

SEBASTIEN

M3 - Report on Data Sources

Milestone Lead	CMCC
Milestone due date	2022/11/30
Status	FINAL
Version	V1.0
Project	SEBASTIEN



DOCUMENT INFORMATION

Title	M3 Report on Data Sources
Agreement	INEA/CEF/ICT/A2020/2373580
Action	2020-IT-IA-0234
Creator	Alfredo Reder (CMCC), Francesco Renzi (Nature4.0)
Milestone Description	Brief report on the data sources identified
Means of verification	Brief report on the data sources identified is shared with the Agency
Contributors	Alfredo Reder (CMCC), Francesco Renzi (Nature4.0), Paola Mercogliano (CMCC)
Requested deadline	M11
Reviewer	Alessandro D'Anca (CMCC), Marco Milanesi (UNITUS)

Introduction

This Milestone reports and shares with the Agency the data sources identified in the Deliverable *D2.2 List of suitable data sources and of newly acquired data*, suitable for developing Services in the framework of the SEBASTIEN project. Below, a summary of the Deliverable is first given, highlighting the datasets and data sources identified for weather-climate information, spatial information and animal welfare; then, a first attempt to link the datasets with the Services (Activity 6) to be developed during the SEBASTIEN project is described.

Summary of D2.2: List of suitable data sources and of newly acquired data

The Deliverable *D2.2 List of suitable data sources and of newly acquired data* aims to identify the existing data sources and ways for their access and extraction in developing SEBASTIEN applications and services. It moved from the Deliverable *D2.1 List of indicators/indices to be proposed to stakeholders* who identified a trial list of indicators for bioclimate, territorial and animal segments. Specifically, the goal of D2.2 is to identify High-Value Datasets coming from multi-sources and multi-thematic portals (Italian Open Data Portal, SIAN, SINANET, SCIA, ISTAT, EAA, Copernicus Land Monitoring Service, Copernicus Climate Change Service, Copernicus-linked observatories, Copernicus Open Access Hub, LEO open data portal and OIE-WAHIS portal) suitable to derive these indicators and support services development (Activity 6). In addition, many other input data used in SEBASTIEN come from geospatial and non-geospatial datasets leveraging previous experiences of project partners (e.g., HIGHLANDER and LEO projects).

The Deliverable identifies three types of datasets, i.e., Datasets containing weather and climate information, Datasets containing territorial information and Datasets containing animal welfare information. In general, these data sources are under open-access licences and usable via API. The document represents a living database, which can be updated if additional datasets are developed and released in the following months.

Finally, the Deliverable consists of a main descriptive text and is accompanied by an Annex in which specific features of the datasets are summarised and reported.

Datasets and data sources identified

Datasets containing weather and climate information

Historically, there has been a clear separation between weather and climate predictions, albeit both exploit similar numerical tools. Weather prediction refers to the prediction of daily weather patterns

from a few days up to about two weeks in advance. Climate prediction refers to predicting climate fluctuations averaged over a season and beyond. However, convergence is occurring, stimulated by the growing realisation that weather and climate occur on a temporal and spatial scales continuum (see Figure 1). Consistent phenomena on a range of scales along this continuum lead to predictability on sub-daily scales, weeks, months, years, decades and beyond.

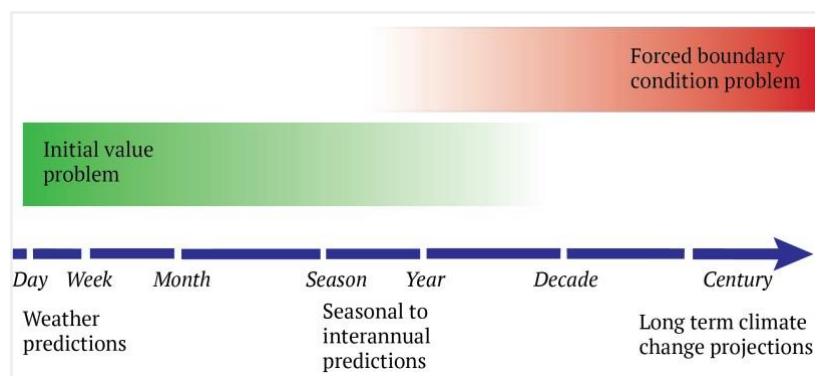


Figure 1: Differences between weather and climate predictions

Depending on the time scale considered, it is possible to categorise the available weather and climate information into:

Observations - They are crucial to understanding the past and current features of a climate system; they include data from a variety of instrumental data records, ranging from historical weather observations to the latest measurements from space.

Climate reanalysis - They combine past observations with models to generate consistent time series for a large set of climate variables for recent and current climate; they are among the most-used datasets in geophysical sciences.

Weather forecasts - They are made by collecting quantitative data on the current state of the atmosphere in a given area and using a scientific understanding of atmospheric processes to predict how the atmosphere will evolve there. There are different types of weather forecasts: nowcasting (a detailed analysis and description of the current weather and then forecasting for a period from 0 to 6 h), short-range forecasting (will last 1-2 days), medium range forecasting (generally covers a period ranging from about three days to seven days in advance; fairly accurate but with a certain margin of error especially after three days) and long-range forecasting (typically provides information about expected future atmospheric and oceanic conditions, averaged over periods of one to three months; they are not very accurate).

Seasonal forecasts - They combine outputs from several state-of-the-art seasonal prediction systems from providers in Europe and elsewhere to estimate weather statistics on monthly and seasonal time scales, which places it somewhere between conventional weather forecasting and climate forecasting.

Climate projections - They give projections of future climate for different scenarios for concentrations of greenhouse gases, aerosols and other atmospheric constituents based on outputs from multiple global and regional climate models.

Table 1: List of datasets containing weather and climate information identified in D2.2

Datasets containing weather and climate information				
Observations	Climate Reanalysis	Weather Forecasts	Seasonal Forecasts	Climate Projections
E-OBS gridded SCIA-ISPRA gridded	ERA5 ERA5-Land UERRA CERRA VHR-REA_IT (reanalysis-based) (from Highlander)	COSMO 2I NCEP-GFS 0.25 ARW-WRF	C3S Seasonal forecast data	EURO-CORDEX ITALY8km-CM (RCP8.5) VHR-PRO_IT (RCP8.5) (from Highlander)

For each Dataset, the Annex to the deliverable D2.2 reports the following information: *ID; Name; Category; Data type; Coordinate System; Horizontal coverage; Horizontal resolution; Vertical coverage; Vertical resolution; Temporal coverage; Temporal resolution; List of Variables; File format; Dimensions; Update frequency; DOI; Licence; Access*

Datasets containing territorial information

The second category of datasets identified for the scopes of the SEBASTIEN project encapsulates indicators to describe spatial information in terms of elevation, slope, aspect, soil type and land cover. The first three are derived directly or indirectly from information coming from digital elevation models, while specific datasets must be considered for soil type and land cover. Based on this initial screening, the spatial information required as input is the digital elevation model, the soil type dataset, and the land cover classification. Such data can be retrieved from the Copernicus Land Monitoring Service and specific Italian databases.

In short, the datasets can be grouped into:

Digital Elevation Model (DEM) - It represents the bare ground (bare earth) topographic surface of the Earth, excluding trees, buildings, and other surface objects. It means the altitude (in metres) and is the base to derive slope and aspect. More specifically, a Digital Surface Model (DSM) also captures the natural and built/artificial features of the environment.

Soil Information - Soil is a natural resource that can be classified into different soil types, each with distinct characteristics that provide growing benefits and limitations. Therefore, identifying the type of soil is paramount to support the healthy growth of plant life, as well as to derive specific features from selected soil parameters such as organic carbon, pH, water storage capacity, soil depth, cation

exchange capacity and clay fraction, total exchangeable nutrients, lime and gypsum contents, sodium exchange percentage, salinity, textural class, and grain size.

Land Cover - It represents a key source as it provides spatial information on different types (classes) of physical coverage of the Earth's surface, e.g., forests, grasslands, croplands, lakes, and wetlands. This information is paramount for investigating morphological and functional changes occurring in terrestrial ecosystems and the environment. Land cover change may affect different ecosystem services, resulting in loss of biodiversity, disruption of the hydrological cycle, increase in soil erosion, microclimatic discomfort and runoff.

Table 2: List of datasets containing territorial information identified in D2.2

Datasets containing territorial information		
Digital Elevation Model	Soil Information	Land Cover
Digital Elevation - Global 30 Arc-Second Elevation (GTOPO30) Copernicus Land Monitoring Service - EU Digital Elevation Model (EU-DEM) TINITALY/01	Harmonized World Soil Database v 1.2 (HWSD) European Soil Database v2.0 SoilGrids250m	CORINE Land Cover 2018 ESA CCI Land Cover map LUCAS

For each Dataset, the Annex to the deliverable D2.2 reports the following information (when possible): *ID; Name; Category; Data type; Coordinate System; Horizontal coverage; Horizontal resolution; Vertical coverage; Vertical resolution; Temporal coverage; Temporal resolution; List of Variables; File format; Dimensions; Update frequency; DOI; Licence; Access*

Datasets for animal welfare indicators

The last category of datasets identified for the scopes of the SEBASTIEN project includes information and indicators to describe the so-called “animal welfare”. In short, the datasets can be grouped into:

LEO (Livestock Environment Opendata) – It is a single database developed in the LEO project¹ containing all information related to livestock in Italy to support and improve the quality of this sector, reducing the impact on animal wellness and the environment. Currently, the database consists of 86 datasets divided into 6 areas: Climate data (ClimData), Precision Livestock Farming (PLFData), Laboratory data, Wellness, Genetic data, Collected data.

¹ The LEO project was born in 2017 in response to the request of the MiPAAF (Ministry of Agriculture, Food and Forestry Policies) to link all useful information for the Italian livestock sector in a single and open-access database. Its leadership is entrusted to the AIA (Associazione Italiana Allevatori, one of the SEBASTIEN partners). Other partners are IZSAM – BDN, Istituto Spallanzani – Banca Dati fertilità, IZSUM, Bluarancio, some research institutions (i.e., Università Cattolica di Piacenza, Università della Tuscia, Università di Palermo, ConSDABI), and the main Italian livestock databases (i.e., Banca dati Nazionale – BDN; Banca dati del sistema allevatori – SIAll e Banca dati della Fertilità Maschile).

OIE-WAHIS (OIE World Animal Health Information System) - It is a comprehensive database through which information on the animal health situation is reported and disseminated worldwide. OIE-WAHIS data reflects the information for a total of 205 diseases of which 16 bovine-specific, 15 regarding sheep and goats and other 27 affecting multiple terrestrial species. OIE-WAHIS is divided into three main parts: the early warning system (animal disease events), the monitoring system (six-monthly reports) and additional information collected annually (annual reports). The data are spatially located at the province level, updated every two hours, and released under an open licence.

Sentinel datasets – They are open datasets of satellite images that could be used to analyse the status of grazing fields. In particular, Sentinel missions, launched by the European Space Agency (ESA) to support the Copernicus project, rely on a constellation of satellites to meet revisit and coverage requirements, providing robust data sets for Copernicus services.

IoT datasets -They are datasets already developed in previous experiences or in production for the SEBASTIEN project that are based on IoT sensors for real time monitoring. Specifically, we make use of data collected during the HIGHLANDER project from a sensor network for animal wellbeing. Such a network will be enlarged during SEBASTIEN and the data collected from it will be used.

Table 3: List of datasets for animal welfare indicators identified in D2.2

Datasets for animal welfare indicators			
LEO (Livestock Environment Opendata)	OIE-WAHIS (OIE World Animal Health Information System)	Sentinel datasets	IoT datasets
Climate data (ClimData) Precision Livestock Farming (PLFData) Laboratory data Wellness Genetic data Collected data	Early warning (animal disease events) Monitoring (six-monthly reports) Annual reports	Sentinel-1 Sentinel-2	Highlander datasets Sebastien IoT sensors

For each Dataset, the Annex to the deliverable D2.2 reports the following information (when possible): *ID; Project Name; Category; Horizontal coverage; Horizontal resolution; Temporal coverage; Temporal resolution; List of Variables; File format; Update frequency; Licence; Access; Group dataset available*

First attempt to link the Datasets with the Services

The services that the SEBASTIEN project will provide are currently being defined, and this Section presents a first attempt to link the datasets with the developed Services . This information may however be updated in the coming weeks.

It reports in brief the aim of each Service and the category of inputs that is expected to be considered for its development. Datasets containing territorial information have not been reported as they represent the starting point for each Service.

Service 1: Coping with environmental stressors for breeds	
DESCRIPTION	
The Service will support livestock farming by matching breeds, as well as by identifying potential breed improvements for different breeds, towards their acclimation, acclimatisation and adaptation to environmental conditions and production needs. It will make use of climate data (from monitoring and modelling) to characterise the farms’ environments in the recent and future periods, plus -omics data, phenotypic data (from monitoring sensors and projects) and other farms/associations’ owned datasets.	
PRELIMINARY INPUT DATA	
Datasets containing weather and climate information	Observations, climate reanalysis and climate projections
Datasets for animal welfare indicators	LEO database

Service 2: Intensive farming risk management under climate extremes	
DESCRIPTION	
The Service aims to alert about the approaching (on the short-term) or projections (on the long-term) of dangerous environmental circumstances for cattle in the farms (due to indoor but also outdoor conditions) causing discomfort and loss of reproduction and/or production (milk yield and milk quality) performances and change in behaviour, all that assessed by integrating phenotypic and -omics data already archived and/or newly collected thanks to synergy with HIGHLANDER and projects of partner AIA.	
PRELIMINARY INPUT DATA	
Datasets containing weather and climate information	For short-term: observations and weather forecast. For long-term: seasonal forecast and/or climate projections
Datasets for animal welfare indicators	IoT datasets and LEO database

Service 3: Extensive farming management and feed availability
--

DESCRIPTION	
The Service will exploit indicators/indices about the phenological stage and greening of the naturally vegetated or managed areas (grassland/pasture land) to feed livestock heads when conducted outdoor, guiding in scheduling/updating daily grazing and, more indirectly, to detect possible overgrazing effects/damages on lands with cascading effects like soil degradation and erosion.	
PRELIMINARY INPUT DATA	
Datasets containing weather and climate information	Outdoor observations and weather forecasts
Datasets for animal welfare indicators	Satellite data; Farm specific information; Pasture features

Service 4: Livestock farming under risks from combined abiotic and biotic factors	
DESCRIPTION	
The Service aims to provide updated risk maps for parasites and diseases' spreads, combining the hazards due to the presence of parasites/diseases in the surroundings, the exposure of animals due to environmental conditions favouring parasites/diseases, and the vulnerability due to the general animal health status and adaptation capacity.	
PRELIMINARY INPUT DATA	
Datasets containing weather and climate information	Seasonal forecasts and climate projections
Datasets for animal welfare indicators	Disease georeferenced data