GEOESSENTIAL ERA-PLANET

Essential Variables workflows for resource efficiency and environmental management

Deliverable 7.2
GEOEssential Indicators Toolbox

Creator	CREAF
Contributors	Ivette Serral, Joan Masó (CREAF); Gregory Giulani (UNIGE); Mattia Santoro (CNR); Paolo Mazzetti (CNR)
Creation date	January 08. 2020
Due date	Feb 28. 2020
Last revision date	January 08. 2020
Status	Draft
Type	Report
Description	This deliverable describes the GEOEssential Toolbox used in GEOEssential and based on the VirtualLab. Some examples of application are described as well.
Right	Public
Language	English
Citation	GEOEssential. 2020. GEOEssential Indicators Toolbox. Deliverable 7.2
Grant agreement	ERA-PLANET No 689443

Table of Contents

Introduction

The GEOEssential toolbox is the solution developed within the project to generate indicators useful in the retrieval of EVs. The solution adopted is based on the CNR Virtual Laboratory (VLab). This platform is specially designed for generating knowledge targeted to policy-makers (EVs, indicators and indices) from open and interoperable access to data by running workflows. This was not specifically foreseen in the DoA, but it has been proved the most suitable one as provides publishing, harmonized discovery, access and visualization of resources together with the running workflows and the user interaction through GUI (Portal) and APIs. GEOEssential also benefits from the fact that CNR, developers of the VLab, are partners of the project and main responsible of this task. Moreover, VLab was utilized in several GEOSS Platform demonstrations and proofs of concepts, including the last EuroGEOSS Sprint to Ministerial activity which was presented at the last GEO-XVI Plenary meeting in Canberra. The VLab solution will allow to connect the workflow results to the GEOEssential Dashboard (more details on this connection will be described on D7.5 GEOEssential Dashboard and report).

This deliverable consists in a complete description on the VLab platform technology and a brief explanation of its applicability within GEOEssential on land degradation indicators (more details on showcases will be shown in D7.4 EVs and policy indicators toolbox).

VLab solution

The Virtual Laboratory solution (VLab) (Santoro, Mazzetti and Nativi 2019) was conceived to work in the data-to-knowledge framework defined in GEOEssential D1.1 (Mazzetti, Santoro i Nativi 2018), supporting the knowledge generation process depicted in Figure 1.

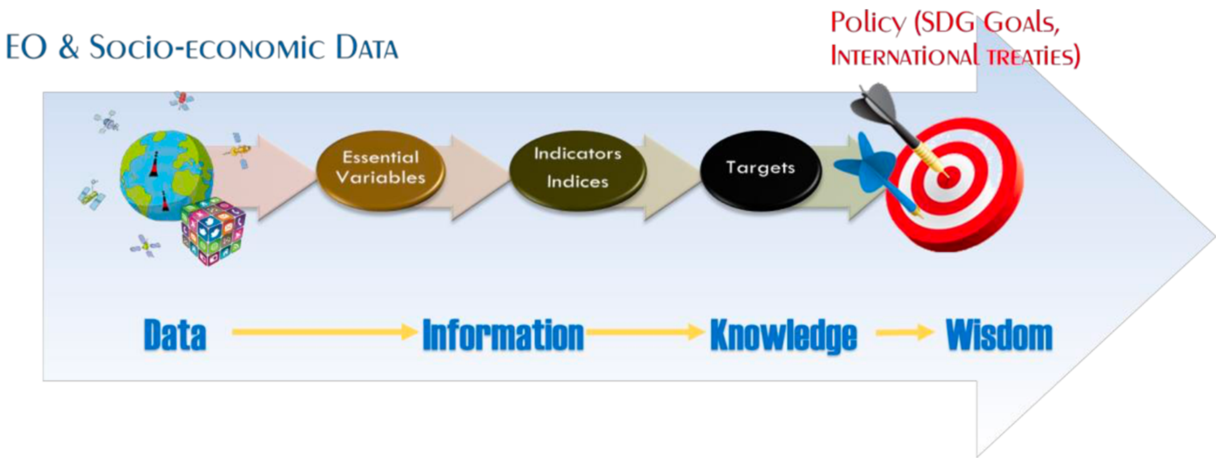

Figure 1. The GEOEssential general scenario: from Data to Knowledge. Source: (Mazzetti, Santoro i Nativi 2018)

The process adopts a step-by-step approach for implementing a data-to-knowledge transition: a set of EVs is generated from observational data (i.e. EO and socio-economic data); the generated EVs are then further processed, generating one or more indicators which capture the state of investigated phenomenon. Such indicators can be used by decision makers to assess the fulfilment of a specific policy goal by comparing them with one or more policy targets ((Lehmann, et al. 2019); (Nativi, Santoro, et al. 2019)).

A key role in the above process is played by environmental models and workflows, which are used for generating both EVs and indicators. VLab is a cloud-based platform aiming to support the activity of environmental scientists in sharing and re-using their models/workflows specifically addressing the Science-to-Information Technology barrier (Santoro, Nativi i Mazzetti 2016). The main challenges addressed by VLab are:

- Minimization of interoperability requirements in the process of model sharing (i.e. to simplify as much as possible the process of publishing and sharing a model for model developers).
- Support multiple programming languages and environments (it must be possible porting models developed in different programming languages and which use an arbitrary set of libraries).

Figure 2. Figure 2 - VLab High-level Overview

Figure 2 depicts a high-level overview of VLab architecture. In brief, VLab must be able to: access the source code of a model, transfer the code to a computing platform, access the required input data and ingest it into the model execution environment, run the model on the computing platform and finally save the generated output data. To implement its business logic, VLab utilizes different resource types (data, source code and computing platforms) which are made available on the Web by external providers. As far as data, VLab exploits the brokering technology provided by the GEO DAB (S. Nativi, P. Mazzetti, et al. 2015) in order to be able to discover and access the required data for the model execution. As far as, model source code, VLab utilizes the Docker technology to support multiple programming languages and execution environments in different cloud platforms. Finally, as far as computing platforms, VLab can run on multiple clouds including: Amazon Web Service (AWS), Copernicus DIAS (ONDA, Sobloo, Creodias) and the European Open Science Cloud (EOSC).
VLab addresses three different types of users:

- Modelers: providers of a model; they can use VLab to publish and share a model.
- Application developers: they use VLab RESTful APIs to develop a web application based on VLab available models.

- End users: they access VLab-enabled web applications to run models available on VLab and visualize/download the results.

Publishing a Model

The model publication procedure is the following:

(i) Share the model source through a code sharing tool (e.g. GitHub[1]);
(ii) share a docker image for the model execution; and
(iii) provide the VLab convention files.

This procedure is documented online[2] along with an "end to end" example which modelers can use to publish and execute a simple model on VLab.
A simple web application (Figure 3) has been developed for modelers to simplify the model publication procedure. This web application is available at https://vlab.geodab.org/.

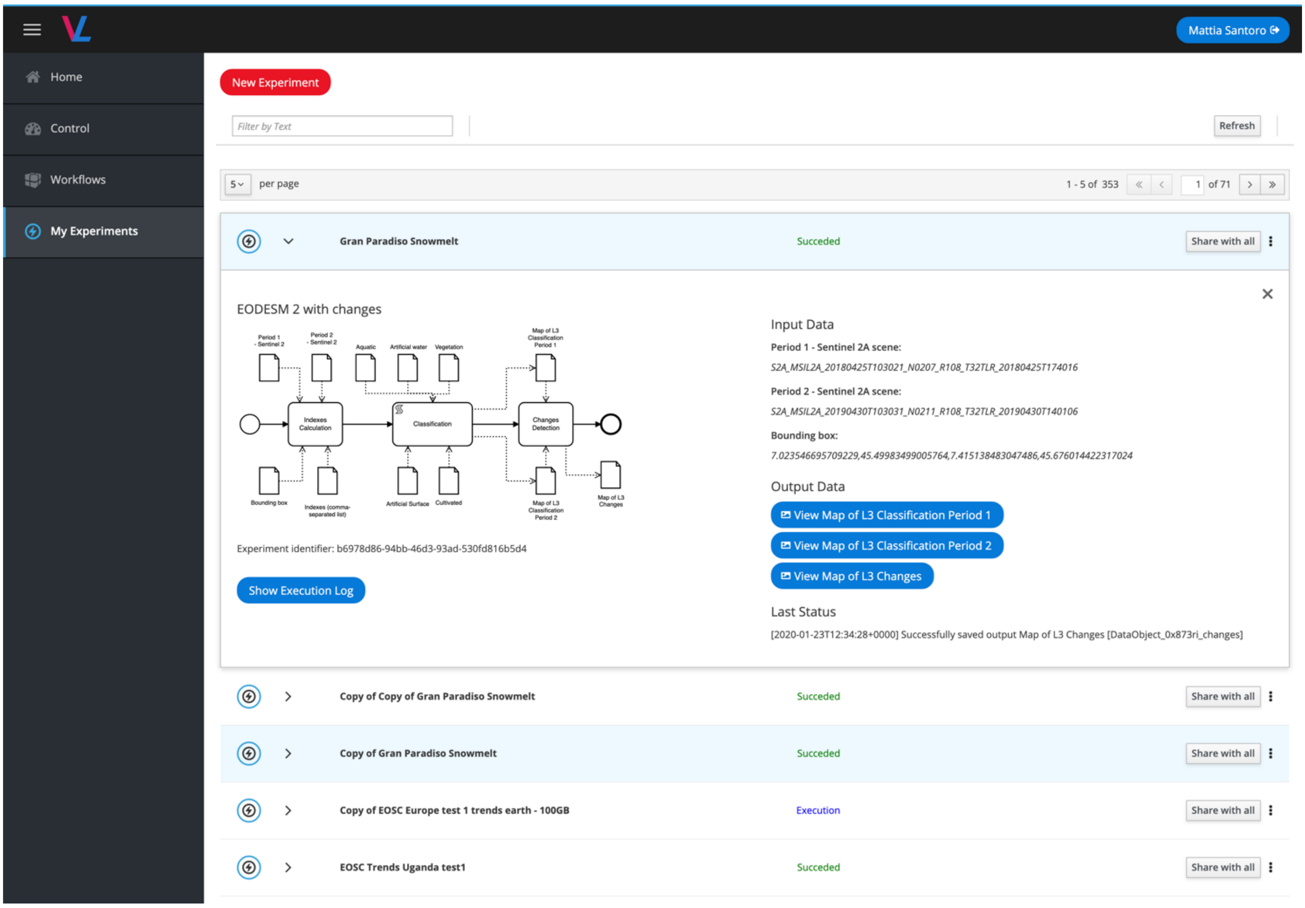

Figure 3. VLab Web Application for Model Publication, Sharing and Testing

[1] https://github.com/
[2] https://confluence.geodab.eu/pages/viewpage.action?pageId=16580641

Currently, about 20 models have been published on VLab and are publicly accessible. As far as programming languages, the following are utilized by the available models: Python, R[3], FORTRAN, NetLogo[4], Matlab[5], NCL[6].

VLab APIs

VLab pusblishes RESTful APIs which are utilized by client applications for end users. The VLab APIs allow to:

- Validate an existing source code repository for its use in VLab platform;
- Publish a new model from a valid source code repository;
- Discover available models/workflows along with their description, both in text and BPMN (OMG, 2011) formats;
- Execute available models, utilizing different cloud platforms for the execution.

VLab APIs are documented online at http://vlabapi.geodab.org/.

Besides the VLab web application for model publication and testing (**Error! Reference source not found.**), VLab APIs are utilized by the GEOEssential dashboard. At the moment, the GEOEssential dashboard utilizes VLab APIs to retrieve the output of the executed models. Next developments will enhance the interaction with the GEOEssential dashboard to allow users to request the execution of the models directly from the GEOEssential dashboard.

Showcase on land degradation

The showcase presented here is based on land degradation to derive SDG 15.3.1 indicator. Through some input indicators/EVs and the workflow executed in the VLab, 4 main outputs are obtained to monitor land degradation and try to achieve land degradation neutrality (LDN) considered as the state whereby the amount and quality of land resources remains stable or increases within specified temporal and spatial scales (Gonzalez-Roglich, et al. 2019).

Indicator calculation

Land degradation is defined by the United Nations as *"the reduction or loss of the biological or economic productivity and complexity of rain fed cropland, irrigated cropland, or range, pasture, forest and woodlands resulting from a combination of pressures, including land use and management practices"* (United Nations Statistical Division 2018). Total land area is the total country area without considering inland waters surfaces (e.g., major rivers and lakes). It is expressed in hectares or km^2. The indicator represents the portion, in percentage, of degraded land over total land area, derived from a binary classification of land condition (i.e., degraded or not degraded) measured by three sub-indicators:

[3] https://www.r-project.org
[4] https://ccl.northwestern.edu/netlogo/
[5] https://mathworks.com/products/matlab.html
[6] https://www.ncl.ucar.edu

(1) *Land cover and land cover changes*, to determine the possible diminution of ecosystem services that are valuable in a local or national context ((Burkhard, et al. 2009); (Ban, Gong i Gini 2015); (Andrew, Wulder i Nelson 2014)).
(2) *Land productivity trends*, to assess land productive capacity and health changes. It can indicate changes in ecosystem functioning and decreasing trends can suggest that land is degrading ((Cowie, et al. 2018); (Oehri, et al. 2017)).
(3) *Carbon stocks (above and below ground) trends*, to evaluate soil quality associated with nutrient cycling and its stability and structure with direct implications for water infiltration, soil biodiversity, vulnerability to erosion, and ultimately the productivity of vegetation, and in agricultural contexts, yields ((Stumpf, et al. 2018); (Hengl, et al. 2017)).

The quantification of the indicator is based on the evaluation of changes in the sub-indicators to derive the proportion of land that is degraded. The three sub-indicators are complementary and responsive to different degradation elements. To determine whether a given land unit is degrading, stable or improving the One Out, All Out (1OAO) principle is used. If one of the sub-indicators value is negative, then the land unit is considered as degraded. This is a precautionary rule considering that stable or improved land condition in any sub-indicators cannot counterbalance the degradation effects in the others (Figure 4).

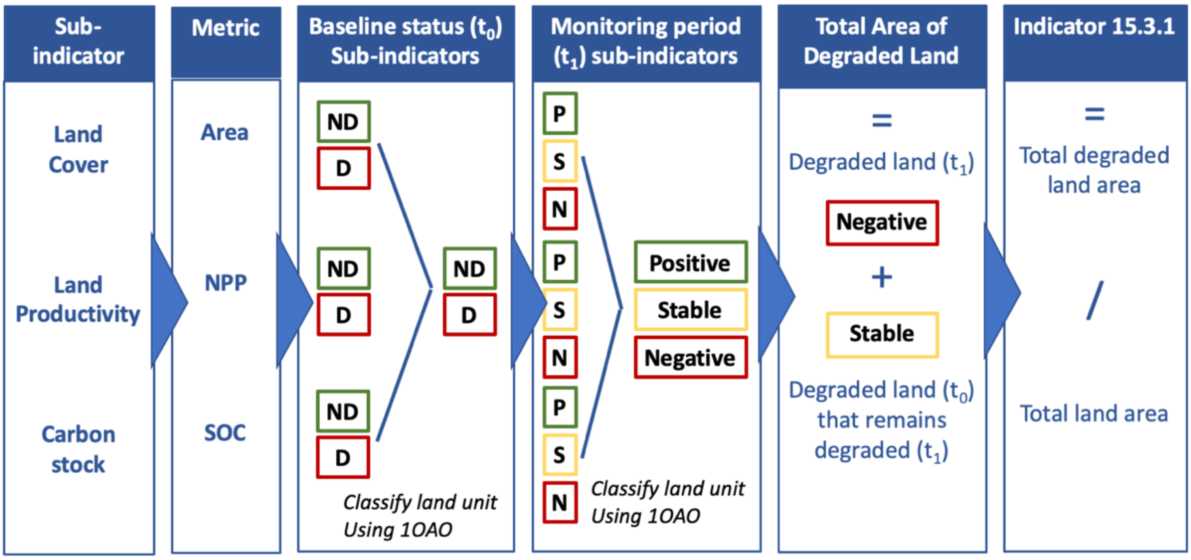

Figure 4. Schematic representation of the methodology for generating the sub-indicators and calculating the SDG 15.3.1 indicator (adapted from (United Nations Statistical Division 2018))

Land degradation is generally context-specific and therefore makes it difficult for a single indicator to capture the full complexity of land state and condition (Gilbey, et al. 2019). However, the sub-indicators are sufficiently robust to address changes in different relevant ways such as understanding relatively fast changes with land cover or productivity trends while apprehending slower changes through carbon stocks ((Cowie, et al. 2018); (Gonzalez-Roglich, et al. 2019)). These sub-indicators are widely accepted to monitor major factors and driving variables reflecting the capacity to deliver valuable ecosystem services (Bojie, et al. 2015). Their definition and methodology for calculation are recognized as technically and economically feasible for systematic observation ((Bojinski, et al. 2014); (Pereira, et al. 2013)). The indicator should be derived generally from standardized and comparable national official data sources. However, due to their nature, these sub-indicators can be derived from satellite EO as well as geospatial data from regional and global data repositories and can replace, complement or enhance national official data sources after validation by national authorities (UNCCD 2017).

System architecture and implementation

Currently, the reference implementation to help countries to monitor degraded land is Trends.Earth (Gonzalez-Roglich, et al. 2019). It is a QGIS plugin working in combination with the Google Earth Engine (GEE) to facilitate data preparation, processing and visualization for generating both the sub-indicators and the final SDG indicator 15.3.1 (Gorelick, et al. 2017)). The objective of this model is to help countries in analyzing data and preparing their reporting (e.g., plot time-series of sub-indicators, maps, graphics) in a format that is directly aligned with the UNCCD's Performance Review and Assessment of Implementation System (PRAIS), which is the reporting portal of LDN. The default datasets provided with Trends.Earth are indeed coarse in scale, but they are also amongst the most consistent globally, and this tool is desgined to use any dataset at any scale. Even if this tool greatly facilitates to production of the SDG indicator, it has some shortcomings:

(1) It mostly relies on regional and global datasets derives from EO data. However, it lacks greater spatial and longer temporal resolution to better capture the dynamics of land degradation at national scale ((Pasquarella, et al. 2016); (Pettorelli, et al. 2018)). This limits the scalability (e.g., different scales from national to global) of the model.
(2) It is largely dependent of the GEE platform. Therefore, enhancing flexibility can be valuable allowing to access different data sources and different processing platforms.
(3) Finally, it translates EO data in useful information. However, it does not provide any knowledge. Having a possibility to create some knowledge would be valuable.

Consequently, to tackle these issues, it is interesting to develop workflows to address the need for trusted sources of data, EVs and information to monitor the progresses made on environmental conditions towards policy targets (Lehmann, et al. 2019).

To facilitate access and integration of scientific models and their outputs, it has been decided to use a Model Web approach (Nativi, Mazzetti i Geller 2013) suggesting four architectural and policy principals for implementation (Mazzetti, et al. 2016):

(1) *Open Access:* noticeably, support the documentation, publication, and sharing of models and algorithms.
(2) *Low entry barrier:* reduce entry barriers for both resources' providers and users.
(3) *Service-driven approach:* in particular, models and algorithms access are provided by online services to enhance machine-to-machine interoperability.
(4) *Scalability:* facilitates the use of increasingly large volume and variety of data –i.e. Big Data requirements.

In addition, and closely related to these principles, an important pattern to be applied is the *separation of concerns* –i.e. separating a computing process into distinct sections, so that each section addresses a separate concern. By adopting the principles and the implementation pattern, the software components (along with their services) dealing with Data, Information, and Knowledge scopes were separated carrying out three sub-systems that generate, respectively: the sub-indicators, the SDG composed indicator, and the online services to access results together with the visualization tools to explore them as maps and graphs (Figure 5).

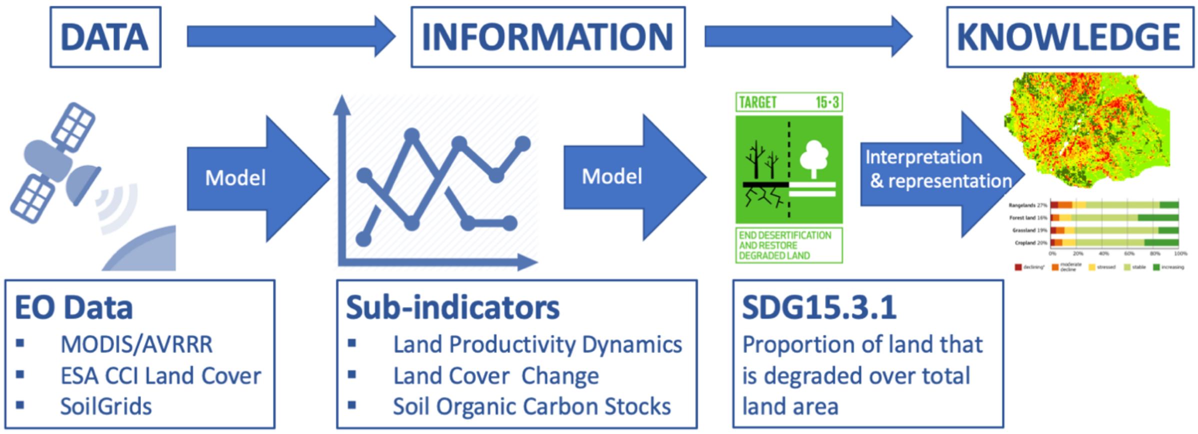

Figure 5. Products and services generated by the proposed workflow

Applying the separation-of-concern pattern brought a set of benefits, including: (a) the Trends.Earth model is now published as a Model Web service; (b) presently, the sub-indicators can be generated by using different data sources and processing platforms; (c) scalability and flexibility of the entire value-chain was improved, introducing, for example, the possibility to parallelize the model execution and address important challenges raised by Big Data. This implementation schema allows users to compute each sub-indicator separately in a spatially explicit manner under the form of raster maps that is then integrated into a final indicator map, producing at the same time a table with results reporting areas potentially improved, stable or degraded over the area of interest. This enables the use of different data sources, such as the Copernicus Open Access Hub (previously known as Sentinels Scientific Data Hub) and/or the Copernicus data and Information Access Services (DIAS), the Global Earth Observation System of Systems (GEOSS), or national data infrastructures such as Data Cubes ((European Commission 2018); (Craglia, et al. 2017); (Shushanik, et al. 2019); (Giuliani, et al. 2017)) (Figure 6).

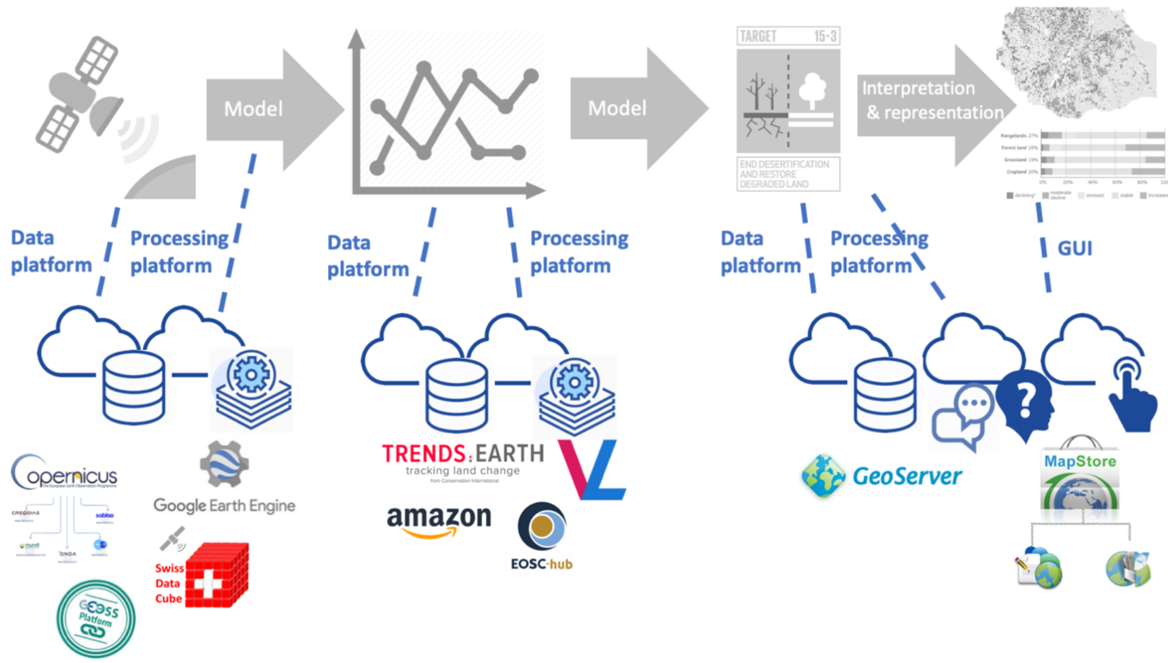

Figure 6. The proposed value-chain process

Model outputs are then published using GeoServer (https://geoserver.org), an open source web server designed to publish geospatial data using widely adopted standards (e.g., Web Map Service (WMS), Web Coverage Service WCS)) advanced by the Open Geospatial Consortium (OGC) (Giuliani, et al. 2011). It allows to share, access, and use data in an interoperable and standardized way, facilitating data access, exchange, and integration.

Results are then organized in a Dashboard environment that allows users visualizing and exploring model outputs in a concise and comprehensible way. Dashboards are currently gaining a lot of interest as a tool for facilitating users' interaction with complex sets of data and information (van Ginkel Kees, et al. 2018) and can potentially help decision-makers or practitioners to better understand an issue (Fegraus, et al. 2012).

To create a dashboard, it has been decided to use MapStore (https://mapstore.geo-solutions.it) that is an open source web-based application conceived to produce, manage and securely share maps, mashups, and dashboards using resources published following OGC standards. It provides users with common standard geoportal functionalities (e.g., map visualization, data discovery, and spatial analysis) allowing to find, view and query geospatial data and integrate multiple data sources into a single map. In addition, it allows to create dashboards using widgets such as maps, statistical charts, tables, and text boxes.

The main output of these models is a pixel-based over the entire country showing area that are considered either degraded, stable or improved according to the SDG15.3.1 indicator. This output can be then aggregated at various administrative levels as well as providing the national value. Further, all this information is integrated in a dashboard providing a consistent and comprehensible one-(web) page document allowing to explore the issue. This is exemplified by the following example (Figure 7).

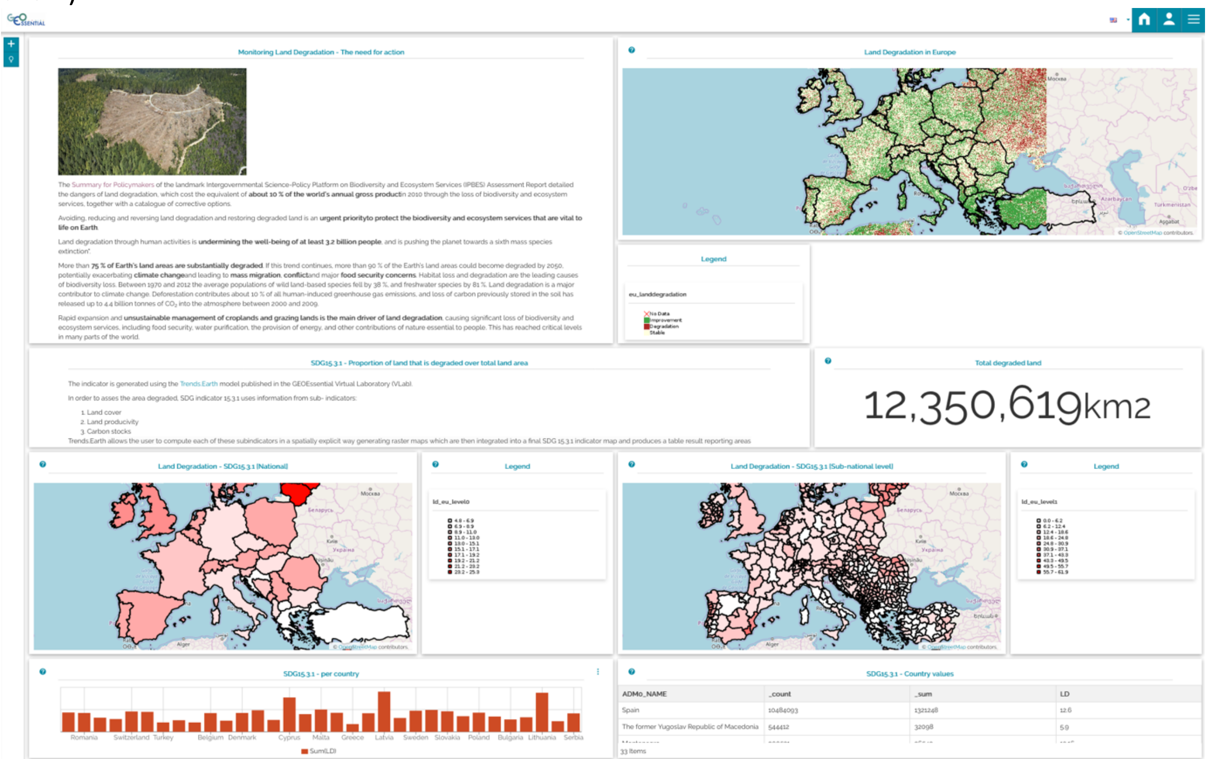

Figure 7. SDG15.3.1 dashboard at European scale (https://geoessential.unepgrid.ch/mapstore/#/dashboard/4)

This dashboard enables to dynamically explore the pixel-based map as well as the aggregated indicator at national and sub-national levels. Maps, graph, table, and counters are synchronized and

dynamically updated according to the zoom level. It allows visualizing only the information on countries that are visible in the current zoom level. Additionally, some text is provided to explain the issue of land degradation and how the indicator is calculated. The dashboard aggregates and provides a possible interpretation of the generated information: i.e. the current status and the temporal changes of the land degradation index. Providing such a meaning to the worked-out information, the process generates knowledge.
Through the dashboard, users access the knowledge on land degradation status directly provided by the value of the index. They also access further knowledge through the generation of aggregated indices and through visual analysis of spatial distribution through the visualized maps.
All the generated information and knowledge are exposed with well-recognized interfaces such as OGC standards for efficient discovery, access and use. This makes data Findable, Accessible, Interoperable and Reusable (FAIR)((Stall, et al. 2019); (Wilkinson, et al. 2016)) and greatly facilitates the sharing of data, information and knowledge and contributing to major initiative like GEOSS.

More information can be found in: Giuliani G., Mazzetti P., Santoro M., Nativi S., Van Bemmelen J., Colangeli G., and Lehmann A. (2020) **Knowledge generation using satellite Earth Observations to support Sustainable Development Goals (SDG): a use case on Land Degradation**, International Journal of Applies Earth Observation and Geoinformation, *in press*.

Glossary

API	Application Programming Interface
EO	Earth Observation
DAB	Discovery and Access Broker
EV	Essential Variables
GEE	Google Earth Engine
GEO	Group on Earth Observations
GEOSS	Global Earth Observation System of Systems
GUI	Graphical User Interface
LDN	Land Degradation Neutrality
PRAIS	UNCCD's Performance Review and Assessment of Implementation System
SDG	Sustainable Development Goals
VLab	Virtual Laboratory

Bibliography

Andrew, M. E., M. A. Wulder, i T. A. Nelson. «Potential Contributions of Remote Sensing to Ecosystem Service Assessments.» *Progress in Physical Geography*, 2014: 328–53.

Ban, Y. F., P. Gong, i C. Gini. «Global Land Cover Mapping Using Earth Observation Satellite Data: Recent Progresses and Challenges.» *Isprs Journal of Photogrammetry and Remote Sensing*, 2015: 1–6.

Bojie, Fu, Liwei Zhang, Zhihong Xu, Yan Zhao, Yongping Wei, i Dominic Skinner. «Ecosystem Services in Changing Land Use.» *Journal of Soils and Sediments*, 2015: 833–43.

Bojinski, Stephan, Michel Verstraete, Thomas C. Peterson, Carolin Richter, Adrian Simmons, i Michael Zemp. «The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy.» *Bulletin of the American Meteorological Society*, 2014: 1431–43.

Burkhard, Benjamin, Franziska Kroll, Felix Müller, i Wilhelm Windhorst. «Landscapes' Capacities to Provide Ecosystem Services – a Concept for Land-Cover Based Assessments.» *Landscape Online*, 2009: 1–22.

Cowie, Annette L., et al. «Land in Balance: The Scientific Conceptual Framework for Land Degradation Neutrality.» *Environmental Science & Policy*, 2018: 25–35.

Craglia, Max, Jiri Hradec, Stefano Nativi, i Mattia Santoro. «Exploring the Depths of the Global Earth Observation System of Systems.» *Big Earth Data*, 2017: 1–26.

European Commission. «The DIAS: User-Friendly Access to Copernicus Data and Information.» 2018.

Fegraus, E. H., et al. «nterdisciplinary Decision Support Dashboard: A New Framework for a Tanzanian Agricultural and Ecosystem Service Monitoring System Pilot.» *IEEE Journal of Selected Topics In Applied Earth Observations And Remote Sensing*, 2012: 1700–1708.

Gilbey, Ben, Jonathan Davies, Graciela Metternicht, i Chris Magero. «Taking Land Degradation Neutrality from Concept to Practice: Early Reflections on LDN Target Setting and Planning.» *Environmental Science & Policy*, 2019: 230–37.

Giuliani, G., et al. «Sharing Environmental Data through GEOSS.» *International Journal of Applied Geospatial Research*, 2011: 1–17.

Giuliani, Gregory, et al. «Building an Earth Observations Data Cube: Lessons Learned from the Swiss Data Cube (SDC) on Generating Analysis Ready Data (ARD).» *Big Earth Data*, 2017: 1–18.

Gonzalez-Roglich, Mariano, et al. "Synergizing Global Tools to Monitor Progress towards Land Degradation Neutrality: Trends.Earth and the World Overview of Conservation Approaches and Technologies Sustainable Land Management Database." *Environmental Science & Policy*, 2019: 34–42.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, i Rebecca Moore. «Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone.» *Remote Sensing of Environment*, 2017: 18–27.

Hengl, Tomislav, et al. «SoilGrids250m: Global Gridded Soil Information Based on Machine Learning.» *PLOS ONE*, 2017.

Lehmann, Anthony, et al. «GEOEssential – Mainstreaming Workflows from Data Sources to Environment Policy Indicators with Essential Variables.» *International Journal of Digital Earth*, 2019: 1–17.

Mazzetti, P, M Santoro, i S Nativi. «GEOEssential Deliverable 1.1: Knowledge Services Architecture.» 2018.

Mazzetti, Paolo, et al. «Integration of Data and Computing Infrastructures for Earth Science: An Image Mosaicking Use-Case.» *Earth Science Informatics*, 2016: 1–18.

Meyer, D., i M. Riechert. «Open Source QGIS Toolkit for the Advanced Research WRF Modelling System.» *Environmental Modelling & Software*, 2019: 166–78.

Nativi, S, M Santoro, G Giuliani, i P Mazzetti. «Towards a knowledge base to support global change policy goals.» *International Journal of Digital Earth*, 2019: 1–29.

Nativi, S, P Mazzetti, M Santoro, F Papeschi, M Craglia, i O Ochiai. «Big Data Challenges in Building the Global Earth Observation System of Systems.» *Environmental Modelling & Software*, 2015: 1–26.

Nativi, Stefano, Paolo Mazzetti, i Gary N. Geller. «Environmental Model Access and Interoperability: The GEO Model Web Initiative.» *Environmental Modelling and Software*, 2013: 214–28.

Oehri, Jacqueline, Bernhard Schmid, Gabriela Schaepman-Strub, i Pascal A. Niklaus. «Biodiversity Promotes Primary Productivity and Growing Season Lengthening at the Landscape Scale.» *Proceedings of the National Academy of Sciences*, 2017.

Pasquarella, Valerie J., Christopher E. Holden, Les Kaufman, i Curtis E. Woodcock. «From Imagery to Ecology: Leveraging Time Series of All Available Landsat Observations to Map and Monitor Ecosystem State and Dynamics.» *Remote Sensing in Ecology and Conservation*, 2016.

Pereira, H. M., et al. «Essential Biodiversity Variables.» *Science*, 2013: 277–78.

Pettorelli, Nathalie, et al. «Satellite Remote Sensing of Ecosystem Functions: Opportunities, Challenges and Way Forward.» *Remote Sensing in Ecology and Conservation*, 2018: 71–93.

Santoro, M, P Mazzetti, and S Nativi. "The VLab: a cloud-based platform to share and execute scientific models." *Geophysical Research Abstracts.* Vienna: EGU, 2019. 12360.

Santoro, M, S Nativi, i P Mazzetti. «Contributing to the GEO Model Web implementation: A brokering service for business processes.» *Environmental Modelling & Software*, 2016: 18-34.

Shushanik, Asmaryan, et al. «Paving the Way towards an Armenian Data Cube.» *Data*, 2019.

Stall, Shelley, et al. «Make Scientific Data FAIR.» *Nature*, 2019.

Stumpf, Felix, Armin Keller, Karsten Schmidt, Andreas Mayr, Andreas Gubler, i Michael Schaepman. «Spatio-Temporal Land Use Dynamics and Soil Organic Carbon in Swiss Agroecosystems.» *Agriculture, Ecosystems & Environment*, 2018: 129–42.

UNCCD. «Good Practice Guidance - SDG Indicator 15.3.1.» 2017.

United Nations Statistical Division. *SDG Indicator 15.3.1 - Metadata.* 2018. https://unstats.un.org/sdgs/metadata/.

van Ginkel Kees, C. H., Arjen Y. Hoekstra, Joost Buurman, i Rick J. Hogeboom. «Urban Water Security Dashboard: Systems Approach to Characterizing the Water Security of Cities.» *Journal of Water Resources Planning and Management*, 2018.

Wilkinson, Mark D., et al. «The FAIR Guiding Principles for Scientific Data Management and Stewardship.» *Scientific Data*, 2016.