D6.4 Data Management Plan

(initial)

Funded by
the European Union

Document Identification			
Related WP	**WP6**	**Document reference**	**D6.4**
Status	**Final**	**Submissions date**	**30/06/2025**
Related deliverables	**N/A**	**Dissemination Level**	**PU**
Lead Author	**Tomas Mildorf (UWB)**	**Version**	**1.0**
Co-Authors	**Monika Mundilova (UWB)**	**Reviewers**	**All WP leaders**
Keywords	**data management, FAIR, open**	**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**	**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

Table of Contents

List of Tables

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, nd 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 observations, and artificial intelligence (AI)-driven analytics. The project supports the development of harmonised monitoring systems across the EU and China, contributing to the refinement of Essential Biodiversity Variables (EBVs) and Essential Climate Variables (ECVs), and ultimately providing evidence-based recommendations for policy development.

Purpose of Data Collection/Generation/Re-use

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

- **Support AI model development** for biodiversity and climate monitoring across diverse ecosystems.
- **Populate harmonised indicators** (e.g. EBVs and ECVs) used in scenario modelling, environmental assessments, and policy recommendations.
- **Enable cross-scale validation** between EO data, sensor networks, and ground-based observations.
- **Facilitate EU–China collaboration**, creating interoperable data structures and protocols.
- **Feed into use cases and case studies** conducted across different biogeographic regions in Europe.

The generated data will underpin technical deliverables, scientific publications, and policy briefs, ensuring long-term reuse within and beyond the project consortium.

Overview of Datasets to Be Reused

Table 1 summarises datasets that are expected to be reused within the BioClima project.

Table 1: Datasets to be reused

Dataset / Source	Type	Provider	Purpose
Copernicus EO data (Sentinel 1, 2, 3)	Satellite imagery (multispectral, radar, thermal)	ESA / Copernicus	Monitoring land cover, vegetation, climate variables
EuropaBON datasets	Biodiversity observation data, indicator baselines	EuropaBON	Input to EBV methodology, indicator benchmarking
GBIF (Global Biodiversity Information Facility)	Species occurrence records	GBIF	Species distribution modelling
In-situ sensor data (EU networks)	Climate and environmental	National authorities /	Ground-truthing EO data and validating AI models

Dataset / Source	Type	Provider	Purpose
	monitoring (e.g. temperature, humidity, CO_2)	environmental agencies	
Public climate model outputs (e.g. Copernicus C3S)	Climate reanalysis and projections	ECMWF / C3S	Downscaling climate trends for case studies
Existing case study baseline data	GIS layers, socio-ecological data, land cover	Local and regional authorities, pilot partners	Contextual inputs for selected pilot areas

Overview of Datasets to Be Generated

New data generated by BioClima are included in Table 2.

Table 2: Datasets to be generated

Dataset / Output	Type	Generated by	Purpose
AI-derived indicators	Maps and indicators of EBVs/ECVs	Neuralio, LU, UT-ITC, SYKE	Support for environmental reporting and monitoring
Training and test data for AI	Annotated image sets, feature vectors	NTNU, ICCS, IDC	Used in supervised learning for remote sensing analytics
Use case data collections	Observational data from pilot regions	Pilot partners	Validation of AI models and participatory assessments
Scenario datasets	Outputs from climate and biodiversity models	IIASA, SYKE, LU, UT-ITC	Evaluation of policy impacts and climate-biodiversity synergies
Harmonised EO + ground observation products	Integrated raster/vector datasets	All partners	Interoperable monitoring outputs across pilot sites
Stakeholder engagement records	Workshop results, feedback, and participatory data	pilot partners	Used for the co-design and evaluation of decision-support tools

Volume and Nature of Data

The estimated total data volume ranges from **1–10 TB**, depending on the resolution and frequency of EO data, modelling outputs, and AI training datasets. Most data will be **non-sensitive**, with potential exceptions for biodiversity data (see below) and participatory datasets containing personal information, which will be anonymised and stored securely in compliance with GDPR.

Sensitive biodiversity data - While most datasets in BioClima are expected to be non-sensitive and openly shareable, certain biodiversity-related data—such as the exact

locations of endangered species, nesting sites, or habitats under legal protection—may be considered sensitive. Disclosure of such data could pose risks to conservation efforts or lead to unintended negative impacts, such as poaching or habitat disturbance. Therefore, any biodiversity datasets identified as sensitive will be subject to controlled access procedures, generalisation (e.g., spatial masking or resolution reduction), or restriction in line with conservation best practices and relevant national or international regulations. These safeguards will be implemented during data processing and prior to dissemination to ensure that open science goals are balanced with ecological responsibility.

Intended Reuse and Long-term Value

The datasets developed or enhanced in BioClima are expected to support:

- **Future research** on climate-biodiversity linkages
- **Policy frameworks** such as the EU Biodiversity Strategy and Green Deal
- **Monitoring networks** like EuropaBON and Copernicus
- **Cross-border cooperation** between EU and Chinese environmental institutions
- **Open science initiatives**, subject to ethical and licensing conditions

Detailed measures to ensure reuse, including data documentation, licensing, and hosting, will be presented in Section 2 (FAIR Data).

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 Zenodo[1]. 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

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

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.

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

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

This ensures compatibility with existing tools and platforms used by researchers, environmental authorities, and other stakeholders.

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 Formats[2], November 2015) to ensure future compatibility. These

[2] https://dans.knaw.nl/en/file-formats/

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)

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.

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).

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

3. Other Research Outputs

Apart from the datasets described in the preceding sections, the BioClima project will generate additional research outputs that contribute significantly to achieving the project's overall goals. BioClima will rigorously apply the **FAIR principles** across all categories of outputs, including software, methodologies, models, documentation, and communication materials.

We will utilise robust platforms such as **GitHub** or **GitLab** for collaborative development, version control, and hosting of software tools. An appropriate **open-source licence** (e.g., MIT, GPL, Apache 2.0) will be selected on a case-by-case basis to clarify how others can use and adapt the tools.

This section summarises the key categories of other research outputs and outlines how they will be managed.

Software Tools and Platforms

BioClima will develop several software components and analytical tools to support the integration of **Earth Observation (EO) data**, **in-situ monitoring**, and **AI-based modelling** for biodiversity and climate applications. These tools may include:

- Data pre-processing scripts and model training pipelines
- APIs and services for accessing harmonised indicators
- Web-based visualisation components integrated into the **BioClima Platform**

Where feasible, development will follow standard practices in code quality, documentation, testing, and version control. Software will be:

- Published in open repositories (e.g. GitHub)
- Archived with stable versions and **DOIs via Zenodo**
- Licensed under **permissive open-source licences** to maximise reuse and interoperability

Methodologies and Models

BioClima will generate several analytical methodologies and models, particularly in the areas of:

- Integration and harmonisation of EO, in-situ, and socio-ecological data
- AI/ML models for environmental variable estimation, anomaly detection, and predictive monitoring
- Assessment frameworks for Essential Biodiversity Variables (EBVs) and Essential Climate Variables (ECVs)

These will be documented in **technical deliverables**, peer-reviewed publications, and **workflow documentation** shared via the BioClima Platform. Where appropriate, **semantic descriptions** (e.g. in OWL or JSON-LD) may be used to improve model interoperability and reuse.

Training and Educational Materials

To support capacity building and broader uptake of project outcomes, BioClima will develop and share a variety of training resources:

- Manuals and guidelines for applying AI to EO-based biodiversity and climate monitoring
- Training slides, webinars, and tutorials used in project workshops
- Self-paced online content and multimedia resources for stakeholders and the public

These materials will be made openly accessible through the **project website**, **Zenodo**, and other educational platforms.

Project Reports and Policy Documents

The project will produce a wide range of textual outputs including:

- Deliverables, technical reports, and white papers
- Policy briefs tailored for decision-makers and regulatory audiences
- Synthesis documents and guidelines for harmonised monitoring systems

All public reports will be clearly written, version-controlled, and published on the **project website**, **Zenodo**, and shared via **OpenAIRE** where possible.

Multimedia and Promotional Materials

To support outreach and stakeholder engagement, BioClima will produce:

- Videos, infographics, slide decks, and digital flyers
- Visual storytelling assets explaining the importance of EBVs, ECVs, and integrated monitoring
- Dissemination and conference materials

All such materials will be published in accessible formats under open licences (e.g. **Creative Commons Attribution**), and stored in project communication repositories.

Standards and Guidelines

Where applicable, BioClima will contribute to the development or application of standards, particularly in relation to:

- Metadata and data exchange (e.g. **ISO 19115**, **OGC APIs**, **INSPIRE** guidelines)
- Semantic alignment of biodiversity and climate variables
- Harmonisation of monitoring practices across EU and Chinese contexts

All developed standards or best-practice documents will be aligned with **internationally recognised frameworks** and shared with relevant standardisation bodies and communities, including the OGC as one of the BioClima partners.

4. Allocation of Resources

The BioClima consortium has allocated appropriate resources—both in terms of personnel and infrastructure—to ensure the effective implementation of data management, FAIR principles, and long-term accessibility of research outputs. Data-related activities are embedded across multiple work packages and coordinated centrally to support the full data lifecycle: from collection and generation to sharing, reuse, and preservation.

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:

- **UWB (Scientific Coordinator)**: Oversees the Data Management Plan (DMP), FAIR data compliance, and liaison with the European Commission on data issues.
- **Plan4all**: Maintains and operates the **BioClima Platform**, which hosts harmonised datasets, APIs, metadata catalogues, and AI tools.
- **Work Package Leaders**: Supervise data generation, curation, and documentation within their respective thematic and technical domains.
- **Task-level Data Stewards**: Designated individuals in pilot and technical tasks who manage data quality, standardisation, and metadata entry.

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 activities are primarily embedded in the personnel budgets of Work Package 1 (Project Management), Work Package 2 (Harmonisation), and all technical work packages generating or processing data (WP3–WP6). Specific activities funded include:

- Preparation and maintenance of the Data Management Plan and updates (incl. final version at project close),
- Operation and maintenance of the **BioClima Platform** (hosting, APIs, metadata catalogue, integration).
- Development and curation of metadata and licences.
- Data publication and documentation for internal and external access.
- Long-term preservation in open repositories (e.g. Zenodo, GBIF).

Each partner is responsible for covering the costs associated with the data they generate, including preparation of metadata and documentation, publication to shared platforms, and where applicable, anonymisation or preprocessing for open access.

No separate cost is anticipated for Zenodo or OpenAIRE publishing, as these services are free for Horizon Europe projects. Additional efforts related to licensing, version control, or interoperability (e.g. OGC services) are covered under the technical implementation tasks.

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

Data security is a key priority for the BioClima project. Given the diverse nature of the data involved—including environmental, geospatial, Earth Observation (EO), AI-derived outputs, and potentially personal data from stakeholder engagement—appropriate technical and organisational measures will be applied to ensure **integrity, confidentiality, and availability** throughout the project lifecycle.

Data Protection Officers

Where legally required under the GDPR, partner institutions have appointed Data Protection Officers (DPOs) responsible for advising on data protection obligations, monitoring compliance, and serving as contact points for data subjects and supervisory authorities. For organisations not mandated to appoint a DPO, designated staff members will oversee data protection responsibilities within the scope of the BioClima project. All partners processing personal or sensitive data will ensure appropriate data protection oversight, including risk assessments and secure handling of stakeholder-related or ecologically sensitive information.

Secure Storage and Infrastructure

All project data will be stored on **secure, GDPR-compliant platforms**, including institutional servers, trusted cloud services, and the **BioClima Platform** maintained by Plan4all. The platform will apply **role-based access control**, system monitoring, and routine security updates. Public datasets will also be archived in trusted repositories such as **Zenodo**, which ensure long-term preservation and certified infrastructure security.

Access Control and Authentication

Access to internal data (including pre-publication results and restricted datasets) will be **limited to authorised personnel**. Strong authentication methods, such as secure passwords and role-specific permissions, will be implemented. Internal data access will follow the **principle of least privilege**, and access logs will be maintained where feasible.

Data Encryption

Sensitive data will be encrypted **at rest and in transit** using industry-standard protocols (e.g. **AES-256**, **TLS/SSL**). This includes partner data exchanges, field data uploads, and repository transfers.

Data Minimisation and Pseudonymisation

If personal or stakeholder-related data is collected (e.g. in surveys or interviews), BioClima will apply the **principle of data minimisation**, collecting only what is necessary. Pseudonymisation or anonymisation will be applied at the earliest stage possible to reduce risks of re-identification.

GDPR Compliance and Legal Basis

All processing of personal data will comply with the **General Data Protection Regulation (Regulation (EU) 2016/679)**. Informed consent will be obtained where required, and data subjects will be informed of their rights. The legal basis for any data processing will be documented and auditable.

Data Transfer Considerations

In cases of data transfer between EU and non-EU countries (e.g. with Chinese partners), **appropriate safeguards** will be applied. If necessary, **Standard Contractual Clauses (SCCs)** will be used to ensure lawful data transfers under GDPR.

6. Ethics

BioClima is committed to conducting all research in accordance with the highest ethical standards, as required by Horizon Europe and international frameworks. Ethical considerations are embedded across all stages of the project, from data collection and processing to stakeholder engagement, publication, and long-term data preservation. All partners are bound by their institutional codes of ethics and must comply with relevant EU legislation, including the General Data Protection Regulation (GDPR), the Charter of Fundamental Rights of the European Union, and applicable national laws.

As part of the project's commitment to responsible research, a dedicated Work Package 7 (WP7) – *Ethics Requirements* – was added during the Horizon Europe grant agreement negotiation phase. This work package is specifically designed to ensure that all ethical requirements outlined by the European Commission are systematically addressed throughout the project lifecycle. WP7 will oversee ethics-related procedures, documentation, and compliance, working closely with other WPs to monitor risks, implement safeguards (e.g. informed consent, data minimisation), and support the internal ethics advisory mechanism described above.

In cases where personal data is collected—for example, through stakeholder interviews, participatory assessments, or surveys—BioClima will ensure that only strictly necessary information is gathered and processed. Clear and informed consent will be obtained from all participants, who will be informed of the purpose of data use, their rights under GDPR, and any intended data sharing. All personal data will be pseudonymised or anonymised as early as possible, and the principles of data minimisation, purpose limitation, and storage limitation will be applied consistently across all relevant tasks.

Given the project's collaboration with Chinese partners and the evolving legal landscape around data transfers, BioClima will proactively address ethical and legal concerns related to cross-border data flows. China's Data Security Law (DSL) and Personal Information Protection Law (PIPL) impose specific restrictions on the export of certain types of data, including environmental and scientific data. To remain compliant, BioClima will classify all datasets involving Chinese partners to identify any data that may be subject to export controls or government approval. If necessary, Standard Contractual Clauses (SCCs) or other legal instruments will be used to govern international transfers. In cases where data cannot be exported from China, alternative strategies—such as on-site processing or federated analysis—will be explored to ensure continued collaboration without violating national legislation.

Privacy and Environmental Considerations in Monitoring Activities

The BioClima project acknowledges the need to address potential privacy and ethical implications arising from the use of remote monitoring technologies such as camera

traps, drones, and satellite imagery. These technologies, while essential for collecting environmental and biodiversity data, may inadvertently capture images of individuals or private properties in monitored areas. To mitigate such risks, BioClima will apply strict protocols for image anonymisation, data minimisation, and purpose limitation, ensuring that no personally identifiable information is processed without legal basis and informed consent where applicable.

In addition to human privacy, the project is also sensitive to the potential impact of monitoring technologies on wildlife. Camera traps and drones, in particular, can cause disturbance to animal behaviour if not used appropriately. Therefore, all field activities involving these tools will follow established ethical guidelines and best practices to minimise harm to animal populations, including limiting exposure time, using non-invasive placement, and consulting relevant wildlife authorities or conservation experts when planning deployments.

Artificial Intelligence

BioClima places particular emphasis on the ethical use of **Artificial Intelligence (AI)** in environmental monitoring and decision support. The project will develop and apply AI models to analyse Earth Observation (EO) data, in-situ measurements, and other spatial datasets for detecting trends, predicting environmental risks, and estimating Essential Biodiversity and Climate Variables. While these technologies provide powerful tools for scientific insight and operational efficiency, their deployment raises important ethical considerations related to transparency, accountability, and bias.

To address these, BioClima will ensure that all AI models are developed and used in accordance with the **EU Ethics Guidelines for Trustworthy AI**, which outline key principles such as human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability. The project will avoid black-box modelling where possible and will prioritise **explainable AI (XAI)** approaches, particularly in cases where outputs may influence policy decisions or conservation priorities.

Datasets used for AI training and inference will be scrutinised for **potential biases**, gaps, or regional imbalances that could compromise model fairness. Any limitations in model applicability—such as geographic scope or data quality constraints—will be documented alongside the models themselves. Where AI tools are made publicly available, their training datasets, pre-processing steps, evaluation metrics, and decision logic will be clearly described, supporting transparency and reproducibility.

Moreover, BioClima acknowledges that the outputs of AI models, particularly in biodiversity and climate contexts, may influence public perception, land management, and resource allocation. Therefore, all AI-supported decisions will remain under **human**

control, and local stakeholders will be involved in the interpretation and validation of AI-generated results in pilot sites. The project will also provide capacity-building materials to ensure users understand the scope, strengths, and limitations of AI outputs.

By embedding these safeguards into its design and implementation, BioClima aims to foster **trustworthy, human-centric AI** that contributes to evidence-based decision-making without undermining individual rights or social equity.

7. Conclusions

The BioClima project is fully aligned with Horizon Europe's principles for open science, responsible data sharing, and FAIR (Findable, Accessible, Interoperable, and Reusable) data management. This Data Management Plan outlines a robust, flexible, and transparent framework for managing the wide range of data and research outputs that will be generated or reused across the project's activities. From Earth Observation and in-situ datasets to AI models, software tools, and stakeholder-driven results, BioClima integrates data management practices into the scientific, technical, and policy dimensions of the project.

The consortium has identified clear roles and responsibilities for data stewardship, ensured that adequate resources are allocated for managing data securely and ethically, and adopted recognised infrastructures such as Zenodo and the BioClima Platform for long-term preservation and access. Particular attention has been given to ethical issues, AI transparency, and international data-sharing constraints, including those related to emerging legal frameworks in partner countries such as China.

This plan will be updated regularly as the project evolves, reflecting new datasets, methods, legal requirements, and user needs. BioClima views the Data Management Plan not as a static document but as a living strategy to ensure that all research outputs achieve maximum scientific, societal, and environmental impact.