DELIVERY NOTE FOR DATA PRODUCTS AND DATA PRODUCTS DESCRIPTION

EARTH OBSERVATION FOR HIGH IMPACT MULTI-HAZARDS SCIENCE (EO4MULTIHAZARDS)

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

Code:	EO4MULTIHA-GMV-DeliveryNote_D3.3&D3.4
Version:	2.0
Date:	19/05/2025
Internal code:	GMV 25621/24 V2/25

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 2-1).

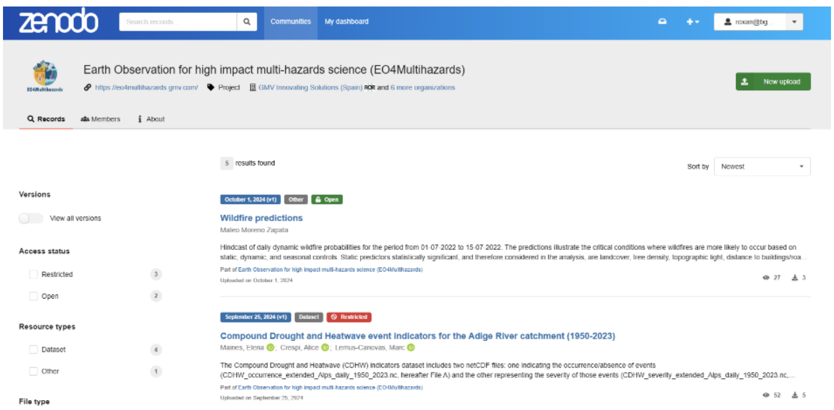

Figure 2-1: EO4Multihazards community in Zenodo.

To enhance the accessibility of datasets produced in the project Science Cases via Zenodo, both consistent data description structure and dataset names are used. Consistently and concisely describing datasets across different data providers is crucial for ensuring clarity, interoperability, and ease of use for the research community. Table 2-1 indicates the structure used in Zenodo.

Table 2-1: Science Cases datasets description.

Field name	Example description
Science Case Name	Name and focus area of the Science Case (e.g., *Multi-Hazards in the Downstream Area of the Adige River Basin*).
Dataset Name/Title	Name or descriptive title of the dataset (e.g., *DBSCAN 3D Clusters of SPEI-90 days Values – Italian NUTS3 (ITH31, 32, 34, 35, 36, 37), 1950–2023*).
Dataset Description	Detailed description of the dataset, including its source, contents, purpose, resolution, etc. (e.g., *Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm output based on the daily Standardized Precipitation Evapotranspiration Index (SPEI) with a 90-day timescale for elevation lower than 1500 m a.s.l. applying the threshold SPEI-90 days ≤ -1.*)
Key Methodologies	Description of the main methodologies used in generating the dataset. (e.g., *The DBSCAN algorithm included in the scikit-learn package in Python environment was used to detect the spatio-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 spatio-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, spatio-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*)
Temporal Domain	Time range covered by the dataset (e.g., *1950–2023*).
Spatial Domain	Geographic area covered by the dataset, including spatial resolution and coordinate system if applicable (e.g., *Italian Provinces identified by the NUTS3 codes ITH31, ITH32, ITH34, ITH35, ITH36, ITH37 considering elevation lower than 1500 m a.s.l.*).
Key Variables/ Indicators	Key metrics or variables measured or represented in the dataset (e.g., *Spatio-temporal clusters of dry/drought events*).
Data Format	Format of the dataset files (e.g., *CSV, netCDF, GeoTIFF, shapefiles*).
Source Data	Input datasets or sources used for analysis (e.g., *E-OBS dataset, Copernicus Land Service Products, NASA GPM Precipitation Data, SPOT Imagery Classification*).
Accessibility	Hosting platform and access details, including URL and DOI (e.g., *Zenodo, https://zenodo.org/records/13785996*).
Stakeholder Relevance	Potential use cases or applications of the dataset for stakeholders (e.g., *Both the SPEI index and its adoption as an input in the DBSCAN algorithm used to identify drought spatio-temporal clusters represent a key step to identify the spatio-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).
Limitations/ Assumptions	Key limitations, assumptions, or exclusions related to the dataset (e.g., *Excludes areas above 1500m a.s.l. due to limited data reliability*).
Additional Outputs/ Information	Other related data, models, or deliverables linked to the dataset (e.g., *Daily SPEI-90 Days Values - Elevation Below 1500 m a.s.l.*).
Contact Information	Name and affiliation of the dataset owners (e.g., *Maraschini, Margherita (CMCC Foundation - Euro-Mediterranean Center on Climate Change*)

Please ensure that information regarding restrictions, identifiers, related work, funding, creation date, software used, and associated references are captured separately within the Zenodo structure, rather than being included in the metadata table above.

The official funding acknowledgment, which should be copied into the "Additional Details > Funding" section on the Zenodo page, is as follows:

"EO4MULTIHAZARDS (Earth Observation for High-Impact Multi-Hazards Science), funded by the European Space Agency and launched as part of the joint ESA-European Commission Earth System Science Initiative."

The datasets are publicly available to users registered on Zenodo. However, access may be temporarily restricted if the data is being prepared for publication. After the publication, the datasets will be made accessible for download.

The datasets should be curated by the owner organization following the provided guidelines. The structured description should be reviewed for accuracy, including formatting and consistency. GMV and BGS will conduct an internal review prior to GMV authorizing the dataset for online publication. Dataset owners are encouraged to provide as much relevant information as possible in the metadata to give users a clear and comprehensive understanding of the dataset's content, context, and intended use.

3. DATASETS UPLOADED

The Science Cases are currently finalised. The materials currently uploaded to Zenodo are the final versions of the data or data products in each Science Case.

The authorship, descriptions, and metadata of the resulting products are detailed in the records available in the Zenodo repository. A short description of existing datasets and products developed in each Science Case together with the associated Zenodo link and reference is listed below.

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/14859795

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 provides daily Standardized Precipitation Evapotranspiration Index (SPEI) values computed using the E-OBS gridded dataset, with a spatial resolution of 0.1°, for the period 1950–2023. E-OBS is a land-only dataset for Europe based on station observations from European National Meteorological and Hydrological Services (NMHSs) and other institutions (Copernicus CDS link).

SPEI quantifies anomalies in the climatic water balance (precipitation minus potential evapotranspiration), accounting for both temperature and precipitation inputs (Vicente-Serrano et al., 2010). It is widely used for drought assessment, especially in agriculture, as it reflects both reduced precipitation and increased evapotranspiration driven by higher temperatures. In this dataset, SPEI is calculated at a 90-day timescale (seasonal scale) using the Hargreaves-Samani formulation for PET, implemented via the 'SPEI' R package (Beguería & Vicente-Serrano, 2023; GitHub link). Values above 1500 m a.s.l. are excluded to focus on low and mid-elevation areas, based on E-OBS orography.

SPEI serves as a robust meteorological indicator for drought events and can enhance Earth Observation-based analysis by providing a reliable basis for event identification. These events can then be refined using high-resolution EO datasets (e.g., for soil moisture or land surface temperature analysis).

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

A complementary dataset includes DBSCAN-based spatio-temporal clusters of drought events derived from the same SPEI-90 dataset, limited to Italian NUTS3 provinces ITH10, ITH20, ITH31, ITH32, ITH33, ITH34, ITH35, ITH36, and ITH37 (below 1500 m a.s.l.), over the period 1981–2023 period. The clustering is based on days where SPEI-90 ≤ -1, capturing moderate to extreme drought conditions. The SPEI index ranges from +3 (wet) to -3 (dry), with -1 commonly used as a threshold for drought onset.

The DBSCAN algorithm from the Python scikit-learn library (link) was used, with the following parameters:

- ε (neighbourhood radius): 20
- r (spatial-temporal ratio): 4
- μ (minimum points per cluster): 20

These values were chosen for their physical significance and refined by comparing results with documented historical drought events (e.g., newspapers, technical reports, regional council resolutions).

The dataset is accessible at the URL https://zenodo.org/records/15212462

A third dataset contains DBSCAN-based clusters of hot weather events, also over the Italian NUTS3 provinces ITH31–37. In this case, clustering is based on daily maximum temperatures (Tmax) exceeding the 90th percentile of the 1991–2020 daily climatology for at least three consecutive days. The 90th percentile was calculated using a centred 15-day window around each calendar day.

DBSCAN was again applied with the following parameters:

- ε (neighbourhood radius): 20
- r (spatial-temporal ratio): 10
- μ (minimum points per cluster): 10

These parameters were selected through an iterative process informed by historical heatwave records (from newspapers, regional bulletins, and technical documentation).

The dataset is accessible at the URL https://zenodo.org/records/15212520

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 Monitoring Land Service, NASA, or Copernicus Climate Change Service (ERA-5 Land 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 D3.1 v2).

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

The Science Case in the Caribbean region presents records on landslides and landslide susceptibility maps, floods used as inputs of hazard models, and drone imagery over the region of interest used as input for 3D models. A landslide susceptibility assessment was generated for the current situation (2025). As the basis for the landslide susceptibility assessment a data layer was generated that represents homogeneous terrain units. Initially an attempt was made to generate these automatically, using the r.slopeunits application, which generates Terrain Units using a set of parameters from a Digital Elevation Model. To compute the susceptibility, we opted for a terrain unit partition and for the implementation of statistical models, in combination with expert-based mapping. These models learn from past events (and specifically from past landslide occurrences) to find patterns with respect to a set of predisposing factors. On the basis of these patterns a prediction is then made on the expected unstable locations in the future. To assess the susceptibility, we used a statistical model known as binomial Generalized Linear Model or Logistic Regression. The map includes the following classes: Very low, Low, Moderate, High, and Very high.

The dataset is accessible at the URL https://zenodo.org/records/15182981

Flood depth maps were made for all watersheds in Dominica for return periods of 5, 10, 20, 50 and 100 years and the flood extend maps for Tropical Storm Erika and Hurricane Maria. The flood maps were generated using the FastFlood tool (http://fastflood.org) which makes use of global data sets (such as SoilGrids, WorldCover, ERA5, etc.).

The dataset is accessible at the URL https://zenodo.org/records/15180875

3D models for Pichelin and Coulibistrie were generated from nadir drone images using photogrammetric techniques employed by the software Pix4D. The raw drone images are not available for privacy rules. The generated 3D models are available for two areas:

The Pichelin dataset is accessible at the URL https://zenodo.org/records/15353099

The Coulibistrie dataset is accessible at the URL https://zenodo.org/records/15224568

Results of the stakeholder questionnaire focused specifically on the tools developed, in part, in the use case of Multi-hazards in the Caribbean SIDS for rapid hazard modelling were also provided.

The dataset is accessible at the URL https://zenodo.org/records/15212825

E. SCIENCE CASE 5: MULTI-HAZARDS IN SENEGAL

The Science Case in Senegal presents records on droughts, heatwaves and fires. These datasets are currently available to project partners but restricted to the public pending the publication of the results.

The first dataset contains gridded data on SPEI-90 days over Linguere area of Senegal. Droughts were computed with SPEI-90, with daily precipitation form CHIRPS (1981-2023) and daily maximum, average and minimum air temperature from ERA5-Land. Potential evapotranspiration (PET) was computed with the Hargreaves equation from the SPEI R package. Water balance, the difference between precipitation and PET, was aggregated to 90-day rolling sums and Z-scores were computed from distributions of values from day of year (43 points). Days with Z-scores below or equal to –1 were marked as droughts.

Spatio-temporal DBSCAN was conducted with Python packages st_dbscan https://github.com/eren-ck/st_dbscan. Spatial proximity (epsilon 1) was set to 0.5 (0.5 degree), temporal proximity (epsilon 2) was set to 1.5 (1.5 days); min number of samples was set to 30.

The dataset is called DBSCAN 3D Clusters of SPEI-90 days – Linguere, Senegal, 1981-2023 and it is accessible at the URL https://zenodo.org/records/15212446

Another dataset contains gridded data on heatwaves over Linguere area of Senegal. Heatwaves were defined as at least 3-day-long events on which daily maximum air temperature exceeded the above the 95th percentile of the 15-day moving average daily maximum air temperature. Daily maximum air temperature at 2 meters was downloaded from ERA5-Land dataset.

Spatio-temporal DBSCAN was conducted with Python packages st_dbscan https://github.com/eren-ck/st_dbscan. Spatial proximity (epsilon 1) was set to 0.5 (0.5 degree), temporal proximity (epsilon 2) was set to 1.5 (1.5 days); min number of samples was set to 30.

The dataset is called DBSCAN 3D Clusters of Heatwaves – Linguere, Senegal, 1950-2023 and it is accessible at the URL https://zenodo.org/records/15212719

A third dataset contains gridded data on co-occurrence of droughts and heatwaves (CDHW) over the whole area of Senegal. Droughts were defined as events with Z-scores of SPI-6 below or equal to -1. 6-month standardized precipitation index (SPI-6) was computed from CHIRPS monthly precipitation data.

Heatwaves were defined as at least 3-day-long events on which daily maximum air temperature exceeded the above the 90th percentile of the 31-day moving average daily maximum air temperature. Two heatwave events separated by only one day were merged into a single event. Daily maximum air temperature at 2 meters was downloaded from ERA5-Land dataset.

CDHW is a spatial and temporal intersection of droughts and heatwaves. The uploaded dataset has values 1 in cell with CDHW and 0 otherwise.

The dataset is called Compound Drought and Heatwave (CDHW) Events – Senegal, 1981-2023 and it is accessible at the URL https://zenodo.org/records/15211948

The final dataset contains gridded data on Fire Probability Index (FPI) over Tambacounda area of Senegal. FPI is computed from NDVI, relative humidity and air temperature

The dataset is called Fire Probability Index (FPI) – Tambacounda, Senegal, 1999-2024 and it is accessible at the URL https://zenodo.org/records/15212768

END OF DOCUMENT