DELIVERY NOTE FOR DATA PRODUCTS AND DATA PRODUCTS DESCRIPTION

EARTH OBSERVATION FOR HIGH IMPACT MULTI-HAZARDS SCIENCE (EO4MULTIHA)

Prepared by:	EO4Multihazards Team
Approved by:	Carlos Domenech
Authorized by:	Carlos Domenech

Code:	EO4MULTIHA-GMV-DeliveryNote_D3.3&D3.4
Version:	1.1
Date:	17/10/2024
Internal code:	GMV 25621/24 V2/24

1. INTRODUCTION

Deliverable items D3.3 and D3.4 are associated with WP300, *Advancing Fundamental Scientific Understanding of Multi-Hazards*, and were produced by the Science Case contributors: UC-ITC, CMCC, EURAC, and BGS.

1.1. PURPOSE

The purpose of this report is to serve as the delivery note for the online-hosted deliverables: Data Products (D3.3) and Data Products Description (D3.4). It provides a comprehensive account of the dataset repository where the results from the Science Cases developed under the EO4Multihazards project are stored.

1.2. SCOPE

This report covers the datasets produced within WP300 and details the hosting locations on the Zenodo platform, which is managed by CERN (European Organization for Nuclear Research) as part of the European OpenAIRE project.

2. ZENODO

The EO4Multihazards project team has decided to use [Zenodo]() to store the results of the four Sciences Cases being developed in this activity.

Zenodo is a general-purpose open repository where researchers can deposit and share their research outputs, such as:

- Research papers
- Data sets
- Research software
- Reports
- Other digital artefacts

It's a free and open-access platform that helps researchers make their work more discoverable, citable, and reusable. Zenodo provides a persistent digital object identifier (DOI) for each submission, making it easier to reference and track the work.

Key benefits of using Zenodo:

- Open Access: All content is freely accessible.
- Preservation: Zenodo ensures long-term preservation of research outputs.
- Citability: Each submission receives a DOI for easy citation.
- Discoverability: Zenodo makes research outputs easily searchable.
- Versatility: It can accommodate a wide range of research materials.

The project team has created a public community in Zenodo called EO4Multihazards (https://zenodo.org/communities/eo4multihazards) where the Science Case researches can upload their results and provide them with a DOI for citation.

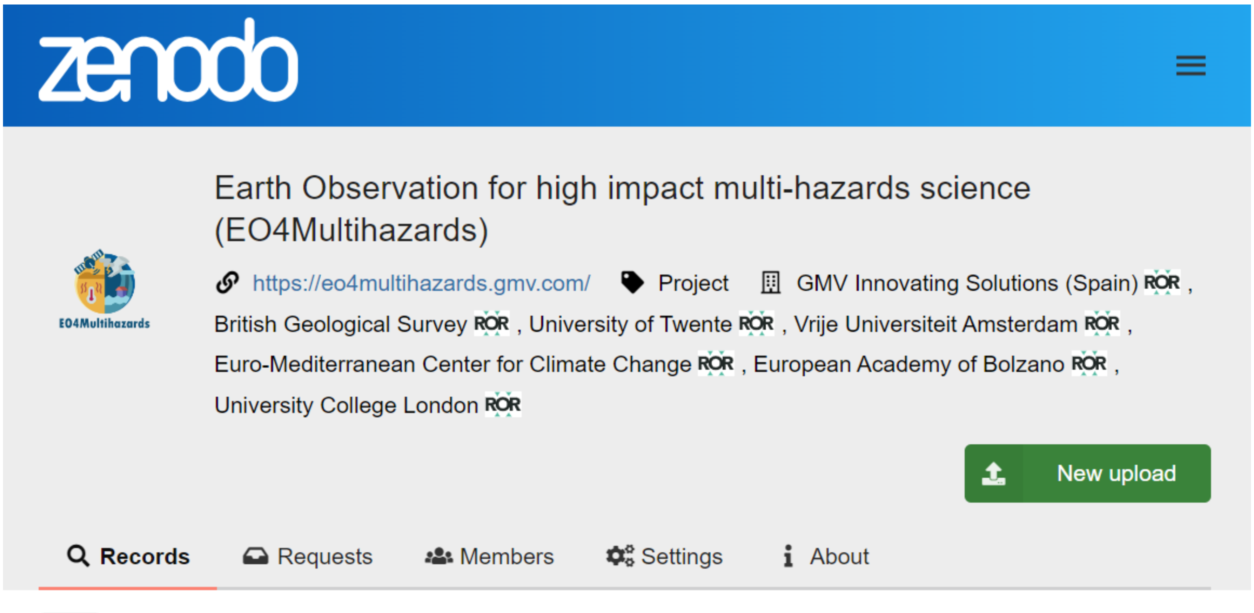

Figure 1 EO4Multihazards community in Zenodo

3. DATASETS UPLOADED

At the time of writing, each Science Case is at a different stage of development. The Science Cases will continue to evolve over the coming months, and the materials currently uploaded to Zenodo represent preliminary versions of the results.

The authorship, descriptions, and metadata of the resulting products are detailed in the records available in the Zenodo repository.

A. SCIENCE CASE 1: HOT-DRY COMPOUND EVENTS IN THE UPPER ADIGE RIVER BASIN (ITALIAN ALPS)

Analysis for the upper Adige River Basin on Compound Drought and Heatwave (CDHW) indicators has developed a dedicated repository. The Compound Drought and Heatwave (CDHW) indicators dataset includes two netCDF files: one indicating the occurrence/absence of events (CDHW_occurrence_extended_Alps_daily_1950_2023.nc, hereafter File A) and the other representing the severity of those events (CDHW_severity_extended_Alps_daily_1950_2023.nc, hereafter File B). A CDHW event is determined by the co-occurrence of drought and heatwave conditions over at least 60% of the Adige catchment area. The indicators were derived using the E-OBS 0.1ºx0.1° daily gridded dataset (v29.0e), covering the period from 1950 to 2023, and the spatial domain [7.10, 44.10, 15.30, 49.10] (min longitude, min latitude, max longitude, max latitude in WGS84, EPSG:4326).

A drought period is identified as a sequence of consecutive months with a negative Standardized Precipitation Index (SPI), starting with the first month where the SPI-6 (6-month timescale) falls below -1. Heatwaves are defined as periods of at least three consecutive days where the daily maximum temperature (TX) exceeds the 90th percentile for that specific calendar day, determined using a 31-day running mean centred on the day under evaluation and considering all values from 1950 to 2023. When two or more periods of consecutive exceedances are separated by one day with TX below the threshold, they are considered as a single heatwave occurrence and the day below the threshold is included in the event duration.

In File A, for each day in a CDHW event the grid cells where both drought and heatwave conditions are detected are flagged as "1". If the compound condition is not met, the cell is flagged as "0". In File B, the daily CDHW severity (dimensionless) is calculated for each day in the compound event as the product of the standardized daily TX over the days of the event and the absolute value of the SPI in the corresponding month. The calculation of the CDHW severity is like the one proposed by Mukherjee and Mishra (2021), but with the percentiles used in the standardization of TX varying with the day of the year.

Lastly, the list of CDHW events affecting the Adige catchment between 1950 and 2023 is provided in a csv file, including the start and end dates, the percentage of the area affected, and the total severity of the event (dimensionless). The total severity of the event is defined as the average of the CDHW severities of all grid cells in the Adige catchment experiencing the CDHW conditions. The total severity of the CDHW event at each grid cell is calculated as the sum of the daily severities during the event by considering only the days flagged as "1" in File A.

The dataset is accessible at the URL https://zenodo.org/records/13839123

Another dataset regards the wildfire predictions. This dataset provides hindcast of daily dynamic wildfire probabilities for the period from 01-07-2022 to 15-07-2022. The predictions illustrate the critical conditions where wildfires are more likely to occur based on static, dynamic, and seasonal controls. Static predictors statistically significant, and therefore considered in the analysis, are landcover, tree density, topographic light, distance to buildings/roads. Dynamic predictors are mean annual precipitation, mean annual temperature and day of the year, and have been combined dynamically to find the optimal time window to describe the wildfire occurrence i.e., the temperature on the observed day and the cumulative precipitation of 30 days before observation. Direct anthropogenic factors are not considered in the analysis.

The dataset is accessible at the URL https://zenodo.org/records/13865655

B. SCIENCE CASE 2: MULTI-HAZARDS IN THE DOWNSTREAM AREA OF THE ADIGE RIVER BASIN

A repository contains the daily SPEI (Standardized Precipitation Evapotranspiration Index) calculated using the E-OBS daily gridded data with a spatial resolution of 0.1° for the period from 1950 to 2023. E-OBS is a land-only European dataset based on observational data from stations of the European National Meteorological and Hydrological Services (NMHSs) or other data holding institutions (https://cds.climate.copernicus.eu/datasets/insitu-gridded-observations-europe?tab=overview). SPEI described the anomalies of the balance between precipitation and potential evapotranspiration with respect to the average conditions (Vicente-Serrano et al., 2010), considering both temperature and precipitation as inputs. It is commonly used to assess drought, e.g. in agriculture, as it reflects both precipitation deficits and the increased water demand of crops (represented by the potential evapotranspiration, ET, in the index computation). Daily SPEI is here calculated at a time scale of 90 days (3 months). The SPEI calculation is based on the 'SPEI' R package by Beguería and Vicente-Serrano (2023) (https://github.com/sbegueria/SPEI) with the potential evapotranspiration derived from Hargreaves-Samani formulation. Based on E-OBS orography, SPEI values for altitudes above 1500 m a.s.l. are excluded from the current analysis to focus on low and mid-elevation areas only.

SPEI therefore represents a comprehensive index to identify the drought hazard (dry event), while also considering temperature in the computation. Moreover, it can foster the use of Earth Observation when coupled with event identification methods by providing a robust meteorological-based event identification. This can be then further refined using higher spatial resolution EO data capable of capturing subtle spatial patterns (e.g., soil moisture as influenced by droughts).

The dataset is provided over the [7.10, 44.10, 15.30, 49.10] spatial domain (min longitude, min latitude, max longitude, max latitude in WGS84, EPSG:4326).

The dataset is accessible at the URL https://zenodo.org/records/13778103

Another repository contains the DBSCAN 3D clusters from SPEI-90 days values for Italian NUTS3 ITH31, 32, 34, 35, 36, 37 from 1950 to 2023. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm output (csv format) based on the daily Standardized Precipitation Evapotranspiration Index (SPEI) with a 90 days' timescale for elevation lower than 1500 m a.s.l. (https://doi.org/10.5281/zenodo.13778103) applying the threshold SPEI-90 days ≤ -1.

The SPEI is calculated using the E-OBS daily gridded dataset with a spatial resolution of 0.1° for the period from 1950 to 2023, as the difference between precipitation and potential evapotranspiration (Vicente-Serrano et al., 2010). E-OBS is a land-only European dataset based on observational data from stations of the European National Meteorological and Hydrological Services (NMHSs) or other data holding institutions (https://cds.climate.copernicus.eu/datasets/insitu-gridded-observations-europe?tab=overview). The SPEI considers both temperature and precipitation as inputs, and is commonly used to assess agricultural drought, as it reflects both precipitation deficits and the increased water demand of crops (represented by the potential evapotranspiration, ET, in the index computation). It therefore represents a comprehensive index to identify the drought hazard (dry event), while considering also temperature. A timescale of 90 days is used to detect the seasonal drought conditions. The SPEI index ranges from 3 to -3, with negative values representing precipitation deficit conditions. Here, the threshold value of -1 was selected to capture a wide severity range of drought events, from moderate to extreme.

The DBSCAN algorithm included in the scikit-learn package in Python environment (https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html) was used to detect the spatial-temporal clusters of droughts, taking the SPEI index as input. Three parameters guide the DBSCAN clustering procedure: the neighbourhood parameter (ε), which defines the search radius around a point (a SPEI index value); the spatial-temporal ratio (r), which controls the importance of spatial distance relative to temporal lag when computing the Euclidean distance between data points; the density threshold parameter (μ), representing the minimum number of neighbouring SPEI pixels required for a point to be considered as a core point (a point representing a suitable point to generate a new drought cluster).

The selected parameter values are: neighbourhood parameter (ε) = 30, spatial-temporal ratio (r) = 3 and density threshold (μ) = 10. These parameters were selected based on their physical significance and then refined through the comparison with the drought historical events retrieved from newspapers, official technical reports and regional council resolutions.

Both the SPEI index and its adoption as an input in the DBSCAN algorithm used to identify drought spatial-temporal clusters represent a key step to identify the spatial-temporal footprints of hazard events. The identification of spatial patterns enables a greater understanding of hazard dynamics, it can be coupled with other hazard footprints (e.g., heatwaves) and fosters the use of EO data by providing a robust meteorological-based event identification to be then further refined by the use of higher spatial resolution EO data capable of capturing subtle spatial patterns (e.g., soil moisture as influenced by droughts, or land surface temperature as a response to different land uses during hot events).

The spatial domain of the dataset is represented by the Italian Provinces identified by the NUTS3 codes ITH31, ITH32, ITH34, ITH35, ITH36, ITH37 considering elevation lower than 1500 m a.s.l. The dataset covers the period from 1950 to 2023. The following dataset is the raw output of DBSCAN.

The dataset is accessible at the URL https://zenodo.org/records/13785996

C. SCIENCE CASE 3: MULTI-HAZARDS IN SOUTHEAST UK REGION

For the UK Science Case, the trends analysis was based on calculating relative changes between relevant temporal periods and did not result in the production of any new datasets.

The EO-derived climatic and environmental parameters used for pattern recognition analysis are raw datasets or existing products sourced from Copernicus Land Service Products, NASA, or via Google Earth Engine (ERA-5 Land ECMWF reanalysis dataset). The initial threshold values obtained based on the trends analysis for specific hazard events do not constitute new datasets per se but can be used in the production of new datasets (see sections 5.5 and 5.7.2 in D3.1v0).

D. SCIENCE CASE 4: MULTI-HAZARDS IN CARIBBEAN SIDS

The Science Case in the Caribbean region presents records on landslides, precipitation, maps used as inputs of hazard models and drone imagery over the region of interest. For the Carribean study-case, an analysis of open and proprietary satellite-based dataset was used to facilitate the setup and evaluation of physically based multi-hazard models. These allow for qualification and quantification of spatial-temporal multi-hazard patterns. These form a crucial input into the general hazard and risk assessment workflow.

Landslide inventories are included, which are part of the observational data. These are mostly based on high-resolution satellite imagery. Mapping these, and finally distributing them as shapefiles, allows for calibration and validation of the models. See also https://reliefweb.int/map/dominica/dominica-landslides-and-floods-triggered-huricane-maria-18-september-2017

An example of a landslide on Dominica after Hurricane Maria (2017) observed by commercial satellite data (MAXAR).

The precipitation data from NASA Global Precipitation Mission (GPM) and the Maps for model input are a collection of geotiff rasterized dataset that contain physical parameters concerning surface and subsurface landscape. These include soil properties (texture, e.g. clay, silt, sand), parameters (saturated hydraulic conductivity of the soil), land use (classified), elevation (height of the landscape using space and local based sources.

A collection of datasets, including soil, land use, and channel data, SOILGRIDS, and SPOT image classification, all prepared for model input; this dataset is intended for calibrating and validating flood and landslide modelling

The stakeholder questionnaires particularly relating to the tools developed partly by this project on rapid hazard modelling. Stakeholder Engagement survey and Stakeholder Survey Results prepared and implemented by Sruthie Rajendran as part of her MSc Thesis Twin Framework For Decision Support In Flood Risk Management supervised by Dr M.N. Koeva (Mila) and Dr B. van den Bout (Bastian) submitted in July 2024

Finally, the drone images were captured using a DJI drone of part of the Study area in February 2024. The file comprises three different regions: Coulibistrie, Pichelin and Point Michel. The 3D models for Coulibistrie were generated from the nadir drone images using photogrammetric techniques employed by the software Pix4D. The image Coordinate System is WGS 84 (EGM 96 Geoid0), but the Output Coordinate System of the 3D model is WGS 84 / UTM zone 20N (EGM 96 Geoid). The other two folders contain only the drone images captured for the Pichelin and Point Michel regions. The dataset is used with other datasets to prepare and create the digital twin framework tailored for flood risk management in the study area.

The dataset is accessible at the URL https://zenodo.org/records/13834495

END OF DOCUMENT