This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059548.

OPEN EARTH MONITOR

MS4 Open-Earth-Monitor Computing Engine and In-situ Data 1st release and Usability Assessment

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059548.

Document control page

Project	Open-Earth-Monitor (OEMC)
Project, full title	A cyberinfrastructure to accelerate uptake of environmental information and help build user communities at European and global levels
Project number	101059548
Project start	June 1st 2022
Milestone	MS4 Open-Earth-Monitor computing engine and in-situ data 1st release and usability assessment
Work Package	WP2 User-driven system design and FAIR workflows WP3 Open-Earth-Monitor computing engine WP4 Open-Earth-Monitor in-situ O&M usability tools WP8 Communication, dissemination and collaboration
Description of Milestone	Software released via GitLab; tutorials and documentation published and available via the project website
Version	Final
Responsible authors	Simone Sabbatini and Leandro Parente
Contributors	OGH and CMCC
Date of delivery	30 November 2023
Due date of Milestone	30 November 2023
Type	
Language	English
Rights	
Status	( ) In progress ( ) For review (x) Approved
Dissemination level	Public

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059548.

Table of Contents

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059548.

Executive Summary

In-situ datasets are key for the successful realization of the goals of the OEMC project. The back-end components Open-Earth-Monitor computing engine and in-situ O&M data service are the core functional components of the cyberinfrastructure that will ultimately support producing the most accurate and most complete and consistent analysis-ready data. The integration of in-situ datasets with earth-observations on a cloud-based solution is a critical target of the OEMC project. Eight categories have been identified amidst the project, each cared about by a dedicated task in the WP4. An assessment of the usability was carried out by scouting available datasets from the different categories, and, after checking their characteristics, a subset was selected to be included in the first release. A STAC catalog with this subset was created, which can be accessed here: https://stac.openlandmap.org/ . Having all the datasets listed and described in the same tool is a crucial effort in terms of standardization and harmonization, and the starting point to foster accessibility and usability of the datasets provided. More in-situ datasets will be added to the catalog as soon as their compatibility and integration will be assessed and carried out.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059548.

Introduction

In-situ data definition has an extended meaning in the project: it includes data measured “on the ground” by dedicated instrumentation working continuously, or by campaigns carried out in the field by specialized operators and teams. Citizen science observations are also relevant as in-situ datasets. LIDAR campaigns and airborne measurements can be included as well in the in-situ category. In addition to the intrinsic value of point observations, the importance of in-situ datasets also spans from their integration with EO to their use as training points for modeling and calibration, to validation of maps.

Eight big categories of in-situ datasets have been identified: 1. GHG fluxes; 2. Forest biomass; 3. Marine, terrestrial, landscape biodiversity; 4. Ocean, seas and coastal waters; 5. LC/LU and soil; 6. Automated ground networks; 7. Citizen science; 8. In-situ/gridded harmonization. Investigation and gathering of datasets in each category is the responsibility of a task inside the WP4. Datasets in the same categories are often heterogeneous in terms of different characteristics: spatial and temporal scales, spatial and temporal resolutions, collection methods, granularity. The data format can be vectorial (points) or as ASCII files. For that a high degree of heterogeneity is present, that can affect the usability of these datasets within the OEMC project. Even if all the datasets follow open-access criteria and FAIR principles, some tools for ensuring accessibility and usability in the OEMC project are needed.

This started with gathering info on the available datasets from each of the tasks, investigating their characteristics to understand their degree of readiness in terms of accessibility and FAIRness, and finally selecting the more suitable to be included in the first release. An effort to harmonize their critical info in the same format preceded the insertion of these datasets on the OGH STAC catalog, under the vectorial option: https://stac.openlandmap.org/ . Having all the datasets listed and described in the same tool is crucial in terms of standardization and harmonization, and the starting point to foster accessibility and usability of the datasets provided.

Many of the datasets gathered existed before the start of the OEMC project, or their collection started and is currently in progress. Other relevant in-situ datasets will be added to the list if they will be produced in the future.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059548.

In-situ Data 1st release

A short description of the datasets included in the first release is reported.

Global Ecosystem Dynamics Investigation (GEDI) analysis-ready height measurements

Access link: https://stac.openlandmap.org/l2a.gedi/collection.json

Description: High resolution laser ranging observations of the 3D structure of the Earth. Millions of points with diversity of variables covering latitudes up to 50 degree north

- WP4 task: 4.4 - biodiversity
- dataset documentation: Eegholm, B., Wake, S., Denny, Z., Dogoda, P., Poulios, D., Coyle, B., ... & Blair, B. (2019, August). Global Ecosystem Dynamics Investigation (GEDI) instrument alignment and test. In Optical Modeling and System Alignment (Vol. 11103, pp. 53-70). SPIE.
- dataset duration / period: 2019-2022+
- dataset temporal resolution: regular every year
- dataset spatial scale: global
- link to spatial characteristics table / map: https://lpdaac.usgs.gov/products/gedi02_av002/
- observation Reference Area: Point area
- owner Institution Code: UMD
- database references: https://gedi.umd.edu/data/download/
- connected OEMC use case 1: Global topographic and hydrological service
- connected OEMC use case 2: Global monitoring system for livestock and grasslands / pastures
- data format: vectorial product
- Other data format:
- Metadata attached: Yes, with a standard scheme
- License: CC-BY
- recorded By: Tom Hengl
- email: tom.hengl@opengeohub.org

Ocean Biogeographic Information System (OBIS)

STAC link: https://stac.openlandmap.org/obis/collection.json

Description: OBIS is a global open-access data and information clearing-house on marine biodiversity for science, conservation and sustainable development

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059548.

- WP4 task: 4.5 - ocean
- dataset documentation:
- dataset duration / period: 1876-2022
- dataset temporal resolution: not regular
- dataset spatial scale: global
- link to spatial characteristics table / map: https://obis.org/data/access/
- observation Reference Area: Point area
- ownerInstitutionCode: Unesco
- database references: https://geobon.org/downloads/PDF/OBIS_MBON-whatis_FEB2017.pdf
- connected OEMC use case 1: Scale-dependency of "potential" marine biodiversity distribution patterns a national and European scales
- connected OEMC use case 2: Development of EU-biodiversity monitor
- data format: ASCII file (*.txt, *.csv, etc.)
- Other data format:
- Metadata attached: Yes, with a standard scheme
- License: CC-BY
- recorded By: Francesco De Leo
- email: francesco.deleo@cnr.it

FLUXNET: GreenHouse Gases Fluxes Dataset

Access link: https://stac.openlandmap.org/fluxnet/collection.json

Description: Half-hourly eddy covariance fluxes and their quality flags, when present, in most cases comprehensive of storage fluxes and footprint information, calculated by the station teams and/or ICOS ETC. Most relevant variables in the dataset are: carbon dioxide (CO2) flux, sensible heat flux, H2O molar fraction, latent heat flux, carbon dioxide (CO2) storage flux, eddy covariance momentum flux, friction velocity, net ecosystem exchange, CO2 mixing ratio

- WP4 task: 4.2 - fluxes
- dataset documentation:
- dataset duration / period: from 1 to >25 years, depending on the site
- dataset temporal resolution: 30 minutes
- dataset spatial scale: global
- link to spatial characteristics table / map: https://docs.google.com/spreadsheets/d/1Je3fb4hh0hvPJGDgF5y8aTDfGy4hqvOSlqVyY2PvnLw/edit?usp=sharing; https://drive.google.com/file/d/1EWbsBVvysk9sCD1nDN23ZwPTzcXRecTd/view?usp=share_link
- observation Reference Area: 10-1000 m
- ownerInstitutionCode: ICOS+AmeriFlux+other

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059548.

- database references: https://data.icos-cp.eu/portal/#%7B%22filterCategories%22:%7B%22project%22:%5B%22icos%22%5D,%22level%22:%5B1,2%5D,%22stationclass%22:%5B%22ICOS%22%5D%7D%7D; http://gaia.agraria.unitus.it/home/; https://fluxnet.org/
- connected OEMC use case 1: SIF-based high spatial resolution GPP flux estimations
- connected OEMC use case 2:
- data format: ASCII file (*.txt, *.csv, etc.)
- other data format:
- metadata attached: Yes, with a standard scheme
- license: CC-BY
- recorded By: Simone Sabbatini
- email: simone.sabbatini@cmcc.it

Geo-wiki Drivers of Tropical Forest Loss

Access link: https://stac.openlandmap.org/geowiki.forest.loss/collection.json

Description: Dataset contains 1,150,000 unique locations in the tropics identifying drivers of forest loss (derived from Global Forest Watch map) between 2008 and 2019. Data were collected using Geo-Wiki and is currently hosted by IIASA

- WP4 task: 4.8 - citizen
- dataset documentation: Laso Bayas, J.C., See, L., Georgieva, I. et al. Drivers of tropical forest loss between 2008 and 2019. Sci Data 9, 146 (2022). https://doi.org/10.1038/s41597-022-01227-3
- dataset duration / period: 2008-2019
- dataset temporal resolution: regular every year
- dataset spatial scale: Intercontinental (tropics)
- link to spatial characteristics table / map:
- observation Reference Area: Point area
- owner Institution Code: IIASA
- database references: https://doi.org/10.1038/s41597-022-01227-3
- connected OEMC use case 1:
- connected OEMC use case 2:
- data format: ASCII file (*.txt, *.csv, etc.)
- other data format:
- metadata attached: Yes, with a non standard scheme
- license: CC-BY
- recorded By: Carmelo Bonannella
- email: carmelo.bonannella@opengeohub.org

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059548.

sPlotOpen: An environmentally-balanced, open-access, global dataset of vegetation plots

Access link: https://stac.openlandmap.org/veg.plot/collection.json

Description: Vegetation plots (n = 95,104) recording cover or abundance of naturally co-occurring vascular plant species within delimited areas. sPlotOpen contains three partially overlapping resampled datasets (c. 50,000 plots each), to be used as replicates in global analyses. Besides geographical location, date, plot size, biome, elevation, slope, aspect, vegetation type, naturalness, coverage of various vegetation layers, and source dataset, plot-level data also include community-weighted means and variances of 18 plant functional traits from the TRY Plant Trait Database.

- WP4 task: 4.4 - landscape diversity
- dataset documentation: Sabatini, F. M., Lenoir, J., Hattab, T., Arnst, E. A., Chytrý, M., Dengler, J., ... & Wagner, V. (2021). sPlotOpen–An environmentally balanced, open‐access, global dataset of vegetation plots. Global Ecology and Biogeography, 30(9), 1740-1764.
- dataset duration / period: 1888--2015
- dataset temporal resolution: Not regular
- dataset spatial scale: Global
- link to spatial characteristics table / map: https://onlinelibrary.wiley.com/cms/asset/29811ced-a5c8-4f1e-bfce-702b2f8f7425/geb13346-fig-0001-m.jpg
- observation Reference Area: Point area
- owner Institution Code: iDiv
- database references: https://doi.org/10.25829/idiv.3474-40-3292
- connected OEMC use case 1:
- connected OEMC use case 2:
- data format: ASCII file (*.txt, *.csv, etc.)
- other data format:
- metadata attached: Yes, with a non standard scheme
- license: CC-BY-4.0
- recorded By: Martijn Witjes
- email: martijn.witjes@opengeohub.org

Ground-Based Observations for Validation (GBOV) of the Copernicus Global Land Service

Access link: https://stac.openlandmap.org/gbov/collection.json

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059548.

Description: The GBOV service provides multiple years of high quality in-situ measurements (88 sites) to validate 7 core land products (Top-of-canopy reflectances, Surface albedo, fAPAR, LAI, fCover, Land Surface Temperature and Soil Moisture)

- WP4 task: 4.4 - biodiversity
- dataset documentation: Song, R., Kharbouche, S., & Muller, J. P. (2019, January). Ground-Based Observations for Validation (GBOV) of Copernicus Global Land Products. In Geophysical Research Abstracts (Vol. 21).
- dataset duration / period: 2014–2020
- dataset temporal resolution: not regular
- dataset spatial scale: global
- link to spatial characteristics table / map:
- observation Reference Area: Point area
- ownerInstitutionCode: ACRI-ST
- database references: https://land.copernicus.eu/global/gbov/
- connected OEMC use case 1:
- connected OEMC use case 2:
- data format: ASCII file (*.txt, *.csv, etc.)
- other data format:
- metadata attached: I don't know
- license: I don't know which of these licenses apply to my data
- recorded By: Tom Hengl
- email: tom.hengl@opengeohub.org

Geo-wiki ground observations of land cover

Access link: https://stac.openlandmap.org/geowiki.lc/collection.json

Description: About 50,000 ground observations of land cover / land use which was used to generate global land cover maps and similar.

- WP4 task: 4.6 - LULC
- dataset documentation: Fritz S, See L, Perger C, McCallum I, Schill C, Schepaschenko D, et al. (2017) A global dataset of crowdsourced land cover and land use reference data. Scientific Data. 4:170075. doi:10.1038/sdata.2017.75.
- dataset duration / period: 2010–2020
- dataset temporal resolution: not regular
- dataset spatial scale: global
- link to spatial characteristics table / map:
- observation Reference Area: Point area
- owner Institution Code: IIASA
- database references: https://doi.org/10.1038/sdata.2017.75

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059548.

- connected OEMC use case 1: Land Degradation Neutrality tool
- connected OEMC use case 2:
- data format: ASCII file (*.txt, *.csv, etc.)
- other data format:
- metadata attached: Yes, with a non standard scheme
- license: Restricted access
- recorded By: Tom Hengl
- email: tom.hengl@opengeohub.org

Federated Computing Engine

The OEMC computing engine is responsible for serving the monitor applications, use-cases and building blocks of the project, providing different levels of Earth Observation (EO), in-situ and reference data for a multi-purpose, scalable and computation efficient analysis. At this milestone, several functionalities were implemented by the project partners in the **OpenEO**, **SITSA**, **XCube**, **EO-Learn** and **scikit-map** solutions, enabling for example applications for land potential estimation and biological species distribution mapping. Specifically about the integration with OpenEO, the project is currently working to deploy a stand-one service for SITSA that will be further integrated as OpenEO backend. Additionally, a preliminary design was defined for converting Open-EO response formats into EOPatch, a core object of EO-Learn.

As most of these solutions are inline with I, the project is working to federate them with the support of the Copernicus Data Space Ecosystem (CDSE), aiming to contribute with:

1. Analysis-Ready and Cloud-Optimized (ARCO) data provision (Ex. GLAD Landsat ARD-2 imagery),
2. Customized and reproducible processing pipelines,
3. Feedback on scalability and production functionalities of the ecosystem.

Means of Verification

OpenEO:

- Git repository: https://github.com/Open-EO/openeo-processes-dask
- Video recording: https://youtu.be/EwFaJjaf5bc
- Tutorial(s):
    - https://github.com/Open-EO/openeo-python-client/tree/master/examples/notebooks
    - https://github.com/Open-EO/openeo-python-client/blob/master/docs/machine_learning.rst
    - https://github.com/EO-College/cubes-and-clouds/blob/main/lectures/3.1_data_processing/exercises/31_data_processing.ipynb

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101059548.

SITSA:

- Git repository:
  - SITS: https://github.com/e-sensing/sits
  - Openeosits: https://github.com/Open-Earth-Monitor/openeosits
- Video recording: https://drive.google.com/file/d/1gf-geY5QxWt0tMIyZ-cEtcuLcxIEmiv6 (It will be soon available in TIB AV-Portal)
- Tutorial(s): https://e-sensing.github.io/sitsbook

XCube:

- Git repository: https://github.com/dcs4cop/xcube
- Video recording: https://youtu.be/0b-H96beEQg
- Tutorial(s): https://xcube.readthedocs.io/en/latest/examples

EO-Learn:

- Git repository: https://github.com/sentinel-hub/eo-learn
- Tutorial(s):
  - http://eo-learn.readthedocs.io
  - https://github.com/sentinel-hub/eo-learn-examples

Scikit-map:

- Git repository: https://github.com/openlandmap/scikit-map
- Video recording: https://youtu.be/QorMipL2OG8
- Tutorial(s): https://github.com/openlandmap/scikit-map/blob/main/docs/notebooks