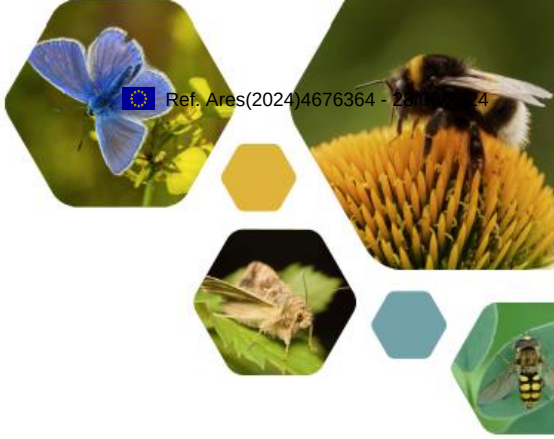

WildPosh

D8.3 Initial Data Management Plan

28/06/2024

Lead beneficiary: UMONS

Author/s: Sara Reverté Saiz, Denis Michez, Jordan Benrezkallah

Funded by the European Union

Prepared under contract from the European Commission
Grant agreement No. 101135238
EU Horizon Europe Research and Innovation Action

Project acronym:	WildPosh
Project full title:	Pan-European assessment, monitoring, and mitigation of chemical stressors on the health of WILD pollinators
Project duration:	01.01.2024 – 31.12.2027 (48 months)
Project coordinator:	Prof. Denis Michez, University of Mons (UMONS)
Call:	HORIZON-CL6-2023-BIODIV-01-1
Deliverable title:	Initial Data Management Plan
Deliverable n°:	D8.3
WP responsible:	WP8
Type of the deliverable:	Report
Dissemination level:	Public
Lead beneficiary:	UMONS
Due date of deliverable:	M6
Actual submission date:	28.06.2024

Deliverable status:

Version	Status	Date	Author(s)
1.0	Draft	13.06.2024	Reverté Saiz, Michez, Benrezkallah UMONS
1.1	Review	19.06.2024	Novkova PENSOFT
1.0	Final	28.06.2024	Reverté Saiz, Michez, Benrezkallah UMONS

Table of Contents

Preface

This document is a deliverable for the WildPosh project, funded under the European Union's Horizon Europe Research and Innovation Action under grant agreement No. 101135238.

The aim of this document is to present the WildPosh Initial Data Management Plan, which sets the strategy to be followed to collect, organise, use and share the data obtained throughout the project. The document describes the types of data that will be collected throughout the project, the data management process, the data repository selected, compliance with the FAIR principles and relevant regulations (particularly involving the GDPR), ethics, data security, partnership with the sister project PollinERA and the open science strategy to be followed.

This document is a living document, being constantly updated where relevant changes should be applied during the evolution of the data management process. Updates will provide history of changes to locate easily the differences between versions.

Summary

WildPosh is a multi-actor, transdisciplinary project whose overarching mission and ambition are to significantly improve the evaluation of the risk to wild pollinators of pesticide exposure and enhance the sustainable health of pollinators and pollination services in Europe.

The Data Management Plan defines the strategy followed by the project in terms of the data collection, the management of all the data collected, generated and used during the project, along with the long-term strategies for ensuring the use of the data in the future by other platforms and projects. The report consists of the following sections:

Section 1. Introduction. This section contains an overview of the document.

Section 2. Data summary. This section contains information about the kind of data that will be generated and collected, its sources, the data formats to be used, along with the purpose of the data collection and how it aligns with the objectives of the project.

Section 3. Data management. It provides the workflow of the management of the data, including information about what is the repository selected (EU Pollinator Hub) and its characteristics and the process of data validation that will be followed.

Section 4. FAIR data. This section provides detailed information about how the FAIR principles will be followed across the project.

Section 5. Allocation of resources and responsibilities. This section provides information about the responsibilities of the different partners in the project, and how the resources are allocated to data management and publications in terms of budget and time.

Section 6. Data security. This section provides all the details available up to this moment of the security to be followed to protect the data, both on the website of the project and the selected repository.

Section 7. Ethics. This section provides information of the compliance with the relevant regulations and legislations regarding the collection, storage and use of sensitive information (personal data) and the transfer of genetic material and information across countries.

Section 8. Collaboration with PollinERA. This section provides the terms of the sharing of information with the sister project.

Section 9. Open Science. This section provides details about the compliance with the requirements on open access of the data and the publications, and the licenses that are envisaged.

List of abbreviations

CC-BY 4.0	Creative Commons Attribution License 4.0
CO	Coordinator
DM	Data Manager
DMP	Data Management Plan
DOI	Digital Object Identifier
DPO	Data Protection Officer
EC	European Commission
EFSA	European Food Safety Authority
EU	European Union
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
OECD	Organisation for Economic Co-operation and Development
PCC	Project Coordination Committee
WP	Work Package

1 Introduction

WildPosh is a large and ambitious project, where the data is at the centre of it. The data used and collected, once analysed, will sustain the results and be the base for the communication and dissemination of results over the scientific, social and political spheres to maximise the impact of the project. Thus, an extraordinary effort is done and will be done throughout the course of the project to streamline the processing of data collection, data sharing, data manipulation, data analysis and ensuring data security and ethics. The WildPosh project will implement the already available FAIR methods throughout its lifetime and ensure the availability and transparency of the data after the end of the project.

During the project WildPosh will create huge quantities of information of different kinds, including data tables, maps, models, interviews, etc. We depend on existing data from previous studies, previous and current projects. This is a very connected project in which the WPs are strongly linked to each other and there is an important data exchange between WPs.

The data needs to be reused, both during the lifetime of the project and after the end of the project. We want to make this data available to the community to make it the most useful and easy to find and to navigate to ensure its impact in science, policy and society. We need to ensure and maximise that the data is interpreted into strategies and policies.

We will also deliver a Mid-term DMP (M24) and a Final DMP (M47).

2 Data summary

A lot of different data will be created during the WildPosh project. It will produce data through in vitro, in vivo, in silico methods via laboratory, semi-field, field, and modelling approaches, and personal data through questionnaires and interviews. WildPosh will also collect data from the literature and obtain data from other research projects. All the data generated and obtained through the project will be centralised on an integrative database. All the data collected needs to have a purpose and a direct link to fulfilling the objectives of the project, particularly the personal data.

2.1 Types and formats of data

As WildPosh is a large and complex project, several types of data will be collected during the project. The databases created will contain files of the following types (non-exclusive list):

Table 1: Types of data that will be collected during the project, along with a description of what it will be and the expected formats.

Type of data	Description	Format
Metadata	The metadata of all datasets will be provided based on Darwin Core and Ecological Metadata Language. The list of minimum required items to be found on the metadata is specified in section 4.1.	.xlsx
Data from field, laboratory, semi-field and modelling approaches	New experimental data will be collected during WPs 1 to 5.	.docx, xslx, .pdf, .shp
Protocols	Protocols will be prepared on multiple WPs, plus protocols will be obtained from external sources during the project. They may contain videos as a support.	.docx, .mp4, .pdf
Literature review	Review on the available methodologies and literature will be performed during WP1, WP3 and WP4. Data will be extracted from those literature sources.	.docx, .xslx
Historical data	Background data on use of pesticides in the different European countries will be collected during WP4.	.docx, .xslx, .pdf
Interviews and surveys	Interviews and questionnaires are planned for farmers and stakeholders during WP1.	.docx, .xlsx, .pdf, Qualtrics
Pictures	Pictures of experimental settings or protocols, of meetings or events, or of interviews with people can be collected.	.jpg, .tiff, .png
Video and audio recordings	Video and audio recordings of experimental settings or protocols, of meetings or events, of	.mp4, .mp3

	interviews with people, can be collected.	

Even though some of the data formats selected do not come from open source software, they are still the most commonly used formats and they can be opened with open access software.

The project website repository will include research outputs as well as unconventional project outputs such as policy briefs, policy recommendations, factsheets, inventories, case studies and data management plans.

2.2 Purpose of the data generation or re-use

To fulfil the goals of the WildPosh project, data will be collected through the different WPs:

- WP1 aims to understand the sources and routes of exposure of the pollinators to pesticides in the environment. For this purpose, we will obtain data from a network of sites along four European countries representing for eco-climatic regions in Europe (continental, UK; mediterranean, Spain; continental, Germany; boreal, Estonia). We will sample different environmental matrices (mud, water, plant tissues, pollen, nectar), interview farmers, obtain data on landscape and pesticide use around the farms.
- WP2 aims to understand the effects of exposure to single pesticides (single exposure) and their mixtures on wild pollinators as novel models in laboratory and semi-field experiments. For this purpose, we will evaluate the variability among wild insect pollinator species in their sensitivity to pesticides, define the extent to which semi-field sensitivity mirrors sensibility under laboratory conditions, and develop new OECD protocols for testing wild insect pollinators in pesticide risk assessment.
- WP3 aims to understand the “omics” of pesticide responses in wild pollinators. For this purpose, we will elaborate MALDI-MS molecular fingerprints of fat bodies and haemolymph, assess the overall stress response to pesticides on the peptidome/proteome dysregulation, define diagnostic transcriptional signatures that can be used to predict sensitivity to pesticide exposure of pollinators in the field and decode the molecular machinery underlying the response of pollinators to pesticides.
- WP4 aims to expand current datasets and perform in silico modelling for risk assessment. For this purpose, we will compile a comprehensive distributional and trait database of European pollinators, identify and collate data on pesticide exposure and effects, as well as on other stressors able to amplify the adverse effects on model species, improve the existing in silico prediction methodologies for toxicity endpoints relevant to pollinators and build an open-source curated database which will include information on pollinator traits and distribution, on pesticides, and on other stressors for model species.
- WP5 aims to develop an integrated systems-based risk assessment. For this purpose, we will review current approaches of risk assessment and provide strategies for improvement integrating lethal and sublethal effects of single and multiple pesticides including interactions, develop landscape models integrating

pesticide exposure and hazard, develop environmental scenarios for pesticide risk and mitigation models and develop an integrated open-access tool for a systems-based risk assessment.

- WP6 aims to assess the effectiveness and feasibility of mitigation measures. For this purpose, we will identify effective response options to reduce pesticide risks to wild pollinators, synthesise WildPosh project findings and external knowledge, develop good practice guides for practitioners to mitigate the impacts of pesticides on wild pollinators and engage in science-policy dialogues to inform national and international policy on the development of mitigation measures.

2.3 Expected size of the data generation or re-use

There is no formal estimation of the size of the data generated during the project. It is difficult to predict due to the complex nature of the collected and generated data, which will include structured data, pictures, videos, models, and websites. The chosen data repository (section 3.1) will store the final version of all data generated by the project, while internal documents and non-structured data can be stored on the project website.

2.4 Origin of the data (old and new data)

The project will create new data and will collect data available on the literature, along with collecting information from other ongoing projects where the data has not been published yet, as an input for all the models that will be developed during WildPosh.

Origin of new data:

- A significant part of the data produced during WildPosh will be obtained through experimental laboratory manipulations, or fieldwork. We will compile empirical data on pollinator sensitivity and sensibility to pesticides, pollinator distribution and traits, and environmental data on different matrices.
- New results will be generated through the Environmental Risk Assessments performed during the project.
- New data will be obtained through interviews and questionnaires performed to stakeholders.

Also, this project is building on the findings of previous projects, or other active projects:

Project	**Type of data**	**Purpose**
Poshbee https://poshbee.eu/	Data from experimental manipulations of different pollinator species and pesticide screening on the environment will be obtained from this project.	The data can be directly used for analysis within WildPosh, and will also help generate the updated protocols for field and laboratory experimentation.

Safeguard https://www.safeguard.biozentrum.uni-wuerzburg.de/	The information available so far on distribution and functional traits of bees, hoverflies, butterflies and moths will be obtained from this project.	All this data is going to feed the Tasks 4.1 and 4.2 of WildPosh and will be useful for pesticide risk modelisation.
Data published in literature	All published data related to pollinator sensitivity, risk assessments, pesticide use, presence of pesticides in the environment, will be collected.	It will be collected for meta-analysis, including Task 4.3, and to feed the models.
PollinERA and other projects from the same call https://pollinera-horizon.eu/	PollinERA will generate data from experimental manipulations from pollinators exposed to pesticides, field data about pesticide distribution, and ERA models.	We will exchange data between projects to maximise the potential of the models of both projects.
Data from many different hub / structures (e.g. EU Pollinator Hub).	Public databases will be screened for data potentially useful for the project.	All the available data will feed the databases within WP4 and the models in WP5.

3 Data management

Data management is the process where the datasets are organised, checked, stored, shared, and provided of metadata and where the rules and procedures of use of the data are decided.

The original data will be stored by each producing partner and published in the selected repository. Consortium members are responsible for storing all the information needed to replicate the databases, analysis or results during all phases of the project (original data, scripts, models, etc).

The different steps of data management encompass the following:

- **Meta-data creation.** Every dataset must be accompanied by a metadata for its traceability. Details of what are the minimum contents of the metadata are detailed in section 4.1.
- **Data upload.** Data must be uploaded on the selected repository. Uploading the data on the repository for the accessibility of other consortium members should be done as soon as possible, before the official publication date of the dataset.

- **Data analysis.** The data uploaded on the repository will be used by different partners of the project to fulfil the scientific objectives of WildPosh.
- **Data download.** Data download will follow a request procedure. Consortium members that want to download datasets from other partners from the repository must send an official request through the repository with the intended purposes of use of the data.
- **Data sharing and publishing policy.** The data sharing and publishing policy must be agreed by the PCC before starting the upload of datasets on the repository.

Data will be stored in the repository prior to its publication to permit an anticipated sharing timeline within the consortium. Versioning of datasets and documents must be accompanied by a justification and an explanation of the changes done to the previous version. The goal is to ensure traceability of the changes. Task 4.4 of the project, led by UMONS, corresponds to the development of the global database, data curation and collection. This task will encompass all the data collected through the different WPs. During this task, trainings to the consortium about data management will be performed.

3.1 Data repository

EU Pollinator Hub has been decided as the repository for the project. This is a young repository developed during the last 5 years with the support from the EFSA. This hub was created to improve and centralise the efforts done on pollinator health and protection, and it is focused on data related to pollinators and the agricultural sector. It aims to become the central place to go for stakeholders involved in pollinator health, making it the optimal place to maximise the impact of the datasets. It provides open access to metadata, secured authentication for users, a streamlined procedure for data access and upload, versioning will be available, and a secured request procedure for data download.

The main arguments in favour of the use of EU Pollinator Hub are:

- The layout of this Hub has been designed to be user-friendly and intuitive.
- This hub is already made, optimising the use of resources since we don't have to create our own.
- They have an IT support, that will help us during the entire process of database creation, upload, maintenance, and storage. The different protocols of how to use the different functionalities of the repository are provided on their website.
- It is possible to allow different roles for different users. This means that different users have access to different parts of the data or different datasets. They can work as quality assessors and reviewers of certain parts that they have been granted access. This way it is controlled who can do what and allow access according to expertise/role.
- They are actively working on developing versioning and version control inside the Hub.
- Users are informed about the license of use of every dataset.

Under this hub, the security of the databases is ensured, and the rights of the owners respected. Each database is stored separately with the owners clearly acknowledged, and

the access to the database is granted in concordance with the dissemination level stated by the owners. The database can be freely downloadable or it can be only downloadable upon request. This situation is helpful particularly during the development of the project, when running versions of the databases may be circulated within the consortium through the hub but not ready yet to be shared widely (particularly when they are used in a publication in preparation).

As this hub is a long-lasting resource, it will also be useful as a long-term storage for the data of the project. Data will only be introduced once, and no change or effort will need to be taken at the end of the project.

3.2 Validation of the data

To ensure that data is in the correct format, we need to force people to encode data correctly. This function is not currently available inside the Hub but we will work together with the developers to implement such function. It may be necessary to create an external application that checks the data format to grant access to upload.

There would be 3 validation steps:

1. Technical validation of the structure of the dataset.
2. Peer review of the content of the dataset.
3. Double check of the clean version.

The data extension that can be uploaded is .csv, a non-proprietary, standard and widely used format. At present, it is not possible to upload other data file extensions, but this could be done on request, if necessary. The hub currently accepts scripts and is recommended to store the scripts to perform the statistical analysis together with the data.

4 FAIR principles

WildPosh will adhere strictly to the EC (2022) Guidelines on Open Access to Scientific Publications, the HORIZON Programme Guide 2022 Version 2.0105 and to the EC guidelines on FAIR data management in Horizon Europe. The data quality and format standards selected will ensure easy future data sharing, combination, interoperability and analysis of the original data and the results. The goal is also to ensure the preservation and the long-term accessibility to the data, metadata and project outputs.

The standard protocol will be the creation of the data, followed by an internal sharing process inside the consortium following a request procedure and finally the publication. Other project partners that are not the creators of the database need to request access to the database in the repository. Sharing with external parties before the data is published needs to be done with the agreement of the CO.

4.1 Findable

All datasets will be findable. The EU Pollinator Hub will create a Unique Identifier for each dataset, which is unique to a dataset version, linked to a URL (both the Unique Identifier and the URL never change). Rich metadata will be produced for each database to ensure that the data produced is easy to find. Keywords will be associated to the database to further enhance their findability. The selected data repository allows for flexibility in terms of quantity and structure of the metadata associated to the different datasets, as there will be data in multiple formats. File versioning (under development in the EU Pollinator Hub) will allow for users to find the adequate version for their purposes.

To facilitate the findability, a template for the names of the files to be uploaded in the repository will be created. Metadata for each dataset must contain at least the following information: author(s), description or abstract, year of publication in the repository, title of the dataset, data repository, unique identifier, access date or version, language, metadata language, license of use, date of metadata creation, hierarchy level, character encoding, format version, keywords, validation process, embargo period (if any), grant project name, project acronym, project number.

Even if the information contained in the dataset is not available due to restrictions in the use of the data (specially datasets under embargo or containing personal data), metadata will be available in all cases. The metadata will be always freely available and findable, even when the access to the data is protected with restricted user access.

4.2 Accessible

The level of dissemination of the different datasets will follow the statement: “As open as possible, as closed as necessary”. We will pursue a full open access to the entire database, but we will maintain the datasets private while they are being manipulated, cleaned and curated. Most datasets will be made freely available during the project or at the end of it. If there is any dataset that there is a special reason to keep it private after the end of the project (for example, personal data), it will not be made public. If some datasets are planned to be part of a scientific publication in progress at the end of the project, an embargo on the datasets will be envisaged. For those datasets, access might be granted under request, made through the repository. This procedure will ensure the identification of the person making the request, and all the process will be traceable. The maximum embargo period envisaged for these cases is 3 years. If the owners of the dataset need an embargo longer than 3 years, they need to justify the reasons and provide the estimated publication date.

The data will be hosted on the EU Pollinator Hub. Datasets will be uploaded by their creators as soon as they are ready to be shared within the consortium, before the actual publication date. The selected repository is accessible via all types of software and electronic devices, maximising the access of current and future users of the databases. In this repository the data is structured, the ownership is recognised, and the adequate level of openness is associated. The datasets will be made available long-term, with no

envisaged expiration date. The access to the datasets in the repository will follow the standard protocol from the repository.

The metadata will be publicly available of all datasets. It will be made publicly available under CC-BY 4.0 Creative Commons. This has the goal of further enhancing the impact of the project through communication of the activities of the project and transparency, and potentially improving the scientific collaboration of the project.

Along with the data and metadata, in the EU Pollinator Hub we will publish the scripts to make the analysis that were done during the publications derived from the project. The tools for risk assessment will be open-source tools (WP5) under licenses that follow Open Source Initiative criteria.

4.3 Interoperable

Interoperability of research outputs will follow format and vocabulary standards already set during the PoshBee project and descriptor categories by EFSA financed platform EU Bee Partnership Prototype Platform.

The formats that are selected for the data are standard cross-platform and easy to use by most of the operative systems. The selected data formats are specified in Table 1. By using the most commonly used file formats, we ensure the easy merging of databases from different origins. Partners are allowed to use standards and common methodologies used on their respective fields of science for the creation of their data and metadata.

4.4 Reusable

The ultimate goal is that all the data collected and generated through the project is available to the community and can be of utility to other projects or entities. The final toolbox produced by the project will also be free for access and use.

Data and metadata comply with community standards, to ensure other users can automatically start using it without having to work on it. A special effort will be done during the process to ensure that the formats used are the most suitable for the community and that the datasets made available are fully cleaned, processed and curated. It will be associated to relevant attributes to make it easy to find, understand and reuse. Information on the source of origin, methods on how the data was taken, processing of the data, and every detail necessary for future users in terms of making the data reusable will be provided. All this information will be provided on the metadata associated to the dataset.

All data will be made accessible, except personal data. Thus, all data that can be traceable or potentially be linked to the identity of the data subject, will be anonymised wherever possible before making it available. When anonymisation is not possible, the data will not be made publicly available.

Relevant datasets will be published as data papers in open access journals with DOI, to ensure scientific credit for creators and owners. To make them more accessible, published versions of the articles will be deposited on the Open Access Infrastructure for Research in Europe (OpenAIRE).

5 Allocation of resources and responsibilities

Making the data public will incur certain costs, through i) publication fees in open access journals and ii) fees for depositing the data in the repositories. The payment of the publication fees in open access journals has already been foreseen during the preparation of the GA and is included in the budget of the different partners as a function of the number of articles they are expected to release from their tasks. In the case these funds are not sufficient or more publications are published, green open access will be pursued, or other sources of funding will be searched.

Data management in WildPosh is coordinated through WP8, where UMONS is WP leader. UMONS will oversee informing the project partners of the decisions on data management. The CO is ultimately responsible for the compliance with the DMP during all phases of the project.

The Data Manager (DM) is in charge of conceptualising the data workflow and data management processes to be followed during the project. The estimate for the DM is to spend around 18 person-month in developing the templates, developing the WildPosh database on the EU Pollinator Hub, assisting partners in preparing to upload their files, providing technical support and organising training for the partners on data management. The DM is ultimately responsible for quality control of the data to be uploaded to the repository.

The Data Protection Officer (DPO) is in charge of overseeing the compliance with GDPR during all phases of the project and by all members of the consortium manipulating personal or sensitive data. The DPO is responsible for providing advice to the CO about the obligations in terms of data collection, processing and management under the GDPR. The DPO is the contact point within the project for data subjects and is in charge of following the requests and inquiries related to the processing of their personal data. The DPO is designated by the CO.

WP leaders are responsible for ensuring the datasets coming from their WPs respect the guidelines provided.

Task leaders are responsible for ensuring that the data generated through their tasks is complete and provided in an adequate format, including the metadata. They should also oversee the process of data storage on the EU Pollinator Hub.

All people in the consortium or working on data of the WildPosh project are responsible for managing the data they create or manipulate according to the DMP. Data collectors

are ultimately responsible for complying with the relevant regulations and legislations, particularly when involving the processing of personal data.

6 Data security

All of the project data will be stored on the EU Pollinator Hub's servers, located in Slovenia, so within European Union borders. All the documents of the project (templates, deliverables, reports, images, audio) will be hosted on the website of the project, which ultimately are hosted by Pensoft's servers, in Sofia, Bulgaria, EU. Both EU Pollinator Hub and Pensoft servers ensure a backup process. The EU Pollinator Hub site is hosted on a managed Virtual Private Server (VPS), using encrypted disks. All data is replicated daily internally and twice a week on an external server for backup purposes. Similar practices will be followed by the partners when working through their institutional repositories during the project. Partners are responsible for the data security before uploading it to the EU Pollinator Hub or the Pensoft servers.

Each user account in the WildPosh website is verified and validated by the Data Manager (DM). The Project Manager will verify the users' account in the WildPosh team in the EU Pollinator Hub. In principle, only WildPosh partners will be allowed to have accounts on those platforms. For external users, e.g. project reviewers, their account will be discussed within the PCC about its necessity and be approved individually.

7 Ethics

WildPosh will collect sensitive information during its research activities. For those cases particularly, relevant ethics and legal requirements apply. Principally, the sensitive information collected by the project belong to two main categories:

- Genetic information of wild pollinator specimens, and importantly, the transfer of this information between countries (within EU and outside the EU, as partners of the project include universities from UK and Serbia).
- Personal data. During the activities of the project, collection of personal data of stakeholders is envisaged through interviews, questionnaires, and others (potentially pictures, videos or audio recordings). The legal and ethical framework through which we can collect personal and sensitive data is being prepared. This has to be done before we can start collecting personal data, as the people from which personal data needs to be collected needs to be fully aware of their rights, the purpose of the use of their data, and they need to sign an informed consent before their data can be collected.

In D9.1, we will appoint an External Ethics Advisor for the project who will ensure that the project does comply with all relevant ethic principles and all regional and national regulations and legislations. Compliance with the legislations and regulations applying to sensitive data collection, manipulation and use will be demonstrated on deliverables D9.2 and D9.3.

7.1 Personal data

Particular attention of data security will be put on personal data. All processing involving personal data will strictly follow GDPR along with other applicable legislation, and consent of data manipulation and sharing will always be asked in advance. Before start collecting data from them, they will be informed on their rights and informed about the people in charge of processing their data and the consent. The rights of the data subjects (individuals providing personal data) in regards of their personal data are:

- **Right to be informed.** Data subjects have to be informed at all times of what data is planned to be collected, for what purpose, how it is going to be stored and used, always in terms that they can understand.
- **Right to erasure.** Data subjects can request the partial or total deletion of their personal data at any time, and without any sort of penalisation or negative consequences.
- **Right of access.** Data subjects can access all their personal data.
- **Right to rectification.** Data subjects can make a request for the modification of their personal data at any time.
- **Right to restrict processing.** Data subjects can decide the purposes for which they want their personal data to be used and restrict their use to certain purposes or to restrict it from any processing altogether. This does not invalid the processing of the data prior to the restriction.
- **Right to complaint.** Data subjects have the right to file a complaint with a supervisory authority.
- **Right to data portability.** Data subjects can ask for their personal data in a common, easy to use, format.
- **Right to share the minimum information necessary to fulfil the project goals.** Only the necessary data will be collected, nothing more, it will be only used during the framework of the project and in the minimum necessary to accomplish the projects objectives.

The consent form will include information about the project, how the data will be used, which data they allow to be used (and re-used) and under which conditions, data storage, the kind of data formats that may be collected: audio recordings, video recordings, photographs, transcriptions of interviews, translations, etc. The data collected and used every time will be the minimum necessary to fulfil the research purposes. Only research purposes are legitimate uses for the personal data collected during the WildPosh project, and the data must only be used for the exact purposes for which it was collected. The products made with the personal data collected may include anonymised data.

The different datasets will remain separated, to ensure the privacy and easier control the use for the different purposes allowed for each dataset. Each dataset will be controlled by the person/institution who collected the data. People manipulating personal data will do so accurately, securely, confidentially and anonymously, and always ensure that the subjects rights and wishes are respected. These legal requirements need to be fulfilled by

all partners of the project, and potential future subcontractors that would need to work on these data on the framework of the scientific activities of the project.

Personal data will be anonymised before being uploaded to the EU Pollinator Hub. The storage of personal data needs to be done in a confidential manner on servers located inside the EU (no Microsoft, Google or similar servers are allowed), to ensure the applicability of the European laws. However, the partners of the project from outside the EU (UK and Serbia) must be granted access if necessary for fulfilling the project objectives. This will be done in compliance with the ethical requirements and the relevant national, European and international legislation.

For socio-economic data (specially for farmers, and other stakeholders of the project), particular effort will be done in order to ensure that the answers cannot be linked to the participants. In case of online surveys, cyber-security will be ensured via encrypted connections, the data will be stored on the PENSOFT servers, and IP-addresses of the participants will be treated according to relevant national and European legislation. The anonymised data has no limited storage time (except for the maximum conservation time), but the time will be determined by the prevalent legislation, the informed consent of participants and the time needed for the project to use those data for scientific activities. To ensure the respect of the privacy of the owners of the land that will be sampled during the study, the information of the coordinates of the sampled sites along with the landscape and climate characteristics of those sites will not be made public. The EU Pollinator Hub has a function to blur spatial reference systems thus anonymising coordinate information within the datasets that is or may become personal data and thus sensitive to the compliance with the GDPR.

7.2 Transfer of genetic information

It is envisaged during the project to collect genetic information of wild insect pollinator species (bees, hoverflies, butterflies, and moths). Also, with the purposes of data analysis the data will be transferred between the different partners in different countries performing the different tasks (some partners oversee collecting the information in the field, some partners oversee performing the DNA sequencing, some oversee performing the annotations of the genomes). During the movement of specimens across countries, the Nagoya protocol will be complied with. Particularly so when sending samples to the UK and Serbian partners (outside the EU).

8 Collaboration with PollinERA

WildPosh and PollinERA will have separate DMPs, but we will do them in synchrony in order to ensure the correct supply of data between Work Packages and projects. The data of both projects will be sent to the same repository: EU Pollinator Hub.

We should keep the PollinERA consortium updated regularly about the current version of the DMP.

We will maximise the reusability and interoperability of the datasets between and within projects. Both projects decided to use the Darwin Core as an international standard, and expand as necessary with extra columns. We will then merge our templates with those of PollinERA to improve data compatibility between the two projects.

9 Open Science

WildPosh will adhere rigorously to the EC (2022) Guidelines on Open Access to Scientific Publications and the HORIZON Programme Guide 2022 Version 2.0105.

The WildPosh Open Science Pilot will start with the open access publication of the project Description of Work in the Research Ideas and Outcomes (RIO) journal (organised by PENSOFT). Similarly to Pilots already launched for other EU projects, unconventional research outputs, such as policy briefs, policy recommendations, factsheets, inventories, case studies and data management plans, will be added to the collection. This will ensure that all project outputs are published openly, with a stable DOI assigned, and comprehensively collected in one place.

We will favour transparency of the publication process preferentially choosing open peer-review journals for WPs producing and analysing data (WP1-6). WildPosh will ensure reproducibility by providing extended documentation on the methodologies employed as well as the raw data of individual studies. The citizen science methodology will be a key point for the involvement of farmers in providing information on pesticide application (WP1) and for the identification of appropriate response options to reduce pesticide risks (WP6).

WildPosh will publish results under the Creative Commons Attribution 4.0 International (CC-BY 4.0) (Gold open access). In case this is not possible, Green open access will be pursued. We will pursue the distribution of the tools for risk assessment in open-source tools (WP5) under licenses that follow Open Source Initiative criteria. As a rule, data and software will also be published as data papers and software description papers in appropriate journals. The selected repository allows the attribution of a license to each entire dataset or to attribute different licenses to the components of a dataset (tables) according to SOP 017 (Rubinigg, Dataset integration, 2024).

9.1 Data publishing policy

In collaborative synthesis studies, credit to the data suppliers will consist of co-authorship. Only results validated by the whole consortium will be publicly released on the “Open data” and on the project repository. A beneficiary that intends to disseminate its results must give at least 15 days advance notice to the other beneficiaries (unless agreed otherwise), together with sufficient information on the results it will disseminate.

9.2 Data ownership and intellectual property rights

Intellectual Property Rights will be regulated as part of the Consortium Agreement, but publication of the data under Creative Commons Attribution 4.0 International (CC-BY 4.0) will be pursued, where attribution is required. Other open data licenses that can be used are Creative Commons CC0 and the Open Data Commons Public Domain Dedication and License (ODC-PDDL).

10 Conclusions

This document will be distributed within the WildPosh and PollinERA consortia. We will make sure all partners are informed of its content and that they comply with the different elements of the DMP during the lifetime of the project and beyond it. This document will be updated throughout the life of the project with two released updates through the D8.4 (M24) and D8.5 (M47) of the project.