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

OPEN EARTH MONITOR

MS7 Open-Earth-Monitor Computing Engine and In-situ Data 2nd 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	MS7 Open-Earth-Monitor computing engine and in-situ data 2nd 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	Daniel Thiex (Simone Sabbatini and Leandro Parente for the 1st release)
Contributors	OGH, INPE, CMCC, Brocmann Consult GmbH and CMCC
Date of delivery	28 February 2025
Due date of Milestone	30 November 2024
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.

For the second and now updated report (MS7 deliverable), the reporting of in-situ data follows the same structure as in the first report (MS4: *Open-Earth-Monitor Computing Engine and In-Situ Data 1st Release and Usability Assessment*). The metadata is hosted in a newly created STAC catalog specifically designed for in-situ data (https://stac.earthmonitor.org/), as vector/point data has different metadata requirements compared to EO raster data.

At the time of writing, only the initial iteration of datasets has been added, with additional datasets (and potentially more, depending on further use case development) to be incorporated in the coming weeks and months of the project.

Within the federated compute solutions, we continue to build upon the existing and previously listed compute options while also integrating new functionalities and resources developed since the first report.

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.earthmonitor.org/insitu_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

In-situ Data 2nd release

A short description of the datasets included in the 2nd release is reported.

Global Pasture Watch: Grassland classes reference samples

Access link: https://stac.earthmonitor.org/insitu_gpw_grassland.samples/collection.json

Description: Reference samples (established by Feature Space Coverage Sampling-FSCS) comprising 2.3M points visually classified using Very High Resolution imagery (Google Maps and Bing Maps). The ARIS data (including metadata) is hosted in Zenodo.

- WP4
- dataset documentation: Leandro Parente, Lindsey Sloat, Vinicius Mesquita et al. Annual 30-m maps of global grassland class and extent (2000–2022) based on spatiotemporal Machine Learning, 21 October 2024, PREPRINT
- dataset duration / period: 01.04.2004 - 31.12.2022
- dataset temporal resolution: annual
- dataset spatial scale: global
- link to spatial characteristics table / map: https://zenodo.org/records/14035457/files/00-preview.png?download=1
- observation Reference Area: Point data
- owner Institution Code: OGH
- database references: https://doi.org/10.21203/rs.3.rs-4514820/v3
- connected OEMC use case 1: Global EO data to support monitoring livestock and grasslands / pastures
- connected OEMC use case 2: /
- data format: Geopackage (.gpkg)
- other data format: /
- metadata attached: Yes
- license: CC-BY-4.0
- recorded By: Leandro Parente
- Email: leandro.parente@opengeohub.org

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

OEMC Hackathon 2023: EU Land Cover Classification Dataset

Access link: https://stac.earthmonitor.org/insitu_oemc.hackathon_lc/collection.json

Description: Stratified sampling of the ground-truth data provided by LUCAS Survey and overlaid with 416 raster spatial layers. The target land cover considered level-3 classes from the harmonized legend, resulting in 72 classes distributed over 5 years (2006, 2009, 2012, 2015, 2018). Dataset used in OEMC Hackathon 2023. The ARIS data (including metadata) is hosted in Zenodo.

- WP4
- dataset documentation: Parente, L., Witjes, M., & Tomislav, H. (2023). OEMC Hackathon 2023: EU Land Cover Classification Dataset (Version v1) [Data set].
- dataset duration / period: 01.01.2006 - 31.12.2018
- dataset temporal resolution: every 5 years
- dataset spatial scale: Europe
- link to spatial characteristics table / map: /
- observation Reference Area: Point data
- owner Institution Code: OGH
- database references: https://doi.org/10.5281/zenodo.8306554
- connected OEMC use case 1: EO data and automated mapping for EU soil observator
- connected OEMC use case 2: /
- data format: Comma-separated values (.csv)
- other data format: /
- metadata attached: Yes
- license: CC-BY-4.0
- recorded By: Leandro Parente
- Email: leandro.parente@opengeohub.org

OEMC Hackathon 2023: Global FAPAR Modeling Dataset

Access link: https://stac.earthmonitor.org/insitu_oemc.hackathon_fapar/collection.json

Description: Monthly averaged FAPAR (Fraction of Absorbed Photosynthetically Active Radiation) values aggregated by each ground station and overlaid with 32 raster spatial layers. FAPAR represents the fraction of the incoming (photosynthetic active) radiation that is absorbed by vegetation, and is given in the range 0-1. Dataset used in OEMC Hackathon 2023 The ARIS data (including metadata) is hosted in Zenodo.

- WP4
- dataset documentation: Parente, L., Hackländer, J., & Hengl, T. (2024). OEMC Hackathon 2023: Global FAPAR Modeling Dataset (including raster data) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13874505

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

- dataset duration / period: 01.04.2014 - 31.12.2020
- dataset temporal resolution: monthly
- dataset spatial scale: global
- link to spatial characteristics table / map:/
- observation Reference Area: Point data
- owner Institution Code: OGH
- database references: https://doi.org/10.5281/zenodo.13874505
- connected OEMC use case 1: Land Degradation Neutrality tool to support LDN initiative
- connected OEMC use case 2: Biodiversity monitoring and reporting tools for IDH
- data format: Comma-separated values (.csv)
- other data format: Shapefile (.sh)
- metadata attached: Yes
- license: CC-BY-4.0
- recorded By: Leandro Parente
- Email: leandro.parente@opengeohub.org

FAPAR EML reference samples

Access link: https://stac.earthmonitor.org/fapar.eml/collection.json (will be published in the catalog within the next months)

Description: Training points (3,132,000 unique space-time points) generated from a stratified random sampling design. The 12,500 spatial points were overlaid with time series data of 2000 - 2021 of GLASS FAPAR V6 and of the covariate layers used for training the FAPAR ensemble machine learning model.

- WP4
- dataset documentation: Julia Hackländer, Leandro Parente, Yu-Feng Ho et al. Land potential assessment and trend-analysis using 2000–2021 FAPAR monthly time-series at 250 m spatial resolution, 08 October 2023, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-3415685/v1]
- dataset duration / period: 01.04.2000 - 31.12.2021
- dataset temporal resolution: monthly
- dataset spatial scale: global
- link to spatial characteristics table / map: https://s3.eu-central-1.wasabisys.com/stac/prev/veg.plot.jpg
- observation Reference Area: unique space-time points
- owner Institution Code: OGH
- database references: https://doi.org/10.21203/rs.3.rs-3415685/v1
- connected OEMC use case 1: Land Degradation Neutrality tool to support LDN initiative
- connected OEMC use case 2: Biodiversity monitoring and reporting tools for IDH
- data format: Comma-separated values (.csv)

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

- other data format: Shapefile (.sh)
- metadata attached: Yes
- license: CC-BY-4.0
- recorded By: Julia Hackländer
- Email: julia.hacklaender@opengeohub.org

Biomass reference data for global space-based data validation

Access link: https://stac.earthmonitor.org/agbref/collection.json (will be published in the catalog within the next months)

Description: Processed biomass plot data to provide reference database for comparison with space-based biomass estimates. Original data sources are multiple.

- WP4 task: 4.3 - biomass
- dataset documentation: Julia Hackländer, Leandro Parente, Yu-Feng Ho et al. Land potential assessment and trend-analysis using 2000–2021 FAPAR monthly time-series at 250 m spatial resolution, 08 October 2023, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-3415685/v1]
- dataset duration / period: Different years 2000-2020
- dataset temporal resolution: not regular
- dataset spatial scale: regional
- link to spatial characteristics table / map: /
- observation Reference Area: 100-1000 m
- owner Institution Code: ESA CCI biomass project / GFZ
- database references: https://climate.esa.int/en/projects/biomass/
- connected OEMC use case 1: Large-area estimation of forest carbon emissions
- connected OEMC use case 2: Tropical deforestation monitoring and characterisation tool
- data format: ASCII file (*.txt, *.csv, etc.)
- other data format: /
- metadata attached: /
- license: Restricted access
- recorded By: Martin Herold
- Email: herold@gfz-potsdam.de

LIDAR survey of Trento province (Trentino), Italy

Access link: https://stac.earthmonitor.org/agbref/collection.json (will be published in the catalog within the next months)

Description: Processed biomass plot data to provide reference database for comparison with space-based biomass estimates. Original data sources are multiple.

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

- WP4 task: 4.3 - biomass
- dataset documentation: Julia Hackländer, Leandro Parente, Yu-Feng Ho et al. Land potential assessment and trend-analysis using 2000–2021 FAPAR monthly time-series at 250 m spatial resolution, 08 October 2023, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-3415685/v1]
- dataset duration / period: Different years 2000-2020
- dataset temporal resolution: not regular
- dataset spatial scale: regional
- link to spatial characteristics table / map: /
- observation Reference Area: 100-1000 m
- owner Institution Code: ESA CCI biomass project / GFZ
- database references: https://climate.esa.int/en/projects/biomass/
- connected OEMC use case 1: Large-area estimation of forest carbon emissions
- connected OEMC use case 2: Tropical deforestation monitoring and characterisation tool
- data format: ASCII file (*.txt, *.csv, etc.)
- other data format: /
- metadata attached: /
- license: Restricted access
- recorded By: Martin Herold
- Email: herold@gfz-potsdam.de

Intertidal macrospecies

Access link: https://stac.earthmonitor.org/intertidal.macro/collection.json (will be published in the catalog within the next months)

Description: Occurrence and cover data of the main algal, invertebrate and lichen macrospecies typifying intertidal rocky shores surveyed from Scotland to Morocco

- WP4 task: 4.5 - ocean
- dataset documentation: Sinde Mano A L, Monteiro C, Pádua Lima F, Pereira J, Nieto Vilela R A, Seabra R (2024). Biodiversity data from rocky intertidal zones, surveyed from Scotland to Morocco in 2022 and 2023. Version 1.2. CIBIO (Research Center in Biodiversity and Genetic Resources) Portugal. Samplingevent dataset. http://ipt.gbif.pt/ipt/resource?r=intertidal_biodiversity_northatlantic22&v=1.2
- dataset duration / period: 2022-2023 (different years)
- dataset temporal resolution: not regular (monthly)
- dataset spatial scale: Northwestern Atlantic coast, from Scotland to Morocco
- link to spatial characteristics table / map: /
- observation Reference Area: Point area
- owner Institution Code: CIBIO-BIOPOLIS

- database references: http://ipt.gbif.pt/ipt/resource?r=intertidal_biodiversity_northatlantic22&v=1.2
- connected OEMC use case 1: Scale-dependency of "potential" marine biodiversity distribution patterns a national and European scales
- connected OEMC use case 2: Biodiversity monitoring and reporting tool
- data format: Darwin Core Archive
- other data format: /
- metadata attached: Yes (with standard scheme)
- license: CC-BY-4.0
- recorded By: Ana Luísa Sinde Araújo Torres Mano
- Email: alu.sinde.mano@gmail.com

The BioTIME database

Access link: https://stac.earthmonitor.org/biotime/collection.json (will be published in the catalog within the next months)

Description: The database BioTIME is designed especially for scientific synthesis studies with research questions about global biodiversity. We believe that data is valuable and should be made usable. Our database contains tables on species abundances across time and space, as well as important metadata about the taxa, habitat, and sampling methods. This is all made possible by our data contributors.

- WP4 task: 4.4 - biodiversity
- dataset documentation: Sinde Mano A L, Monteiro C, Pádua Lima F, Pereira J, Nieto Vilela R A, Seabra R (2024). Biodiversity data from rocky intertidal zones, surveyed from Scotland to Morocco in 2022 and 2023. Version 1.2. CIBIO (Research Center in Biodiversity and Genetic Resources) Portugal. Samplingevent dataset. http://ipt.gbif.pt/ipt/resource?r=intertidal_biodiversity_northatlantic22&v=1.2
- dataset duration / period: 1874-2016
- dataset temporal resolution: not regular
- dataset spatial scale: global
- link to spatial characteristics table / map: https://biotime.st-andrews.ac.uk
- observation Reference Area: Point area
- owner Institution Code: BioTIME is hosted and based at the University of St Andrews. It grew out of two ERC grants (AdG BioTIME 250189 and PoC BioCHANGE 72744) awarded to Professor Anne Magurran.
- database references: https://onlinelibrary.wiley.com/doi/10.1111/geb.12729
- connected OEMC use case 1: Scale-dependency of "potential" marine biodiversity distribution patterns a national and European scales
- connected OEMC use case 2: /
- data format: ASCII file (*.txt, *.csv, etc.)

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

- other data format: /
- metadata attached: Yes (with standard scheme)
- license: CC-BY
- recorded By: Francesco De Leo
- Email: francesco.deleo@cnr.it

LandPKS crowd-sourced observations

Access link: https://stac.earthmonitor.org/land.pks/collection.json (will be published in the catalog within the next months)

Description: Crowdsourced observations of land cover, soil properties, land degradation; citizen-science based with over 80,000 points as in 2022.

- WP4 task: 4.8 - citizen
- dataset documentation: Herrick, J. E., Urama, K. C., Karl, J. W., Boos, J., Johnson, M. V. V., Shepherd, K. D., … & Kosnik, C. (2013). The Global Land-Potential Knowledge System (LandPKS): Supporting Evidence-based, Site-specific Land Use and Management through Cloud Computing, Mobile Applications, and Crowdsourcing. Journal of Soil and Water Conservation, 68(1), 5A-12A.
- dataset duration / period: 2016-2022+
- dataset temporal resolution: not regular
- dataset spatial scale: global
- link to spatial characteristics table / map: https://storage.googleapis.com/api-docs.landpotential.org/index.html
- observation Reference Area: Point area
- owner Institution Code: USDA-ARS
- database references: https://landpotential.org/data-portal/
- 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 standard scheme)
- license: CC-BY
- recorded By: Tom Hengl
- Email: tom.hengl@opengeohub.org

EEA Air Quality Measurements

Access link: https://stac.earthmonitor.org/air.quality/collection.json (will be published in the catalog within the next months)

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

Description: Hourly measurement values of NO2, SO2, O3, PM10, and PM2.5 from stations reported by EEA member states.

- WP4
- dataset documentation: Sabatini, F.M., Lenoir, J., Bruelheide, H. & the sPlot Consortium (2021) sPlotOpen – An environmentally-balanced, open-access, global dataset of vegetation plots (Version 1.0) [Dataset]. iDiv Data Repository. https://doi.org/10.25829/idiv.3474-40-3292
- dataset duration / period: 2013-2024
- dataset temporal resolution: hourly
- dataset spatial scale: Europe
- 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 standard scheme)
- license: CC-BY-4.0
- recorded By: Martijn Witjes
- Email: martijn.witjes@opengeohub.org

World Cereal Data Module

Access link: https://stac.earthmonitor.org/cereal/collection.json (will be published in the catalog within the next months)

Description: The dataset includes 75M features (polygons/points) including crop types for wheat and maze

- WP4 task: 4.8 citizen
- dataset documentation: Sabatini, F.M., Lenoir, J., Bruelheide, H. & the sPlot Consortium (2021) sPlotOpen – An environmentally-balanced, open-access, global dataset of vegetation plots (Version 1.0) [Dataset]. iDiv Data Repository. https://doi.org/10.25829/idiv.3474-40-3292
- dataset duration / period: 01.01.2016-31.12.2021
- dataset temporal resolution: Not regular
- dataset spatial scale: Global
- link to spatial characteristics table / map: https://worldcereal-rdm.geo-wiki.org/collections/

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

- observation Reference Area: Point area
- owner Institution Code: various organisations (ESA, IIASA, etc.)
- database references: https://worldcereal-rdm.geo-wiki.org/collections/
- connected OEMC use case 1: Crop yield monitoring system for Africa
- connected OEMC use case 2: Tools for Digitalisation of Agriculture in Ethiopia
- data format: ASCII file (*.txt, *.csv, etc.)
- other data format: /
- metadata attached: Yes (with standard scheme)
- license: CC-BY
- recorded By: Steffen Fritz, Milutin Milenkovic
- Email: fritz@iiasa.ac.at, milenkovic@iiasa.ac.at,

Snow cover in the European Alps

Access link: https://stac.earthmonitor.org/alps.snow/collection.json (will be published in the catalog within the next months)

Description: Station observations of snow depth and depth of snowfall

- WP4 task: 4.9 - integration with gridded
- dataset documentation: Matiu, M., Crespi, A., Bertoldi, G., Carmagnola, C. M., Marty, C., Morin, S., Schöner, W., Cat Berro, D., Chiogna, G., De Gregorio, L., Kotlarski, S., Majone, B., Resch, G., Terzago, S., Valt, M., Beozzo, W., Cianfarra, P., Gouttevin, I., Marcolini, G., … Weilguni, V. (2021). Snow cover in the European Alps: Station observations of snow depth and depth of snowfall (v1.3) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5109574
- dataset duration / period: 1971-2021
- dataset temporal resolution: not regular
- dataset spatial scale: Alps
- link to spatial characteristics table / map: https://clirsnow.netlify.app/dash-results/dash-climatology.html#in-situ-snow-depth
- observation Reference Area: Stations
- owner Institution Code: various organisations
- database references: https://zenodo.org/record/5109574#.ZGssfHZByUn
- connected OEMC use case 1: High resolution SWE in selected mountain regions
- connected OEMC use case 2: /
- data format: ASCII file (*.txt, *.csv, etc.)
- other data format: /
- metadata attached: Yes (with standard scheme)
- license: CC-BY-4.0 (if not stated otherwise)
- recorded By: Valentina Premier
- Email: valentina.premier@eurac.edu

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

Drivers of Tropical Forest Loss

Access link: https://stac.earthmonitor.org/trop_forest.loss/collection.json (will be published in the catalog within the next months)

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: Stations
- owner Institution Code: IIASA
- database references: https://doi.org/10.1038/s41597-022-01227-3
- connected OEMC use case 1: Tropical deforestation monitoring and characterisation tool
- connected OEMC use case 2: /
- data format: ASCII file (*.txt, *.csv, etc.)
- other data format: /
- metadata attached: Yes (with standard scheme)
- license: CC-BY
- recorded By: Carmelo Bonannella
- Email: carmelo.bonannella@opengeohub.org

Forest Inventory of Trento province (Trentino), Italy - InfoCarb

Access link: https://stac.earthmonitor.org/info.carb/collection.json (will be published in the catalog within the next months)

Description: Forest Carbon Stocks in the Province of Trento inventory of Trentino based on 150 plots (15 m radius) over 6200 Km2 land area. Estimation of organic carbon pools stored above and belowground.

- WP4 task: 4.3 - biomass
- dataset documentation: Tabacchi G, Gasparini P (2008). L'inventario del carbonio forestale in provincia di Trento: commento sul disegno campionario e i primi risultati. Forest@ 5: 195-200. - doi: 10.3832/efor0530-0050195; Rodeghiero M, Tonolli S, Vescovo L, Gianelle D, Cescatti A, Sottocornola M. INFOCARB: A Regional Scale Forest

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

Carbon Inventory (Provincia Autonoma di Trento, Southern Italian Alps). FOREST ECOLOGY AND MANAGEMENT 259 (6); 2010. doi:10.1016/j.foreco.2009.12.019; Tonolli S. , Salvagni F. (2007). InFoCarb Inventario Forestale del Carbonio della Provincia di Trento, 1-176, Centro ecologia Alpina, Trento.

- dataset duration / period: 2002
- dataset temporal resolution: no regular
- dataset spatial scale: regional
- link to spatial characteristics table / map: /
- observation Reference Area: Plots (radius: 15 m)
- owner Institution Code: FEM - PAT (Provincia Autonoma Trento)
- database references: /
- connected OEMC use case 1: /
- connected OEMC use case 2: /
- data format: ASCII file (*.txt, *.csv, etc.)
- other data format: /
- metadata attached: /
- license: CC BY-NC 3.0
- recorded By: Luca Belelli Marchesini
- Email: luca.belellimarchesini@fmach.it

PATRIS (parameterization of Trentino forest types by rapid forest inventory)

Access link: https://stac.earthmonitor.org/patris/collection.json (will be published in the catalog within the next months)

Description: The dataset consists of relascopic measurements and forest structure parameters (LAI, clumping index, gap fraction by zenith angle range) derived from hemispheric photos over 789 plots in the province of Trento

- WP4 task: 4.3 - biomass
- dataset documentation: Tonolli S., Dalponte M., Neteler M., Rodeghiero M., Vescovo L., Gianelle D., Fusion of airborne LiDAR and satellite multispectral data for the estimation of timber volume in the Southern Alps, Remote Sensing of Environment, Volume 115, Issue 10, 2011,Pages 2486-2498, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2011.05.009.
- dataset duration / period: 2004-2005
- dataset temporal resolution: not regular
- dataset spatial scale: regional
- link to spatial characteristics table / map: /
- observation Reference Area: Plot (relascope)
- owner Institution Code: FEM - PAT (Provincia Autonoma Trento)
- database references: /

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: Large-area estimation of forest carbon emissions
- connected OEMC use case 2: Tools and data for improved biomass estimation
- data format: ASCII file (*.txt, *.csv, etc.)
- other data format: /
- metadata attached: No
- license: CC-BY-NC 3.0
- recorded By: Luca Belelli Marchesini
- Email: luca.belellimarchesini@fmach.it

Zooplankton in BPNS

Access link: https://stac.earthmonitor.org/bpns.zooplankton/collection.json (will be published in the catalog within the next months)

Description: In the framework of the Lifewatch marine observatory a number of fixed stations on the Belgian Part of the North Sea (BPNS) are visited on a monthly or seasonal basis using the RV Simon Stevin. A grid of nine stations covers the coastal zone and are sampled monthly. Eight additional stations, located further at sea, are sampled on a seasonal basis. This dataset contains zooplankton observations in the Belgian Part of the North Sea (BPNS) since 2012. Zooplankton is sampled by vertical WP2 net tows, samples scanned with ZooScanner and identification with plankton analyser software, followed by manual validation.

- WP4 task: 4.5 - ocean
- dataset documentation: Mortelmans J, Goossens J, Amadei Martínez L, Deneudt K, Cattrijsse A, Hernandez F. LifeWatch observatory data: Zooplankton observations in the Belgian part of the North Sea. *Geosci Data J*. 2019; 6: 76–84. https://doi.org/10.1002/gdj3.68
- dataset duration / period: 2015-present
- dataset temporal resolution: /
- dataset spatial scale: Several stations in Belgian part North Sea (BPNS)
- link to spatial characteristics table / map: /
- observation Reference Area: /
- owner Institution Code: FEM - PAT (Provincia Autonoma Trento)
- database references: https://rmets.onlinelibrary.wiley.com/doi/10.1002/gdj3.68
- connected OEMC use case 1: Scale-dependency of "potential" marine biodiversity distribution patterns a national and European scales
- connected OEMC use case 2: /
- data format: Darwin Core Archive
- other data format: /
- metadata attached: Yes (with standard scheme)
- license: CC-BY
- recorded By: Jonas Mortelmans

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

- Email: Jonas.mortelmans@vliz.be

Phytoplankton in BPNS

Access link: https://stac.earthmonitor.org/bpns.phytoplankton/collection.json (will be published in the catalog within the next months)

Description: In the framework of the Lifewatch marine observatory a number of fixed stations in the Belgian Part of the North Sea (BPNS) are visited on a monthly or seasonal basis using RV Simon Stevin. A grid of nine stations covers the coastal zone and are sampled monthly. Eight additional stations, located further at sea, are sampled on a seasonal basis. Samples are taken using a 55µm mesh size Apstein net and fixed in Lugol's iodine solution. In the lab, the samples are processed using a VS-4 FlowCAM model at 4X magnification, size range imaged is 55-300µm. The identification of the image data is done with the use of a classifier and followed by a manual validation step. Since May 2017, this dataset provides micro- and phytoplankton observations, mainly covering diatoms, dinoflagellates and cilliates, for the Belgian Part of the North Sea (BPNS).

- WP4 task: 4.5 - ocean
- dataset documentation: Amadei Martínez L, Mortelmans J, Dillen N, Debusschere E, Deneudt K (2020) LifeWatch observatory data: phytoplankton observations in the Belgian Part of the North Sea. Biodiversity Data Journal 8: e57236. https://doi.org/10.3897/BDJ.8.e57236
- dataset duration / period: 2015-present
- dataset temporal resolution: /
- dataset spatial scale: Several stations in Belgian part North Sea (BPNS)
- link to spatial characteristics table / map: /
- observation Reference Area: /
- owner Institution Code: FEM - PAT (Provincia Autonoma Trento)
- database references: https://bdj.pensoft.net/article/57236/
- connected OEMC use case 1: Scale-dependency of "potential" marine biodiversity distribution patterns a national and European scales
- connected OEMC use case 2: /
- data format: Darwin Core Archive
- other data format: /
- metadata attached: Yes (with standard scheme)
- license: CC-BY
- recorded By: Jonas Mortelmans
- Email: Jonas.mortelmans@vliz.be

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

Global harmonized Land Use and Land Cover Reference Data

Access link: https://stac.earthmonitor.org/bpns.phytoplankton/collection.json (will be published in the catalog within the next months)

Description: In the framework of the Lifewatch marine observatory a number of fixed stations in the Belgian Part of the North Sea (BPNS) are visited on a monthly or seasonal basis using RV Simon Stevin. A grid of nine stations covers the coastal zone and are sampled monthly. Eight additional stations, located further at sea, are sampled on a seasonal basis. Samples are taken using a 55µm mesh size Apstein net and fixed in Lugol's iodine solution. In the lab, the samples are processed using a VS-4 FlowCAM model at 4X magnification, size range imaged is 55-300µm. The identification of the image data is done with the use of a classifier and followed by a manual validation step. Since May 2017, this dataset provides micro- and phytoplankton observations, mainly covering diatoms, dinoflagellates and cilliates, for the Belgian Part of the North Sea (BPNS).

- WP4 task: 4.5 - ocean
- dataset documentation: Amadei Martínez L, Mortelmans J, Dillen N, Debusschere E, Deneudt K (2020) LifeWatch observatory data: phytoplankton observations in the Belgian Part of the North Sea. Biodiversity Data Journal 8: e57236. https://doi.org/10.3897/BDJ.8.e57236
- dataset duration / period: 2015-present
- dataset temporal resolution: /
- dataset spatial scale: Several stations in Belgian part North Sea (BPNS)
- link to spatial characteristics table / map: /
- observation Reference Area: /
- owner Institution Code: FEM - PAT (Provincia Autonoma Trento)
- database references: https://bdj.pensoft.net/article/57236/
- connected OEMC use case 1: Scale-dependency of "potential" marine biodiversity distribution patterns a national and European scales
- connected OEMC use case 2: /
- data format: Darwin Core Archive
- other data format: /
- metadata attached: Yes (with standard scheme)
- license: CC-BY
- recorded By: Jonas Mortelmans
- Email: Jonas.mortelmans@vliz.be

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

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 openEO response formats into EOPatch, a core object of EO-Learn.

In addition to the work on the openEO integration, the relevant new developments in **SITSA** include:

- a) improved support for data access on CDSE;
- b) integration of Sentinel-1, Sentinel-2 and Copernicus DEM-30 into data cubes for multi-source data analytics;
- c) access to the ESA World Cover product;
- d) support for vector data cubes built with spatio-temporal segmentation;
- e) Implementation of the RADD Monitor algorithm to detect new deforestation in Amazonia in support of OEMC use case;
- f) improved tuning methods for deep learning models;
- g) optimization of deep learning performance for classification of large EO data in CUDA and Apple Neural Engine.

Within the **xcube framework**, a compute engine functionality was realized with the development of the xcengine package, which provides a tool for converting scientific use case notebooks into self-contained compute engines while minimizing the effort for the user and boilerplate. The generated compute engines can be operated in two modes: interactively, incorporating an xcube server and a viewer component to provide data visualization and API access; or as OGC EO Application Packages, which provide a standardized interface for integration into OGC-compliant reproducible science workflows. xcengine has already been released as open source software in a usable state (see link below), and is still under active development.

We have released new versions of **eo-learn**, of which version 1.5.0 is the biggest improvement as it contains some important updates to the library. It facilitates the creation of training datasets and ML models by adding support for the zarr data format. This format can be selected as the return data format in the Batch Processing API, and can now be integrated into the ML workflow implemented through eo-learn. A major focus of this release was also to reduce the large number of dependencies, allowing the package to be used in more lightweight docker environments. We also improved support for the ray executor, further strengthening the multiprocessing and scaling capabilities of the package.

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

Subsequent releases focused more on the long-term stability of the package and minor technical updates.

The core functionalities of **scikit-map** were recently re-implemented in C++ and exposed to Python via Pybind, including parallel multi-dimensional numeric operations, raster reading/saving and fast data transposition for optimized ML predictions. These functionalities were instrumental for running time-series reconstructions algorithms on the entire Landsat archive (https://doi.org/10.7717/peerj.18585) and producing global predictions for grassland (https://doi.org/10.1038/s41597-024-04139-6) and Gross Primary Productivity-GPP (https://doi.org/10.21203/rs.3.rs-5587863/v1) at 30-m spatial resolution.

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/3.1_exercises/31_data_processing.ipynb

SITSA:

- Git repository:
  - SITS: https://github.com/e-sensing/sits
  - Openeosits: https://github.com/Open-Earth-Monitor/openeosits
- Video recording: https://av.tib.eu/media/66283
- 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.html
- xcengine git repository: https://github.com/xcube-dev/xcengine

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

EO-Learn:

- Git repository: https://github.com/sentinel-hub/eo-learn
- Zenodo citation: https://zenodo.org/records/13847341
- 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:
    - Build and visualize your own raster data cube: https://youtu.be/QorMipL2OG8
    - Geo-AutoML with Scikit-map: https://youtu.be/NiMiXCzmlQ4
- Tutorial(s):
    - https://github.com/openlandmap/scikit-map/blob/main/docs/notebooks
    - https://gitlab.com/leal.parente/geo-snippets/-/blob/main/data_cubes/OEMC_2024_Monitoring_livestock_and_agricultural_systems.ipynb?ref_type=heads
    - https://gitlab.com/leal.parente/geo-snippets/-/blob/main/modeling/OEMC_2024_Geo_Automl_scikit_map.ipynb?ref_type=heads