BioClima

D1.1 Analysis of Data for Integration

Funded by
the European Union

Document Identification			
Related WP	**WP1**	**Document reference**	**D1.1**
Status	**Final**	**Submissions date**	**31/12/2025**
Related deliverables	**-**	**Dissemination Level**	**PU**
Lead Author	**Emmanouela Prassou (ICCS)**	**Version**	**1.0**
Co-Authors	**Giorgos Tsilimanis (ICCS)** **Ionut Sandric (UB)** **Philip Stanley Mostert (NTNU)** **Dmitry Shchepashchenko (IIASA)**	**Reviewers**	**Michal Kepka (UWB)** **Karel Charvat (P4A)**
Keywords	**Data analysis, Citizen Science, EO, in-situ, UAV, Integration**	**Website**	**http://bioclima.net/**

This document is issued within the frame and for the purpose of the BioClima project. This project has received funding from the European Union's Horizon-RIA Framework Programme under Grant Agreement No. 101181408. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Commission.

Document History			
Version	**Date**	**Editors**	**Changes**
0.1	09/12/2025	ICCS	Initial Draft
0.2	19/12/2025	ICCS	2nd Draft
1.0	31/12/2025	ICCS	Final Version

1. Table of Contents

List of Tables

List of Figures

List of Acronyms

Acronym	Description
AGB	Above-Ground Biomass
AI	Artificial Intelligence
API	Application Programming Interface
AWS	Amazon Web Services
BA	Burned Area
CAMS	Copernicus Atmosphere Monitoring Service
CARD-BS	Terrain-Corrected Backscatter
CAS	Chinese Academy of Sciences
CCI	Climate Change Initiative
CEDA	Centre for Environmental Data Analysis
CEMS	Copernicus Emergency Management Service
CHM	Canopy Height Model
CLMS	Copernicus Land Monitoring Service
CRS	Coordinate Reference System
CSV	Comma-Separated Values
DEM	Digital Elevation Model
DIAS	Data and Information Access Services
DSM	Digital Surface Model
DTM	Digital Terrain Model
DwC-A	Darwin Core Archive
EBV	Essential Biodiversity Variable
EC	European Commission
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
EML	Ecological Metadata Language
ENVISAT	Environmental Satellite
ET	Evapotranspiration
ETM+	Enhanced Thematic Mapper Plus

Acronym	Description
EVI2	Two-band Enhanced Vegetation Index
ESA	European Space Agency
FAO	Food and Agriculture Organization
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
FCover	Fraction of Green Vegetation Cover
FLAM	Wildfire Climate Impacts and Adaptation Model
FSC	Fractional Snow Cover
GAM	Global Forest Model
GBIF	Global Biodiversity Information Facility
GeoBON	Group on Earth Observations Biodiversity Observation Network
GeoTIFF	Geographic Tagged Image File Format
GNDVI	Green Normalized Difference Vegetation Index
HADISD	Hadley Centre Integrated Surface Database
HPC	High-Performance Computing
HR-VPP	High-Resolution Vegetation Phenology and Productivity
HTTPS	Hyper Text Transfer Protocol Secure
ICOS	Integrated Carbon Observation System
IAM	Identity and Access Management
JWT	JSON Web Token
JSON	JavaScript Object Notation
LAI	Leaf Area Index
LLL	Low-Light-Level
LiDAR	Light Detection and Ranging
LST	Land Surface Temperature
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MII	Multispectral Imager
MSI	Multispectral Instrument
NDSI	Normalized Difference Snow Index
NDRE	Normalized Difference Red-Edge Index

Acronym	Description
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NetCDF	Network Common Data Form
NFI	National Forest Inventory
NFS	Network File System
NPP	Net Primary Production
OGC	Open Geospatial Consortium
OIDC	OpenID Connect
PNG	Portable Network Graphic
PPI	Plant Phenology Index
RGB	Red-Green-Blue
RTC	Radiometrically Terrain Corrected
SAR	Synthetic Aperture Radar
S2GLC	Sentinel-2 Global Land Cover
S3	Simple Storage Device
SDG	Sustainable Development Goals
SDGSAT-1	Sustainable Development Goals Satellite-1
SeBMS	Swedish Butterfly Monitoring Scheme
SITES	Swedish Infrastructure for Ecosystem Science
SLSTR	Sea and Land Surface Temperature
SMOS	Soil Moisture and Ocean Salinity
ST	Seasonal Trajectories
STAC	Spatio-Temporal Asset Catalog
TIS	Thermal Infrared Spectrometer
TM	Thematic Mapper
UAV	Unmanned Aerial Vehicle
URL	Uniform Resource Locator
VHR	Very High Resolution
VI	Vegetation Index
WP	Work Package

2. Executive Summary

This deliverable provides a landscape analysis of the ground-based/citizen science, airborne/UAV, and spaceborne data to be used within the BioClima project. It summarises information to be utilised from established biodiversity and climate monitoring initiatives, citizen science platforms, and research infrastructures such as GBIF, GeoBON, SITES, SeBMS, ICOS, and national statistical bodies. It also describes how these sources include species composition, vegetation structure, biomass estimations, forest resource statistics, climate measurements, and other datasets relevant to the BioClima case studies. In parallel, the document outlines Earth Observation datasets that will support modelling and analysis, including Sentinel missions, Landsat and MODIS.

The report further presents an earth observation (EO) monitoring platform and its technical approach, which uses STAC catalogues, S3-compatible object storage, and OIDC-based authentication to harmonise metadata, access, and dataset discovery. It also describes the Geo-Wiki crowdsourcing tool for identifying drivers of forest biomass change, as well as a modular software tool designed in Python to integrate multisensory UAV outputs for the mapping of Essential Biodiversity Variables (EBVs) and Essential Climate Variables (ECVs). Together, these elements reflect the data, resources, and software components developed and adopted in Work Package (WP) 1, "Integrated observations for climate and biodiversity", forming the basis for subsequent activities as described in the BioClima workflow.

3. Introduction

3.1 Purpose of the document

The purpose of this document is to perform a landscape analysis of the ground-based, airborne, UAV and satellite data to be used in the context of the BioClima project by the 7 case studies[1]. It also serves as a summary report from tasks:

- T1.2 “Ground-based biodiversity, and citizen science monitoring network”
- T1.3 “Airborne/UAV biodiversity and climate change monitoring network”
- T1.4 “Spaceborne biodiversity and climate change monitoring network”

Including a software tool designed to integrate multi-sensor UAV outputs for the mapping of EBVs and ECVs.

3.2 Relation to other project work

This document forms part of WP 1 of the BioClima project, which aims to define the conceptual and technical framework for the observation, standardisation, and interoperability of biodiversity and climate-related data. Deliverable 1.1 is a summary report of:

T1.2 Ground-based biodiversity and citizen science monitoring network

This task focuses on developing a tool to enable the extraction, harmonisation and integration of biodiversity data from major sources such as Global Biodiversity Information Facility[2] (GBIF), GeoBON[3], citizen science platforms and natural history collections. Building on the standards defined in T1.1, the work will ensure that both data and associated metadata are accessible in structured, interoperable formats suitable for T1.5, “EU-China joint observation and demonstration platform for biodiversity”, and downstream use across other WPs.

T1.3 Airborne/UAV biodiversity and climate change monitoring network

Task 1.3 focuses on developing a user-orientated tool for generating high-resolution EBVs and ECVs products at a local scale through the fusion of multi-sensor UAV data. The tool will address key challenges related to sensor calibration, data standardisation, and the integration of heterogeneous UAV data streams. Building on the metadata framework established in T1.1 “Standards for observational systems of biodiversity and climate change in terrestrial ecosystems”, all outputs will incorporate comprehensive quality assurance information covering sensor specifications, acquisition context, and full processing history in order to ensure transparency, reproducibility, and robust data interoperability. The resulting structured UAV-based products will be used in T1.5 as well as in the analytical, modelling and integration activities carried out in WP2 and WP3.

[1] https://bioclima.net/case-studies/european-case-studies2/
[2] https://www.gbif.org
[3] https://geobon.org/

T1.4 Spaceborne biodiversity and climate change monitoring network

This task develops a cloud-based architecture that automates and scales EO data pre-processing by leveraging cloud platforms like Copernicus DIAS and the Copernicus Data Space Ecosystem. Through a unified API and interoperability standards, the system will streamline access and processing of EO datasets, directly supporting Task 1.5 and enabling downstream use in WP2 and WP3.

In addition, a Geo-Wiki-based crowdsourcing tool will be used to identify the drivers of forest biomass change. Citizen scientists and experts will classify gain and loss factors using remote-sensing inputs, producing driver-specific statistics for the EU and China. These outputs will inform analysis in WP3 and serve as inputs for WP4.

3.3 Structure of the document

The document is structured to guide the reader through the different types of data and tools that will be used in the BioClima project. It starts with an Executive Summary that gives a brief overview of the work, followed by an introduction describing the purpose of the deliverable and how it connects to other project tasks. The main body of the report is organised into three core data categories. Chapter 4 focuses on ground-based biodiversity and citizen science networks, presenting the in-situ measurements and community-driven data that support the project. Chapter 5 covers spaceborne monitoring, outlining the BioClima EO platform, its access and metadata standards, and the satellite datasets that will be used. Chapter 6 introduces the crowdsourcing tool based on Geo-Wiki and explains how visual interpretation will help characterise drivers of forest biomass change. Chapters 7 and 8 present the airborne/UAV monitoring activities, describing the sensors, collection sites and the software tool developed to integrate UAV data. The document closes with some conclusions connecting all the data types with key EBVs and ECVs and a list of references supporting the work.

4. Ground-based biodiversity, and citizen science monitoring network

4.1 Information on species' composition

One of the fundamental types of in-situ data commonly used across biodiversity research and considered among the BioClima case studies is georeferenced information on species composition collected at sampling sites. These data are typically collected in a variety of ways and thus often contain different metrics of occupancy, most notably information on the presence and absence, abundance, and biomass of species.

Ideally, we would like structured data following a standardised sampling protocol and known sampling effort to use in analyses[4]. However, collecting this data is expensive and tends to be geographically restricted. With the rapid advancements in digital technology and the rise of citizen science initiatives, the 21st century has been associated with a significant increase in the volume and diversity of biodiversity data, which has thereby allowed researchers to conduct analyses on larger scales than previously possible.

4.1.1 Citizen science initiatives

Biodiversity data can be accessed through numerous open-access platforms which are primarily driven by community initiatives. Notable examples of these platforms include iNaturalist[5] and eBird[6], both of which allow users to easily submit species observations and identify sightings through the use of mobile phone applications and web forms.

However, the Global Biodiversity Information Facility[7] (GBIF), which is an international network and data infrastructure, is the most comprehensive biodiversity data aggregator, currently containing over 3.5 billion occurrence records from over 118,000 datasets from across the globe. The resources GBIF provides have been invaluable for allowing researchers, conservationists and policymakers to understand our natural world and develop effective conservation strategies.

Information from GBIF can be easily accessed and downloaded directly through their website. However, toolboxes, which have been created to offer a programmatic interface to the GBIF API, are readily available to assist users with their data requests. These interfaces are most notably implemented through the *R* package *rgbif*[8], and the *Python* library *pygbif*[9].

[4] https://www.cell.com/trends/ecology-evolution/fulltext/S0169-5347(19)30255-1?sf225734247=1
[5] https://www.inaturalist.org
[6] https://ebird.org/home
[7] https://www.gbif.org
[8] https://www.gbif.org/tool/81747/rgbif
[9] https://www.gbif.org/tool/OlyoYyRbKCSCkMKIi4oIT/pygbif-gbif-python-client

The Swedish Butterfly Monitoring Scheme[10] (SeBMS) is one of the datasets considered in the BioClima case studies that is hosted on GBIF, which represents a standardised monitoring scheme for collecting information on butterflies across Sweden. The data is collected by volunteers visiting sampling sites 3-7 times per year, between the beginning of April and the end of September. Information on butterflies is collected using two different sampling methodologies: the fixed-route Pollard walk transects and point site counts. Through these two methods, researchers are able to evaluate yearly fluctuations in butterfly counts, as well as assess species compositions across the country.

While GBIF serves as a major hub for expanding biodiversity research, the data it compiles comes from a multitude of diverse sources, each with its own data collection methodologies, resolutions and collection objectives. Moreover, presence-only data, which is most prominent on GBIF, is known to be affected by a myriad of different collection biases[11]. This therefore means that practitioners should take care when using these data products and should properly account for the heterogeneous collection methodologies and sampling biases within each dataset by applying appropriate frameworks, rather than naively aggregating the records together by reducing the data to the lowest possible denominator – usually presence-only records[12].

A statistical framework based on modelling heterogeneous datasets in a single model has recently been introduced in the field of ecology [13]. This framework is based on an inhomogeneous Poisson point process for the actual distribution, which serves as a common underlying state for different datasets collecting different metrics of species occupancy. Then, different observation models are defined for each dataset, which account for the different data collection processes and biases within each. Simulation studies suggest that if the biases are correctly accounted for in integrated species distribution models, these models perform better than single-dataset models, providing more accurate and robust estimates[14].

Most data shared on GBIF are published using the Darwin Core Archive format (DwC-A), developed by the Biodiversity Information Standards (TDWG), which is a simple and flexible framework for compiling biodiversity data coming from heterogeneous sources[15]. The data comes in a compact and standardised package containing interconnected comma-separated value (CSV) files, thereby allowing data publishers to share data using common terminology. The central idea behind the data format is that the data files are arranged in a "star schema" format, with one core data file surrounded by any number of optional extension data files. The required core data file contains the key biodiversity data. GBIF currently supports three biodiversity data types as the basis for a core file: occurrence data – information concerning evidence of an occurrence, checklist data – information concerning taxa or taxon concepts, and sampling event data – information concerning evidence of an occurrence with details on sampling effort.

[10] https://www.gbif.org/dataset/be77e203-486c-4651-91b9-8347968b728c
[11] https://academic.oup.com/biolinnean/article/115/3/522/2440472#82259368
[12] https://nsojournals.onlinelibrary.wiley.com/doi/full/10.1111/ecog.05146
[13] https://www.cell.com/trends/ecology-evolution/fulltext/S0169-5347(19)30255-1?sf225734247=1
[14] https://nsojournals.onlinelibrary.wiley.com/doi/full/10.1111/ecog.05146
[15] https://ipt.gbif.org/manual/en/ipt/latest/dwca-guide

Furthermore, optional extension files may be included, which allows data providers to include more detailed information that relates to the core file. These extensions are typically stored in separate CSV files and are linked to the core file through shared identifiers. Finally, DWC-A also contains a descriptor metadata file describing how the files in the archive are organised and a resource metadata document based on the Ecological Metadata Language[16] (EML) describing information about the dataset itself.

4.2 Ground-based monitoring for climate data

The Integrated Carbon Observation System (ICOS) is a European research infrastructure that has been set up to monitor and provide standardised and robust data describing carbon fluxes and greenhouse gas concentrations across 16 European countries collected from around 180 measurement stations[17]. The measurement stations operate in three distinct domains: atmosphere, which observes greenhouse gases as well as the drivers responsible for the exchanges of greenhouse gases, and ocean observations, which observes carbon levels on the ocean surface.

The infrastructure uses the raw data to develop a multitude of different data products, which can easily be downloaded through the data portal provided on their website or through tools such as their Python library[18]. ICOS's role in the BioClima project is to provide measurements on carbon flux, soil moisture and evapotranspiration in the Arctic region to be used for model development and validation.

4.3 Ground-based monitoring for calibrating satellite data

Across BioClima case studies, in-situ data is predominately used as a means to calibrate and validate the data collected by satellite systems, which will thereby improve the accuracy and reliability of their results. Notably, the European Union's space program, Copernicus, relies on using in-situ data to enhance their satellite products[19]. These in-situ data are collected through several varied monitoring infrastructures, such as weather stations, buoys, and research stations, many of which are operated by institutions and agencies at a national level[20], to support their different monitoring services.

Furthermore, ground-based monitoring for calibrating satellite data is applied in the Swedish Infrastructure for Ecosystem Science (SITES) program, which is a program focused on facilitating access to field stations for novel research by providing standardised and open-access data collected over long time periods. The program has 9 field stations across Sweden which focus on capturing information on terrestrial and freshwater ecosystems. Ground-based spectral data collected by SITES to capture seasonal and inter-annual variations in vegetation conditions across Sweden[21] will be used in the BioClima case studies as a means for ground truth for vegetation growth dynamics. This data is obtained through research stations equipped with phenological cameras, which are used primarily for

[16] https://eml.ecoinformatics.org
[17] https://www.icos-cp.eu/
[18] https://www.icos-cp.eu/data-services
[19] https://insitu.copernicus.eu
[20] https://insitu.copernicus.eu/state-of-play/data-providers
[21] https://www.fieldsites.se/thematic-programs/sites-spectral

comparing and validating data from fixed spectral sensors and UAVs in addition to phenological modelling at the site-level. Data from SITE is available through their open-access data portal provided on their website[22].

4.4 Miscellaneous ground-based monitoring

A considerable amount of the ground-based data used in the BioClima case studies could be classified as miscellaneous data, provided by researchers and government agencies who collect data for a specific purpose in mind. Despite this, reusing these data through appropriate data integration will be vital in addressing the problems defined in the case studies.

An example of these data could be information on the total roundwood removals and drain[23], and information on wood consumption[24] provided by the Natural Resources Institute Finland[25] (Luke). The origin of the data is primarily from forest industry companies and all significant users of wood across Finland, collected both by Luke directly and through other relevant organisations, such as the Finnish Forest Industries Federation. Luke then completes validation to ensure relevancy and accuracy and statistical processing to derive the appropriate statistics. The derived statistics may be accessed and downloaded through a database available on Luke's website.

A further example includes information on forest removals, forest country stock and stocks of deadwood on forest floors provided by the Food and Agriculture Organization (FAO) of the United Nations, which collected information on Forests from its member nations every 5 years through the use of Forest Resource Assessments (FRAs)[26]. FRAs are collected by correspondents on a national level, who compile and collect information on forests using a standardised set of definitions and methodologies to ensure consistency. These data are available from the FAO through a user-friendly application programming interface.

Information on soil health, such as soil moisture, organic carbon content and nutrient profiles, is curated by the PREPSOIL project[27], implemented within the EU soil mission. PREPSOIL makes use of citizen scientist initiatives to expand and enhance soil health monitoring, collecting data through the use of a mobile phone application[28]. In this context, research institutes and universities launch "soil quests", which then prompt citizen scientists through notifications to survey a specific region by submitting images and other relevant information of the soil. These data will be used within the BioClima project to understand the effect of different types of vegetation on soil moisture.

---

[22] https://data.fieldsites.se/portal/

[23] https://www.luke.fi/en/statistics/total-roundwood-removals-and-drain/documentation-of-statistics-total-roundwood-removals-and-drain

[24] https://www.luke.fi/en/statistics/wood-consumption/documentation-of-statistics-wood-consumption-0#statistical-processing

[25] https://www.luke.fi/en

[26] https://openknowledge.fao.org/items/090d2fbb-32a6-412b-a3b8-1ce5c5905df2

[27] https://prepsoil.eu/

[28] https://prepsoil.eu/news/prepsoil-mobile-app-live

The Hadley Centre Integrated Surface Database (HADISD) is a comprehensive dataset maintained by the UK Met Office's Hadley Centre, which provides sub-daily, detailed and long-term information on meteorological observations, such as temperature, precipitation and pressure data, collected from a network of over 10,000 weather stations globally[29]. Extensive quality controls are implemented within this dataset to remove stations and individual observations with excessive issues. Information from this dataset may be obtained through the Centre for Environmental Data Analysis's (CEDA) archive, which is a repository for obtaining atmospheric and earth observation data[30]. The BioClima project will make use of this data to monitor local conditions and to understand the comparison of influence that climate has on vegetation. Finally, forest increment, referring to the growth of trees per calendar year, is a vital parameter considered in assessing the sustainability of using forest resources. However, increment data are typically compiled using numerous different metrics and definitions by the National Forest Inventories (NFI) of European countries[31]. Nevertheless, work has been completed to integrate measurements of different increment metrics to obtain a harmonised and complete metric, in terms of gross annual increment and net increment and natural losses, with associated uncertainty across Europe. These products come in the form of a shapefile and will be used in the BioClima project as a growth constraint for model calibration in the analysis of forest growth.

[29] https://www.metoffice.gov.uk/hadobs/hadisd/
[30] https://archive.ceda.ac.uk/
[31] https://www.nature.com/articles/s41597-023-02868-8#Sec7

5. Spaceborne biodiversity and climate change monitoring network

5.1 BioClima EO data landscape analysis

The BioClima project will integrate a wide range of Earth Observation datasets to support the analysis of vegetation dynamics, habitat types, forest structure, hydrological variables, and climate change impacts. These datasets include optical and radar images from Sentinel-1 and Sentinel-2 missions, thermal observations from Sentinel-3, multispectral time series from Landsat, and global products from MODIS. Together, these data provide high spatial resolution and temporal coverage, enabling applications such as vegetation phenology and productivity assessment, habitat mapping, soil moisture monitoring, and climate trend analysis. Data products from SDGSAT-1 satellite could also potentially be utilised to complement these observations.

5.1.1 Sentinel-1

The Sentinel-1 mission provides data from a dual-polarisation C-Band Synthetic Aperture Radar (SAR) instrument. Each product has one of three resolutions (10, 25 or 40 m), 4 band combinations (VV, HH, VV+VH, HH+HV) and three instrument modes. For land monitoring applications the most suitable product is the Interferometric Wide with dual polarisation. VH is more sensitive to change in vegetation density and structure, while the polarisation ratio VH/VV is an important index for vegetation phenology or water content.

Sentinel-1 data will be used in combination with Sentinel-2 optical observations within the ALIANCE[32] fusion model to support a detailed mapping of vegetation changes. In addition to vegetation indices derived from Sentinel-2, Sentinel-1 data will be used to quantify and evaluate changes in vegetation condition and to assess forest degradation.

Finally, timeseries data from the Sentinel-1 mission will be used within the BioClima project to extract hydrological variables and estimate soil moisture.

5.1.2 Sentinel-2

Sentinel-2 is a Copernicus Earth observation mission operated by the European Space Agency (ESA), consisting of two identical satellites providing high-resolution multispectral imagery of the Earth's surface. The mission carries a Multispectral Instrument (MSI) with 13 spectral bands covering the visible, near-infrared, and shortwave infrared regions, with spatial resolution of 10 m, 20 m, and 60 m

[32] https://annals-csis.org/Volume_39/drp/pdf/5446.pdf

and a revisit time of approximately 5 days at the equator. Sentinel-2 data are widely used for land cover mapping, vegetation monitoring, water resources and biodiversity-related applications.

Within the BioClima project, the High-Resolution Vegetation Phenology and Productivity (HR-VPP)[33] product suite provided by Copernicus Land Monitoring Service (CLMS) will be used to assess plant trait diversity and evaluate the ecosystem functional diversity across the European Arctic. The HR-VPP dataset offers high spatial resolution (10 m x 10 m) and high temporal frequency, and it is derived from the Sentinel-2A and Sentinel-2B satellite constellation, which provides a revisit time of approximately 5 days.

The HR-HPP product suite consists of three main product groups:

- **Vegetation Indices (VI)** comprise near real-time products that characterise vegetation through four indices:
  1. Leaf Area Index (LAI): It measures the one-sided green leaf area over a unit of land.
  2. Fraction of Absorbed Photosynthetically Active Radiation (FAPAR): Is the fraction of incoming solar radiation absorbed for photosynthesis by a photosynthetic organism.
  3. Normalized Difference Vegetation Index (NDVI): It quantifies green vegetation. It normalises green leaf scattering in near-infrared wavelengths with chlorophyll absorption in red wavelengths.
  4. Plant Phenology Index (PPI): It is used to track canopy green foliage dynamics and is derived from the radiative transfer equation.
- **Seasonal Trajectories (ST)** provide annual, smoothed and gap-filled PPI time series at a regular 10-day interval. They are generated after the end of the growing season through the application of a smoothing and interpolation algorithm to raw VI data.
- **Vegetation Phenology and Productivity (VPP)** include 13 annual metrics, produced from the ST products, that describe the timing and magnitude of the vegetation growth cycle. These metrics include the PPI value and date at the start, end and peak of the season, the length of the growing season and indicators of seasonal and total productivity.

In addition to the indices mentioned above, the following indices will be exploited for assessing ecosystem functioning, vegetation performance and management intensity:

- **Normalized Difference Water Index (NDWI)**: It is used to monitor changes in water content. As water bodies strongly absorb light in the visible-to-infrared electromagnetic spectrum, NDWI uses green and near-infrared bands to highlight water bodies.
- **Fraction of Green Vegetation Cover (FCover):** corresponds to the fraction of ground covered by green vegetation. Practically, it quantifies the spatial extent of the vegetation.

Sentinel-2 time series data will be used to monitor changes in agriculturally affected biotopes and vegetation status. These datasets include multispectral imagery with 10 m, 20 m and 60 m spatial resolution bands, as well as all the derived vegetation indices mentioned above. These datasets, in combination with fusion models, will be applied to map changes in vegetation and calculate relevant

[33] https://land.copernicus.eu/en/products/vegetation?tab=overview

indices for assessing vegetation condition. Sentinel-2, in combination with Sentinel-1 and airborne laser scanning data, will be utilised to map habitat types and identify main tree species.

5.1.3 Sentinel-3

In the context of the BioClima project, Sentinel-3 Sea and Land Surface Temperature Radiometer (SLSTR) data will be utilised to support vegetation phenology and productivity assessments. The LST product provides surface temperature measurements in Kelvin, at a spatial resolution of 1000 m with twice-daily temporal coverage. These observations will be compared against the model's outputs to evaluate vegetation's seasonal dynamics and productivity patterns. In addition, Sentinel-3 LST will be used together with derived evapotranspiration (ET) estimates to monitor surface temperature and assess vegetation performance by extracting relevant hydrological variables.

5.1.4 Landsat

Landsat-5 Thematic Mapper (TM) provides 6 multispectral bands at 30 m spatial resolution, along with a thermal band at 120 m. Landsat-7 ETM+ offers 6 multispectral bands at 30 m, a 15 m panchromatic band, and a thermal band at 60 m resolution. Landsat-8 (OLI/TIRS) expands the spectral range with 8 multispectral bands at 30 m and a 15 m panchromatic band. Landsat-9 follows the same spectral and spatial configuration as Landsat-8. All Landsat missions operate with a 16-day revisit cycle.

Landsat data will be used within the project to support both habitat-level vegetation assessments and long-term environmental change monitoring. The multispectral reflectance and vegetation index products derived from Landsat bands will enable the detection of selected habitat types, providing insights into vegetation conditions. The long historical record of Landsat missions, combined with the thermal bands, will be utilised to map spatiotemporal changes in biotopes and vegetation status, as well as to monitor climate-related trends.

5.1.5 MODIS

The Moderate Resolution Imaging Spectroradiometer (MODIS), onboard the Terra and Aqua satellites, provides near-daily observations of the Earth's surface, with a revisit frequency of 1-2 days. MODIS acquires data in 36 spectral bands ranging from visible to thermal infrared, with spatial resolutions of 250 m, 500 m, and 1000 m depending on the band.

The MODIS/Terra+Aqua Land Cover Dynamics (MCD12Q2)[34] data product is derived from the 2-band Enhanced Vegetation Index (EVI2). It provides global land surface phenology metrics at yearly intervals from 2001 to 2021 and at a spatial resolution of 500 m. In the context of the BioClima project, these vegetation phenology data will be used in combination with Sentinel-3 observations for model validation. The MOD10A1/Snow Cover Daily Global (500 m)[35] product will also be used to monitor temporal changes in snow cover across the study areas. It provides daily information on snow cover, snow albedo, fractional snow cover (FSC) and associated quality assessment layers. Snow detection is

[34] https://www.earthdata.nasa.gov/data/catalog/lpcloud-mcd12q2-061
[35] https://nsidc.org/data/mod10a1/versions/61

based on a mapping algorithm that uses the Normalized Difference Snow Index (NDSI) combined with additional spectral and classification tests.

5.1.6 SDGSAT-1

SDGSAT-1 is the first satellite developed to support the implementation of United Nations 2030 Agenda for Sustainable Development Goals (SDGs) and also the first Earth science satellite developed by the Chinese Academy of Sciences (CAS).

Designed to meet the needs of monitoring global SDG indicators, SDGSAT-1 is equipped with three types of payloads: a Thermal Infrared Spectrometer (TIS), a Multispectral Imager (MII), and a Low-Light-Level (Glimmer) Imager (LLL). The satellite has a revisit cycle of approximately 11 days. SDGSAT-1 operates in different modes depending on the time of day. During the daytime, it uses a thermal infrared + multispectral mode, while at night it operates with thermal infrared + glimmer mode. Single-sensor observation modes are also available. Through these modes SDGSAT-1 is capable of collecting three types of Earth surface data within a 24-hour period.

In terms of spatial resolution, the Thermal Infrared Spectrometer provides a resolution of 30 m, while the Multispectral Imager provides a resolution of 10 m. The Glimmer Imager features an innovative RGB band design, with a spatial resolution of 10 m for the panchromatic band and 40 m for the RGB bands.

The SDGSAT-1 satellite provides on-demand thermal, multispectral and low-light-level data products that could be considered within the scope of biodiversity and climate-related analyses for the BioClima project.

5.1.7 Processed datasets

Within the framework of the BioClima project, two harmonised forest datasets – forest area and forest biomass maps (at 100 m spatial resolution) – provided by Avitabile et al.[36][37] will be used to support model calibration and spatial analysis. The **forest area map** provides a bias-adjusted representation of forest extent across Europe, ensuring close agreement with reference statistics at the administrative scale. The original Copernicus Forest Type 2018 map[38] was converted to a forest/non-forest classification and subsequently adjusted to accurately represent forest cover for each administrative unit. This map will be used to define forest area masks for years from 2020 onwards, delineating the spatial extent of forests.

---

[36] https://www.nature.com/articles/s41597-023-02868-8
[37] https://figshare.com/collections/Forest_Biomass_dataset_for_Europe_-_Supplementary_data/6787140/6
[38] https://land.copernicus.eu/en/products/high-resolution-layer-forests-and-tree-cover/forest-type-2018-raster-10-m-100-m-europe-3-yearly

The **forest biomass dataset** provides a harmonised map of forest biomass density (t/ha) and total biomass stock (t), also bias-adjusted and masked using the forest area map. The forest biomass map was derived from the original ESA Climate Change Initiative (CCI) biomass map of 2020[39] and subsequently corrected for systematic differences to align with sub-national reference statistics. It captures patterns of biomass distribution, which will be used for model calibration.

For the calibration and validation of the wildfire climate impacts and adaptation model (FLAM)[40] , the **global burned area (BA)[41][42]dataset** developed under the ESA Climate Change Initiative (CCI) Fire Disturbance project (Fire_CCI)[43] will be used. The FIREECCI50 and FIRECCI51 datasets were generated from MODIS red and near-infrared reflectance and thermal anomaly data and provide a high-resolution record (~250 m) of global burnt areas. They include the full time series from 2000 to 2017 derived from the Terra MODIS archive. The burnt area detection algorithm that was utilised was based on monthly composites of daily images, combining temporal and spatial proximity to active fires.

ERA5 [44] is an atmospheric reanalysis dataset of the global climate, provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), covering the period from January 1940 to present. It provides hourly estimates of a wide range of atmospheric, land and oceanic climate variables. ERA5 uses a state-of-the-art numerical weather prediction model to assimilate a variety of observations, including satellite and ground-based measurements. Within the BioClima project, ERA5 will be used to monitor climate changes in the study areas from 1986 onwards.

CORINE Land Cover[45] is a pan-European land-cover dataset developed under the Copernicus Land Monitoring Service. It provides a consistent overview of how land is used and covered across Europe, including forests, agricultural areas, urban zones, wetlands, and water bodies. The dataset is derived from the analysis of satellite imagery, mainly from Sentinel and Landsat missions, and is commonly used as a reference layer for land-cover and land-use change studies, habitat mapping, and environmental assessments.

5.1.8 CDSE Platform and EO Interoperability

The Copernicus Data Space Ecosystem (CDSE) is an open platform offering free, immediate access to a broad array of data and services from the Copernicus Sentinel missions (Sentinel-1, -2, -3, -5P) and additional sources. It provides easy access to key datasets from the Copernicus Land, Marine, Atmosphere, and Emergency Monitoring Services. CDSE provides flexible access and search capabilities that allow users to efficiently find and retrieve data. Users can filter datasets by spatial extent, time

[39] https://catalogue.ceda.ac.uk/uuid/af60720c1e404a9e9d2c145d2b2ead4e/
[40] https://iiasa.ac.at/models-tools-data/flam
[41] https://catalogue.ceda.ac.uk/uuid/9c666602b89e468493e1c907a4de62ff/
[42] https://catalogue.ceda.ac.uk/uuid/f1c9c7aa210d4564bd61ed1a81d51130/
[43] https://essd.copernicus.org/articles/10/2015/2018/
[44] https://cds.climate.copernicus.eu/datasets/reanalysis-era5-complete?tab=overview
[45] https://land.copernicus.eu/en/products/corine-land-cover

range, cloud cover, and collection type and access them through a web interface, API, or Spatiotemporal Asset Catalog (STAC)[46]-compliant catalogue.

The Spatiotemporal Asset Catalog is an Open Geospatial Consortium (OGC) standard designed to standardise the description, organisation and querying of spatiotemporal geospatial data. It provides a common framework for managing Earth Observation assets across a wide range of data types, platforms and applications.

The CDSE platform includes datasets from the **Copernicus Land Monitoring Service (CLMS)**, allowing users to access both near real-time and historical data through standard interfaces. CLMS products are made available via S3 or NFS protocols. The service offers three types of coverage: global, pan-European, and local, along with imagery and reference data (IAR).

The **Copernicus Emergency Management Service (CEMS)** delivers geospatial data and imagery to support timely decision-making during emergencies. It continuously monitors Europe and the globe for early signs or real-time evidence of disasters. CEMS products are generated from satellite, remote sensing, in-situ, and modelled data. One key offering is Rapid Mapping, which provides geospatial information within hours or days of a request to assist emergency response. These datasets are categorised by emergency type, including floods, fires, earthquakes, epidemics, humanitarian crises, industrial accidents, storms, volcanic activity, and mass movements.

The **Copernicus Atmosphere Monitoring Service (CAMS)**, implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF), provides continuous, free, and open access to data and information on atmospheric composition. CAMS offers current observations, forecasts, reanalyses, and consistent retrospective data records for recent years. In addition to raw satellite imagery, the CDSE provides a wide range of derived products that add value for analysis and applications. These include indices and processed datasets such as NDVI (Normalized Difference Vegetation Index), LAI (Leaf Area Index), surface reflectance, cloud masks, and land cover classifications. Products are available in standard geospatial formats, including GeoTIFF, NetCDF, and JPEG/PNG for quick-look previews, and are accompanied by metadata describing spatial extent, acquisition date and time, sensor, processing level, and quality indicators. It offers a wide range of complementary data sources. These include high-resolution satellite imagery from various commercial and institutional providers, as well as curated datasets from different Copernicus Services:

- Soil Moisture and Ocean Salinity (SMOS)
- Medium Resolution Imaging Spectrometer (MERIS) – ENVISAT
- Landsat – 5, -7, -8, -9
- Sentinel-1-related products such as RTC (Radiometrically Terrain Corrected), CARD-BS (Terrain-Corrected Backscatter), and Orbits
- Sentinel-2-based global mosaics, and land cover for Europe and Poland (S2GLC)
- Digital Elevation Models (COP DEM and SRTM DEM)

[46] https://www.ogc.org/standards/stac/

5.2 BioClima EO monitoring platform

This chapter outlines how the BioClima EO monitoring network will ingest data from multiple sources and maintain consistent metadata standards throughout the entire process. A key element to this approach will be the use of the STAC catalogue, which not only enables efficient discovery and management of diverse datasets but also acts as a metadata harmonisation layer. By applying a common structure for spatial, temporal and descriptive metadata elements, STAC ensures that all datasets are described in a uniform way. As a result, users can access and work with data from any source through a single, coherent and interoperable interface.

To meet the demanding requirements for storing, accessing, and discovering EO data, the platform incorporates an S3-compatible object storage system as a fundamental part of its infrastructure. This choice is guided by an assessment of interoperability needs across the platform's services, with emphasis on high availability, horizontal scalability, and straightforward integration. Adopting S3-compatible storage follows established industry practice, as the S3 API has become the standard for cloud-native data management and is widely used in leading EO platforms such as the Copernicus Data Space Ecosystem.

Complementing the storage backend, a publicly accessible STAC API (implemented using stac—fastapi) provides a unified framework for indexing, harmonising, and discovering all datasets stored in the object repository. The STAC specification forms the first layer of metadata alignment, ensuring that EO datasets share a common spatial, temporal and metadata description. This enables users and applications to search and filter geospatial assets based on consistent metadata without needing to interact directly with the underlying storage structure.

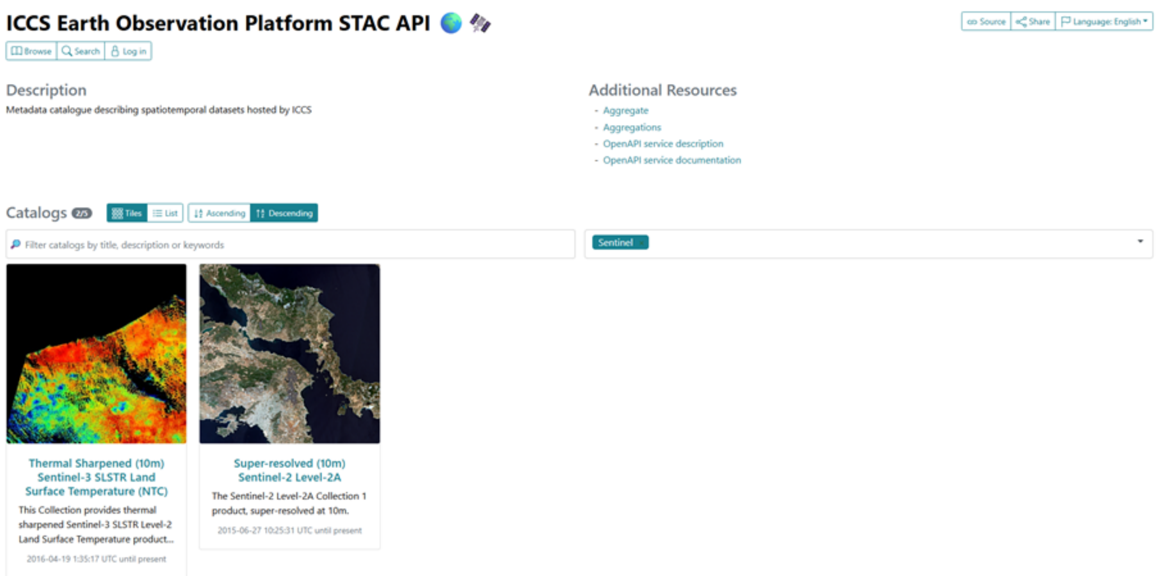

Figure 1: STAC Browser

Figure 2: STAC API

Building on the STAC API, we have integrated the STAC Browser to offer a user-friendly web interface for navigating and inspecting available datasets. As shown in Figure 1, the browser presents each dataset with its associated metadata and preview imagery. Through this interface, users can easily browse collections, inspect metadata tables, and interact with the data catalogue. The browser supports spatial and temporal filtering, allowing users to define an area of interest directly on the map or specify data ranges to refine their search. In addition, users can retrieve asset URLs for direct download or extract STAC item references for programmatic use, enabling straightforward integration into their workflows.

5.3 BioClima EO platform data identification/authorisation

An authentication/authorisation layer has been integrated into the platform to safeguard sensitive datasets and support controlled access. This system is built on Keycloak, using the OIDC standard to manage user identities and permissions in a unified way across all platform components, including the S3 storage service, the STAC API, and the STAC Browser.

The platform follows a strict "default-deny" policy – only explicitly public datasets are visible to unauthenticated users, while restricted resources remain completely hidden unless the user has the appropriate permissions. These permissions are assigned through Keycloak roles, which map directly to specific STAC collections or S3 buckets.

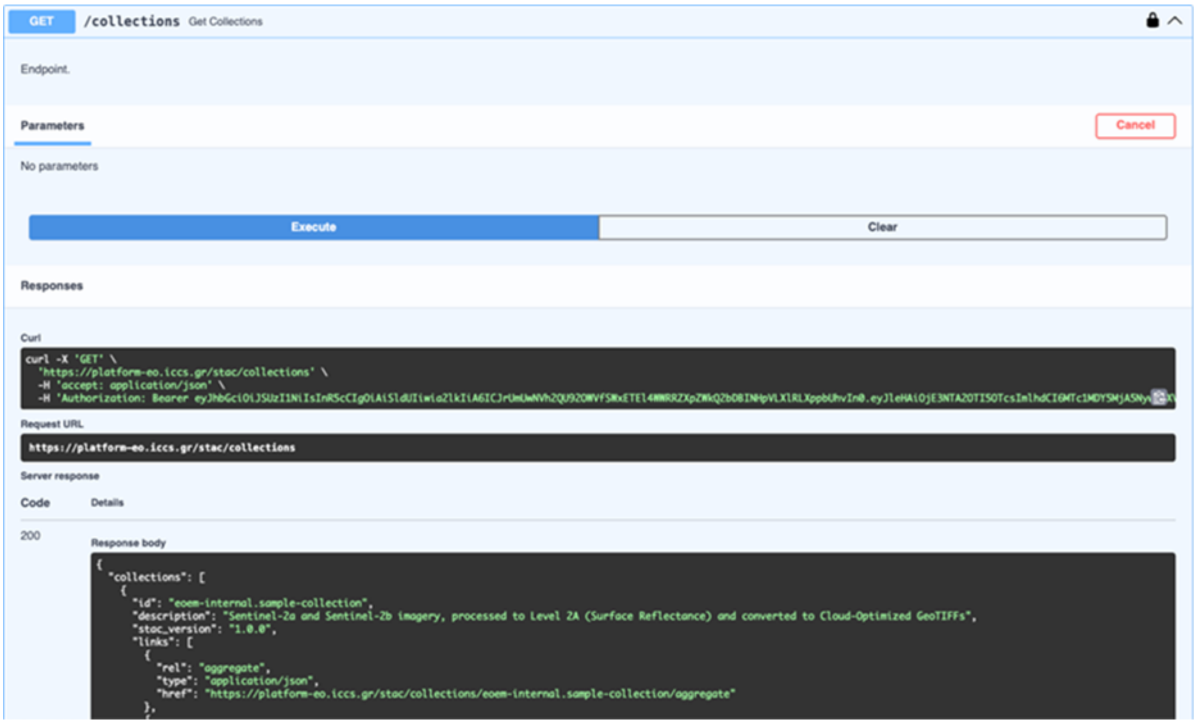

Figure 3: Authenticated request to STAC API

To access data in the object store via the S3 API, users can use the AssumeRoleWithWebIdentity endpoint. This process allows a user to exchange a JWT token for temporary S3 credentials, providing secure, time-limited access to the stored datasets (Figure 4)

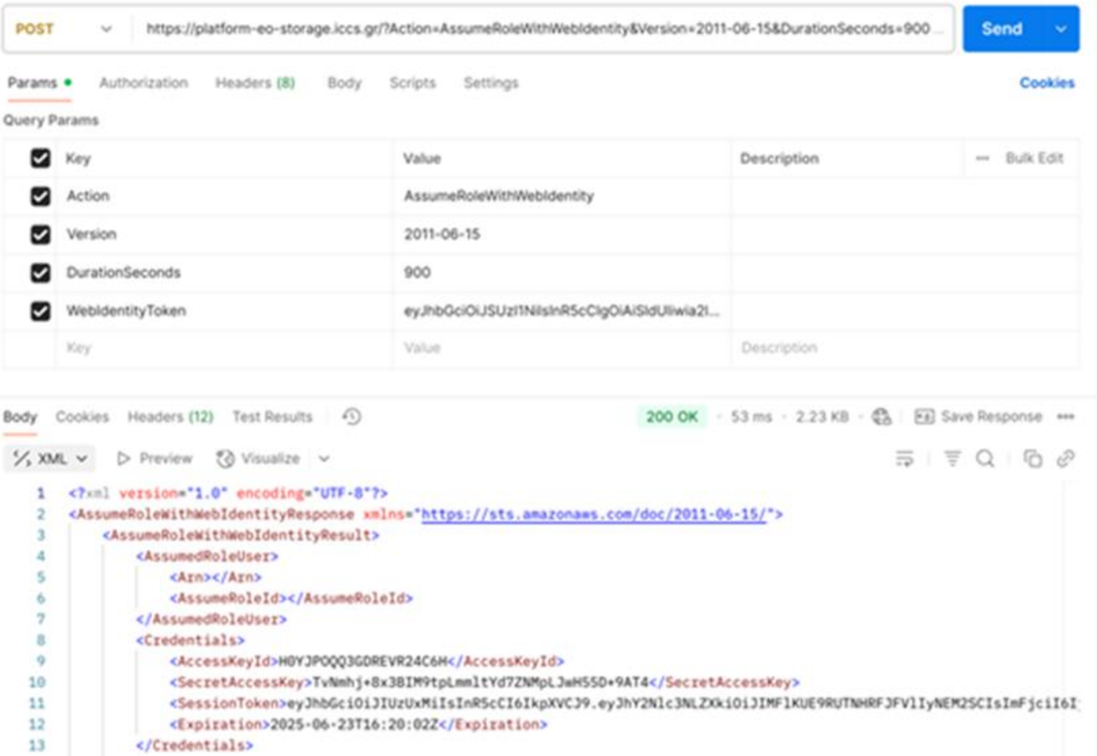

Figure 4: Request to get temporary S3 credentials

6. Crowdsourcing tool based on Geo-Wiki

6.1 Background and Motivation

Understanding the drivers behind changes in forest biomass is essential for effective land use and climate policy, as well as for monitoring habitat health and development. While remote sensing can quantify where biomass is increasing or decreasing, it does not inherently explain why these changes occur. The Geo-Wiki platform[47] provides a scientifically validated, scalable crowdsourcing framework that bridges this gap by enabling expert- and citizen science-based visual interpretation of satellite imagery and vegetation indices. It has been successfully applied to detect drivers of deforestation[48], land cover change[49], and forest management[50] at global and regional levels.

In this project, we will apply Geo-Wiki to identify and map the drivers of forest biomass change in the EU and China, leveraging recent developments in remote sensing (e.g., ESA CCI Biomass, Sentinel imagery) and participatory data collection approaches. This activity will support biomass and biodiversity monitoring, as well as forest management strategies, by providing spatially explicit, driver-specific statistics and insights.

Our objectives include:

- Identify and classify the main drivers of forest biomass gain and loss over the past decade (2015–2023) in the EU and China.
- Quantify the contribution of specific drivers (e.g., afforestation, harvesting, fire, storm, urban expansion) across spatial and ecological gradients.
- Validate satellite-based biomass change maps through visual interpretation and crowd-based verification.
- Generate a harmonised dataset of biomass change drivers to support downstream modelling and policy assessments.

6.2 Methodological Approach

6.2.1 Data Inputs

[47] https://www.geo-wiki.org/

[48] Laso Bayas, J.C., See, L., Georgieva, I. et al. Drivers of tropical forest loss between 2008 and 2019. Sci Data 9, 146 (2022). https://doi.org/10.1038/s41597-022-01227-3

[49] Fritz S., McCallum I., Schill C., Perger C., See L., Schepaschenko D., van der Velde M., Kraxner F., Obersteiner M. (2012) Geo-Wiki: An online platform for improving global land cover. Environmental Modelling & Software. V.31, P. 110-123. https://doi.org/10.1016/j.envsoft.2011.11.015

[50] Lesiv, M., Schepaschenko, D., Buchhorn, M. et al. Global forest management data for 2015 at a 100 m resolution. Sci Data 9, 199 (2022). https://doi.org/10.1038/s41597-022-01332-3

We will use the upcoming version of ESA CCI Biomass maps (v7), which provide aboveground biomass (AGB) estimates at 100 m resolution for 2015 and 2023. These maps will serve as the baseline for identifying areas of significant biomass gain or loss.

To ensure representative coverage and statistical robustness, a stratified random sampling approach will be applied. Sampling will consider: (1) Magnitude of biomass change (increase or decrease); (2) Geographic distribution (ecoregions, administrative units); (3) CCI map quality flags; (4) Known hotspots of change (e.g., afforestation in China).

Each sample location will be visually interpreted using time series for very high-resolution (VHR) satellite imagery (e.g., Google Earth, Bing Maps) through the Geo-Wiki interface.

6.2.2 Crowdsourcing Campaign

We will organise a targeted crowdsourcing campaign through Geo-Wiki, involving:

- Citizen scientists and trained volunteers, guided by detailed classification protocols.
- Experts in forestry and land use, particularly for ambiguous or complex areas.

Participants will interpret biomass change locations and select from a predefined list of potential drivers, including:

- Afforestation/reforestation
- Harvesting or thinning
- Natural disturbances (fire, windthrow, drought)
- Land-use change (e.g., agriculture or urban expansion)

The campaign will include training materials, a quality control system, and feedback loops to ensure data consistency and reliability.

6.2.3 Output Generation and Analysis

Crowdsourced interpretations will be aggregated to produce:

- Driver-specific statistics at national, regional, and ecological zone levels.
- Uncertainty metrics based on interpretation consensus and known data quality issues.

These outputs will serve as validation and contextual layers for forest modelling (e.g., G4M) and help attribute observed biomass dynamics to anthropogenic or natural causes and contribute to habitat and biodiversity monitoring.

7. Airborne/UAV biodiversity and climate change monitoring network

7.1 UAV Biodiversity & Climate Monitoring Network in Romania – Overview

The UAV monitoring network consists of four core sites (Bahna, Oglanic, Histria, MicroRezervație (Constanta)), repeatedly surveyed between June and October 2025 using a combination of Matrice 350, DJI Phantom 4 Multispectral, and DJI Mavic 3 Multispectral platforms. Across these sites, the network integrates LiDAR (L2), thermal (H20T), and multispectral (P4M, M3M) sensors to jointly monitor:

- Turtle presence, detection and identification, especially in aquatic and wetland habitats.
- Vegetation structure and condition, including canopy height, density, and habitat heterogeneity.
- EBVs related to ecosystem extent, vegetation structure, and species/habitat distribution.
- ECVs proxies such as land surface temperature and microclimatic patterns in and around turtle habitats.

7.1.1 Platforms and Sensors in Support of EBVs and ECVs

Across the network, the following platform–sensor combinations are used:

- Matrice 350 + L2 LiDAR + H20T thermal
    - Provides 3D canopy and terrain structure, enabling EBVs related to vegetation height, vertical complexity, and habitat structure (e.g., basking sites, sheltering vegetation, and access corridors for turtles).
    - H20T thermal imagery supports ECV-related indicators, including land/surface temperature patterns, thermal refugia along water–land interfaces, and microclimatic gradients that influence turtle behaviour and habitat suitability.
    - DJI Phantom 4 MultispectralProduces multispectral reflectance and vegetation indices (e.g. NDVI), supporting EBVs such as ecosystem extent, vegetation condition, productivity, and habitat type differentiation (emergent vs submerged vegetation, open bare soil vs vegetated banks).
    - Helps to delineate potential turtle nesting areas, feeding zones, and movement corridors based on vegetation composition and ground cover.
- DJI Mavic 3 Multispectral (starting from September 2025 at Histria)
    - Adds high-quality multispectral mapping over complex wetland and coastal systems, further refining EBVs related to habitat mosaics, vegetation zonation, and transitional habitats that are critical for turtle populations.

Together, these sensors allow joint derivation of structure (LiDAR), condition and composition (multispectral), and thermal environment (H20T) – exactly the combination needed to link turtle presence and behaviour with vegetation structure and local climate conditions.

7.1.2 Site-Level Roles in the Monitoring Network

Bahna (2025-07-12 → 2025-09-30)

- Platforms: Matrice 350, DJI Phantom 4 Multispectral
- Sensors: L2 LiDAR, H20T, P4M multispectral

At Bahna, repeated LiDAR, thermal, and multispectral flights provide a structurally and spectrally rich baseline for inland turtle habitats. L2 LiDAR is used to derive canopy height models and micro-topography, supporting EBVs related to vegetation structure and potential refuge sites (shrubs, riparian belts, forest edges). H20T thermal surveys capture daily and seasonal thermal patterns along water bodies and adjacent land, informing ECV-related metrics (e.g., land–water temperature contrast, thermal stress periods). Multispectral P4M flights complement this with vegetation indices and habitat maps, allowing mapping of basking areas, nesting substrates, and feeding vegetation relevant to turtle ecology.

Eselnița (2025-09-30)

- Platform: Matrice 350
- Sensors: L2 LiDAR, H20T

At Eșelnița, the study site is designed as a controlled experimental environment to systematically assess sensor performance and flight configuration effects for turtle detection and habitat characterisation. Repeated UAV flights are conducted at multiple altitudes using LiDAR, thermal, and RGB sensors over a research centre (belonging to the University of Bucharest) enclosure where turtle presence, location, and body size are precisely known. This controlled setting enables quantitative evaluation of detection limits, spatial resolution thresholds, and sensor complementarity. LiDAR acquisitions support detailed micro-topography and enclosure structure mapping, providing a stable geometric reference for assessing positional accuracy across flight heights.

Histria (2025-06-21 → 2025-10-02)

- Platforms: Matrice 350, DJI Phantom 4 Multispectral, DJI Mavic 3 Multispectral
- Sensors: L2 LiDAR, H20T, P4M multispectral, M3M multispectral

Histria represents the most complex, multi-sensor coastal/wetland node in the network. Repeated flights with LiDAR, thermal, and two different multispectral platforms support a rich set of EBVs and ECVs:

- LiDAR: fine-scale elevation and vegetation structure, including reed beds and shrubs that define habitat mosaics used by turtles.
- Multispectral (P4M, M3M): mapping of aquatic and emergent vegetation, water quality proxies (via spectral indices), and habitat types along lagoon margins and canals.
- Thermal (H20T): detection of thermal heterogeneity in shallow waters, mudflats, and terrestrial basking areas.

These combined datasets enable turtle detection and habitat identification (e.g., basking individuals on logs or rocks in high-resolution RGB/thermal), as well as robust EBVs for ecosystem extent, vegetation condition, and habitat connectivity, plus ECV proxies for surface temperature patterns in coastal wetlands.

MicroRezervație (Constanta) (2025-06-20)

- Platforms: Matrice 350, DJI Phantom 4 Multispectral
- Sensors: H20T, P4M multispectral

MicroRezervație is a compact, intensively monitored site focused on fine-scale habitat characterisation for turtles in a semi-controlled or highly managed environment (micro-reserve / urban nature setting). Thermal imagery (H20T) is used to locate preferred basking and resting micro-sites, while multispectral imagery captures vegetation structure and health in and around ponds and channels. EBVs here focus on local habitat configuration and vegetation structure, while ECV-related indicators track microclimatic conditions that can be linked to turtle activity patterns and potential thermal stress.

Oglanic (2025-07-11 → 2025-09-29)

- Platforms: Matrice 350, DJI Phantom 4 Multispectral
- Sensors: L2 LiDAR, H20T, P4M multispectral

Oglanic extends the inland monitoring network with a mixture of structural, spectral, and thermal data similar to Bahna. LiDAR-based EBVs (canopy height, vegetation roughness, and terrain features) are complemented by multispectral EBVs (habitat types and vegetation vigour) and thermal ECV proxies (surface temperature patterns).

7.2 Data needs for EBVs and ECVs estimation in the case study in Crete - Greece

The estimation of EBVs and ECVs enables further analysis of the structure and function of the semi-arid Mediterranean pine forest ecosystem in south-west Crete. More specific research foci that has been identified can be summarised as:

- Investigate the link between drought-induced Land Surface Temperature (LST) hotspots with Leaf Area Index dynamics
- UAV-LiDAR sub-pixel analysis of Sentinel-2 for improved Above Ground Biomass (AGB) estimation
- Assessing heatwave impacts on ecosystem productivity Net Primary Productivity (NPP) using satellite LST.

Apart from the EO satellite data presented in chapter 5, the following UAV sensors and platforms will be utilized:

- DJI M600 + RIEGL miniVUX-2UAV
    - Provides 3D canopy and terrain structure, enabling estimation of metrics such as vegetation height, canopy delineation, LAI, and Fraction of vegetation in plots. These metrics help further in quantification of AGB
- DJI Mavic 3 Multispectral (Including RGB imagery)
    - Adds high-quality multispectral mapping over the forest that is linked with photosynthetic activity and forest health. The RGB sensors are used to derive ground data of the tree mortality on the plot level that will be compared with ECVs.
- DJI Mavic 3 Thermal
    - Ultra-high-resolution data offer the quantification of LST at very fine scale. This offers the opportunity to scale up and compare LST from UAV and satellite data, and it is also an indicator of water stress in drought-impacted ecosystems.

7.3 Contribution to EBVs and ECVs

Across all sites, the UAV network is designed to link structural, spectral, and thermal information to the ecology of turtles and the dynamics of vegetation:

- EBVs:
    - Ecosystem extent and fragmentation: derived from multispectral classification (water, emergent vegetation, shrub, forest, agricultural land, bare soil).
    - Vegetation structure: canopy height and vertical complexity from LiDAR;
    - Species and habitat distribution proxies: habitat types and microhabitat features that are strongly associated with turtle presence (e.g., basking substrates, nesting sites, sheltered inlets).
- ECVs (proxied by UAV data):
    - Surface/land temperature from H20T: describing thermal regimes at water–land interfaces, heat islands, and cool refugia in vegetation.
    - Local microclimate patterns in structurally complex habitats

8. Software tool for integrating Multi-sensor UAV products in mapping EBVs and ECVs

8.1 Overall Design and Implementation Approach

The software tool has been implemented as a modular Python package, ***uav_biodiv***, designed to integrate heterogeneous UAV data streams (LiDAR, multispectral, RGB, and thermal) and derive **Essential Biodiversity Variables (EBVs)** and **Essential Climate Variables (ECVs)** at very high spatial resolution.

The implementation follows three guiding principles:

1. **Modularity and separation of concerns** – core functionalities are grouped into focused submodules (ingestion, fusion, EBV/ECV derivation, AI analytics, validation, and high-level orchestration).
2. **Extensibility** – the package exposes clear extension points to incorporate additional sensors, indices, EBV/ECV indicators, and machine learning models.
3. **Reproducibility** – configurations, processing steps, and outputs are explicitly parameterised and will be documented or version-controlled at the project level.

The tool is implemented in Python, using widely adopted geospatial and machine learning libraries (e.g., rasterio, geopandas, numpy, and scikit-learn), which allows easy integration into existing research workflows and HPC environments.

8.1.1 Package Structure

The uav_biodiv package is organised into the following submodules:

- config.py – project-level configuration and metadata management.
- ingestion.py – data ingestion and harmonisation utilities.
- fusion.py – multi-sensor data fusion and vegetation index computation.
- ebv_ecv.py – derivation of Essential Biodiversity Variables (EBVs) and Essential Climate Variables (ECVs).
- ai.py – machine learning and AI analytics.
- validation.py – accuracy assessment, uncertainty proxies, and report generation.
- pipeline.py – high-level orchestration of end-to-end workflows.
- sensors/ – sensor-specific helpers (e.g., LiDAR DSM/DTM to CHM, DJI Mavic 3M multispectral).

The root __init__.py file exposes the main classes so that a typical workflow can be written as:

From uav_biodiv import UAVProjectConfig, UAVMonitoringPipeline

This structure allows the same code base to support multiple case studies and monitoring sites by changing only the project configuration and the choice of workflow.

8.1.2 Project Configuration and Metadata Management

The **UAVProjectConfig** class (config.py) centralises all project-level parameters:

- project name and base data directory,
- spatial reference system (CRS),
- optional paths to reference DEM/DTM,
- arbitrary metadata about sensors, acquisition campaigns, or site-specific parameters.

The configuration can be saved and reloaded from JSON, ensuring that processing runs are fully traceable and reproducible. This is particularly important when multiple sites or campaigns are processed in a consistent way (e.g., different forest stands, protected areas, or time steps).

8.1.3 Data Ingestion and Harmonisation

The **DataIngestionHarmonization** class (ingestion.py) manages the ingestion of UAV products and auxiliary datasets:

- **Raster data** (orthomosaics, multispectral cubes, thermal images, DSM/DTM) are loaded using rasterio.
- **Vector data** (plots, training polygons, habitat boundaries) are loaded using geopandas.

A dedicated harmonise_rasters() method provides a central place to implement:

- reprojection to a common CRS,
- resampling to a consistent resolution,
- clipping to a shared spatial extent.

In the current implementation, this method logs the operations and serves as a template where project-specific harmonisation routines (e.g., using rioxarray or rasterio.warp) can be plugged in. This design allows the same interface to be used for different sensor configurations and resolutions, while keeping harmonisation logic in a single, maintainable location.

8.1.4 Multi-Sensor Data Fusion

The **MultiSensorFusion** class (fusion.py) implements the integration of structural and spectral information:

- A generic stack_rasters() method prepares multi-band analysis-ready rasters from several inputs (e.g., RGB + NDVI + CHM), which can later be extended to perform real band stacking on disk.

- The compute_vegetation_indices() method computes standard vegetation indices such as NDVI from NIR and Red bands.
- The fuse_lidar_and_spectral() method merges LiDAR-derived metrics (e.g., canopy height) and spectral indices into a single feature dictionary that can be used as input to machine learning models or indicator derivation.

A **concrete implementation** is provided for **DJI Mavic 3M multispectral data**:

- load_dji_m3m_multispectral() reads a 4-band multispectral orthomosaic (G, R, Red-Edge, NIR).
- compute_dji_m3m_indices() derives key indices for forest applications, including NDVI, GNDVI and NDRE, which are sensitive to chlorophyll content and canopy condition.

These functions encapsulate sensor-specific handling while exposing only high-level structures (band arrays and indices), simplifying their reuse in different workflows.

8.1.5 EBV and ECV Indicator Derivation

The **EBVECVDerivation** class (ebv_ecv.py) provides generic and forest-specific routines for indicator computation, with EBV and ECV expanded explicitly in the documentation:

- **EBV – Essential Biodiversity Variables**
- **ECV – Essential Climate Variables**

1. **Generic indicators**
    a. derive_ecosystem_extent() computes the area (in m²) of each habitat or land-cover class from a classified raster, given a class mapping and the pixel area. This directly supports **EBVs related to ecosystem extent and fragmentation**.
    b. derive_canopy_structure_metrics() calculates simple canopy metrics (mean height, standard deviation, 95th percentile) from a canopy height model (CHM), optionally constrained by a spatial mask (e.g., forest areas only). These metrics support **structural EBVs**, such as canopy height and vertical complexity.
    c. derive_lst_anomalies() estimates land surface temperature (LST) anomalies relative to a reference field or scene mean, serving as a basic **ECV related** indicator for thermal stress.
2. **Forest-specific EBV example**
    a. forest_health_from_indices() combines NDVI and NDRE thresholds to produce a binary forest health mask (healthy vs. unhealthy/non-forest). This is a simple but effective way to translate spectral information into a **forest condition EBV**, which can be further refined or calibrated using field plots and expert knowledge.

The design intentionally separates **indicator logic** from **data access**, making it straightforward to extend with additional EBVs and ECVs as the project evolves (e.g., deadwood proxies, understory exposure, or snow cover dynamics).

8.1.6 AI and Machine Learning Analytics

The **AIAnalytics** class (ai.py) encapsulates the machine learning components:

- train_rf_classifier() trains a **Random Forest** model (scikit-learn) on a feature matrix derived from multi-sensor inputs (e.g. CHM, NDVI, NDRE) and a corresponding label vector (habitat classes, health categories).
- predict_classes() and predict_probabilities() provide a standard interface for inference on new data.
- feature_importances() exposes model-derived importance scores, supporting both model interpretation and feature selection.

This module is deliberately generic: the same interface can accommodate more advanced models (e.g., gradient boosting, convolutional neural networks, or transformers) while remaining backward-compatible. It thus acts as a **plug-in layer** for AI models that operate on the fused multi-sensor feature space.

8.1.7 Validation, Uncertainty and Reporting

The **ValidationReporting** class (validation.py) provides utilities for model evaluation and result summarisation:

- The confusion_matrix() method computes class-wise counts and overall accuracy for a classification task, providing a first-order measure of performance for habitat maps, forest health maps, or other categorical outputs.
- The summary_report() method assembles a human-readable textual summary combining:
    - EBV metrics (e.g., ecosystem extents by habitat class),
    - canopy structure statistics,
    - classification accuracy and confusion matrix.

In the report, sections are labelled explicitly as:

- **[EBV] Essential Biodiversity Variables – Ecosystem Extent (m²)**
- **[EBV] Essential Biodiversity Variables – Canopy Structure Metrics**

This makes the link between code and EBV/ECV concepts transparent for deliverables and documentation.

8.1.8 Sensor-Specific Helpers (LiDAR and DJI Mavic 3M)

The sensors subpackage contains helper functions tailored to particular sensor types:

- sensors/lidar.py implements **compute_chm_from_dsm_dtm()**, which derives a canopy height model (CHM = DSM − DTM) from LiDAR-derived DSM and DTM rasters and writes the result as a GeoTIFF. This CHM is the main structural input for forest EBVs such as canopy height and vertical complexity.

- sensors/dji_m3m.py is reserved for more detailed DJI Mavic 3M handling should additional sensor-specific logic be required (e.g., radiometric correction, band alignment). Currently, DJI-specific operations are implemented in MultiSensorFusion, but this module is ready for future extensions.

This separation keeps the core package sensor-agnostic but allows optimisation and special handling when needed for specific platforms.

8.1.9 High-Level Orchestration: Forest EBV/ECV Workflow (LiDAR + DJI M3M)

The **UAVMonitoringPipeline** class (pipeline.py) orchestrates the end-to-end processing. A concrete method, **run_forest_dji_m3m_lidar_example()**, demonstrates how the modules are combined in a realistic forest monitoring scenario:

1. **Input data**
    a. LiDAR DSM and DTM (GeoTIFF).
    b. DJI Mavic 3M multispectral orthomosaic (4 bands: G, R, RE, NIR).
    c. A habitat classification raster (e.g., 1 = Forest, 2 = Grassland, 3 = Bare Soil).
2. **Processing steps**
    a. **CHM derivation**: DSM and DTM are processed via compute_chm_from_dsm_dtm() to generate a CHM raster, which is then read into memory for metric computation.
    b. **Spectral index computation**: the Mavic 3M multispectral data are loaded using load_dji_m3m_multispectral(), and indices (NDVI, GNDVI, NDRE) are computed via compute_dji_m3m_indices().
    c. **Forest health mask (EBV)**: NDVI and NDRE are combined using forest_health_from_indices() to obtain a simple binary forest health map (healthy vs. unhealthy/non-forest), representing a forest condition EBV.
    d. **Ecosystem extent EBV**: the habitat raster is used in derive_ecosystem_extent() to compute class-specific extents (in square meters) using the pixel area derived from the raster geotransform.
    e. **Canopy structure EBV**: the CHM is masked to forest pixels (habitat class 1), and derive_canopy_structure_metrics() is used to estimate mean canopy height, variability, and upper percentile height.
    f. **Machine learning-based classification**: a feature matrix is constructed by flattening CHM, NDVI, and NDRE, and a Random Forest classifier is trained using train_rf_classifier() to predict habitat classes. Predicted labels are compared to the reference habitat raster using confusion_matrix() to assess performance.
    g. **Reporting**: all metrics (ecosystem extent, canopy structure, classification results) are consolidated into a textual summary via summary_report(), which provides a compact, interpretable overview of the forest EBVs and related ECV indicators derived from the UAV data.
3. **Outputs**
    a. CHM raster for structural analysis.
    b. Index rasters (NDVI, GNDVI, NDRE) for spectral condition.

c. Forest health mask (binary EBV).
d. Quantitative metrics of ecosystem extent and canopy structure (EBVs).
e. Accuracy and confusion matrix for the classification task.
f. A summary report (plain text) suitable for documentation and project deliverables.

This example demonstrates how the package can be used to **operationalise** the extraction of EBVs and ECVs from multi-sensor UAV campaigns in forest environments. The same pipeline can be adapted to other ecosystems (e.g., wetlands, grasslands, agricultural landscapes) by changing the input data and indicator functions while reusing the same core infrastructure.

8.1.10 Extensibility and Future Work

The current implementation deliberately focuses on a robust, extensible foundation:

- New sensors (e.g., thermal cameras, hyperspectral imagers) can be added as separate helper modules in sensors/.
- Additional EBV and ECV workflows (e.g., deadwood detection, understory mapping, snow cover metrics, microclimate proxies) can be implemented as new methods in EBVECVDerivation.
- More advanced AI models (e.g., convolutional networks or transformer-based architectures) can be integrated into the AIAnalytics class while preserving its high-level interface.

9. Conclusion

The deliverable presents a comprehensive and structured overview of the datasets and tools mobilised across the BioClima case studies, covering EO products, in-situ and citizen science measurements, UAV-based observations and harmonized derived datasets. By bringing together spatially extensive satellite observations with high-resolution airborne data and high frequency ground-based measurements, the project establishes a robust, multi-scale observational foundation for biodiversity and climate analysis.

Data Type	Description	Linked EBVs/ECVs
Earth observation	Satellite imagery, remote sensing-derived vegetation indices, land cover and land use maps	EBVs: Ecosystem structure, ecosystem extent, habitat condition ECVs: Land cover, vegetation, surface temperature
Airborne / UAV	LiDAR, multispectral, RGB and thermal	EBVs: Ecosystem structure, habitat conditions, species/habitat distribution (proxies)
In-Situ & Citizen Science	Field surveys, species occurrence and abundance data, phenology observations, crowdsourced interpretations	EBVs: Species distribution, population abundance, phenology
Sensor/ IoT	Soil moisture sensors, air quality sensors, weather stations	ECVs: Soil moisture, temperature, precipitation EBVs: Habitat condition
Derived / processed data	Harmonized maps, biomass and forest area products, habitat suitability layers, anomaly metrics.	EBVs: Ecosystem structure, ecosystem extent, habitat condition ECVs: Vegetation indices, hydrological and climate variable

Table 1: BioClima Data types and the related EBCs/ECVs

The integration of these datasets directly supports key EBVs such as species distribution, population abundance, phenology, and ecosystem structure, as well as ECVs including temperature, precipitation, soil moisture, vegetation indices, and land surface characteristics. Derived products, such as habitat suitability maps and environmental anomaly metrics, further translate raw observations into actionable insights for monitoring and decision-making.

By systematically linking collected and derived data to established EBVs and ECVs, the project ensures interoperability, consistency, and comparability across pilot sites. This alignment provides a solid foundation for predictive modelling, conservation planning, policy support, and the long-term assessment of biodiversity and climate interactions, reinforcing the value of integrated environmental monitoring.

10. Bibliography

Laso Bayas, J.C., See, L., Georgieva, I. et al. Drivers of tropical forest loss between 2008 and 2019. Sci Data 9, 146 (2022). https://doi.org/10.1038/s41597-022-01227-3

Fritz S., McCallum I., Schill C., Perger C., See L., Schepaschenko D., van der Velde M., Kraxner F., Obersteiner M. (2012) Geo-Wiki: An online platform for improving global land cover. Environmental Modelling & Software. V.31, P. 110-123. https://doi.org/10.1016/j.envsoft.2011.11.015

Lesiv, M., Schepaschenko, D., Buchhorn, M. et al. Global forest management data for 2015 at a 100 m resolution. Sci Data 9, 199 (2022). https://doi.org/10.1038/s41597-022-01332-3