DELIVERY NOTE FOR DATABASE ALGORITHMS AND SERVICES, HOSTED ONLINE

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

Prepared by:	Andrea Vianello, Bartolomeo Ventura
Approved by:	Carlos Domenech
Authorized by:	Carlos Domenech

Code:	EO4MULTIHA-GMV-DeliveryNote_D2.2&D2.3
Version:	1.1
Date:	10/10/2024
Internal code:	GMV 25537/24 V2/24

1. INTRODUCTION

The WP200 is being developed with the primary objective of creating a multi-hazards event database. As per SoW, the database's main goals are the following:

- Gather and harmonize an extensive suite of satellite, in-situ data, campaign data, relevant statistics for the four case studies (from existing national, regional, and local platforms), existing climatology, model results.

- Develop an attribute-based schema to track and store different climate-driven events related to the multi-hazard topic.

- Generate a relational database as a basis to perform algorithms development, validation activities and scientific analyses required in the project.

- Develop a project database publicly available and open for uploads to authorized contributors.

These tasks entail the development of code for the web application and scripts to automate data collection and harmonisation, with the objective of setting up the Multi-Hazards Event database.

The database is intended to provide information about events and links to available datasets for the study of specific events of interest. Furthermore, the availability of event information allows the initiation of studies on event connections.

The SoW included an additional objective: Provide metadata in a standard format to describe all the collected datasets and Digital Object Identifier for citation. This objective is not considered because we cannot redistribute collected datasets. We were granted with permission only to visualize them. Instead, we provide links to the provider's data portal for further details.

1.1. PURPOSE

The purpose of this report is to provide a detailed account of the code repositories and services developed for the EO4Multihazards project, specifically focusing on the deliverable items hosted online: Events Database v.01 (D2.2) and Database Services v.01 (D2.3). It serves as a delivery note, documenting how the code and scripts produced within WP200 are stored, maintained, and made accessible for use, at this point of time, by project partners. The report also highlights the processes and platforms used to host, manage, and visualize the project's event-related data.

1.2. SCOPE

This report covers the methods and locations for hosting the code and scripts generated within WP200 on the GitLab platform managed by Eurac Research. It details the repository structure for algorithms that automate the population of the EO4Multihazards Database with event data, leveraging the EMDAT API, and managing data updates through a CI/CD process. Additionally, the report addresses the development of a Django-based web application for visualizing event information. The scope includes both the event database's core functionalities and the user interface for interacting with event data, while discussing access restrictions and future plans for public reproducibility of the code.

2. CODE REPOSITORY

This chapter provides information about the hosting of algorithms produced to fill the EO4Multihazards Database with events information, and the code produced for the visualization service developed in the WP200.

For the hosting of these deliverables, we use the GitLab repository platform managed by Eurac Research available at this URL: https://gitlab.inf.unibz.it

The Database is the core of the EO4Multihazards web portal to provide visualization of events and related information, collecting datasets from different trusted repositories.

In the Eurac GitLab repository, we release project-related code, either publicly or with restricted access for authorized users. For the EO4Multihazards project, we have opted to keep the code restricted to project partners. However, after removing any sensitive information that could compromise credentials, a public version of the repository will be made available on GitHub to support reproducibility once the project is completed.

Partners and ESA reviewers can sign up to the Eurac GitLab repository at the URL https://gitlab.inf.unibz.it/users/sign_in?redirect_to_referer=yes and ask the WP200 leader for read permission. Please follow these instructions:

External users can sign up with their social media or google account. After signing up they will receive an email with instructions on how to enable their account from noreply@unibz.it. Please check your spam folder.

2.1. ALGORITHMS REPOSITORY

This repository contains the scripts written in the framework of the WP200. The repository is configured to automatize the download of the latest data available in the EMDAT database, and also the last fire events from the EFFIS database.

The automatization is managed via the CI/CD process, which is scheduled at a precise time of the day, and lets the code download new data, harmonise them, and upload them to the database. The data are then visible via a web app: cssprocapi01.eurac.edu. At the state of the art of this document, this web app is still in beta version.

Furthermore, it is worthwhile noting that we developed the script using a beta version of the EMDAT API, and an old WFS service provided by EFFIS portal. Therefore, any future updates of these services can affect the correct execution of our pipeline. For this reason, further adjustments of the pipeline scripts must be planned.

System maintenance should be guarantee by additional funding (to be discussed) for: DB and web application hosting on cloud or Eurac network, monitoring, security system updates, and pipeline updates in case of external changes.

The URL of the GitLab EO4Multihazards repository is the following:

https://gitlab.inf.unibz.it/CSS-DEV/projects/eo4multiha/eo4multiha_db

2.2. SERVICE REPOSITORY

This repository contains the source code of the web application developed to provide the visualization service for the data contained in the EO4Multihazards database.

The application is accessible at the URL https://cssprocapi01.eurac.edu/ and allows users to interact with content to filter events of interest and visualize all related information we collected.

This URL is provisional and will be replaced by the one agreed upon the client in the next months.

The application has been developed using the Django framework. It allows the web application to run queries on the database and visualize results on a web browser.

Due to the Data policies of the harvested events information, we can provide download service only for the dataset about useful dataset to study events.

The application will share the following datasets:

- Events information and location
- List of useful datasets related to the events
- List of events related to a specific event

The portal will provide link to full metadata in ISO compliant format. Additionally, these metadata will be automatically harvested by the GEOSS catalogue (https://www.geoportal.org) to reach a larger audience.

On the contrary, for the useful datasets listed in the portal, we generally are not the providers and responsible, and for this reason we cannot provide metadata but only link to the official portal or metadata, when available.

URL of the repository is the following:

https://gitlab.inf.unibz.it/CSS-DEV/projects/eo4multiha/eo4multiha_dev/-/tree/main/eo4multiha_dev?ref_type=heads

The application code can be released as open source to collaborate with partners for further development and to allow reproducibility.

3. FINAL CONSIDERATION

Next year, the EO4MULTIHAZARDS system will be finalised and operational. It can be considered as a useful application for the study of multiple hazards in specific areas, suggesting filtered useful datasets, mainly from remote sensing platforms, available for the selected area to further investigate the relationship between events, their impacts and their antecedent state. This portal will be a prototype to be scaled up in the near future, integrating additional datasets and providing new functionalities on the portal.

The existing databases that we are querying to download event information are not huge databases, but when you combine them, they are. The pipeline for download automatically datasets do not have any problem if the number of databases or their size increase, because they are studied to download only the new records.

Of course, there are challenges that need to be considered when scaling the system. It can affect the usability of the interface, slow down queries or affect the visualisation of event features on the main map due to their overlap. This challenge is partially solved in the main map on the home page, which pre-counts the number of vents in each NUTS polygon and provides filtering tools for the events.

The more data sets there are, the more time it takes to run queries, but this increases the interest in the portal itself. Another challenge to keep in consideration, is to guarantee there are no duplicates of events, when collecting different data sources.

Further improvements will be made in the next year of the project to speed up the browsing of the portal and its functionalities. The final report will also include possible ideas for scaling the system and further improvements.

END OF DOCUMENT