OBSGESSION

Observation of ecosystem changes for action

Data management Plan

Funded by the European Union

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

UK Research and Innovation

DOCUMENT TRACKS DETAIL

Project acronym	**OBSGESSION**
Project title	Observation of ecosystem changes for action
Starting date	01/01/2024
Duration	48 months
Call identifier	HORIZON-CL6-2023-BIODIV-01
Grant Agreement No	101134954

Deliverable Information	
Deliverable number	D7.2
Work Package number	WP 7
Deliverable title	Data Management Plan (DMP) for the OBSGESSION Project: Observation of Ecosystem Changes for Action
Lead beneficiary	Finnish Environment Institute (Syke)
Authors	Maria Söderholm (Syke), Maria Hällfors (Syke)
Contributors	Susana Baena (UNEP-WCMC), Claire Brown (UNEP-WCMC), Marcel Buchhorn (VITO), Jorrit Scholze (BC)
Due date	30/06/2024
Actual submission date	17/06/2024
Type of deliverable	DMP – Data Management Plan
Dissemination level	Public

Revision table		
Version	Description	Date
1	Version provided by the Syke coordination team	18/03/2024
2	Updated version following the information provided by the WPs	22/5/2024
3	Updated draft after final review	14/6/2024

LIST OF ACRONYMS

Acronym / Abbreviation	**Meaning / Full text**
ACDD	Attribute Convention for Data Discovery
AGA	Horizon Europe Annotated Model Agreement
BIEN	Botanical Information and Ecology Network
CBD	Convention on Biological Diversity
CC0	Creative Commons Zero licenses
CC-BY	Creative Commons Attribution licence
DMP	Data management plan
DOI	Digital Object Identifier
EML	Ecological Metadata Language
EO	Earth observation
Euro+Med	Euro+Med PlantBase
Elter	European Long Term Ecosystem Research
EVA	European Vegetation Archive

FAIR	Principles of Findable, Accessible, Interoperable and Re-usable data
GA	Grant Agreement
GDPR	General Data Protection Regulation
GBIF	Global Biodiversity Information Facility
GIVD	Global Index of Vegetation-Plot Databases
sPLOT	Global Vegetation Database
HANDLE	The Handle system is a non-commercial identifier resolution system
HE	Horizon Europe funding programme
iNat	Refers to iNaturalist that provides tools for collecting, providing and analysing observations of biodiversity.
LUCAS	Land Use and Coverage Area frame Survey
IPBES	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
IPR	Intellectual property rights
ORCID	Open Researcher and Contributor ID
PID	Persistent identifier
Pl@ntNe	A citizen science project for automatic plant identification through photographs and based on machine learning.
ROR	Research Organization Registry
TRY	Plant Trait Database
STAC	SpatioTemporal Asset Catalog
WCVP	World Checklist of Vascular Plants
WP	Work Package

1. Executive Summary

The OBSGESSION project (Observation of Ecosystem Changes for Action, https://obsgession.eu/) aims to enhance terrestrial and freshwater biodiversity monitoring and policy by new approaches for integrating data and modelling, and by developing science-based solutions. This data management plan (DMP) provides details about the data to be collected, generated, and processed in OBSGESSION, as well as other outputs. It also outlines the key principles of data management practices, as required in the Horizon Europe Programme Guideline. The project is committed to following the policies of open science and research as the foundation for all activities, including data management. Data management in the project will adhere to the FAIR principles, ensuring that data is Findable, Accessible, Interoperable and Reusable. The plan also specifies that other outputs besides data will be managed in line with these principles whenever applicable.

The project will utilise a variety of data and combine biodiversity data from multiple sources. The types of data included in the OBSGESSION project are experimental, observational, statistical, and qualitative data, such as interviews. The DMP outlines the overall data management practices for the duration of the data lifecycle within the framework of the FAIR principles and provides details on planned practices and data-specific considerations. This includes, among other things, producing metadata and other documentation related to the data. The plan also presents practices for ensuring data interoperability, reusability, and publication. Relevant metadata standards have been identified to make EO data interoperable. In addition, the DMP describes the In-Situ database, OpenEO platform and Data Cubes as tools and mechanisms for implementing interoperability and (re)usability of EO and in-situ data.

2. Project description

The OBSGESSION project supports cooperation between the European Commission and the European Space Agency to develop remote sensing of biodiversity.

OBSGESSION is jointly funded by the EU's research and innovation program Horizon Europe, the UK Research and Innovation (UKRI) and The State Secretariat for Education, Research and Innovation (SERI) of Switzerland. The project, which started in January 2024, involves 11 partner organisations.

The project is coordinated by the Finnish Environment Institute (Syke). The other organisations involved are:

Partners:

- Brockmann Geomatics Sweden AB (BG)
- Brockmann Consult GmbH (BC)
- Lund University (ULUND)
- The French National Centre for Scientific Research (CNRS)
- Pensoft Publishers (PENSOFT)
- Stichting Wageningen Research (WR)
- The Flemish institute for technological research (VITO)

- University of Twente (UTWENTE)

Associated partners:

- The University of Zürich (UZH)
- United Nations Environment Programme World Conservation Monitoring Centre (UNEP-WCMC)

The OBSGESSION project will help address the challenges of declining freshwater and terrestrial biodiversity through a multi-faceted approach aimed at integrating existing methods and technologies for assessing and predicting biodiversity change and its drivers. Its results will be used to help inform efforts like the Convention on Biological Diversity (CBD), Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) and the European Green Deal and EU Biodiversity Strategy for 2030, and provide state-of-the-art ecological models on biodiversity change on a case study and EU-level.

3. Introduction

This deliverable is the first Data Management Plan (D7.2) of the OBSGESSION project. The DMP describes the research data to be collected, /generated and utilised, as defined at this stage of the project. Additionally, it outlines the data management practices being followed throughout the data lifecycle, encompassing other outputs produced and their management. As the project progresses, further details on data and other outputs will be specified, and management practices will be implemented. The template in Appendix 1 is used to describe data, other outputs, and related FAIR practices. The data management of the project adheres to the requirements of the Horizon Europe programme, following principles of good scientific conduct and ethical guidelines. In addition, the data management practices follow the FAIR principles, ensuring that the data produced by the project is Findable, Accessible, Interoperable and Re-usable.

The plan will be updated throughout the project to reflect the data utilised and other outputs developed in the project, as well as to supports ongoing data/information management activities.

4. Identification and summary of data and other outputs

4.1. Data

Observational data from iNat, GBIF, PlantNet, EVA, LUCAS

Observational data for the project will come from a variety of sources:

- The Global Biodiversity Information Facility (GBIF): an international network and data infrastructure, providing open access to comprehensive data encompassing all forms of life on Earth. Utilizing common standards, best practices, and open-source tools, GBIF facilitates the sharing of species occurrence information from diverse sources, including museum specimens, DNA barcodes, and smartphone photos.
- The World Checklist of Vascular Plants (WCVP) database offers a global consensus view of vascular plant taxonomy, incorporating nomenclatural data from the International Plant

Names Index (IPNI) and taxonomic data from an international collaborative program. It encompasses flowering plants, conifers, ferns, clubmosses, and firmosses.

- The Botanical Information and Ecology Network (BIEN) specializes in consolidating data related to plant distribution, abundance, and traits, with a primary focus on predicting and addressing the impacts of climate change. BIEN's database integrates georeferenced observations from various sources, including specimens, inventories, and checklists, along with measurements of species-level traits such as size, growth form, and leaf characteristics.
- The European Vegetation Archive (EVA) functions as a centralized repository for vegetation-plot observations primarily from Europe and adjacent regions. EVA includes ReSurveyEurope, a database of repeated records from the same plots, and collaborates closely with other global initiatives such as the Global Index of Vegetation-Plot Databases (GIVD), the Global Vegetation Database (sPlot), the Plant Trait Database (TRY), and the Euro+Med PlantBase (Euro+Med).
- European Long Term Ecosystem Research (Elter) network provides standardized data, services, and training across over 500 sites in Europe and adjacent regions, focusing on sustainable solutions to societal challenges.
- Land Use and Coverage Area frame Survey (LUCAS) conducts harmonized surveys across Europe to collect data on land cover and use, estimating areas occupied by different types based on observations from numerous sample points.
- iNaturalist functions as an online social network for sharing biodiversity information, contributing observations of various organisms worldwide, including plants, animals, and fungi, recorded at specific times and locations.
- Pl@ntNet, a citizen science project for automatic plant identification through photographs, provides reliable data accessible on the GBIF website, leveraging machine learning for identification purposes.

Interview, workshop and survey data

The results of the policy landscape review will be verified through either interviews or surveys. Decision makers at the European Community level and at the pilot level who implement the policies will be targeted. Questions will focus on understanding how policies are implemented, monitored and reported and what needs for EBV and other technical developments in the project. The results of the interviews or surveys will be used to strengthen the policy landscape review. The raw data, along with personal information will not be made public.

4.2. Other outputs – Tools and mechanisms for processing data

In-situ database

In-situ database using for example PostgreSQL, extended with capabilities such as PostGIS will be created for storage of biodiversity data to enable efficient management and analysis of spatially referenced datasets. We will utilize existing biodiversity databases, like the Global Biodiversity Information Facility (GBIF) or the Integrated Taxonomic Information System (ITIS). This integration allows for comprehensive analyses combining diverse datasets. The accessibility level of each dataset in the database will be discussed and decided at a later stage, depending on factors such as whether it is freely available, subject to restrictions, or exclusively accessible within the project

(D2.2, M24). This consideration ensures fair access and proper handling of data in accordance with relevant policies and agreements.

OpenEO platform

The OpenEO platform (https://openeo.org/) to be adopted by the project provides a standardized interface and infrastructure to access and process EO data (D2.4, M40). One of the key goals of the openEO standard, is to support FAIR principles and open science. The implementation in the Copernicus dataspace / Terrascope data space makes it easier to comply with these principles, by incorporating these principles in the implementation, so that users are automatically a step closer to generating FAIR-compliant open data.The generated output data directly fullfills some of the main FAIR principle standards:

- F2 Rich metadata: openEO generates rich STAC metadata, that includes processing info, complete raster metadata, band information, etcetera.
- R1.2 Detailed provenance: In result metadata derived-from links link back to all input products to provide provenance.
- R1.3 use of domain relevant (meta)data standard: openEO generates STAC metadata, so this one is included by default. For the data formats, it supports well known options such as Cloud optimized Geotiff, netCDF with CF conventions, GeoParquet, and many more.

With respect to open science, the main benefit of openEO is that the workflows can be stored in a standardized notation, in the form of openEO 'process graphs'. This gives scientists a novel way of exchanging algorithms, without having to exchange a complex code base. The key element is that openEO code or process graphs are often a lot easier to understand, because much of the boilerplate logic is handled by the backend rather than by the user code. This also has consequences for replicating work: the same process graph can be executed on different backends, or evaluated against different datasets. This allows to evaluate whether an algorithm is broadly applicable, or only works in a very specific environment.

Data cubes

The spatial datasets in OBSGESSION are represented in data cubes, a technology that enables the storage and analysis of multidimensional spatial data (https://openeo.org/documentation/1.0/datacubes.html). By adopting the EBV-Cube standard, biodiversity researchers can benefit from a standardized approach to organizing and sharing biodiversity data in a format that supports interoperability, reproducibility, and integration with other scientific datasets and analysis tools. This standardization facilitates collaborative research efforts and enables more comprehensive analyses of biodiversity patterns and trends across different spatial and temporal scales.

The EBV-Cube standard builds upon existing standards and best practices in the field of biodiversity informatics, including the use of netCDF (Network Common Data Form) as the underlying data format. NetCDF is a widely used data storage format for multidimensional scientific data, providing efficient storage and retrieval capabilities for large datasets. In addition to the netCDF format, the EBV-Cube standard incorporates metadata standards such as the Attribute Convention for Data Discovery (ACDD) and the Ecological Metadata Language (EML). These metadata standards help ensure that the biodiversity datacubes are properly documented, making it easier for users to understand the content and context of the data.

The software framework xcube is specifically designed for working with EO data cubes, which are multidimensional arrays containing spatiotemporal information derived from satellite imagery and other sources. While xcube is not explicitly tied to biodiversity data, it can be used to handle and analyze biodiversity data in the form of data cubes, including those conforming to the EBV-Cube standard. Xcube is compatible with the netCDF and Zarr format, allowing users to work with data cubes stored in this format seamlessly. This compatibility means that biodiversity data cubes conforming to the EBV-Cube standard can be stored and managed using Zarr, and then easily loaded into xcube for analysis.

The accessibility level of each data cube will be discussed at a later stage, depending on factors such as whether the cubes are freely available, subject to restrictions, or exclusively accessible within the project. This consideration ensures fair access and proper handling of data in accordance with relevant policies and agreements.

5. FAIR data

5.1 Making data findable, including provisions for metadata

Data will be provided with comprehensive metadata throughout the entire data collection/generation process in OBSGESSION to describe the research data. Metadata and other standards are essential for facilitating data discovery, identification and understandability during the project and for future reuse. Metadata are stored within the data using appropriate data storage solutions and is regularly updated to ensure accuracy when archiving and publishing the metadata and data. Similar metadata provision practices will be applied to other outputs whenever appropriate.

Metadata are produced in accordance with metadata standards and domain-specific practices. The following metadata standards and conventions are identified as relevant for the data used in the project. These standards are utilised in the project tools (see 4.2). The application of these standards in other contexts will be explored and agreed upon during the project:

The Attribute Convention for Data Discovery – ACDD
https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery_1-3
ACDD is a general framework but particularly applicable to EO and environmental monitoring data facilitating data discovery, integration, and analysis across diverse domains and disciplines. It defines standards to document metadata attributes such as variables, spatial and temporal coverage, units and coordinate reference systems.

The Ecological Metadata Language standard – EML
https://eml.ecoinformatics.org/
The use of EML is supported by the fact that its features are especially suitable for describing, sharing and reusing research data in the fields of ecology and environment. For example, it provides a standardised framework to document study design, sampling methods, variables, data quality and provenance.

The SpatioTemporal Asset Catalogs specification – STAC specification

https://stacspec.org/en
The STAC specification is a common language for describing geospatial information. The metadata attributes cover information about data such as, spatial and temporal extent, data properties, links to related resources, provenance and quality and annotations.

The Climate and Forecast Metadata Conventions – CF Conventions
https://cfconventions.org/
The CF Conventions are a set of standard conventions and guidelines for encoding metadata in NetCDF files. NetCDF (Network Common Data Form) (https://en.wikipedia.org/wiki/NetCDF) is a data format commonly used for storing and exchanging scientific data. The CF Conventions define standardized metadata attributes and conventions for describing the content, structure, and interpretation of data stored in NetCDF files. These conventions include guidelines for, for example, coordinate variables, grid mappings, metadata attributes for variables and coordinate systems.

If standards or specific practices are not available, or at the beginning of data collection and processing, temporary solutions should be used to produce metadata. In such cases, we will develop instructions and, for example, a metadata template. These instructions and template will guide the production of relevant information on the data, aligning with the information covered by general data repositories. This information will encompass descriptive, administrative, and structural details about the data. Where applicable, metadata attributes from the domain-specific standards mentioned above will be included in the template. Below is a list of basic metadata attributes, along with additional information on why metadata production is crucial for implementing the FAIR principles in the OBSGESSION project:

1. Descriptive metadata such as
    - Names of authors/ creators and ORCIDs
    - Title of the data.
    - Keywords, topics, themes (from standard sources) ensuring findability, interoperability and understandability of the data.
    - A brief description of the data making data understandable and helping users evaluate whether the data meets their needs. The description includes for example, what the data is about, what its origin is, how it was produced and processed.
    - Year when data was created / period to which the data relates.
    - A persistent identifier, such as DOI and HANDLE, is assigned to data when it is made publicly available in the repository/archive. The persistent identifier identifies the data (and a specific version of it), facilitates citation of the data and as a link helps to locate the data.
2. Administrative metadata such as:
    - Information about the provenance/origin of data/ owner of the data (organisations' ROR IDs).
    - License information about the data and associated metadata to define the conditions of re-use is crucial both to increase both access to data and increase re-use of data.
    - Access information meaning the level of openness/ publicity, if necessary, i.e., whether the data is open access, embargoed, whether it is restricted or whether it must be requested.
    - Information about the funder, Horizon Europe, co-funders and funded project (grant project name, acronym and number).
3. Structural metadata such as:
    - Information and links to relevant publications and other outputs that are relevant for understanding and validating the data.

Appropriate organisation of data and file naming conventions, and version control are also employed to enhance the findability of the data throughout the project lifecycle and when it is deposited in a data repository or other dedicated service.

5.2 Making data accessible

In OBSGESSION, data and other outputs produced by the project will be made available according to the principles of open science, 'as open as possible, as restricted as necessary', which means that they will be made publicly available to the extent that legitimate interests and restrictions are respected. For example, data containing personal data such as interviews will not be disclosed without anonymising.

Data that can be opened will be made publicly available in open repositories/archives. Wherever possible, data will be published in a trusted repository[1] to ensure open access in line with the requirements of the HE and GA. When selecting a repository, especially the following accessibility options required by HE guidelines and AGA will be met:

- Data will be published preferably through domain specific (data type specific) repositories endorsed by the research communities. If this is not possible, international general-purpose repositories, such as the Zenodo repository that meet the key criteria for trusted repositories will be used.
- The data is given a persistent identifier (e.g., a PID such as DOI or HANDLE).
- Metadata is openly available and licenced under the CC0 license.
- Data that can be provided openly available is published under CC-BY license.
- In the repository
    - data and metadata are accessible through a free and standardized access protocol.
    - the long-term availability of data is ensured.

The data that can be made publicly available will be published as soon as possible during the project and at the latest when the project is completed. However, the timing of data publication is chosen to allow sufficient time for data processing, quality assurance, analysis and drawing conclusions. Where possible, data relating to peer-reviewed publications will be published at the time of the article publication via the data repository. All project outputs and results are also presented through various communication channels, for example the project's web page and the project social media channels (more detailed information in the project's Communication, Dissemination and Exploitation strategy (D6.2)).

5.3 Making data interoperable

In OBSGESSION we will implement the following key practices to ensure the interoperability, consistency and integrity of data and other outputs we generate:

- Standardised software, tools and methodologies will be used.
- Controlled vocabularies and/or ontologies will be used.

[1] The trusted repositories, for example (1) assign persistent identifiers to data and other outputs (e.g., DOIs, handles, etc.). (2) facilitate data re-use in the short- and long-term, (3) give explicit information about their policies, which define their services, (4) provide broad and ideally open access to content free at the point of use, (5) respect legal and ethical limitations. See in detail the characteristics: AGA trusted repository definition (pp. 283-284).

- We either utilize ready-made or develop protocols.
- Guidelines and instructions will be produced where applicable.
- Metadata produced will be understandable to human users and machine actionable. Section 5.1 presents the relevant standards we have identified. Section 5.1 presents the relevant standards we have identified.
- Protocols and standards are used to structure the data so that other researchers can assess and reproduce the data. Where available, the community data standards are referenced.
- Links to other data-related resources are added to the metadata/documentation to enrich the information about the data and provide the data user with contextual information about the data.
- Outputs are published using open or commonly used formats wherever possible or according to standards used within the given community.

Data and other outputs are stored and shared in different formats depending on their type and the software used to produce them. However, their interoperability and hence (re)usability is ensured by using preferably formats that are non-proprietary. For sharing data, we can also use formats that are commonly used, if they are relevant for the new end users of the data. Already during the project, data will be stored in a format that is accessible to other project partners. For instance, when original data or analysis results are generated using software, outputs are produced in non-proprietary formats, whenever feasible.

5.4 Increase data re-use

OBSGESSION will facilitate the re-use of data first, by licensing the data, and other outputs where applicable, under the most permissive license. The metadata of data will be licenced under the CC0 license and data under CC-BY license. For code, one of the most permissive licenses is the MIT license. Second, documentation will be provided to facilitate the understandability and validation and thus re-use of data and other outputs, where applicable. Section 5.1 presents the relevant standards we have identified. How the documentation will be provided in OBSGESSION depends on the data(sets)/output. In some cases, documentation (metadata) is embedded, i.e., it is provided as a part of data file, making it self-describing. In other cases, additional documents, such as read-me files, are attached to the data. Additional documentation will be provided if the metadata of data and/or the data structure itself is not sufficient to ensure a smooth re-use of the data. In general, the additional documentation includes, for example, a detailed methodological description of how the data was collected and analysed, information on the methods used to collect and process the data, description of the variables, and more general information about the project and purposes for which the data was produced. The use of the software and tools used will also be documented to ensure the reusability of the data. For the source code, the documentation practices depend on the platform/service where the code is developed.

6. Data security and storage during the project

The data collected and generated for the project, as well as the supporting metadata and documentation is stored in a distributed manner on the servers of partner organisations, or any other adequate services chosen and developed/adopted in the project. In general, the ICT services of the host organisations are responsible for providing secure storage and backup services for those working on the project. In line with good and secure practice, backups are mainly done automatically on a regular scheduled basis. In order to ensure data security within the project, it

will be identified where password-protected services should be used to be able to restrict access to authorised individuals only.

Teams, the Microsoft SharePoint cloud-based collaboration platform provided by Syke ICT, will serve as a general tool for managing, storing, and sharing documents. Research data will be managed with Microsoft Teams only if it's technically feasible, comply with ethical guidelines and privacy legislation, and promotes joint data management.

7. Ethical and legal considerations

In OBSGESSION before sharing or making data available, all results are protected by clearing copyright, privacy, and confidentiality issues. The ownership and intellectual property aspects of data and other outputs can lead to restrictions on their use. In general, the ownership and Intellectual property rights (IPR) belong to the party that has generated the data and other outputs. The ownership of data is recognized according to the 'rights to first publication/authorship' rule.

The project has agreed on procedures on how to proceed before publishing the data. Publication of data and results must be approved by the concerned partners. The data and peer-reviewed publications are published in accordance with HE guidelines under licenses that preserve sufficient intellectual property rights of the authors/creators. Where data originates from third parties, the intellectual property rights and terms of use of data in question will be respected.

The OBSGESSION project involves human participants as volunteers for citizen science and as participants of the workshops. When generating and processing personal data, all project partners follow the research practices and ethical guidelines applicable to the data concerned. Prior to collecting personal data, the procedures to be followed in partner countries for ethical assessment practices will be clarified. If it turns out that an ethical assessment needs to be conducted, we will agree on common practices. For example, that we will adhere to the ethical assessment practices of the country where they are the most stringent. Data management practices regarding human participation in research and eDNA data are based on safeguarding the rights of individuals. We comply with applicable legislation, such as the EU's and each country's own data protection laws. This means, among other things, that (1) we minimise the collection of personal data, ensuring that only information necessary for achieving the project's objectives is collected, (2) participants receive comprehensive information about the project, the processing of personal data, and their rights, (3) research data containing personal data is stored and shared securely, with access restricted to authorized individuals only and (4) data containing personal data is anonymised before publication, ensuring that no personally identifiable information is disclosed.

In all its activities, including in data generation and collection, OBSGESSION will pursue an equal representation of women and men, as well as groups with broader social differences, such as age ethnicity and health. The role of open science practices, in general, are seen as a central means to make science inclusive regarding gender and minority aspects.

8. Data management resources and responsibilities

The joint guidelines and practices for data management are developed together and described in the DMP and in the Project Management Guide (D7.1). Major decisions on agreeing on common practices are decided on by the Steering Committee.

As coordinator, Syke has an overall responsibility to ensure that management of data follows good scientific practices, valid legislation, and the funder's requirements. However, each partner ensures that the data collection and/or generation, for which it is responsible, is carried out in compliance with common data management standards and good scientific practice. Syke is responsible for completing and updating the DMP by ensuring the partners contribute and agree to the joint DMP. Data protection officers of each organisation will guide in issues concerning personal data protection.
At the WP level, the overall responsibility lies with the WP lead. The WP lead ensures the adequate management of the data, e.g., the data curation and quality assurance. They also provide support for data and/or output specific questions under their responsibility, for example provide their expertise to support others in the collection and/or use of these data.

The costs associated with data management (including processing and storage) are included in the project budget. The day-to-day data management practices are mainly integrated into research activities and included in the project budget as salary costs. Also, data gathering and/or collection service costs are included in the budget. However, we have not identified any costs related to the purchase of equipment or facilities. We use equipment, services and facilities that are either free or provided by the partner organisations. For example, the use of existing data, data repositories and archives, which are used to make data and other outputs openly available, are free of charge. In addition, Syke offers Microsoft Teams/Sharepoint for storing, managing and sharing documents at no direct cost to the other project partners.

Appendix 1. Template for Describing Data and Data Management Practices

1. Data Summary	
Title of the data	
Description of the data	Write a descriptive abstract summarising the content and context of the data. ▪ Describe the nature of the data, including its type, relevant species, habitats, geographical areas and the target group(s) of interviews, if applicable. ▪ Specify any special instruments or methodology utilised to generate the new data. ▪ indicate whether source data will be utilised in the production of the new data. ▪ State the purpose behind the collecting or generating the data. ▪ Outline the potential beneficiaries or users of the data. ▪ Specify the file formats. It is recommended to use non-proprietary or widely adopted formats, ensuring the ability to convert to open formats for interoperability. ▪ Detail the origin or provenance of the existing / source data. ▪ Specify if the use of existing / source data is based on an agreement.
2. FAIR data	
Making data findable	Specify the metadata standards to be applied:
	If there are no metadata standards, outline what metadata will be created to enable data discovery:
	Outline other approaches towards discoverability (adding keywords, (file)naming and version control conventions):
Making data accessible	Specify the repository / service where the data and associated metadata, code and documentation will be available. Note that the repository must provide persistent identifier (e.g., DOI) to the data. If the open availability is not possible or the re-use of data must be restricted or even the data must be kept closed, provide the reason:
	If the opening of data (embargo) in a repository must be delayed, provide the reason:
	Specify what software tools or method are needed to access or read the data (if any). Is documentation about the software needed to access the data included? Is it possible to include the software (e.g., in open source code) into the repository as part of the data or as a reference to the data?
	If the source data underlying the project's produced data is published in a repository, provide the link and/or the name of the service:
Making data interoperable	Specify what data and metadata vocabularies, standards, methodologies or protocols will be used to promote data interoperability:

	Specify whether you are able to use standard vocabulary to allow inter-disciplinary interoperability. If not, will you provide mapping to more commonly used vocabularies/ ontologies?
	Specify if here are discipline/domain specific practices that promote interoperability that will be used?
Increase data re-use	Specify what documentation will be provided and how to validate data analysis and facilitate data re-use and understanding (e.g., readme files with information on methodology, data cleaning, analyses, variable definitions, units of measurement, etc.):
	Describe data quality assurance processes to be followed:
	Specify the other outputs that are needed to validate and reuse the data. Describe tools and instruments, any other research outputs, algorithms, code, protocols, models, and how access to them will be organised:
	Specify the terms of use for the restricted/closed data produced by the project. By default, the data produced by the project should be licensed under the CC-BY license. However, due to considerations such as intellectual property rights (IPR) and privacy, some data access must be restricted or kept closed. Outline the arrangements for how this closed/restricted data can still be used by others, if applicable (e.g., based on a contract or upon request):
3. Data security and storage during the project	
Describe the security and storage solutions for collecting, storing, and sharing data (e.g., data recovery/backup, secure storage/archiving, and authorization where necessary). Additionally, specify how the data will be shared if other partners need access to it:	
4. Ethical and legal considerations	
Specify any legal and ethical issues that may impact data sharing, unless these have been covered above or in the ethics review, DoA, or related deliverables. In practice, such issues may relate to IPR, the EU's GDPR, or other privacy-related legislation, as well as the processing of personal data or any sensitive data:	
5. Description of other outputs and their management practices. These include e.g., digital outputs such as protocols, software, models and code (if code and model are related to specific data, perhaps reasonable to describe in the context of that data) and physical outputs such as samples.	
Provide a name for the output and a brief description. Additionally, specify the management practices (FAIR principles, where applicable) that will be applied to the output:	
6. Other issues	
Outline briefly any national, domain specific or other relevant procedures for data management to be employed:	