



Highlander

High performance computing
to support smart land services



Highlander

High performance computing
to support smart land services

D3.1 Catalogue of background and foreground data suitable for the Open Data Portal

| | |
|----------------------|-------------|
| Deliverable Lead | Arpae |
| Deliverable due date | 2021/31/03 |
| Status | DRAFT/FINAL |
| Version | V1.0 |



DOCUMENT CONTROL PAGE

| | |
|--------------------|--|
| Title | D3.1 Catalogue of background and foreground data suitable for the Open Data Portal |
| Creator | Arpae |
| Publisher | Highlander Consortium |
| Contributors | Davide Cesari, Cinzia Caroli |
| Type | Report |
| Language | EN |
| Rights | Copyright "Highlander Consortium" |
| Audience | Public |
| | Restricted |
| Requested deadline | M12- postponed to M18 based on the request made in the Contingency plan |

Index

| | | |
|-----|---|----|
| 1 | Executive Summary..... | 4 |
| 2 | Introduction | 5 |
| 3 | Background datasets..... | 6 |
| 4 | Foreground datasets | 7 |
| 5 | Hosting and publishing the data | 8 |
| 5.1 | Background datasets | 8 |
| 5.2 | Foreground datasets | 8 |
| 6 | Critical issues..... | 9 |
| 6.1 | Datasets with limitations of use..... | 9 |
| 6.2 | Legal definitions in ECMWF agreements | 10 |
| | TECHNICAL ANNEX..... | 12 |

1 Executive Summary

This document contains a review of the background and foreground datasets involved in the Highlander project, examines their mutual interaction, and gives guidelines on how to make these datasets available to the community and the project's outcome DapOS'es reusable by the end users.

2 Introduction

Datasets in input to the DApOS have been divided into background and foreground datasets.

The background datasets can further be divided into datasets generated by a project partner prior to the project or under other funding, and datasets provided by third-party organisations.

The foreground datasets are generated by the partners during the Highlander project and under Highlander funding.

The output of each DApOS is, in general, considered a dataset as well. These are not treated explicitly in this document.

3 Background datasets

Tree Talker Fire, Animal Talker, Tree Talker: these are background datasets produced by project partners, and they are available under the CC BY 4.0 license.

Satellite images SENTINEL 2: these are external datasets owned and distributed by the Copernicus organisation; they are distributed under an open data license that is compatible with the CC BY 4.0 license.

High-resolution aerial data, LIDAR aerial data: these are external datasets owned by Provincia di Trento, they are available under older versions of the CC BY licenses; this is not ideal from the point of view of the protection of data-owner rights, but it does not impair the usability and the openness of the data themselves and of derived works.

Climatic requirements of native and invasive forage species in pastures: these are data owned by the University of Torino; at the time of writing, CIA obtained to make them freely available to the Highlander partners for the project purposes.

DON and AFLATOXIN values on common wheat and corn: these data, owned by CIA, are freely available only to the Highlander partners for the project purposes. Due to privacy issues, they are averaged over municipality areas to not identify single farmers.

4 Foreground datasets

Here we will examine the foreground datasets which are input to the DApOS'es

VHR-REA_IT: this dataset, produced by CMCC in Highlander's framework, contains a climatic model simulation based on ECMWF ERA5 reanalysis, downscaled at a grid step of 2.2 km; the data are not open but freely available to everybody for research purposes and to Highlander partners for all the project purposes.

VHR-PRO_IT: this is similar to the previous dataset but based on different boundary conditions; the terms of use are the same.

Statistical downscaling of ERA5: this dataset, produced by ECMWF within the project, is also based on the ECMWF ERA5 reanalysis; it is open with a CC BY 4.0 license.

Downscaled sub-seasonal forecasts: ECMWF also produces this dataset, but, since it is based on ECMWF forecasts, it requires data available only under the ECMWF proprietary license for real-time forecast data and, although it does not allow the reconstruction of the original datum, it may undergo restrictions for public use due to restrictions of the input dataset; according to the license terms, the data become free and open (CC BY 4.0) only after the expiration of the forecast verification time.

5 Hosting and publishing the data

5.1 Background datasets

The background datasets **Tree Talker Fire**, **Animal Talker**, **Tree Talker** have an open data license and they are produced by project partners. They may be updated in (near) real-time with new data, so it would be reasonable to store them and make them available to the public, besides to the DApOS'es, on the project data portal.

Other background data **Satellite images SENTINEL 2**, **High-resolution aerial data**, **LIDAR aerial data** are produced and hosted by third-party entities and publicly available with an open data license and with standard web access, so it would be not appropriate to mirror them in the project portal if not internally for technical reasons.

Finally, the background datasets **Climatic requirements of native and invasive forage species in pastures** and **DON and AFLATOXIN values on common wheat and corn** are only available to the project partners, and they are required only in the DApOS'es where the owner (CIA) is involved; so it is probably not strictly necessary to make them available in the data portal unless requested by other DApOS'es.

5.2 Foreground datasets

Coming to the foreground datasets, **VHR-REA_IT**, **VHR-PRO_IT** and **Statistical downscaling of ERA5**, are suitable for hosting and publication on the data portal, with the additional consideration that, for the first two datasets, an informative page will have to be shown to the user for acceptance, specifying that the data are going to be used for research purposes only. The dataset **Downscaled sub-seasonal forecasts** is central to the project because it is the primary prognostic dataset that will be updated in the future with new data. It feeds most of the DApOS'es, so it is crucial to host it in the data portal. However, due to licensing issues connected to the input required for this processing, it is not clear whether this dataset can be publicly distributed with an arbitrary licence or it can only be made available to the Highlander partners to generate project results (DApOS'es).

6 Critical issues

In terms of data policy, the DApOS'es can be represented as a process generating a derived work of the input data. Thus, willing to keep the result of the DApOS'es open and reusable, possibly also for commercial purposes, the rules for derived works have to be applied to derive the terms of use of a DApOS output.

According to previous experience, in this respect, the Creative Commons CC BY 4.0 public licence is one of the most suitable because the restrictions that it imposes do not impair the free usability of the result, unlike, for example, licenses with a "Share alike" clause as CC BY-SA and Odbl, which may limit the applicability of the result or prohibit the creation of derived works if data having different licenses belonging to these categories are combined.

However, some of the input datasets are not open for various reasons so that these particular cases will be examined here in detail.

6.1 Datasets with limitations of use

The following datasets are declared to be freely available only to Highlander partners and freely available to other partners only for research purposes:

VHR-REA_IT, VHR-PRO_IT, Climatic requirements of native and invasive forage species in pastures, DON and AFLATOXIN values on common wheat and corn.

This means that if the DApOS'es making use of them are willing to make their output publicly available and usable for commercial purposes, the algorithms used in the DApOS'es must make sure that the original datasets or relevant parts of them cannot be retrieved from the DApOS result.

The dataset **Downscaled sub-seasonal forecasts** is based on real-time ECMWF seasonal forecasts, which are widely available to the public only under a costly agreement. During the project, ECMWF grants the use of these data for free for the sole project purposes under an ad-hoc research agreement (for a "Research or educational user"). However, suppose the relevant DApOS'es should continue producing and updating their output based on these real-time data even after the end of the project, the project must take care to either purchase these data in the future as a "Service Provider/Broadcaster" or obtain them from the Italian National Meteorological Service (which is the official distributor of ECMWF products in Italy) free of charge under a "NHMS Non-commercial agreement". In the absence of these input

data, the Downscaled sub-seasonal forecasts cannot be computed in real-time, and the dataset will not be timely available to the DApOS'es.

The **Downscaled sub-seasonal forecasts** dataset is defined as “**Non-retrievable Value Added Service**”, according to the terminology used in ECMWF agreements. Consequently, DApOS results based on these data also fall under the same category, thus they can be used only for research purposes during the project time, and their future (free) availability to the public also for commercial purposes may depend on the type of agreement which will rule the provision of these data. So, in case data will be obtained under a costly agreement as a Broadcaster, it will be possible to distribute the project result also for commercial purposes, however the costs involved by this solution could be unacceptably high; conversely, in the case that the Highlander project output is recognised as strategic for the national development, the data may be obtained at no cost under a NHMS Non-commercial agreement from the Italian National Meteorological Service, however, in this case, there may be limitations in the use of the output for commercial purposes.

6.2 Legal definitions in ECMWF agreements

The relevant definitions in the ECMWF standard agreements are the following:

Non-retrievable Value Added Services: Any VAS from which the original Products cannot be retrieved or reverse engineered without significant technical effort and/or expense.

Value Added Services (VAS): all meteorological services which are:

- derived from one or more Products; and
- specifically conceived for the needs of users; and
- made available under specific licence conditions.

Broadcaster/Publisher: Those users who use Products from the ECMWF Catalogue or Value Added Services in order to provide a service by means of Broadcasting or publishing in any form.

NMHS Non-Commercial: A National Meteorological and Hydrological Service that receives the Products for fulfilling national governmental obligations that are related to the protection of life and property and carried out on a non-commercial basis.

Service Provider: Those users who acquire Products from the ECMWF Catalogue in order to supply Value Added Services under specific licence conditions to a third party clearly identified and known to the Service Provider.

If the Licensee is defined [...] as a **Service Provider/Broadcaster**,

the Licensee is authorised to use on a non-exclusive basis the Products detailed in the Schedule of Products:

- for its own internal purposes;
- to Distribute Retrievable and Non-retrievable Value Added Services to End Users;
- to Broadcast Non-retrievable Value Added Services, subject to the payment of the Additional Internet Broadcast Fee when disseminating Graphics Based on ECMWF Products via the Internet;
- to Distribute Non-retrievable Value Added Services to Broadcasters solely for the purpose of Broadcasting, subject to the payment of the Additional Internet Broadcast Fee by the Licensee when the Broadcaster disseminates Graphics Based on ECMWF Products via the Internet;

If the Licensee is defined [...] as a **Research or Educational User**, the Licensee is authorised to use on a non-exclusive basis the Products detailed in the Schedule of Products only for the Research Project or Educational Use described in the Scope of Use.

If the Licensee is defined [...] as a **NMHS Non-Commercial**, the Licensee is authorised to use on a non-exclusive basis the Products detailed in the Schedule of Products:

- for fulfilling national governmental obligations related to the protection of life and property, and
- for Research Projects and Educational Use,

provided that any such activity is carried out strictly on a non-commercial basis. This licence does not grant any rights for the commercial exploitation of the Products. Should the Licensee wish to exploit Products provided under this licence for commercial purposes, the Licensee must obtain an undiscounted licence before commencing any commercial exploitation.

TECHNICAL ANNEX

HighlanderD3.1-Technical Annex- DApOS vs dataset - Sheet1.pdf