Butterfly

DATA MANAGEMENT PLAN, version 1

Contributing Authors: Laura Drivdal, Jorrit H. Poelen, Lorraine Balaine, Mirella Miettinen, Ugo Mendes Diniz, Nicholas James Balfour, Maria Clara Castellanos, Jeff Ollerton, Nicola Gallai, Ivan Meus, Noa Simon Delso, Jeroen van der Sluijs

Funded by
the European Union

V

Prepared under the contract from the European Commission,
Grant Agreement No. 101181930
EU Horizon Europe Research and Innovation Action

Project acronym:	Butterfly
Project full title:	Mainstreaming pollinator stewardship in view of cascading ecological, societal and economic impacts of pollinator decline
Project duration:	48 months
Project coordinator:	Dr. Jeroen P. van der Sluijs, University of Bergen
Call:	HORIZON-CL6-2024-BIODIV-01-3
Deliverable title:	Data Management Plan
Deliverable No.:	D9.1
WP responsible:	WP9 -
Deliverable type:	Report
Dissemination level:	
Lead beneficiary:	UiB
Recommended citation:	
Deliverable due date:	M6
Submission date:	30.08.2025

Deliverable status:

Version	Status	Date	Author(s)
1.0	Draft		Laura Drivdal, Jorrit Poelen
1.1	Review	29.08.2025	Reviewers: Noa Simon Delso

1.2	Final		

Contents

1 Introduction

This document outlines the Data Management Plan (DMP) for project Butterfly. The DMP is a **'living' document** that will be reviewed and updated over the course of the project. Additionally, we aim to archive versioned snapshots of this document, specifically in the months 24 and 48. As Butterfly is working closely with related projects, we aim to align, reuse, or take inspiration, from the DMPs of our sister project VALOR (Pkhakadze & Stoyanova 2025; Breeze et al. 2025), and projects that we closely collaborate with: SAFEGUARD (Zhang and Steffan-Dewenter 2022), ProPollSoil, and RestPoll (Wintermantel et al. 2024).

This document begins with an overview of data products that are anticipated to be generated (see "Data Index"). Butterfly is a transdisciplinary project, generating different types of data (both quantitative and qualitative) across the various Work Packages and tasks. In section 2 our data overview lists all generated, re-used and curated data products, an important first step to enable data review, referencing and cross-referencing.

In chapter 3, we describe how we aim to use community best practices (e.g., FAIR principles) to facilitate the re-use of this data within, and beyond, the scope of this project ("Data Re-use"). Attention will be given to what kind of metadata the outputs should contain and how and where they will be stored. The internal data review process (described in 1.2 below) is aimed at ensuring that data will be understandable and reusable across disciplines and user-groups. In general, Butterfly is designed to ensure accessible and re-usable data through: a) the EuroAPPA portal which aims to provide user-friendly access for all stakeholders to the most complete, taxonomically-harmonised and well-curated platform of plant-pollinator interactions for Europe and three Overseas Territories/Outermost Regions; and b) the project's co-creation approach in the Living Labs, facilitating 'openness by design', meaning that data creation is a shared venture from the start.

Chapter 4 outlines research outputs other than data, chapter 5 outlines the allocation of project resources for complying with FAIR principles, and chapter 6 summarise our approach to the security of data as well as non-data outputs, such as plans for storage, long-term archiving, and recovery of data, plus how sensitive data will be transferred.

Finally, in chapter 7, we discuss our ethical handling and re-use of personal data (section "Data Ethics"). Project Butterfly is committed to making the data 'as open as possible, as limited as necessary'. Thus, attention will be given to how we deal with data that contains personal information, and how we ensure that all participants are giving informed consent to participate. A general informed consent sheet that can be adapted to each data collection activity in the different countries is included in the appendix.

1.4 Data management workflow and responsibilities

Each partner leading the respective task is responsible for preparing the datasets including metadata. All partners must create, manage, analyse, store and/or share data and/or datasets with respect to the applicable national and international legislation on

data protection. The quality of- and the quality assurance process for these data falls under the responsibility of the task leader, who also should initiate the data review process. In the Consortium agreement, it is stated that the principal investigators and the Data Protection Officer of each beneficiary organization are considered responsible for the DMP actions. Data collectors have the ultimate responsibility of complying with the specifics of the Data Management Plan, as well as with the related GDPR[1] policies and applicable local, government and international laws, regulations and guidelines. However, data reviewers also have a responsibility to contribute to that data is usable and reusable.

To ensure that the datasets generated are usable, reusable, interoperable and understandable across disciplines, the Butterfly consortium aims to set up a system for interdisciplinary review of each other's datasets (see Figure 1). This implies that, if we have ecological data, we should strive to add at least one social or humanities scholar in the internal review, to make sure data is understandable to others outside the experts' field. This process will also increase interdisciplinary understanding in the project. Meetings will be set up approximately four times a year, where process and responsibilities will be discussed and decided.

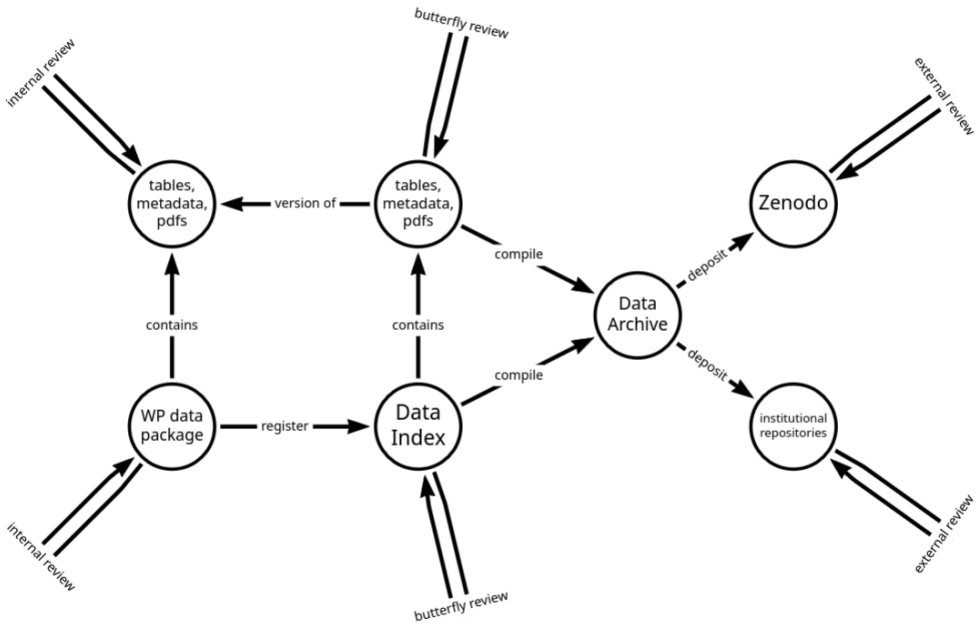

Figure 1: Data Management Process Overview Diagram (draft) generated using https://arrows.app - from left to right the (meta-) data transitions from unreviewed/closed to reviewed/open (as open as possible). In each stage, feedback loops are expected. For instance, on publishing the first Butterfly Archive, we'd get external review notes, which are then recorded and communicated with the data management team as well as the data creator. To take advantage of these feedback cycles, publishing early and often in the project is desired. Otherwise, you'd have review notes that can no longer be incorporated because the project has already completed. The details of this process (e.g., will reviews be Open Access (OA)) will be tested and discussed in the second DMP.

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

2. Data

2.1 Data Summary

Current assumption for data volumes in total is up to ~10TB, but this will be reviewed in the second DMP. The different types of data listed have different origins, as specified in the tables below.

Project Butterfly is fundamentally interdisciplinary and transdisciplinary, where many different types of data will be compiled and generated across various tasks. This includes collecting, generating and reusing a) **ecological data b) biological data, c) human participant data, and d) economic & production data**. Data format will vary according to what data is being generated or reused; however, it will be supported by wide types such as .csv, .xlsx, .mx24, .mp4, and .docx. The different types of data are summarised below, and as they carry different implications for data sharing and ethical approvals, the four major categories of data will be referred to systematically throughout this document.

a) **Ecological data:** The main focus in data the interactions between species of pollinators and the plants whose flowers they visit (but fruit and seed set data will be collected as part of the Butterfly project). The plant- pollinator interaction data will be geo-located using decimal coordinates and elevation, and time-stamped by the day/month/year in which it was collected. In addition, each interaction will be assigned a measure of data 'quality' *sensu (*Ollerton et al. (2025).

b) **Biological data**: Data on pollinators, including DNA sequence data from pollen samples collected during the Living Labs and Overseas Sites field work.

Ecological and biological data will mostly be collected and managed through WP1. The EuroAPPA "one-stop shop" web portal and database for information on plant-pollinator associations across Europe will include **both re-used data and new data**. **T1.2.1.** will review and reuse data from existing databases and mobilise data that is not yet indexed in order to provide a Europe-wide synthesis of information on plant-pollinator databases. **In T1.2.2**, in collaboration with the LL's (WP7), **field-based data** will be collected on pollinators, plant-pollinator interactions, pollinator services and pollinator dependencies. The Field Protocol (WP1) (Diniz et al, 2025), specifies procedures for sampling, sampling efforts required, and standards for labelling physical samples and digital sample data.

c) **Human participant data, quantitative and qualitative**: Participants perceptions, values, belief and opinions, arising from surveys, interviews, workshops, engagement activities, stakeholder input, co-creation and psychological experiments. Both quantitative and qualitative participant data will be collected.

Human participant data will be collected and managed in several WPs and tasks. In WP2, (T2.3 T2.3), citizen perception data will be collected through a)

Population-based surveys in six European partner countries on the 'willingness to pay for the preservation of pollinators' on a national scale, and b) **qualitative workshops** with students specializing in agricultural education. In addition to the national survey, 'willingness to pay' of young people—considered as future key actors in rural territories, will specifically be assessed.

In WP4, (T4.1 and T4.4), quantitative and qualitative data will be collected through five sector-specific Delphi surveys. The data will be generated through a one-round, online real-time questionnaire, targeting 15 experts per sector (food/micronutrients, pharmaceuticals, cosmetics, biomaterials and biomass energy).

In WP6, secondary data will be gathered for the literature synthesis and historical analysis (see table 2). In addition, primary **attitude-behaviour data** will be generated from experimental psychology set-ups**.** Experimental data from controlled psychological experiments will be used to determine how attitudes toward environmental issues and pollinators, social parameters, and training influence how well people adopt and retain pollinator-friendly behaviors. The results of the controlled experiments will be used to develop tools that transfer knowledge about pollinators to everyday citizens (details in Table 1).

In WP7, T7.1 (together with **T6.3**) will generate participant input data from the Butterfly LLs through structured dialogues between stakeholders and participants, field observations of individual practices regarding pollinators and social and environmental interventions**, citizen science participation, interviews, photography, video, document analysis**, and other material produced in the different LL processes. The analysis will provide insights into how stakeholders and individuals produce, use and share knowledge about pollinators. **In addition, qualitative data will be generated in T7.3** Participatory scenario planning and co-creation.

d) Economic and agricultural production data / farm practice data

T2.1 w. WP7 (LLs): Data on agricultural activities, value chains and other economic related to pollinators and ecosystems. Existing digital data sets of e.g. predicted climate and land use and cover under each of the Shared Socio-economic Pathways (IPCC 2023) will be re-used. In addition, data will be collected by each LL-leader through surveys, remote sensing, land uses, pollinator stress, and more.

Table 1: Butterfly data Registry 1): Anticipated 'primary data' to be collected (note that the tables will be elaborated with in the following DMP)

Datasets and origin	Place/sector of collection	WP
Field-based **datasets on pollinators, plant-pollinator interactions, pollinator services and pollinator dependencies**.	All B-Sites, which include the Living Labs (LL) - Region of Murcia (ES), Zeeland (NL), Northern Jutland (DK), Ile-de-France (IDF) (FR), Southern Norway (NO), Milano region	WP1 w. WP7

Baseline data will have the following origins: - 10 pollen samples from every B-/S-pollinator species in each site. Interaction data between plants and pollinators will be obtained via pollen DNA metabarcoding; - 100 flower heads from each B-/S-/I-Plant species. Interaction data will be obtained via eDNA metabarcoding. - 120 minutes of Flower-Insect Timed Counts for each B-/S-/I-Plant species, providing visitation and plant fitness (dependency) data.	(IT) -, as well as non-LL sites, such as the RestPoll Sites and overseas territories (Greenland, Curaçao, and Martinique).	
Student perception data, qualitative. workshops with students specializing in agricultural education. 100 students targeted per country	France, Ireland and Greece with the support of the ENTER Network	WP2
Human perceptions data, quantitative: Survey on willingness to pay at a national scale	Six European partner countries (France, Greece, Ireland, Germany, Norway and Italy) (**PS4**).	WP2
Data on agricultural activities, value chains and other economics related to pollinators and ecosystems. T2.1, asks each LL to collect data on farm structures, key practices, sustainability measures, and market access. For the general survey, data collected include location, years of operation, farm size, key agricultural activities, annual yield, practises such as organic, integrated pest management, water conservation etc), types of fertilisers and pesticides, Vaule chain etc	**Each LL-leader** collects accurate local data through surveys, remote sensing, land uses, pollinator stress, and more..	WP2 w. WP7
Expert perception data: i) raw data sets from five online Delphi questionnaires, ii) automatic reports delivered by the survey software (such as log reports or basic statistics and visualisation of the completed responses in the platform). Raw data sets contain both quantitative (such as % scale and 9-point Likert scale) and qualitative data (arguments).	15 relevant experts in each sector: supply chains for food/micronutrients, pharmaceuticals, cosmetics, biomaterials and biomass energy	WP4
Experimental human participant data from controlled psychological experiments. Recruited participants (sample size TBD) will follow a computer-based program with	Students and employees of the participating universities (Trier University and TUM),	WP6

the goal of testing strategies to facilitate pollinator stewardship (e.g., managing a garden in a computer program with the aim of maximizing pollinator numbers and diversity). Data type: - Personal data (demographics: age, gender, living space, region of residence, etc.) - Experimental data (attitude evaluations, participants' ad hoc knowledge of pollinator-plant interactions, participants' choices within the experimental paradigm, choice outcome, response times, questionnaires, etc.) - To evaluate transfer of learned plant-pollinator interactions to real life, participants may document their efforts via photographs that will have all identifying information removed (blackened, location data and time stamp removed, etc.); photographs will not be shared with others, but will be encoded by several involved experimenters into a quantifiable measure. .	citizens from the local municipalities, as well as participants found via online platforms such as Prolific.	
Living Lab data on practices regarding pollinators and social and environmental interventions, citizen science participation, interviews, photography, video, document analysis, and other material produced in the different LL processes	Region of Murcia (ES), Zeeland (NL), Northern Jutland (DK), Ile-de-France (IDF) (FR), Southern Norway (NO), Milano region (IT),	WP7 T6.3

2.2 RE-use of data

Ecological, economic, legal and policy data will be collected from existing sources and re-used, specifically in WP1, WP4 and WP6.

- **T1.2.1** ("synthesis and mobilization of data sources"), will index and disseminate information contained within **existing databases of biotic interactions** (plant-pollinator networks). Specifically building on the Database of Pollinator Interactions – DoPI (currently focused on the UK), and the Global Biotic Interactions platform – GloBI). This is significant for producing a one-stop shop" (PS3**)** for plant-pollinator interactions. Further, data sources that are currently not indexed by the sources will be mobilized to provide a Europe-wide synthesis of information on plant-pollinator databases with a particular view to targeting lesser-known groups of pollinators.

- **T1.4.1**: Information from existing digital products on the distribution of managed plants and pollinators (such as the EU Crop Map and the Eurostat dataset on main livestock indicators)

- **T1.4.2: digital data sets of predicted climate and land use and cover** under each of the Shared Socio-economic Pathways (IPCC 2023) to produce estimates of plant-pollinator network structure across Europe under various scenarios of human development for 2050/2100

- **T1.4.3**: carry out a systematic review and meta-analysis of the state-of-the-art field experiments in which plant and/or pollinator diversity and/or abundance have been manipulated and an assessment made of the impact on plant-pollinator networks

- **T4.1**: Existing data that will be utilised in step 1 of a Delphi survey protocol includes scientific literature on potential vulnerabilities and tipping points in each supply chain to pollinator loss and possible response options (mitigation and adaptation). Literature search can be complemented by searches on relevant databases (may be different for each supply chain) and other desk research (such as company websites, previous EU or national projects etc.)

- **T5.1:** Data from other WPs and existing sources will be used to categorize pollinator species according to their utility for different stakeholders. This will require information on biogeographic distribution (WP1), EU and national Red List data, and pollinator functional trait databases, such as CropPol."

- **T6.1**: Collect and re-use existing data on human dimensions of pollinator decline from academic literature, including grey literature.

- **T6.2:** In Task 6.2, openly accessible EU legal documents and case law (e.g. from EUR-Lex) or Court of Justice of the European Union, as well as national CAP Strategic Plans of selected countries will be collected and analysed. T**he** Legal analysis done in **Task 6.2** will produce i) a report (D6.2) that includes the results of the analysis on how pollinator conservation has been considered in key pieces of EU legislation, EU's Common agricultural policy (CAP), and national agricultural policy (e.g. in Norway), and how new Regulation (EU) 2024/1991 on nature restoration affects the provisions in these existing instruments; and ii) a report (D6.3) that includes the results of the assessment and recommendations on how the selected instruments developed in WP5 could be integrated into existing legal or regulatory instruments (such as CAP).

- **T6.4** Historical analysis attempts to systematically recapture the complex nuances, the people, meanings, events, and ideas of the past that have influenced and shaped the present. It relies on a wide variety of sources, both primary & secondary including unpublished material. The data and insights from these analyses will then be built together in an overall coherent analysis and synthesis presenting the human, social and historical (past-present-future) aspect of pollinator loss and restoration on micro, meso- and macro level of society.

Table 2: **Butterfly data Registry 2) Anticipated available data to be re-used (secondary data)**

Dataset and type of data	Source	WP
Ecological baseline data on published plant-pollinator networks to be added to The Database of Pollinator Interactions (DoPI)	https://www.sussex.ac.uk/lifesci/ebe/dopi/	WP1
Ecological baseline data from Global Biotic Interactions (GloBI)	https://www.globalbioticinteractions.org/	WP1
Data not currently indexed by existing databases on all pollinators but with special emphasis on **lesser-known pollinators** such as birds, bats, and other insects than bees, wasps, or syrphid flies	TBD, but some suggestions are: Bats: https://www.batbase.org/ Birds, bats, and many other pollination networks (including "lesser known" insects): http://www.ecologia.ib.usp.br/iwdb/resources.html	WP1
EU Crop Map and the Eurostat dataset on main livestock indicators	https://ec.europa.eu/eurostat/web/agriculture/database	WP2
Digital data sets of predicted climate	https://gmd.copernicus.org/articles/9/1937/2016/	WP1
Digital data sets of land use and cover	https://essd.copernicus.org/articles/15/3819/2023/	WP1
Shared Socio-economic Pathways (IPCC 2023) Future Global Climate: Scenario-based Projections and Near-term Information	https://www.cambridge.org/core/books/climate-change-2021-the-physical-science-basis/future-global-climate-scenariobased-projections-and-nearterm-information/309359EDDCFABB031C078AE20CEE04FD	
Scientific literature on potential vulnerabilities and tipping points in each supply chain to pollinator loss and possible response options (mitigation and adaptation).	TBD. Literature search can be complemented by searches on relevant databases (may be different for each supply chain) and other desk research (such as company websites, previous EU or national projects etc.)	WP4
Pollinator functional trait	https://github.com/ibartomeus/OBservData	WP5, T5.1
Literature review on human dimensions	TBD	WP6, T6.1

EU legal documents and case law. National CAP Strategic Plans of selected countries.	EUR-Lex (https://eur-lex.europa.eu/homepage.html) Court of Justice of the European Union (https://curia.europa.eu/jcms/jcms/j_6/en/) CAP Strategic Plans of selected countries (links available: https://agriculture.ec.europa.eu/cap-my-country/cap-strategic-plans_en)	WP6, T6.2
Data for a historical (past-present-future) meta-analysis on human and social determinants and consequences of pollinator loss and restoration covering the **period 1850-2050**	Primary sources in public records & legal documents, minutes of meetings, corporate records, recordings, letters, diaries, journals, drawings, located in university archives, libraries or privately run collections such as local historical society. Secondary sources found in textbooks, encyclopaedias, journal articles, newspapers, biographies and other media.	WP6, T6.4

2.3 Purpose of data generation in relation to objectives

Butterfly has eight specific objectives (SO) presented in Table 1.1 of The DoA. Data generation is specifically relevant in relation to:

SO1: Provide a holistic overview of actionable knowledge on animal pollination ecology and pollination services provided for wild and cultivated plants covering the European continent as well as EU overseas territories.

- For this, it is essential to provide and analyse biological/ecological data on plant-pollinator interaction.

SO3: To comprehensively model and quantify the macro- economic implications of pollinator decline, to model the country-specific economic butterfly effects of dependencies on pollinators, and to provide forward- looking analysis of policy options and scenarios.

- It is vital to collect and analyse economic data for the modelling

SO4: Understand how 5 key biomass supply chains (food/micronutrients, pharmaceuticals, cosmetics, biomaterials, biomass energy) depend on pollination and co-create pollinator restoration options that increase resilience of these supply chains. Promulgate resilience-thinking to businesses beyond Butterfly stakeholders and to EU policymakers.

- To assess potential vulnerabilities and tipping points in each supply chain to pollinator loss and possible response options (mitigation and adaptation), data from experts is generated through five sector-specific Delphi surveys.

SO5: Develop, test and implement transferable tools that enable systematic mainstreaming of proactive pollinator stewardship into key vulnerable sectors through multi-actor co-creation approaches and LLs

- Provide user-friendly interfaces for interacting with the data of the project developed models.

And SO7*: Establishing a test-system of multi-actor communities across sectors to accelerate knowledge transfer and serve as field study sites, multi-actor co-creation of knowledge and solutions, and forum for continuous discussion and networking.*

- The multi-actor dialogues and the co-creation approach implies the collection of feedback and data in workshops and seminars

2.5 Usefulness of data outside the project

BUTTERFLY researchers share new knowledge and data with relevant actors as early in the research process as possible to ensure beneficiaries, particularly the at-risk sectors, benefit from outcomes and learning as it emerges. Beyond BUTTERFLY, the consortium will actively share data and outputs with other initiatives (§1.2.2), including EU-funded projects. We acknowledge that effective knowledge exchange between initiatives can avoid duplicative or unnecessarily competing efforts and instead foster a collaborative culture of effective pollinator restoration research for impact.

Data will also be used to shape the WP5 Decision support tools, maps, and guidelines. A key feature of these tools is that they will inform stakeholders about the risks of pollinator loss for their businesses (data from WP2 and WP3) and assess the impact of the measures on pollinators (task 5.2, task 5.4), within the framework of a global conservation strategy (data from WP1).

Data within the EuroAPPA database will have long-term utility for ecologists, agriculturalists, policy makers, and others interested in the biodiversity of plant-pollinator interactions in Europe and some of the Overseas Territories/Outlying Regions.

In the Butterfly dissemination plan, this the general usefulness of the project is further specified, and it also contains a list of actionable knowledge and tools generated during the Butterfly project and for End Users (see table 3, Butterfly DECE plan, in Delso et al. 2025)

3. FAIR data management

3.1 Making data findable, including provisions for metadata

Every dataset will have a **persistent and unique identifier** throughout the entire project. Depositing datasets in Zenodo will automatically give them a Digital Object Identifier (DOI)

for a record once it is published).

To increase the findability of the data, all generated data will be accompanied by **metadata.** According to the Butterfly Grant Agreement, annex 5, metadata of deposited publications must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following:

- author(s),
- title,
- date of publication,
- publication venue;
- Horizon Europe funding; grant project name, acronym and number;
- licensing terms;
- persistent identifiers for the publication, the authors involved in the action and, if possible, for their organisations and the grant.

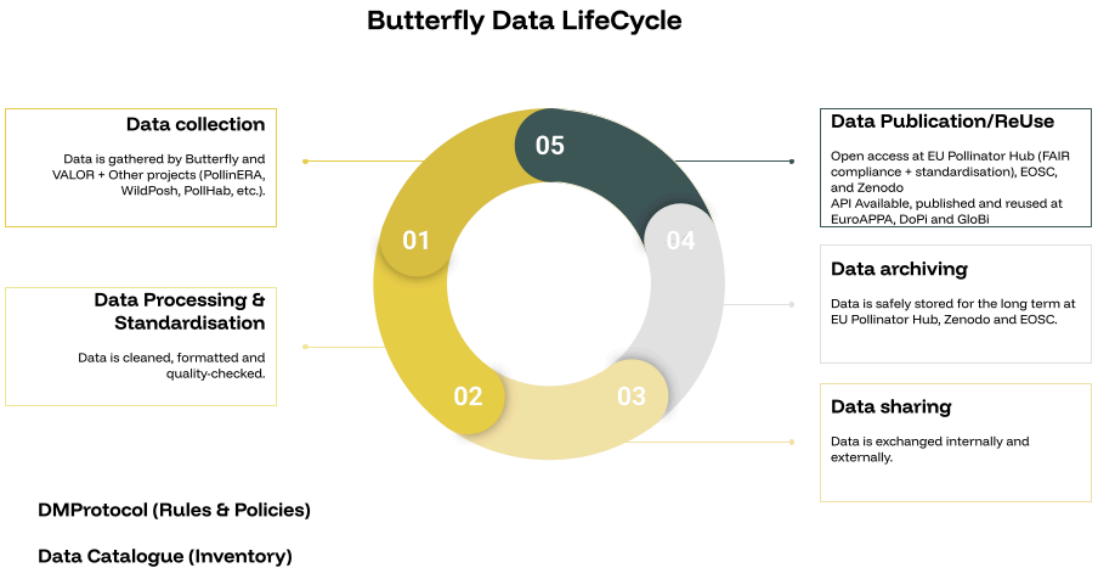

Figure 2**. Butterfly Data LifeCycleSpecific considerations on ecological and biological data metadata**

Good practice in terms of FAIRness for the plant-pollinator interaction data will be informed by the EU-funded WorldFAIR project (Drucker et al. 2024).

The Field Protocol (WP1) (Diniz et al, 2025) specifies the standardised naming and labelling of the physical samples of pollinator specimens pollen loads and flower head (eDNA), as well as the digital data for samples and sampling events to enable data exchange and allow for syntheses. Ecological data sets archived by Butterfly will adhere to the Darwin Core standard vocabulary (Wieczorek ea 2012) with a thorough description of the data generation process and its spatial, temporal, taxonomic, and thematic extent

also adhering to current metadata standards, such as the Ecological Metadata Language standard**.** This ensures that the primary data sources are archived so that they are retrievable long after the project is completed (regardless of the status of EuroAPPA or its constituent databases), adhere to relevant standards to improve interoperability, and are visible to biodiversity data indexing services. In addition, we will improve the visibility of these datasets through the publication of the data through the Global Biodiversity Information Facility (GBIF) and the use of EuroAPPA as a web portal for all ecological deliverables, which will be integrated into the EU Pollinator Hub as well. Similarly, sequence data and their associated annotations generated from the genetic analysis conducted as part of task 1.2.2 will be deposited on an openly accessible sequence database (such as GenBank). Protocols on the archiving of data products generated in WP1 (both the intermediate data products generated as part of the mobilisation of grey literature in T1.2.1 and the new data generated as part of the field campaign in T1.2.2) in the open-access extension of the EuroAPPA repositories.

Specific considerations on metadata for human participant and other data

Publications and datasets will have bibliographic metadata attached. It will be in a standard format and include the terms "European Union (EU)" & "Horizon Europe"; the name of the action, acronym & grant number; publication date, length of the embargo period, if applicable; and a persistent identifier. The metadata will comply with anonymisation processes.

3.2 Making data accessible

Metadata will be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement. It is an overarching aim of Butterfly to make data and results visible and freely accessible and to ensure long term data preservation. Access to research data should be '**as open as possible, but as closed as necessary'**, and here, there are some differences between the types of data.

Ecological and biological data

All ecological data sets that Butterfly will assemble, such as the plant-pollinator network information garnered from the literature review (openly accessible in DoPI) and field campaigns undertaken as part of WP1, will be archived on at least two open-access repositories: Zenodo and EOSC and another platform that specializes in biodiversity data. Biological data (i.e. DNA sequences) will be archived in GenBank. The European Open Science Cloud (EOSC) enables the storage, sharing, processing and reuse of digital research outputs following FAIR practices. EOSC will support the long-term legacy of Butterfly-initiated research, through the preservation of unpublished project knowledge and data.

The EuroAPPA portal will provide open access to all ecological data gathered in WP1. For Butterfly's source code of software for the APIs and R packages (WP1, 5) **GitHub** will be used as a repository for the archival of source code related to digital deliverables. Curated

data from T2.1 on agricultural practices, agricultural systems, value chains, other economic sectors, and ecosystems, will culminate in a Data Repository (D2.1).

Human participant and economic data

In general, all metadata of the data will be opened. Main repositories should be Zenodo, EOSC or disciplinary repository if relevant. However, for human participant data, the **'as closed as necessary'** needs specific consideration, and measures must be taken to accommodate protection of privacy and GDPR. Anonymized and De-identified human participant data can be archived in EOSC and in repositories consortium partners' home countries (e.g., the Norwegian Agency for Shared Services in Education and Research) to ensure transparency and reproducibility of our social science and humanities research in WP2, 4, 6 and 7. **Quantitative** data from the large-scale surveys on citizens' perception and willingness to pay for preserving the pollinators and pollination services (in 6 countries) will be anonymous, and the data can be deposited in Zenodo.

Qualitative data can in some cases be more challenging to anonymize (see ethics Chapter). Thus, the exact extent of openness of the actual datasets can be amended. When in doubt, the consortium will refrain from publishing raw datasets and only report aggregate measures. Data that cannot be anonymised due to practical or technical reasons is excluded from publication to ensure sufficient protection of the fundamental rights and freedoms of the (potentially) affected data subjects. Data that can be curated and be made de-identifiable can be shared more broadly. Decisions will be made on a case-by-case basis to ensure that privacy, anonymity, and confidentiality are not breached by publication of datasets or any other type of publication. Consultation with the relevant Data Protection Offices can be sought during the lifetime of the project.

3.3 Making data interoperable

File formats that are universal, cross platform, open source, with open standard will be applied, such as (xt, pdf, csv, tsv etc.).

In general, in order to ensure that the data is interoperable, the project will set up an interdisciplinary internal review process, see chapter 1.2. As the project progresses and the project consortium has collected data and gained interdisciplinary experience through e.g., the internal data review process, further information on making data interoperable will be outlined in subsequent versions of the DMP.

For **Ecological data**, we will use Darwin Core Terminology (DCT). The DCT is standardised vocabulary for transmitting information about biodiversity in a fully interoperable way. We will follow the vocabulary set out by Salim et al. (2022), specific for pollination interactions and based on the Darwin Core standard. For newly generated field data, interoperability will be ensured by the use of spreadsheet templates provided by Butterfly's field protocol (Diniz et al., 2025), already adapted to the DCT vocabulary. For re-used data, reviews and the DoPI database (built from existing data) will also follow Salim et al, maximising interoperability of data stored in multiple locations.

3.4 Increase data re-use

In order to increase possibilities for data-re use, documentation and metadata is important. Metadata standards are specified above. Further, datasets should be

accompanied by background or contextual information (eg., in Readme files) in order to enhance understanding of the data, its validity, and how it could be re-used, including limitations for re-use (eg very specific contextual data cannot be fully reapplied or reanalysed for completely different contexts - this is particularly relevant for human participant data). In addition, the datasets need to clearly specify the specific data usage licence.

The **internal review process as described in section 1.2**, in addition to ongoing public engagement activities and co-creation workshops, will further increase the refinement of data and project outcomes to make them understandable for third parties.

The Butterfly DECE plan (D8.2), and chapter 2.5 of this document, specifies how data produced in the project could be usable for third parties after the end of the project.

4. Other research outputs

During the project, it is anticipated that 36 deliverables will be produced (see Butterfly DOA Annex 1,, pp 20-23). Most of the outputs in these deliverables are not datasets, and include reports, protocols, tools, models, policy briefs, software and academic publications. All publications will be made available through trusted open-access repositories and linked to the Butterfly website and data portal (EuroAPPA), maximising their reach and long-term usability for researchers, practitioners, and policy actors. The full explanation of how these outputs will be disseminated is provided in Butterfly's Dissemination, Exploitation, Communication and Engagement **(DECE) Plan (D8.2**).

For the purpose of this DMP, the table below indicates how groups of outputs are anticipated to be stored and disseminated in line with a FAIR approach and Open Access best practices.

Table 3: Groups of anticipated "other research outputs"

Output type	Deliverables	Anticipated storage point
Interactive Website	EuroAPPA Atlas (Butterfly's European Atlas of Plant-Pollinator Associations): One-stop shop for pollinator-plant interactions (D1.1, D1.4)	EuroAPPA Website , (integrated with EU Pollinator Hub, GBIF, Zenodo and others)
Decision-support tools / reports	Policy scenario simulation and recommendations (D3.3)	**TBD**
Syntesis reports	Economic valuation of pollination services (D.2.2), Toolbox for resilience thinking (D4.2), Literature review on human	Project website and Zenodo

	dimensions of pollinator decline (D6.1), Butterfly tool introduced to LLs (D7.2),	
Analysis Reports:	Impacts of pollinator shocks (D3.1, D.3.2), Pollinator conservation in EU legislation (D6.2, 6.3), History of human and social determinants of pollinator loss and restoration (D6.4), Territorial diagnosis of Living Labs (D7.1)	Project website and Zenodo
Sector specific reports	Report of Delphi Survey & resilience options for each sector (D4.1)	Project website, and disseminated via sectoral trade associations, such as Cosmetics Europe, EFPIA, AnimalhealthEurope etc
Policy Briefs	D8.3, D8.6	Project website
Pollination Alert maps / Landscape mitigation tools	D5.1, D5.2, D5.3, D5.4	Project Website / EuroAPPA Website
Environmental mitigation tool - impacts of pesticide choices	D5.5	Project Website
Protocols, modelling code, methodology reports, action plans:	Protocol on modelling framework for joint distribution of pollinators and plants (D1.3) Protocol on delphi surveys (D4.1), Reports on Butterfly tools in the LLs (D7.2, D7.3) Action plan for increasing pollination services in LLs (D7.4),	Project website Zenodo GitHub for source codes

	Dissemination, Exploitation, Communication and Engagement plan (D8.2-D8.5) Code documentation on CGE model development (WP3)	
Training resources for agricultural schools, universities, and vocational programmes, exhibitions, etc.	D8.4	TBD
Research publications, scientific conference presentation	Not specified, but anticipated in journals in diverse fields, including pollinator ecology, risk studies, economics, social sciences, law, environmental humanities.	Open-access journal for papers, Zenodo for presentations

5. Allocation of resources

Managing data according to the FAIR brings two overarching types of costs:

1) Fees for depositing data in global data repositories. We have chosen to use Zenodo, which is free of charge for uploading data. .

2) Article processing charges (APC) for publishing data in open access journals.

Each partner of Butterfly is expected to use their budget responsibly and prioritise open access publications. In case of uncleared cost by common generated data, the responsibility and costs will be discussed in the Executive or General meeting

6. Data security

-

Following the UoB storage guide[2], and as mentioned in 3.2, data will be stored in two trusted repositories. This will ensure data recovery if needed.

[2] https://www.uib.no/en/foremployees/153608/storage-guide

Data will be stored and processed on each partner's own harddrives.
To share data, partners will use Nextcloud with password protection.

Complying with GDPR, no personal or identifying data will be stored with response data. Such personal/identifying data will be kept in a separate, password protected location with access only for authorised members of the project team. It will not be shared between partners.

No personal data that is considered sensitive, according to the EC commissions definition[3] will be collected in this project.

7. Ethics

For ethical compliances to be effective, it is important that the main responsibility for ethical considerations is kept close to the actual empirical work. Therefore, and as noted in the introduction, data collectors and task leaders have the ultimate responsibility for complying with local and specific GDPR policies and applicable local, government and international laws, regulations and guidelines. In the following sections, some general and overarching ethical considerations will be outlined.

Ethical issues related to biological and ecological data

Ethics and sample collection: As specified in the WP1 Field protocol, it is the responsibility of every B-Site leader to obtain the collection permits for their sites and to ensure that sample collection and exportation comply with the Nagoya Protocol.

Ethical issues related to human participant data

As summarised in chapter 2 of this DPM, human participant data will be collected in WP2, WP4, WP6 and WP7. In Butterfly's 'Ethics self-assessment' (chapter 4 in the DOA)), the following ethical issues were identified: 1) human participation, 2) personal data collection of data subjects. An initial assessment was included in the GA and summarised below:

> *The consortium will ensure that all necessary procedures are followed, particularly with regard to the signing, collation, and storing of all necessary Informed Consent Forms prior to the collection of any data. All involved stakeholders and citizens will be informed in detail about measures and the consortium will obtain free and fully informed consent*

[3] *The following personal data is considered 'sensitive' and is subject to specific processing conditions:*
- *personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs;*
- *trade-union membership;*
- *genetic data, biometric data processed solely to identify a human being;*
- *health-related data;*
- *data concerning a person's sex life or sexual orientation.*

https://commission.europa.eu/law/law-topic/data-protection/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en

All necessary actions will be taken within the project management and by all beneficiaries to ensure compliance with applicable European and national regulations and professional codes of conduct relating to personal data protection. This will include in particular Directive 95/46/EC regarding data collection and processing, the General Data Protection Regulation (GDPR, 2016/679), and respective national requirements, ensuring legal and regulatory compliance. Ethics considerations will feed into research and data collection protocols used in the project. This will include the collecting and processing of personal data as well as surveys and interviews. For all identified issues, in line with the above standards, ethical approvals will be obtained from the relevant national data protection authorities and/or institutional boards.

In addition to relevant national data protection authorities, the university partners have separate institutional ethics boards or respective national research boards, which will ensure the correct implementation of all human participation and data protection procedures and protocols around social science research. In detail, this includes for Norway (UiB, UiA) the Norsk senter for forskningsdata (Sikt)

(Butterfly 2024, Description of the action (DoA) Part B, pp 34-35)

In order to follow up this summary, the following general measures will be taken:

- For each engagement activity with participants where the purpose is to generate data, an **informed consent** sheet will be provided that specifies the purpose of the data collection, how personal data will be anonymised and stored, etc, and specify participants can decide to withdraw at any point. In the Appendix in this document, a generic informed consent form is provided that can be used, adapted and translated by task leaders and project members who are carrying out the different data collection activities.
- The informed consent sheet must also specify if the data is intended to be open access, including the planned processes of anonymisation/pseudonymisation, and any potential risks to their identification.
- Personal data will be a**nonymised and/or pseudonymised** by the person collecting the data in each country. All identifying information will be removed, including names, addresses, e-mails, phone numbers, IP addresses, and other contact information. As example, In WP4, the anonymity of individual respondents will be ensured, and quantitative results will be presented in aggregated form. Qualitative data (arguments) can be quoted, but the identity of the author will not be used.
- For qualitative research that often relies on rich contextual data and small samples, full anonymisation can be more challenging. Here, **pseudonymisation**[4]

[4] The GDPR defines the term ‘pseudonymisation’ for the first time in EU law and refers to it several times as a safeguard that may be appropriate and effective for the fulfilment of certain data protection obligations, according to the European Data Protection Board (2025). Guidelines 01/2025 on Pseudonymisation.

can be considered as an alternative. Pseudonymization is a de-identification procedure which involves replacing identifiers with pseudonyms or codes, and it offers a balance between protecting participant privacy and enabling meaningful research. As example, in WP7, personal data will be pseudonymised by that each LL assigns a unique ID number to each participant, stores codes in a secure file and replace all names with ID numbers. Only the data file with ID number will be sent to WP2 for analysis. In WP6, in order to protect the personal data of participants, all information collected from participants will be first pseudonymized until data collection is complete and then anonymized. Only anonymized data will be shared between partners outside of participating partners (i.e., UT and TUM).

- For discipline specific data collections, such as psychological experiments in WP6, all experimental procedures will be checked by a local ethics board for compliance with European, national, and local standards.
- As mentioned in chapter 6, no personal data will be shared among partners or transferred from the country where it is collected. Only anonymized data will be shared with other project members for analysis.
- Measures will be taken to avoid questions that provide recognisable data. For example, where information about gender, age or place of birth is not necessary, these questions will be avoided to ensure that participants are not recognizable in the pseudonymised data.
- Data collection will be carried out in different countries and ethics boards or similar must be consulted in each country, as rules and procedures differ between countries.
- To minimise research fatigue among participants, re-use of existing data is encouraged, such as farm data for WP 2.
- For **WP7,** a Memorandum of Collaboration (MoC) will clarify in detail the responsibilities and rights of participants, access to information and results obtained, and processes for resolving issues arising among members. However, **it does not replace informed consent.** When specific activities are set up for human participant data collection, such as a focus group interview, individual interviews, more detailed consent form has to be provided.
- Ethics sessions will be held in the Butterfly's consortium meetings, where training on e.g. anonymisation/pseudonymisation and informed consent procedures will be given and where ethical concerns and experiences can be discussed. This may enable ethical reflexivity and may hinder that ethics is reduced to procedural ethics - to protocols and committee approvals which may perceived as bureaucratic hurdles (Gillam & Guillemin 2018).

https://www.edpb.europa.eu/system/files/2025-01/edpb_guidelines_202501_pseudonymisation_en.pdf

References

Agudelo-Zapata, M. C., Marentes Herrera, E., Taliga, C., Parr, C.S., Cox-Foster, D., Hill, E., Maués, M.M. Agostini, K. Rech, A.R., Saraiva, A. (2024). WorldFAIR (D10.3) Agricultural biodiversity FAIR data assessment rubrics (Version 1). Zenodo. https://doi.org/10.5281/zenodo.10719265

Breeze T. D., Bartomeus I., Damiens F., Kantelhardt J., Kleijn D., Klein A.M., Meinilä J., Schaak H., Sihvonen H., Stoyanova C. and Zugic N.J. (2025). VALOR Standard Operating Procedures. VALOR project deliverable D8.1.

Butterfly (2024b) Description of the action, Part B

Delso, N. S., Ramponi, E., Otalora, S., Froyen A.J, Vlontzos, G., Filipiak, F., Drivdal, L., Miettinnen, M., Ohman Niesen , M. B., Kulan, N., Biella, P., van der Sluijs, J. (2025). D8.2 - Dissemination, Exploitation, Communication and Engagement (DECE) Plan

Diniz, U.M., Myint, A.A., Rasmussen, C., Ollerton, J., Chipperfield, J., Fossøy, F., Leonhardt, S.D. Field protocol (WP1). Description and instructions for the standardization of the sampling design, data collection methods, and sample processing across B-Sites.

Drucker, D. P., Salim, J. A., Poelen, J., Soares, F. M., Gonzalez-Vaquero, R. A., Devoto, M., Ollerton, J., Kasina, M., Carvalheiro, L. G., Bergamo, P. J., Alves, D. A., Varassin, I., Tinoco, F. C., Rünzel, M., Robinson, D., Cardona-Duque, J., Idárraga, M.,

General Data Protection Regulation (GDPR) - Regulation (EU) 2016/679. Regulation - 2016/679 - EN - gdpr - EUR-Lex

Gillam, L., & Guillemin, M. (2018). Reflexivity: Overcoming mistrust between research ethics committees and researchers. *The Sage Handbook of Qualitative Research Ethics, London: Sage, pp263-275*.

Ollerton, J., Taliga, C., Salim, J.A., Poelen, J.H., & Drucker, D.P. (2025) Incorporating measures of data quality into plant-pollinator databases. Journal of Pollination Ecology 38: 151-160

Pkhakadze, N., Stoyanova, C. (2025). Data Management Plan (DMP) following FAIR data principles for collection, storage, evaluation, & exploitation of data during & after the project. VALOR project deliverable D8.3.

Wintermantel, A., Thompson, A., Fornoff, F., Lenohart, S., Prucker, P., Basaran., Z., Kleoftodimos, G., Anderson, G., Rundoff, M., Schweiger, O., …. *Breeze, T.* (2024). *RestPoll Data Management Plan 1.* Available at https://restpoll.eu/publications/?term=deliverables

Zhang, J. and I. Steffan-Dewenter (2022*). Data Management Plan. Deliverable D8.3*

Appendix A: General informed consent form

Butterfly	Insert Insitution Namee

Information letter and consent form

About Project Butterfly

The decline of pollinator populations poses a serious threat to ecosystems and food security, with cascading effects on biodiversity and economic stability. In this context, the EU-funded Butterfly project will strengthen society's ability to anticipate and respond to these challenges. More specifically, it will establish geographically diverse multi-stakeholder communities to collaborate on proactive restoration solutions for pollinators. Through innovative tools and strategic alliances, Butterfly will integrate pollinator management across sectors, ultimately informing EU policies and promoting resilience in vulnerable communities. Read more there: butterfly-europe.eu

About this data collection activity

A minimum of points be filled in by data collector or task leader:

1. Some sentences on the purpose of the activity
2. Format of activity (e.g., interview, workshop, survey, etc.),
3. Duration of activity (e.g., 1 hour),
4. If notes will be taken or the activity will be audio- or video-recorded,
5. How data, notes and/or recordings will be stored, how they will be analysed and when they will be deleted,
6. Anonymisation or pseudonymisation process,
7. Who will have access to raw data and not (eg, analysers in other countries only will get anonymised data)
8. A statement that participation is voluntary and a procedure for withdrawal. Preferably add this: You may withdraw your data at any time up to the data of data processing [date] by submitting an e-mail to the activity leader. Your data will be treated in strict confidence and held on a password protected computer and encrypted cloud storage drive accessible only to [names and affiliations of all data holders].
9. Who to contact for more details.

Consent form:

Name and organization of data collector: (to be filled in by research team).

Name of the research participant:_____

I,: _____ (the research participant), have been informed that:

1. Data is being collected as part of the project Butterfly.
2. Data will be used for scientific analysis, publication and dissemination activities.
3. Data will be anonymized for publication/dissemination purposes.
4. Anonymised data will be analyzed by (insert task leader name).
5. Participation is voluntary.
6. Consent for participation in the project can be withdrawn by contacting the data collector, before (insert date), after which date the data will be anonymised.
7. (If applicable): The conversation will be voice recorded for transcription and will subsequently deleted.
8. Data will be used by the [specify partner institution] and information containing personal identification will not be exchanged.
9. (If applicable): I give permission for the anonymised data I provide to be deposited in an open data repository so it can be shared and used for learning and potentially reused for future research.

Signature: _____ (participant)

Signature: _____ (data collector)

Date _____

Article 13 - EU GDPR: **"Information to be provided where personal data are collected from the data subject"**
1. Where personal data relating to a data subject are collected from the data subject, the controller shall, at the time when personal data are obtained, provide the data subject with all of the following information:
(a) the identity and the contact details of the controller and, where applicable, of the controller's representative;
(b) the contact details of the data protection officer, where applicable; *Please contact Aarhus University at dpo@au.dk*
(c) the purposes of the processing for which the personal data are intended as well as the legal basis for the processing;
(d) where the processing is based on point (f) of Article 6(1), the legitimate interests pursued by the controller or by a third party;
(e) the recipients or categories of recipients of the personal data, if any;
(f) where applicable, the fact that the controller intends to transfer personal data to a third country or international organization and the existence or absence of an adequacy decision by the Commission, or in the case of transfers referred to in Article 46 or 47, or the second subparagraph of Article 49(1), reference to the appropriate or suitable safeguards and the means by which to obtain a copy of them or where they have been made available.
2. In addition to the information referred to in paragraph 1, the controller shall, at the time when personal data are obtained, provide the data subject with the following further information necessary to ensure fair and transparent processing:
(a) the period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period;

(b) the existence of the right to request from the controller access to and rectification or erasure of personal data or restriction of processing concerning the data subject or to object to processing as well as the right to data portability;
(c) where the processing is based on point (a) of Article 6(1) or point (a) of Article 9(2), the existence of the right to withdraw consent at any time, without affecting the lawfulness of processing based on consent before its withdrawal;
(d) the right to lodge a complaint with a supervisory authority;
(e) whether the provision of personal data is a statutory or contractual requirement, or a requirement necessary to enter into a contract, as well as whether the data subject is obliged to provide the personal data and of the possible consequences of failure to provide such data;
(f) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.
3. Where the controller intends to further process the personal data for a purpose other than that for which the personal data were collected, the controller shall provide the data subject prior to that further processing with information on that other purpose and with any relevant further information as referred to in paragraph 2.
4. Paragraphs 1, 2 and 3 shall not apply where and insofar as the data subject already has the information.