Global Earth Monitor

Deliverable 2.6

q

DEMO datasets

The project has received funding from European Union's Horizon 2020 Research and Innovation Programme" under the Grant Agreement 101004112

PREPARATION SLIP			
	Name	**Organisation**	**Date**
Prepared by	Nika Oman Kadunc, Matej Batič	Sinergise	25.08.2022
Reviewed by	Žiga Lukšič	Sinergise	29.08.2022
Approved for submission by	Mariza Pertovt	Sinergise	30.08.2022

EXECUTIVE SUMMARY

This deliverable provides an overview of the data, available within Global Earth Monitor (GEM) for the areas of interest within the project use cases and demonstration activities. Data and data management plays an essential role within GEM and was defined in the first phase of the project to ensure a high level of data quality and accessibility for end users and stakeholders and to enable the application of machine learning techniques.

In general, several types of data are used in GEM. Within this report we categorized the data as Earth Observation (EO) data, weather/climate data, and EO-derived data. We provide an overview of the different data used in the projects use-cases and present a collection of Python notebooks showcasing how to access said data with eo-learn, which is the main building block of the GEM framework.

Keywords:

GEM Data, Data Types, Data Storage, Data Security, Data Publication, Data Sharing

Contractual Delivery Date	31.08.2022
Actual Delivery Date	30.08.2022
Type of delivery	ORDP: Open Research Data Pilot
Dissemination level:	Public

Table of Contents

List of Figures

List of Tables

List of Abbreviations

API	Application Programming Interface
EO	Earth Observation
ESA	European Space Agency
GEM	Global Earth Monitor
ML	Machine learning
MERIS	Medium Resolution Imaging Spectrometer
SatCen	European Union Satellite Centre

1 Introduction

This document is part of the Work Package 2: Data and is dedicated to the description of the data used within Global Earth Monitor (GEM) project use-cases' demonstration purposes. The short document is structured as follows:

Section 2 presents the data used within GEM. It can best be categorized into the following categories: Earth Observation (EO) data, weather/climate data, and other data. These datasets have been (in more detail) reported within the GEM Data Management Plan deliverable (D2.1), nevertheless, the tables Table 1 - Table 6 have been added to this report to make it concise.

Section 3 quickly summarises GEM use-cases, the data used within, and tentative areas of interest for demonstration purposes if available at this time in the project.

In **Section 4** we provide a collection of Python notebooks showcasing how to access data described within this report using eo-learn, the main building block of the GEM framework.

2 GEM data

The data used within GEM project can best be categorized into the following categories:

- Earth Observation (EO) data, provided by Sentinel-Hub services.
- Weather/climate data, provided by meteoblue services.
- Other data (e.g., various datasets like World Cover, Corine, etc.), access provided by Sentinel-Hub services.

These datasets have been in detail reported in the GEM Data Management Plan deliverable (D2.1). The following section just summarises the data and gives a quick overview with links to more details.

EO data, presented in Table 1, are accessible through the Sentinel-Hub services. The table gives information about spatio-temporal availability, and revisit times. Sentinel-Hub services are provided by Sinergise. eo-learn and eo-grow, basis of the GEM platform, presented in D3.3, are tightly interconnected with the service, as it is the main gateway to the EO data.

Weather Data presents a unique approach to combining multiple weather models with machine learning algorithms to provide the best possible weather forecast. Available weather/climate data is delineated in tables Table 2 - Table 5. The data is accessed through API, provided by meteoblue. To facilitate research and innovation using both weather and EO data, eo-learn gateway to the services have been added.

Other data provides valuable information, especially for the intended use cases. This uncategorized data consists of a plethora of inputs that can be used in specific use cases, e.g., for benchmarking of the methods and algorithms, as ground truth/reference data, etc. The raster datasets, available through Sentinel-Hub services are listed in Table 6.

Table 1: EO data

Data collection	Availability - Spatial	Availability - Temporal	Revisit
Sentinel-2 L1C	Whole world	November 2015 →	5 days
Sentinel-2 L2A	Whole world	January 2017 →	3-5 days
Sentinel -1	Whole world	For eo-cloud: October 2014 → For services: January 2017 →	12 days
Sentinel-3 OLCI L1B	Land and coastal areas where solar zenith angle <80 degrees	April 2016 →	< 2 days
Sentinel-3 SLSTR L1B	Land and coastal areas where solar zenith angle <80 degrees	May 2016 →	< 0.9 days
Sentinel 5P L2	Whole world	April 2018 →	daily

Landsat 1-5 MSS Collection 2 Level 1 Data	Global land	LS1: July 1972 → January 1978 LS2: January 1975 →February 1982 LS3: March 1978 → March 1983 LS4: July 1982 → December 1993 LS 5: 1984 → October 1992, and from June 2012 → January 2013 July 1972 → October 1992, June 2012 → January 2013	various (16 – 18 days)
Landsat 4-5 TM Collection 2 Level 1 Data	Global land	LS4: July 1982 → December 1993 LS5: March 1984 → May 2012	16 days
Landsat 4-5 TM Collection 2 Level 2 Data	Global land	LS4: July 1982 → December 1993 LS5: March 1984 → May 2012	16 days
Landsat 7 ETM+ Collection 2 Level 1 Data	Global land	April 1999 →	16 days
Landsat 7 ETM+ Collection 2 Level 2 Data	Global land	April 1999 →	16 days
Landsat 8-9 OLI-TIRS Collection 2 Level 1 Data	Land	February 2013 →	16 days
Landsat 8-9 OLI-TIRS Collection 2 Level 2 Data	Land	February 2013 →	16 days
Envisat Meris	Whole world	from June 2002 to April 2012	3 days
Digital Elevation Model (DEM)	Whole world	Static	/
Copernicus DEM 90	Whole world	Static	/
Copernicus DEM 30	Whole world	Static	/
MODIS	Whole world	24. February 2000 →	daily
Commercial Data – PlanetScope	Whole world	2009 →	daily
Commercial Data – Pleiades	Whole world	December 2011 →	on-demand acquisitions
Commercial Data – Spot	Whole world	September 2012 →	on-demand acquisitions
Commercial Data – WorldView (+GeoEye)	Whole world	2009 →	on-demand acquisitions, archive

Table 2: Weather forecast models

Model	Region	Availability	Resolution	Forecast	Updates	Source
NEMS4	Central Europe	since 2008	4 km, 1h	72h	2 / day	meteoblue
NEMS12	Europe	since 2008	12 km, 1h	168h	2 / day	meteoblue
NEMS2-12	Europe	since 2017	12 km, 1h	168h	2 / day	meteoblue
NEMS-8	Central America	since 2016	12 km, 1h	144h	2 / day	meteoblue
NEMS12	India	since 2014	12 km, 1h	168h	2 / day	meteoblue
NEMS10	South America	since 2015	10 km, 1h	168h	2 / day	meteoblue
NEMS10	South Africa	since 2016	10 km, 1h	144h	2 / day	meteoblue
NEMS8	New Zealand	since 2014	8 km, 1h	144h	2 / day	meteoblue
NEMS8	Japan East Asia	since 2015	8 km, 1h	144h	2 / day	meteoblue
NEMS30	Global	since 1985	30 km, 1h	168h	2 / day	meteoblue
NMM4	Central Europe	since 2008	4 km, 1h	72h	2 / day	meteoblue
NMM12	Europe	since 2009	12 km, 1h	168h	2 / day	meteoblue
NMM18	South America	since 2008	18 km, 1h	144h	2 / day	meteoblue
NMM18	South Africa	since 2009	18 km, 1h	144h	2 / day	meteoblue
NMM18	Southeast Asia	since 2010	18 km, 1h	144h	2 / day	meteoblue
ICON	Global	since 2016	13 km, 3h	180h	1 / day	DWD
ICON7	Europe	since 2017	7 km, 3h	120h	1 / day	DWD
GFS05	Global	since 2008	40 km, 3h	180h	1 / day	NOAA
NAM3	North America	since 2020	3 km, 1h	60h	1 / day	NOAA
NAM12	North America	since 2009	12 km, 3h	84h	1 / day	NOAA
MFGLOBAL	Global	since 2016	40 km, 3h	96h	1 / day	METEO FRANCE
MFEU	Europe	since 2016	11 km, 1h	96h	1 / day	METEO FRANCE
MFFR	France	since 2016	2 km, 1h	36h	1 / day	METEO FRANCE
HIRLAM11	Global	since 2016	11 km, 1h	48h	1 / day	KNMI
HIRLAM7	Europe	since 2018	7 km, 1h	54h	1 / day	Finnish meteorologica l institute
GEM	Global	since 2016	25 km, 1h		1 / day	Environment Canada
GEM15	Global	since 2020	15 km, 3h	168h	1 / day	Environment Canada
GEM2	North America	since 2020	2.5 km, 1h	48h	1 / day	Environment Canada

UMGLOBAL	Global	since 2019	17 km, 3h	144h	1 / day	UK MetOffice

Table 3: Seasonal anomalies forecast models

Model	Model official name	Source
SA-ENSEMBLE		meteoblue
SA-ECMWF	SEAS-5	ECMWF
SA-JMA	CPS-2	JMA
SA-UKMO	HadGEM3-GC2.0	UK MetOffice
SA-METEOFR	System7-v20190301	METEO FRANCE
SA-DWD	GCFS2.1-v20200320	DWD
SA-NCEP	CFSv2-v20110310	NOAA
SA-CMCC	CMCC-CM2-v20191201 / SPS3.5	CMCC
CFS	NCEP CFSv2	NOAA

Table 4: Weather re-analysis models

Model	Region	Availability	Resolution	Updates	Source
ERA5	Global	since 1979	30 km, 1h	monthly	ECMWF
BOM	Australia	since 1901	15 km, 1d	1 / day	BOM

Table 5: Weather measurements data, accessible through meteoblue

Model	Region	Availability	Resolution	Updates
AWEKAS	Germany	February 2020	931 stations	10 min
DWD Climate Open Data Server 10-minutes	Germany		1986 stations, 10 min	up to an hour
DWD Climate Open Data Server daily	Germany		1986 stations, daily	daily
DWD Climate Open Data Server hourly	Germany		1986 stations, 1 hour	daily
DWD points of interests	Germany	January 2020 →	1056 stations	hourly
HydroData CH	Water Bodies in Switzerland	September 2020 →	85 stations	half-hourly
MADIS	Global	January 2005 →	irregular	
Met Office UK	UK	September 2020 →	140 stations	hourly

Metar	Global		irregular	
Meteorological Service of Canada (MSC)	Canada	April 2020 →	982 stations	15 min
MeteoSchweiz	Switzerland	April 2020 →	159 stations	10 min
Pessl-Global	Global		irregular	
Synop				
WMO GSOD	Global	January 1980 →	daily	
WMO ISD	Global	January 1980 →	hourly	

Table 6: Various EO-derived data collections, accessible through Sentinel Hub

Data collection	Availability - Spatial	Availability – Temporal	Revisit
Sentinel-2 L2A 120m Mosaic	Land surface area between 58°S and 72°N	2019, 2020	15 days[1]
Corine Land Cover	Pan-European, French overseas regions and departments	1990, 2000, 2006, 2012, 2018	every 6 years
Corine Land Cover Accounting Layers	EEA39 region	2000, 2006, 2012, 2018	every 6 years
ESA WorldCover	Global land	2020	yearly
Global Land Cover	Global land	2020	yearly
Global Surface Water	Global coverage from longitude 170°E to 180°W and latitude 80°N to 50°S	1984 → 2019, 1984 → 2020	yearly
Global Human Settlements Layer	Global coverage with longitude from 180°W to 180°E and latitude from 72°N to 56°S	2018	static
Sea Ice Index	Longitude from 180°W to 180°E and latitude from 39.23°N to 90°N and 30.98°S to 90°S	2017 → May 2021	none (demo)

1 Data interpolated to regular spatio-temporal grid (static ARD DC)

Water Bodies	Global coverage from longitude -180°E to +180°W and latitude +80°N to -60°S. Depending on the month, some high latitude areas are not available	October 2020 →	monthly
Seasonal Trajectories, 10-daily	Europe	January 2017 →	yearly
Vegetation Indices, daily	Europe	October 2016 → February 2021	daily
Vegetation Phenology and Productivity Parameters Season 1, yearly	Europe	January 2017 →	yearly
Vegetation Phenology and Productivity Parameters Season 2, yearly	Europe	January 2017 →	yearly
Theia Land Cover Map	France	2016-2020	yearly

3 GEM use-cases

Summary of the GEM use-cases, together with the data used within and tentative areas of focus are given in this session. As the several use-cases are still in the phase of development, the details are not final and might change. Please note also that the work done in GEM use-cases will be finalized with the two concluding deliverables, D5.7 – Demonstrator and D5.8 – Validation report.

3.1 Built-up area use-case

For the built-up use case, a process was developed, which can (on an ongoing basis) produce information about new built-up areas on Sentinel-2 120 m mosaic. The drill-down mechanism can then be used to run further models on higher and higher resolution, up to very high resolution (VHR) data for identification of individual buildings.

In the experimentation phase we focused on the area of Slovenia and France, where high quality reference data provided the means to understand the issues of using low resolution inputs. As the process relies on eo-grow framework for scalability we have succeeded producing the results for the whole of Africa.

The built-up use-case relies mostly on EO derived product (Sentinel-2 120 m global cloudless mosaic data cube), and several other datasets (Google Open Building dataset, Global Human Settlements Layer[2], and others).

More details about Built-up area use-case can be found in the deliverable D5.2.

3.2 LC-CMS use-case

LC continuous monitoring service (LC CMS) using eo-learn and eo-grow as the basis of GEM framework will be used to perform a baseline land cover classification on a 6-12 months temporal interval will be developed and deployed in the pilot regions, as well as a subsequent continuous monitoring scheme to assess landscape evolution, based on the combination of likely land cover change trajectories and change detection from short satellite imagery time-series.

The use-case relies on openly available data (Sentinels, LandSat) as input, and on both open and proprietary training/validation datasets.

3.3 Map-making use-case

The main idea of the Map Making use case is that on top of creating "map-ready" land-cover features to update and visually enhance TomTom's map, the extraction performed at 10m resolution, even though insufficient to produce features at the highest zoom levels, can still be leveraged to provide leads at locations where existing map features are likely to have changed. Upon detection of such phenomena, those leads could be provided to editors to update the product accordingly through the employment of GEM's drill-down capabilities. One of the main, and immediate, outcomes in this use case is to enhance TomTom's core water feature set.

[2] https://ghsl.jrc.ec.europa.eu/index.php

The map-making use-case mostly relies on Sentinel-2 data, EO derived products like Global Surface Water[3] and Global Water Bodies[4] datasets, and (private) training/validation dataset collected by TomTom. The processing chain relies on eo-learn tasks and on eo-grow framework for scalability, with specialized pipelines to deal with post-processing tasks for transforming raster data into "map-ready" features that can be directly ingested in TomTom's cartographic master database.

Map-making use-case is still work in progress, details about the use-case will be delivered in D5.4.

3.4 Conflict pre-warning use-case

The Conflict Pre-Warning Map (CPW) aims to merge multiple data sources available in GEM. Geographic data and other open sources of information (e.g., distribution of ethnicities or religion) will provide a map or a report to support political decision-making. A new semi-automated CPW service of extreme flexibility will be developed as a combination of CPW map, and a dedicated decision support system based on weighted variables associated with the input data. Different geospatial data will contribute to CPW, either derived from existing products or generated in the framework of the GEM project:

- Geospatial data (e.g., from the existing Copernicus services, as vegetation indexes series, water bodies), retrieved through SentinelHub services.
- Meteorological data (e.g., precipitation, air temperature, forecasts), retrieved through meteoblue weather api services.
- Data from LC CMS use-case and their subsequent products will be used to detect and analyse potential disruptive changes in land cover.
- Open data (e.g., distribution of ethnicities or religion).
- Very High Resolution (VHR) data.

CPW will directly benefit from the input of GEM products. The LC CMS products and the correlation between different climatic/environmental/thematic variables will be explored to define their incidence in generating a conflict. The datasets will be included in SatCen's GIS system to create a decision support model to define the risk of conflicts (which might be associated to scarcity of resources, natural disasters or impact of drastic changes associated to climate). Alerting capabilities will be employed to generate Warning reports at regional scale when situations match those identified as indicative for upcoming violent outburst. CPW serves as a crown demonstration of higher-level cross-correlated ML processing for decision making support, going beyond typical "one-problem ML challenges".

The CPW will be nominally run on Sahel region. More details will be delivered in D5.5.

3.5 Crop identification use- case

Crop identification use-case relies on a combination of EO and weather data to enable automatic identification of crops. The use case supports operational decisions when managing crops and the monitoring of actual vs. planned or reported agricultural land use (e.g., Common Agricultural Policy monitoring).

Satellite data and weather data come at very different temporal and spatial resolutions: Sentinel-2 constellation nominally provides an observation of a field every 5 days at 10 m spatial resolution, while

[3] https://collections.sentinel-hub.com/global-surface-water/
[4] https://collections.sentinel-hub.com/water-bodies/

weather data has continuous hourly time series at multi-km spatial resolution. Ad-hoc routines to spatially aggregate satellite data at field level and to systematically compose layers of different time discretization series were developed within the use-case, so that each EO is associated with a complete time series (of opportune length) of weather variables at daily resolution. For each field, we extract the time series of the median over field pixels of Sentinel-2 L1C bands, cloud mask and cloud probability using Sentinel Hub's Statistical API. Using meteoblue dataset API, complete time series of daily weather data (NEMS4 model) are then associated to each field observation, following the systematic layer composition approach mentioned above.

For the training data for crop classification the use-case relies on publicly available farmers' claims, e.g., claims for Slovenia, years 2018-2020. Given the encouraging scores so far[5], we aim to perform crop type mapping at least at European scale, thanks to the availability of the EuroCrops data and the cost-effective big data solutions developed during GEM project. The final results will be described in deliverable D5.6.

.

[5] https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLIII-B3-2022/1301/2022/

4 Access to data

To follow best data sharing policies of Open Research Data Pilot, we are providing a series of Python notebooks with code snippets on how data presented in Section 2 can be accessed using eo-learn, main building block of GEM framework.

Notebooks are available at

https://github.com/sentinel-hub/eo-learn-examples/blob/main/GEM-data/introduction.ipynb

and are organised based on the data classification from Section 2. An example for eo-learn task to retrieve data for Sentinel-2 L2A data is shown in Figure 1. For the brevity of this report, we have not included other tasks; they are in any case available on the open repository.

2. Sentinel-2 L2A data collection (Sentinel-2 L2A)

```
s2_l2a_data = SentinelHubInputTask(
    data_collection=DataCollection.SENTINEL2_L2A,
    bands=['B01','B02','B03','B04','B05','B06','B07','B08','B8A','B09','B11','B12'],
    bands_feature=(FeatureType.DATA, 'L2A_data'),
    additional_data=[(FeatureType.MASK, 'dataMask')],
    resolution=10,
    maxcc=0.8,
    time_difference=datetime.timedelta(hours=2)
)
```

```
eopatch_s2_l2a = s2_l2a_data.execute(bbox=roi_bbox, time_interval=["2022-07-03","2022-07-05"])
```

```
plt.figure(figsize=(10,10))
plt.imshow(np.clip(eopatch_s2_l2a.data['L2A_data'][0][...,[3,2,1]] * 2.5, 0, 1), vmin=0, vmax=1);
plt.axis(False);
```

Figure 1: eo-learn task for accessing Sentinel-2 L2A data.

5 Conclusion

This document provides an overview of the data used in GEM use-cases. To follow the principles of research reproducibility, we also provided eo-learn tasks to retrieve the data. Together with eo-grow framework the tasks can be scaled up to create analysis read data cubes over areas of practically any size.

Similarly as shown in this report, the results of the GEM use-cases will be made available through eo-learn where possible, either via Sentinel-Hub Bring Your Own Data capabilities, or via dedicated eo-learn tasks for importing different types of data. As at the time of writing this report, the demonstration capabilities of GEM project are not yet in place, the interim results will be shared on GEM platform.