



# SEBASTIEN

Deliverable Lead	UNITUS
Deliverable due date	2022/12/31
Status	FINAL
Version	V1.0
Project	SEBASTIEN



## DOCUMENT INFORMATION

---

Title	Deliverable 6.1
Agreement	INEA/CEF/ICT/A2020/2373580
Action	2020-IT-IA-0234
Creator	UNITUS
Deliverable Description	Compiled descriptive catalogue of background and foreground data expected and suitable for the Open Portal
Contributors	Nico Sanna, Federica Gabbianelli, Marco Milanesi (UNITUS), Alfredo Reder (CMCC)
Requested deadline	M12
Reviewer	Cinzia Caroli (CINECA), Mario Barbato (UCSC)

## Summary

1. Executive summary	3
2. Introduction	3
3. Datasets	4
3.1 Background datasets	4
3.2 Foreground datasets	5
4. Data Harmonization	6
5. Conclusion	6
6. Bibliography	7

# 1. Executive summary

This document contains a detailed description of the datasets representing the base of information and metadata of the SEBASTIEN project. The datasets will be harmonized into a catalog within T6.2 and finally deployed through the planned four services of T6.3, namely:

*Service 1:* Coping with environmental stressors for breeds

*Service 2:* Intensive farming risk management under climate extremes

*Service 3:* Extensive farming management and feed availability

*Service 4:* Livestock farming under risks from combined abiotic and biotic factors

## 2. Introduction

Datasets in input to the planned four services of SEBASTIEN have been divided into two classes: *background* and *foreground*.

Background datasets can be further divided into:

- (i) datasets generated by a project partner prior to the project or under other funding.
- (ii) datasets provided by third-party organizations.

Conversely, foreground datasets are generated by the partners during the SEBASTIEN project and under Sebastien funding.

## 3. Datasets

Below is the list of datasets used in SEBASTIEN with a brief description and divided in background and foreground classes. For data license policies and other specific information, the deliverable D2.1 (List of indicators/indices to be proposed to stakeholders) and D2.2 (List of suitable data sources and of newly acquired data) and the milestone M14 (Open Data Policy report) could be consulted. The Annex 1 (D6.1\_ANNEX1.xlsx) is a table with the dataset specifications.

### 3.1 Background datasets

#### VHR-REA\_IT

This dataset, produced by CMCC within the Highlander's framework, contains hourly variables for recent climate (January 1981 – December 2020) over Italy obtained by dynamically downscaling the ECMWF ERA5 Reanalysis at 2.2 km grid space resolution. The data are freely available (prior

registration on CMCC DDS portal) for research purposes and to Highlander partners for all the project purposes (License CC BY 4.0). The use for other purposes (and by external end-users) requires an appropriate disclaimer, including references to CINECA and CMCC. For more details, see Raffa et al. 2021.

### **VHR-PRO\_IT**

This dataset, produced by CMCC within the Highlander framework, contains climate projections (Historical, January 1981 – December 2005 + Scenario, January 2006 – December 2070) over Italy obtained by dynamically downscaling the Italy 8 km-CM climate projection (nested from the GCM CMCC-CM) under the IPCC RCP 4.5 and RCP 8.5 scenarios to the same spatial (2.2 km) and temporal (hourly) resolution of VHR-REA\_IT. The terms of use are the same as for VHR-REA\_IT. For more details, see Raffa et al. (under review) and Raffa & Mercogliano (2022).

### **Outdoor forecast weather data**

This dataset contains outdoor weather features (e.g., temperature), provided by Mistral (COSMO2I) at 2 km for the next 2 days forecast for the entire Italian area.

### **Indoor weather data**

This dataset contains indoor temperature and humidity from livestock stables. The data are collected from different sources. For example, ASSONAPA, for ovine, collects the data every 30 minutes; LEO project, for bovine, collects the data every hour.

### **Outdoor weather data (weather ground station)**

This dataset contains outdoor weather observations (e.g., temperature and humidity) from weather ground stations. The source is ARPA (e.g., Dexter for Emilia Romagna) - ISPRA, civil protection or Air Force.

### **Seasonal forecast data**

Seasonal Forecast data from the Copernicus Climate Change Service (C3S). The variables available in this dataset (spatial resolution  $\approx$  100 km) include forecasts created in real-time (since 2017) and retrospective forecasts (hindcasts) initialised at equivalent intervals during 1993-2016.

### **LEO livestock phenotypic data**

This dataset is part of the entire LEO dataset. It contains livestock phenotypic data, such as dairy production data (e.g., quantity of milk; protein, fat and somatic cell content to evaluate the milk quality), and reproduction data (e.g., age of first calving).

### **Genotype and pedigree data**

The data are owned by the breeder association(s) involved in the project, and not open to public use. The genotype and the pedigree information will be used only to obtain service's results.

### **Farm information**

This dataset contains information on each farm involved in the SEBASTIEN project. For example, farm location (longitude, latitude, and altitude), farm structure, and cooling system. This dataset is used to characterize the farms involved in the project.

#### **Satellite data**

Satellite images from SENTINEL 2 (historical for modeling, and recent for application) are external datasets owned and distributed by the Copernicus organization.

#### **IoT animal sensors**

IoT animal sensor data were produced by Nature 4.0 in the HIGHLANDER project. In particular, location, animal micro-environmental data, and accelerometer data were collected.

#### **Parasite and disease data**

This dataset contains information about vectors and diseases affecting livestock animals and influenced by environmental conditions.

## **3.2 Foreground datasets**

#### **IoT sensors on animals and environment**

IoT animal sensors produced by Nature 4.0 were improved for the Sebastien project. This new-generation IoT sensor is able to collect also: animal temperature, blood O<sub>2</sub>, and heartbeat data. Additional data on CH<sub>4</sub>, NH<sub>3</sub>, H<sub>2</sub>S, CO<sub>2</sub>, and particulate matter within some of the farms involved in the project since the beginning of SEBASTIEN through IoT sensors for air quality monitoring.

#### **Pasture features**

This dataset contains all the information about pasture evaluated by the SEBASTIEN project. In particular, pasture location, number of animals using the pasture, biomass quantity and quality, herbal species, among others, are collected.

## **4. Data Harmonization**

Task 6.2 (Open data catalogue) aims to the proper harmonization of generated and used datasets in terms of data (structure, content and grid for geospatial data) and related metadata.

The FAIR (Findability, Accessibility, Interoperability, and Reusability) principles guidelines were published in 2016 with the aim of increasing the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention. The FAIR principles refer to three types of entities: data (or any digital object), metadata (information about that digital object), and infrastructure.

This task involves the definition of a semantic layer compliant with the FAIR principles that will i) homogenize the generated datasets in terms of their specific features, ii) foster interoperability with the Italian and EU Open Data Portals, iii) allow search and discovery activities.

Widely adopted standards will be exploited to highlight key entities, properties and relationships between data such as the DCAT-AP metadata model (as reported in the section 6.2 of the Grant Agreement).

The description of the final data structure will be reported in the deliverable D6.2 (Document describing Portal's architecture), which is the objective of Task 6.3.

## 5. Conclusion

The SEBASTIEN catalogue is an extension of the one developed through the HIGHLANDER Action, by means of the policies established in task 6.1.

## 6. Bibliography

Raffa, M.; Reder, A.; Marras, G.F.; Mancini, M.; Scipione, G.; Santini, M.; Mercogliano, P. VHR-REA\_IT Dataset: Very High Resolution Dynamical Downscaling of ERA5 Reanalysis over Italy by COSMO-CLM. Data 2021, 6, 88. <https://doi.org/10.3390/data6080088>

Raffa, M.; Adinolfi, M.; Reder, A.; Marras, G.F.; Mancini, M.; Scipione, G.; Santini, M.; Mercogliano, P. VHR-PRO\_IT Dataset: Very High Resolution Projections over Italy under IPCC RCP4.5 and IPCC RCP8.5 scenarios. Scientific Data (under review)

Raffa, M., & Mercogliano, P. (2022). Dynamical Downscaling with COSMO-CLM of historical (1989/2005) and future climate (2006/2050) data under scenario RCP4.5 and RCP8.5 at 2.2 km over Italy [Data set]. Fondazione CMCC. <https://doi.org/10.25424/CMCC-J90A-5P12>