



D6.5 Data Management Plan

(interim)



**Funded by
the European Union**

Document Identification			
Related WP	WP6	Document reference	D6.5
Status	Final	Submissions date	30/06/2026
Related deliverables	N/A	Dissemination Level	PU
Lead Author	Tomas Mildorf (UWB)	Version	1.0
Co-Authors	Monika Mundilova (UWB)	Reviewers	Sarka Horakova (Plan4all)
Keywords	data management, FAIR, open	Website	https://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	Change editors	Changes
0.1	03/06/2025	Tomas Mildorf (UWB)	First draft for consortium review
0.2	27/06/2025	Monika Mundilova, Tomas Mildorf (UWB)	Incorporation of reviewer's comments
1.0	30/06/2025	Tomas Mildorf (UWB)	FINAL VERSION TO BE SUBMITTED
1.1	10/06/2026	Monika Mundilova, Tomas Mildorf (UWB), Sarka Horakova (Plan4All)	DMP updated
1.2	15/06/2026	Dubravka Vejinovic	Reviewed by the external ethics advisor

Table of Contents

1. Data Summary	9
Data and Research Outputs Generated by the Project	10
2. FAIR Data	13
2.1. Making Data Findable, Including Provisions for Metadata	14
2.2. Making Data Accessible	15
2.3. Making Data Interoperable	16
2.4. Increase Data Re-Use	16
3. Other Research Outputs	18
4. Allocation of Resources	20
5. Data Security	22
6. Ethics	24
7. Conclusions	25
Annex 1	
Template - Consent Form for Data Collection	26

List of Acronyms

Acronym	Description
AI	Artificial Intelligence
API	Application Programming Interface
CC	Creative Commons
CC0	Creative Commons Zero (Public Domain Dedication)
CC BY	Creative Commons Attribution
DPO	Data Protection Officer
DOI	Digital Object Identifier
DSL	Data Security Law (China)
EBOCC	European Biodiversity Observation Coordination Centre
EBV	Essential Biodiversity Variable
EC	European Commission
ECV	Essential Climate Variable
EML	Ecological Metadata Language
EO	Earth Observation
EOSC	European Open Science Cloud
ESA	European Space Agency
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Reusable
GBIF	Global Biodiversity Information Facility
GDPR	General Data Protection Regulation
GEMET	General Multilingual Environmental Thesaurus

Acronym	Description
GIS	Geographic Information System
INSPIRE	Infrastructure for Spatial Information in the European Community
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
ML	Machine Learning
ODbL	Open Database License
OGC	Open Geospatial Consortium
OWL	Web Ontology Language
PIPL	Personal Information Protection Law (China)
RIA	Research and Innovation Action
SCC	Standard Contractual Clause
SPARQL	SPARQL Protocol and RDF Query Language
TDWG	Biodiversity Information Standards (Taxonomic Databases Working Group)
TLS	Transport Layer Security
UWB	University of West Bohemia in Pilsen
WP	Work Package
WMS/WFS/WCS	Web Map / Feature / Coverage Service (OGC standards)
XAI	Explainable Artificial Intelligence
XML	Extensible Markup Language

Executive Summary

The **BioClima** project, funded under the Horizon Europe Research and Innovation Actions (RIA) framework, seeks to advance integrated climate and biodiversity monitoring through the combination of Earth Observation (EO) data, in-situ measurements, and artificial intelligence (AI)-based analytics. Its overarching objective is to contribute to the refinement of Essential Biodiversity Variables (EBVs) and Essential Climate Variables (ECVs), develop interoperable data infrastructures, and support EU–China collaboration for global environmental monitoring and policy.

This Data Management Plan (DMP) sets out the principles and procedures to ensure that all data and research outputs produced or reused within BioClima are handled according to the FAIR principles while also meeting relevant ethical, legal, and technical requirements.

The DMP outlines:

- The types of data to be generated and reused, including EO imagery, biodiversity records, climate models, sensor data, and stakeholder engagement outputs.
- The use of trusted platforms such as **Zenodo** (via a dedicated BioClima community) and the **BioClima Platform** operated by Plan4all for open data publication, visualisation, and access through APIs.
- The adoption of standardised metadata formats (e.g., ISO 19115, INSPIRE, and Dublin Core) and persistent identifiers (e.g., DOIs).
- Open licensing strategies to maximise reuse while respecting legal and contractual constraints.
- Technical and organisational safeguards to ensure data security, including GDPR-compliant storage, encryption, and role-based access control.
- Ethical measures, especially concerning stakeholder data and the use of AI.

The document also addresses additional research outputs beyond datasets—such as software tools, analytical models, training materials, and policy briefs—and describes how these will be openly shared, licensed, and preserved.

Clear roles, responsibilities, and resources for data stewardship are defined within the consortium, ensuring that data management is an integral and sustained aspect of the project's implementation. This DMP will be updated regularly to reflect new developments, ensuring continued alignment with scientific, ethical, and policy priorities.

By embedding FAIR, open science, and ethics-by-design principles throughout, BioClima's data management strategy supports not only excellent research but also broader societal and environmental impact.

1. Data Summary

The BioClima project aims to improve biodiversity and climate monitoring by integrating Earth Observation (EO) data, in-situ and citizen-science observations, airborne/UAV data, environmental sensor data, modelling outputs and artificial intelligence (AI)-driven analytics. The project supports the development of harmonised monitoring approaches across Europe and China, with a particular focus on Essential Biodiversity Variables (EBVs) and Essential Climate Variables (ECVs), geospatial reporting indicators and policy-relevant decision support.

This Data Management Plan does not reproduce the technical dataset descriptions provided in the project deliverables. Instead, it summarises the main data categories from a data management perspective and identifies where detailed information can be found. The detailed landscape analysis of data sources, tools and processing approaches is provided in Deliverable D1.1, *Analysis of Data for Integration*. The preliminary framework for geospatial reporting indicators, interoperability requirements and relevant standards is provided in Deliverable D1.2, *Biodiversity Specific Geospatial Reporting Indicators – Preliminary Framework*. Communication, exploitation and uptake of the resulting data services and tools are addressed in Deliverable D5.1.

Purpose of Data Collection/Generation/Re-use

Data in BioClima will be collected, reused, and generated to:

- Support integrated biodiversity and climate monitoring across the project case studies.
- Support the estimation, harmonisation and validation of EBVs and ECVs.
- Provide inputs for AI-based and data-fusion approaches.
- Enable cross-scale comparison between satellite, UAV, in-situ, citizen-science and modelled data.
- Support geospatial reporting indicators relevant to biodiversity, climate adaptation and ecosystem condition.
- Contribute to interoperable EU–China monitoring approaches and demonstration activities.
- Underpin scientific outputs, policy recommendations, training materials and exploitable data services.

The DMP focuses on how these data and related outputs will be documented, governed, shared, protected and preserved. Technical descriptions of specific data sources, sensors, workflows and processing tools are maintained in the relevant technical deliverables.

Main Categories of Data Reused and Generated

BioClima works with both reused external data and newly generated project outputs. This DMP summarises these categories solely from a data management perspective. Detailed technical descriptions of sources, tools, processing approaches, and indicator development work are provided in the relevant technical deliverables, especially D1.1 and D1.2.

Data reused by the project

BioClima reuses data from established biodiversity, climate, Earth Observation, environmental monitoring, citizen science, research infrastructure, and case-study sources. These include, for example, satellite and other EO data, biodiversity observation records, environmental and climate measurements, in-situ monitoring data, citizen-science observations, national or regional datasets, and partner-provided case-study data.

The detailed overview of data sources currently identified for integration is provided in D1.1, *Analysis of Data for Integration*. The DMP does not reproduce this inventory. For data management purposes, partners using external datasets are expected to respect the original access conditions, licences, citation requirements, quality limitations and restrictions on redistribution. Where reused data are incorporated into BioClima workflows, sufficient provenance and metadata should be retained to identify the original source, version, licence, processing history and any limitations affecting reuse.

Data and Research Outputs Generated by the Project

BioClima generates new data and research outputs through data harmonisation, modelling, AI-supported analysis, UAV processing, crowdsourcing, platform development, indicator development, standards-related activities, stakeholder engagement, and communication. These outputs may include harmonised datasets, derived EO or UAV products, EBV/ECV-related products, model outputs, AI training or validation datasets, Geo-Wiki outputs, metadata records, software tools, workflows, policy-facing outputs, training materials and documentation.

The maturity and publication status of these outputs will evolve over the course of the project. D1.1 provides the current description of WP1 data and tool outputs, while D1.2 provides the preliminary framework for geospatial reporting indicators, interoperability requirements and standards. As outputs mature, partners will document them using the most appropriate task-level, platform-level, repository-level, or reporting tools, ensuring that provenance, ownership, access conditions, licensing, quality, sensitivity, and preservation requirements are addressed before internal sharing, publication, or long-term archiving.

Data Volume and Technical Characteristics

The data volume remains dependent on the spatial and temporal resolution of EO datasets, UAV acquisitions, modelling outputs and AI training datasets. The current estimate remains in the range of 1–10 TB, with the largest expected volumes associated with EO imagery, UAV data, derived raster products and AI/model training datasets.

The main technical data types expected in BioClima include:

- Geospatial raster data, including EO imagery, UAV products, indices and modelled surfaces.
- Geospatial vector data, including boundaries, observation points, habitats and derived spatial features.
- Tabular data, including species records, environmental measurements and metadata tables.
- Time series data, including climate, environmental and sensor observations;
- Image data from satellite and UAV sources.
- Model and AI outputs, including predictions, scenario results, classifications and validation metrics.
- Software, scripts, notebooks, APIs and workflow descriptions.
- Documentation, reports and training materials.

Sensitive and Restricted Data

Most BioClima datasets are expected to be non-sensitive and suitable for open or controlled sharing, subject to licensing, quality assurance and IPR checks. However, some data may require additional safeguards because of legal, ethical, ecological, confidentiality or international data-governance considerations.

This may include personal data from stakeholder engagement or citizen-science activities, images or observations that could inadvertently capture individuals or private property, sensitive biodiversity data such as exact locations of endangered species or vulnerable habitats, restricted third-party or partner-provided data, and data affected by EU–China or other international exchange requirements.

Such data will be managed according to the principle “as open as possible, as closed as necessary”. Depending on the risk and applicable obligations, safeguards may include anonymisation, pseudonymisation, aggregation, spatial masking, resolution reduction, delayed publication, controlled access, or non-public retention.

Where personal data are collected for the stakeholder database or related engagement activities, the information and consent procedure provided in Annex 1 will be used.

The detailed ethics assessment and recommendations are provided in Deliverable D7.2, *First Ethics Report*. Operational safeguards are further addressed in the Data Security and Ethics chapters of this DMP.

Intended Reuse and Long-term Value

BioClima data and outputs are expected to support future research, environmental monitoring, policy development and operational uptake beyond the project lifetime. Their main reuse value lies in:

- Improving understanding of climate-biodiversity interactions.
- Supporting EBV and ECV estimation and validation.
- Contributing to interoperable biodiversity and climate monitoring workflows.
- Supporting EU and international policy frameworks such as the EU Biodiversity Strategy, the European Green Deal, the Kunming–Montreal Global Biodiversity Framework and climate adaptation objectives.
- Enabling EU–China cooperation on integrated biodiversity and climate observation.
- Supporting future data spaces, observatories and standardisation activities.
- Providing reusable tools, workflows and training resources for scientific and policy communities.

Measures to ensure findability, accessibility, interoperability, reuse, licensing, security and preservation are described in the following chapters of this DMP.

Alignment with European Coordination Initiatives

BioClima is aligned with emerging European initiatives aimed at strengthening biodiversity observation and data coordination, including the proposed European Biodiversity Observation Coordination Centre (EBOCC). EBOCC seeks to provide a long-term governance and coordination structure for biodiversity data and monitoring infrastructures across Europe. BioClima's outputs, including AI-derived EBVs, harmonised EO-in-situ datasets, and monitoring workflows, are designed to be interoperable with and contribute to such future coordination frameworks. Where relevant, BioClima will ensure its data formats, metadata, and indicators are compatible with EBOCC-relevant standards and community practices, facilitating potential integration or support for this evolving European infrastructure.

2. FAIR Data

BioClima is committed to implementing the FAIR principles—ensuring that all data generated, reused, or curated in the project is **Findable, Accessible, Interoperable, and Reusable**. These principles support transparent research, data-driven innovation, and long-term impact through knowledge sharing and reproducibility.

The FAIR data approach in BioClima covers both technical and organisational aspects. It defines how project partners will create, document, store, publish, and preserve datasets and other research outputs during and after the project.

This chapter describes the specific steps the consortium will take to:

- Make all relevant data and metadata findable and traceable through persistent identifiers and cataloguing in trusted repositories.
- Ensure open or restricted access to data in accordance with the principle “as open as possible, as closed as necessary,” respecting licensing, confidentiality, and GDPR obligations.
- Promote interoperability by using standard metadata schemas, vocabularies, and data formats.
- Maximise the potential for data reuse through clear licensing, detailed documentation, and long-term preservation strategies.

BioClima partners recognise that FAIRness is a process rather than a binary state. Accordingly, data management procedures will be reviewed and improved iteratively throughout the project in line with technical developments, user needs, and policy requirements.

The primary tool to be used for sharing open research data generated by BioClima is Zenodo¹. For this reason, a Zenodo PoliRuralPlus community was set up (Figure 1). The community, including its members and uploaded data, can be accessed at

<https://zenodo.org/communities/bioclima>

In addition, geospatial and EO-related resources may be documented, discovered or accessed through project platform components where appropriate, including metadata catalogues, APIs and related services. Detailed technical descriptions of the platform architecture and interoperability framework are provided in D1.1 and D1.2.

¹ Zenodo (<https://zenodo.org/>) is an open-access repository developed under the European OpenAIRE program and operated by CERN. It enables researchers to share and preserve any research outputs in any size, format, and from all fields of science. Zenodo allows users to assign a Digital Object Identifier (DOI) to their data to ensure each dataset is citable and trackable.

2.1. Making Data Findable, Including Provisions for Metadata

The BioClima project is committed to ensuring that all relevant datasets—both reused and newly generated—are easily findable by humans and machines. This is achieved through the use of persistent identifiers, structured metadata, indexing in searchable repositories, and adherence to metadata standards.

Persistent Identification

- All datasets generated and made publicly available by the project will be assigned a **Digital Object Identifier (DOI)**.
- DOIs will be issued via the **BioClima Zenodo community**, ensuring long-term persistence, traceability, and citation.
- Each dataset will be linked to associated publications, deliverables, and project metadata via the EU Funding & Tenders Portal and OpenAIRE.

Metadata Standards and Structure

Each dataset will be described using rich metadata based on well-established standards:

- **ISO 19115/19139** for geospatial data.
- **INSPIRE** metadata profiles for environmental datasets.
- **Dublin Core** for general research outputs.
- **Ecological Metadata Language (EML)** for biodiversity-related datasets.

Searchability and Indexing

Publicly available datasets will be published in:

- The **BioClima Zenodo community**.
- **OpenAIRE**, automatically harvesting Zenodo records.
- **Copernicus DIAS, GEOSS Portal, and GBIF**, where applicable.

Metadata records will be searchable using **standardised keywords and controlled vocabularies**, such as:

- **GEMET** (General Multilingual Environmental Thesaurus).
- **INSPIRE themes**.
- **TDWG vocabularies** for biodiversity data.

For EO and other geospatial assets managed through project platform components, BioClima will also use STAC-compatible metadata where appropriate. STAC provides a common structure for spatial, temporal and descriptive metadata and supports discovery of heterogeneous EO and derived geospatial datasets through a consistent interface. The detailed technical approach is described in D1.1.

2.2. Making Data Accessible

The BioClima project adheres to the principle of making data **“as open as possible, as closed as necessary.”** The consortium ensures that all non-sensitive datasets intended for dissemination are made openly accessible in line with FAIR data principles, the Horizon Europe open science policy, and applicable ethical and legal standards.

Open Access Strategy

The primary repository for harmonised project datasets will be the **BioClima Platform**, operated by **Plan4all**. This platform will:

- Host harmonised datasets, metadata, and services relevant to project use cases and pilots.
- Support visualisation, download, and programmatic access via **standardised APIs** (e.g. OGC WMS/WFS, REST).
- Link each dataset to its original data sources, processing history, and licensing information.

In addition to the platform, all eligible datasets will be published in the **BioClima community on Zenodo**, ensuring:

- Open access under well-defined licenses (e.g. **CC BY**, **ODbL**, or **CC0**).
- Assignment of **Digital Object Identifiers (DOIs)** for traceability and citation.
- Integration with **OpenAIRE** and visibility in the European Open Science Cloud (EOSC).

Interoperable Access via Services and APIs

Harmonised datasets hosted on the **BioClima Platform** will be accessible through **interoperable APIs**, including:

- **OGC-compliant web services** (e.g. WMS, WFS, WCS).
- **REST APIs** for structured access to indicators, time series, and metadata.
- **SPARQL endpoints** for selected semantic data, where applicable.

For platform-hosted EO and geospatial resources, access will be managed in accordance with the technical access model described in D1.1. Public datasets may be made discoverable through open interfaces, while restricted resources will remain subject to authentication and role-based permissions. This supports the DMP principle “as open as possible, as closed as necessary” without requiring all project data to be deposited in open repositories immediately.

2.3. Making Data Interoperable

To achieve high levels of interoperability, the BioClima project will adhere to internationally recognised data and metadata standards, formats, and vocabularies that are endorsed by the research community. For general research data JSON or XML for data formatting. For geospatial data, we will follow the OGC (Open Geospatial Consortium) standards and the INSPIRE directive, which provide frameworks for sharing geographic information and services. These standards ensure that our data is compatible with global systems and can be easily integrated with other datasets and tools.

When uploaded to the repository research data and outputs should only be stored in a preferred file format that conforms to the international standards (based on the KNAW-DANS Preferred Formats², November 2015) to ensure future compatibility. These are some types of formats for long-term preservation of research data that we recommend being used in BioClima:

- Document (.txt; .pdf; .doc; .docx; .odt)
- Spreadsheet (.csv; xls; .xlsx; .ods)
- GIS vector data (ESRI Shapefile, GeoParquet, GeoJSON, PostGIS)
- GIS raster data (.geotiff; .img, NetCDF)
- Database (.csv; .sql; .mdb; .accdb)
- Picture (.jpg; .tif; .png)
- Audio (.wav; mp3)
- Video (.avi; .mp4; .mov)

The interoperability approach will align with the preliminary framework developed in D1.2, which identifies relevant standards and practices for biodiversity, climate, and geospatial reporting data. Where applicable, BioClima will build on OGC, ISO, INSPIRE, TDWG, W3C and related community standards, including API-based access and semantic alignment approaches. The DMP does not duplicate the detailed standards analysis from D1.2 but uses it as the reference framework for future dataset and service-level decisions.

2.4. Increase Data Re-Use

The BioClima project is committed to ensuring that all data outputs—whether reused or newly generated—are accompanied by **clear and standardised licensing information** that enables their broad re-use, while respecting legal, ethical, and contractual constraints.

² <https://dans.knaw.nl/en/file-formats/>

Licensing Strategy

All datasets produced in the project and made publicly accessible will be released under **open licences** that allow reuse, redistribution, and modification.

The choice of licence will depend on the **type of data, degree of sensitivity, and origin**:

- **Creative Commons Attribution (CC BY 4.0)** – Default licence for general research outputs, allowing wide reuse with proper credit.
- **Creative Commons CC0 (Public Domain Dedication)** – Applied to non-sensitive baseline datasets or training datasets, where maximum openness is desired.
- **Open Database License (ODbL)** – Used for spatial datasets where applicable, especially those derived from or compatible with OpenStreetMap.
- **Custom or restricted licences** – Applied only when reusing third-party datasets not covered by open licensing (e.g. public authority data under national access terms).

Before public release, datasets, software, models and other reusable outputs will be checked for applicable third-party licence conditions, IPR considerations, ethical restrictions and sensitivity constraints. Where reuse cannot be fully open, BioClima will provide metadata, documentation, aggregated outputs or controlled-access options where this is legally and ethically appropriate.

Licences will be declared clearly in metadata records, dataset headers, and accompanying README or licence files.

3. Other Research Outputs

In addition to the datasets described in the preceding sections, the BioClima project will generate additional research outputs that contribute **to the project's scientific, technical, policy,** and capacity-building objectives. These may include software, workflows, models, methodologies, documentation, standards-related outputs, training materials, communication materials and policy-facing results.

The DMP does not maintain a detailed inventory of these outputs, as their technical content is documented in the relevant project deliverables and reporting tools. At this stage, examples of such outputs are described in D1.1, D1.2 and D5.1, including platform components, software tools, data-fusion and AI-related methods, interoperability outputs, training materials and exploitation-oriented results.

Software Tools and Platforms

BioClima will develop or adapt software components and analytical tools to support the integration of Earth Observation data, in-situ observations, UAV data, citizen-science inputs and AI-based modelling for biodiversity and climate applications. These may include data pre-processing scripts, model-training workflows, APIs, metadata catalogues, visualisation components, platform services and specialised tools for EBV/ECV-related analysis.

Where feasible, software development will follow standard practices for version control, documentation, testing and reproducibility. Software may be shared through repositories such as GitHub or GitLab and, where appropriate, archived through Zenodo or another suitable repository. Licensing will be decided on a case-by-case basis, taking into account ownership, third-party dependencies, IPR, exploitation plans, security considerations, and ethical or legal restrictions.

Methodologies. Models and Workflows

BioClima will generate methodologies, models and workflows related to EO and in-situ data integration, UAV data processing, data harmonisation, AI/ML analysis, EBV/ECV estimation, scenario analysis, validation and geospatial reporting indicators.

These outputs will be documented through technical deliverables, scientific publications, workflow documentation, platform documentation or other appropriate project records. Where relevant, workflow descriptions, model assumptions, input data, processing steps, validation results and limitations will be documented to support transparency, reproducibility and reuse.

The consortium will maintain provenance information for derived datasets, including references to source datasets, processing workflows, software versions and, where

applicable, AI models used to generate the outputs, in order to support traceability, reproducibility and compliance with FAIR principles.

Training, Educational and Communication Materials

To support dissemination, capacity building and uptake, BioClima will produce training and communication materials such as guidelines, manuals, webinars, tutorials, presentations, e-learning resources, infographics, videos and promotional materials. Public materials will be made available through appropriate channels such as the project website, Zenodo or other dissemination platforms, in line with D5.1.

Standards-Related and Interoperability Outputs

BioClima may also generate standards-related outputs, including metadata profiles, interoperability mappings, validation approaches, semantic resources, best-practice documentation and contributions to relevant standardisation or community processes. The detailed framework for these activities is provided in D1.2 and will be updated through subsequent technical work.

4. Allocation of Resources

The BioClima consortium has allocated appropriate resources—both personnel and infrastructure—to ensure the effective implementation of data management, FAIR principles, and the long-term accessibility of research outputs. Data-related activities are embedded across the work packages that collect, reuse, generate, process, analyse, publish or preserve project data and research outputs.

Roles and Responsibilities

Data management activities are coordinated under the overall leadership of the **University of West Bohemia in Pilsen (UWB)**, which is responsible for ensuring DMP compliance, updates, and integration across the consortium. Specific responsibilities are as follows:

- As Project Coordinator, UWB oversees the Data Management Plan, supports FAIR data compliance, and coordinates reporting to the European Commission on data-management-related issues.
- Plan4all supports the operation and maintenance of project platform components for data publication, visualisation, APIs, metadata catalogues, and access to harmonised data services.
- Work Package Leaders supervise data generation, curation, documentation and reporting within their respective thematic and technical areas.
- Task Leaders and case-study responsible persons manage the practical aspects of data collection, processing, quality control, documentation and transfer within their activities.
- Partners generating or reusing data are responsible for ensuring that relevant metadata, provenance, access conditions, licensing information and sensitivity considerations are documented before sharing, publication or preservation.
- Ethics and data-protection aspects are coordinated with the relevant project management and ethics structures, including consultation with the Independent Ethics Advisor where appropriate.

Zenodo community curation, metadata standard compliance, and training on FAIR practices are coordinated jointly by UWB and Plan4all, with input from partners where needed.

Budget Allocation

Resources for data management are primarily embedded in the personnel and infrastructure budgets of the relevant work packages and tasks, rather than treated as a separate, standalone budget line. This includes resources for:

- Preparation and maintenance of the Data Management Plan, including interim and final updates.
- Operation and maintenance of platform components relevant to data access, APIs, metadata catalogues and integration.
- Preparation and curation of metadata, documentation and licences.
- Data quality control, harmonisation and internal sharing.
- Publication and preservation of appropriate datasets and research outputs in repositories such as Zenodo or other suitable infrastructures.
- Review of access, licensing, IPR, data protection and ethical restrictions before publication.
- Support for long-term accessibility of public research outputs beyond the project lifetime.

Where additional costs arise for storage, publication, long-term preservation, software archiving, or controlled access, these will be handled within the relevant partner, task, or work package budgets in accordance with the Grant Agreement, Consortium Agreement, and project management procedures.

Infrastructure and Tools

To maximise efficiency, BioClima builds on existing, trusted infrastructure:

- **Zenodo (CERN):** For open data publication with DOI assignment and integration into OpenAIRE and EOSC.
- **Plan4all's data infrastructure:** For geospatial data services (WMS/WFS), APIs, metadata catalogues, and the semantic backbone of the BioClima Platform.
- **Institutional repositories:** Partners such as SYKE, LU, and NTNU may also archive outputs in their national or institutional repositories for added redundancy and compliance.

5. Data Security

BioClima will apply appropriate technical and organisational measures to protect project data throughout its lifecycle, from collection and processing to internal sharing, publication and preservation. The level of protection will depend on the type of data, its sensitivity, applicable licences, contractual obligations and legal or ethical requirements.

Project working documents and internal files are stored in the shared project infrastructure described in the Project Management Handbook. Google Drive, hosted and managed by UWB, serves as the central collaboration and document management platform for European partners. Key files may be mirrored on OneDrive to support collaboration with Chinese partners. Partners are expected to avoid storing project documents solely on local or private drives so that relevant files remain accessible, traceable, and backed up.

Open research outputs intended for publication will be deposited in trusted repositories such as the BioClima Zenodo community or other appropriate domain, institutional or project platforms. Platform-hosted EO and geospatial resources will follow the technical access model described in D1.1. Public datasets may be made discoverable through open interfaces, while restricted resources will remain subject to authentication and role-based access control.

The stakeholder database, developed as part of WP5, will contain only the minimum personal data necessary for project communication, engagement, and reporting. Individuals included in the database will be informed about the purpose of processing, data use, retention and their rights through the information and consent template provided in Annex 1. Access to the stakeholder database will be restricted to authorised project staff.

For platform-managed EO and geospatial data, access control will follow the “as open as possible, as closed as necessary” principle. Where applicable, restricted datasets will not be visible to unauthenticated users, and access will be granted only to authorised users or roles. Temporary or time-limited credentials may be used to access protected storage resources when supported by the platform architecture.

Sensitive or restricted data, including personal data, stakeholder records, restricted third-party data, sensitive biodiversity locations and confidential partner-provided information, will be stored only in suitable controlled-access environments. Depending on the risk, safeguards may include pseudonymisation, anonymisation, spatial masking, aggregation, encryption, controlled access, access logging or non-public retention.

Personal data will be processed in accordance with GDPR and relevant national requirements. Partners collecting personal data through stakeholder engagement, citizen-science activities, surveys, workshops or related activities are responsible for

applying data minimisation, appropriate consent procedures, secure storage and controlled access. Identifiable personal data will not be published openly.

Where international data exchange is required, especially in the context of EU-China cooperation, the responsible partners will assess whether the data can be transferred lawfully and securely. Where restrictions apply, appropriate safeguards or alternative approaches, such as controlled access, local processing or federated analysis, will be considered.

Each partner remains responsible for the secure handling of data under its control. Any suspected data security issue, unauthorised access, accidental disclosure, or uncertainty concerning sensitive data must be reported to the relevant Work Package Leader and Project Coordinator. Where appropriate, the Data Protection Officer or designated responsible staff, the Independent Ethics Advisor and relevant institutional support services will be consulted.

6. Ethics

BioClima integrates ethics into project governance, research implementation and data management. Ethical oversight is addressed through WP7 – Ethics Requirements, in coordination with the Project Coordinator, Work Package Leaders, relevant partner responsibilities and the Independent Ethics Advisor.

The detailed ethics assessment for the first reporting period is provided in Deliverable D7.2, *First Ethics Report*. As D7.2 is a dedicated ethics deliverable, this DMP does not reproduce its findings. Instead, the DMP records the data-management implications relevant to ethical compliance and refers to D7.2 for the full assessment and recommendations.

The main ethics-related aspects relevant to the DMP concern personal data, stakeholder and citizen-science engagement, privacy in relation to environmental monitoring technologies, sensitive biodiversity and geospatial data, responsible use of AI-supported workflows, and international cooperation involving non-EU partners.

BioClima follows the principle “as open as possible, as closed as necessary”. Ethical, legal, confidentiality, IPR or ecological restrictions may therefore limit the publication or reuse of some data and research outputs. The practical safeguards for restricted or sensitive data are described in Chapter 5, Data Security.

Where AI-supported methods are used, BioClima will ensure that relevant input datasets, assumptions, limitations and validation approaches are documented as appropriate to the maturity and intended use of the outputs. AI-generated outputs that may inform environmental assessment, policy or management decisions will remain subject to expert interpretation and human oversight. Detailed ethical considerations concerning AI are addressed in Deliverable D7.2 (First Ethics Report) and will inform future updates of this DMP where relevant.

Any new or uncertain ethical issue arising during project implementation will be reported to the relevant Work Package Leader and Project Coordinator. Where appropriate, the Independent Ethics Advisor and relevant institutional data-protection or ethics contacts will be consulted. Subsequent ethics deliverables will continue to inform future DMP updates.

7. Conclusions

This interim Data Management Plan updates the BioClima data management framework on the basis of the project work completed during the first reporting period. It confirms the consortium's commitment to FAIR data management, responsible open science, secure data handling and appropriate long-term preservation of research outputs.

The DMP provides a project-level framework for managing data and other research outputs generated or reused by BioClima. It does not duplicate the technical detail contained in individual deliverables. Instead, it points to the relevant project outputs where detailed descriptions of data sources, tools, workflows, standards, ethics and exploitation pathways are maintained.

The main updates concern the alignment of the DMP with the emerging WP1 data landscape, the BioClima platform and metadata approach, the preliminary interoperability and geospatial indicator framework, the ethics review process, and the practical arrangements for access, security, licensing and preservation.

The DMP will continue to be treated as a living document. It will be updated as datasets, software, models, indicators, platform services and other research outputs mature, and as new legal, ethical, technical or exploitation considerations arise. Future updates will take into account subsequent technical deliverables, ethics reports, project reporting requirements and partner inputs.

Through this approach, BioClima aims to ensure that its data and research outputs are well documented, appropriately protected, reusable where possible, and useful for scientific, policy and societal uptake beyond the lifetime of the project.

Annex 1

Template - Consent Form for Data Collection

The BioClima Project is committed to protecting your privacy and processing personal data in accordance with the General Data Protection Regulation (EU) 2016/679 (GDPR). Detailed information regarding the collection, use, storage, retention periods, recipients, and your rights can be found in the **BioClima Privacy & Data Protection Policy** available on the [BioClima website](#).

Consent Form for Data Collection

[Your Organization's Name]

[Your Organization's Address]

[Your Organization's Contact Information]

Purpose of Data Collection

We are collecting your personal data for the following purposes:

- [List specific purposes, e.g., marketing, service improvement, research, communication regarding project activities, participation in BioClima activities, events, workshops, and pilots]

Data We Collect

We will collect the following personal data:

- [List types of data, e.g., name, email address, phone number, etc.]

Use of Data

Your personal data will be used for the purposes stated above and may be shared with:

- [List third parties if applicable, e.g., marketing partners, service providers]

Retention Period

Your personal data will be retained for:

- [Specify duration, e.g., until the end of the service, for one year, etc.]

Your Rights

You have the right to:

- Access your personal data
- Rectify any inaccuracies in your personal data
- Erase your personal data
- Restrict or object to the processing of your personal data
- Data portability
- Withdraw consent at any time

To exercise these rights, please contact us at [Contact Information].

Withdrawal of Consent

You may withdraw your consent at any time by contacting us at [Contact Information].

Withdrawal of consent does not affect the lawfulness of processing based on consent before its withdrawal.

Contact Information

If you have any questions or concerns about this consent form or our data processing practices, please contact:

Data Protection Officer (if applicable):

Name:

Organization:

Email:

Phone:

Consent

By signing this form, you consent to the collection and use of your personal data as described above.

Name: _____

Signature: _____

Date: _____

Electronic Consent

If you prefer to provide your consent electronically, please tick the box below:

☐ I consent to the collection and use of my personal data as described above.