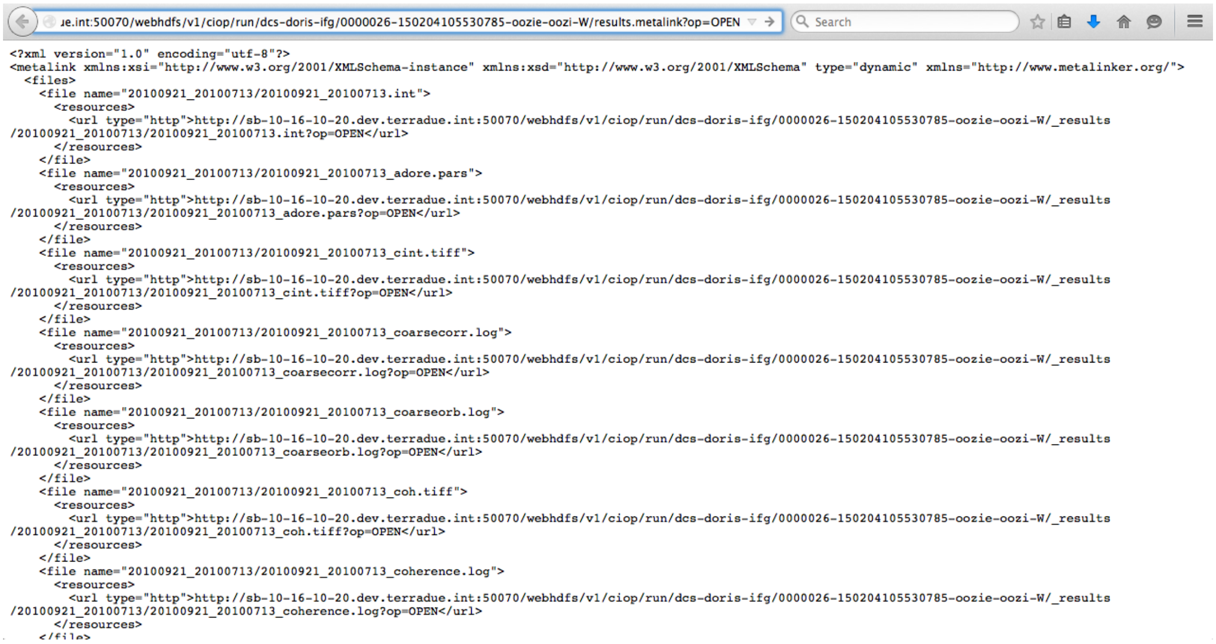
This is made available in two encodings: OGC OpenSearch description encoding and IETF Metalink encoding, for results exploitation.

Figure 18 - GEP WPS process output



```
    <OutputDefinitions/>
    <ProcessOutputs>
      <Output>
        <Identifier
          xmlns="http://www.opengis.net/ows/1.1">result_distribu
tion
        </Identifier>
        <Title
          xmlns="http://www.opengis.net/ows/1.1">Result Files Di
stribution Package
        </Title>
        <Abstract
          xmlns="http://www.opengis.net/ows/1.1"/>
          <Data>
            <ComplexData mimeType="application/xml">
              <Reference href="http://sb-10-16-10-20.dev.terra
due.int:50070/webhdfs/v1/ciop/run/dcs-doris-ifg/0000026-1502041055307
85-oozie-oozi-W/results.metalink?op=OPEN" mimeType="application/metal
ink4+xml"/>
            </ComplexData>
          </Data>
        </Output>
        <Output>
          <Identifier
            xmlns="http://www.opengis.net/ows/1.1">result_osd
          </Identifier>
          <Title
            xmlns="http://www.opengis.net/ows/1.1">OpenSearch D
escription to the Results
          </Title>
          <Abstract
            xmlns="http://www.opengis.net/ows/1.1"/>
            <Data>
              <ComplexData mimeType="application/xml">
                <Reference href="https://geohazards-tep.eo.es
a.int/t2api/proxy?url=http%3a%2f%2fsb-10-16-10-20.dev.terradue.int%2f
sbws%2fwps%2fdcs-doris-ifg%2f0000026-150204105530785-oozie-oozi-W%2fr
esults%2fdescription" mimeType="application/opensearchdescription+xml
"/>
              </ComplexData>
            </Data>
          </Output>
        </ProcessOutputs>
      </ExecuteResponse>
```


Figure 19 - GEP processing results metalinks from WPS output



```
ue.int:50070/webhdfs/v1/ciop/run/dcs-doris-ifg/0000026-150204105530785-oozie-oozi-W/results.metalink?op=OPEN
<?xml version="1.0" encoding="utf-8"?>
<metalink xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://www.w3.org/2001/XMLSchema" type="dynamic" xmlns="http://www.metalinker.org/">
  <files>
    <file name="20100921_20100713/20100921_20100713.int">
      <resources>
        <url type="http">http://sb-10-16-10-20.dev.terradue.int:50070/webhdfs/v1/ciop/run/dcs-doris-ifg/0000026-150204105530785-oozie-oozi-W/_results
/20100921_20100713/20100921_20100713.int?op=OPEN</url>
      </resources>
    </file>
    <file name="20100921_20100713/20100921_20100713_adore.pars">
      <resources>
        <url type="http">http://sb-10-16-10-20.dev.terradue.int:50070/webhdfs/v1/ciop/run/dcs-doris-ifg/0000026-150204105530785-oozie-oozi-W/_results
/20100921_20100713/20100921_20100713_adore.pars?op=OPEN</url>
      </resources>
    </file>
    <file name="20100921_20100713/20100921_20100713_cint.tiff">
      <resources>
        <url type="http">http://sb-10-16-10-20.dev.terradue.int:50070/webhdfs/v1/ciop/run/dcs-doris-ifg/0000026-150204105530785-oozie-oozi-W/_results
/20100921_20100713/20100921_20100713_cint.tiff?op=OPEN</url>
      </resources>
    </file>
    <file name="20100921_20100713/20100921_20100713_coarsecorr.log">
      <resources>
        <url type="http">http://sb-10-16-10-20.dev.terradue.int:50070/webhdfs/v1/ciop/run/dcs-doris-ifg/0000026-150204105530785-oozie-oozi-W/_results
/20100921_20100713/20100921_20100713_coarsecorr.log?op=OPEN</url>
      </resources>
    </file>
    <file name="20100921_20100713/20100921_20100713_coarseorb.log">
      <resources>
        <url type="http">http://sb-10-16-10-20.dev.terradue.int:50070/webhdfs/v1/ciop/run/dcs-doris-ifg/0000026-150204105530785-oozie-oozi-W/_results
/20100921_20100713/20100921_20100713_coarseorb.log?op=OPEN</url>
      </resources>
    </file>
    <file name="20100921_20100713/20100921_20100713_coh.tiff">
      <resources>
        <url type="http">http://sb-10-16-10-20.dev.terradue.int:50070/webhdfs/v1/ciop/run/dcs-doris-ifg/0000026-150204105530785-oozie-oozi-W/_results
/20100921_20100713/20100921_20100713_coh.tiff?op=OPEN</url>
      </resources>
    </file>
    <file name="20100921_20100713/20100921_20100713_coherence.log">
      <resources>
        <url type="http">http://sb-10-16-10-20.dev.terradue.int:50070/webhdfs/v1/ciop/run/dcs-doris-ifg/0000026-150204105530785-oozie-oozi-W/_results
/20100921_20100713/20100921_20100713_coherence.log?op=OPEN</url>
      </resources>
    </file>
```

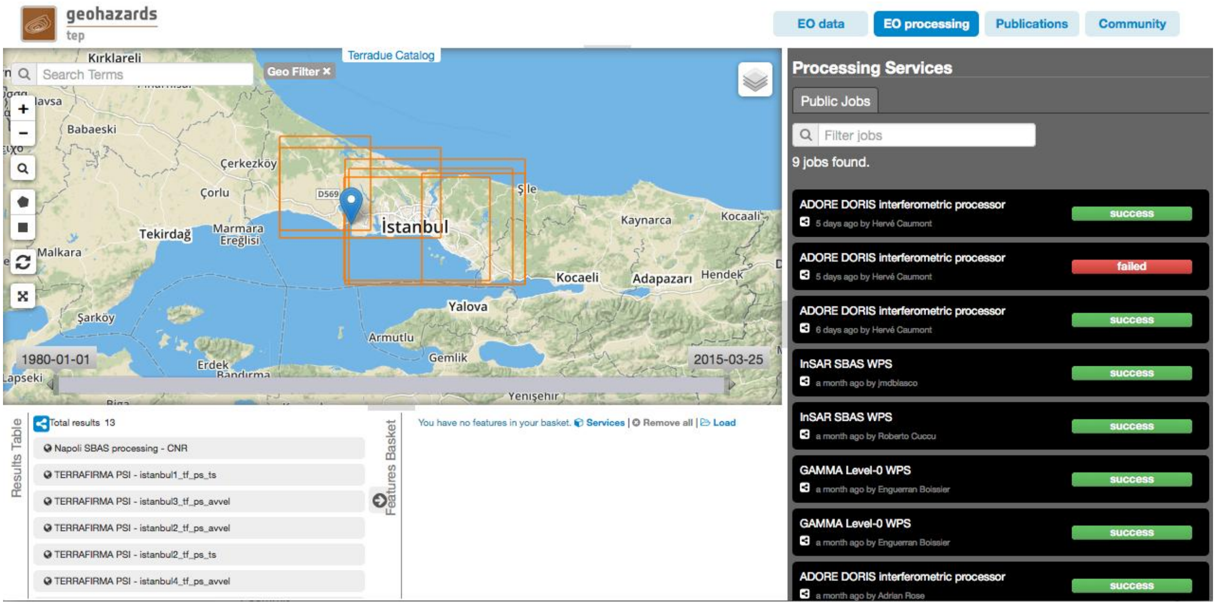
GEP ultimately supports users to share and reproduce processing tasks, analysing and reusing successful jobs, as well as analysing the reasons for the failed ones.

5.4. Discover and access value-added products

Finally, users can share, reuse and reproduce from the platform processing experiments for the definition of new products.

When validated, this process is then applied to the dissemination of community approved value-added products.

Figure 20 - GEP results management



6. Conclusions and next steps

Providing a validated approach and services for sound integration of EO data, products and toolboxes, GEP contributes to connecting the MARsite system to overarching systems like GEOSS and EPOS.

From February to October 2015, ESA is running a GEP validation phase with early adopters, in order to prepare for the pre-operations phase, which starts at the end of 2015. The geohazards community is welcome to apply as early adopters, by committing to run one or several of three scenarios: EO data exploitation, new EO Service development, or new EO Product development. It is an opportunity for the MARsite partners to engage with this process and contribute requirements, use cases and community feedback. ESA edited an Early Adopter Registration Document for exploitation scenarios during the Platform's Validation Phase to support partner's applications. This process has already been adopted by the MED-SUV project.

One key activity to engage with the MARsite partners and community will be the coordination with BRGM in charge of the MARsite Portal and Map Viewer, in order to leverage the interoperability protocols that will allow BRGM to integrate the GEP services and products within the MARsite Portal.