



SAFEGUARD

Data Management Plan

Deliverable D8.3

14 February 2022

Jie Zhang, Ingolf Steffan-Dewenter
University of Würzburg, Germany

Safeguard
Safeguarding European wild pollinators



Prepared under contract from the European Commission

Grant agreement No. 101003476.

EU Horizon 2020 Research and Innovation action

Project acronym	Safeguard
Project full title	Safeguarding European wild pollinators
Start of the project	September 2021
Duration	48 months
Project coordinator	Ingolf Steffan-Dewenter Julius-Maximilians-Universität Würzburg https://www.safeguard.biozentrum.uni-wuerzburg.de/
Deliverable title	Data Management Plan
Deliverable n°	D8.3
Nature of the deliverable	Report
Dissemination level	Public
WP responsible	WP8
Lead beneficiary	Julius-Maximilians-Universität Würzburg (UWUE)
Citation	Zhang, J. & Steffan-Dewenter, I. (2021). Data management plan. Deliverable D8.3 EU Horizon 2020 Safeguard Project, Grant agreement No 101003476.
Due date of deliverable	Month 6
Actual submission date	Month 6

Deliverable status:

Version	Status	Date	Author(s)	Reviewer(s)
1.0	Draft	16 February 2022	Jie Zhang UWUE	Ingolf Steffan-Dewenter UWUE
2.0	Draft	21 February 2022	Jie Zhang UWUE	Ingolf Steffan-Dewenter UWUE
3.0	Draft	25 February 2022	Jie Zhang UWUE	PCC members/ All PIs
4.0	Final	28 February 2022	Jie Zhang UWUE	Ingolf Steffan-Dewenter UWUE

The content of this deliverable does not necessarily reflect the official opinions of the European Commission or other institutions of the European Union.

Table of contents

Summary	4
List of abbreviations	4
1. Data Overview	5
1.1. Data collection	5
1.2. Data types and formats	6
1.3. Data size	6
2. FAIR data management	7
2.1. Making data findable, including provision for meta data	7
2.2. Making data accessible	8
2.3. Making data inter-operable	8
2.4. Data re-use	9
3. Allocation of resource	13
4. Data security	14
5. Ethical aspect	14
Appendix: Safeguard data sharing and publishing policy	17

Summary

This report describes the data management procedures applied in Safeguard for secure and sustainable management and exchange of research data, in accordance with the GDPR and principles of FAIR data management (Findable, Accessible, Inter-operable, Re-usable). We first describe the overall infrastructure of how data is managed in Safeguard, followed by the sources, types and estimated size of the research data to be used and collected in Safeguard. Safeguard upholds each pillar of the FAIR principle using the provisions and capacities of an online database tailored to the project needs for research data to be transferred to the Biodiversity Data Centre (BDC) or another recommended European pollinator data repository after the project end, as well as clear open access policies. We also describe the allocation of resources to data management in Safeguard, and the measures taken to ensure data security and compliance with ethical data principles.

During the Safeguard funding period, all research data collected or generated will be stored in the online Safeguard database (www.safeguard.biozentrum.uni-wuerzburg.de), which will be hosted and maintained by the University of Würzburg (UWUE). Personal data will be anonymized before upload. This database accepts access from only Safeguard partners via validated user accounts. For external reviewers, an account is approved case by case. Access of research data requires a request-grant procedure provided by the database, where only authors have the control to grant access. At the end of the funding period, validated research data will be migrated to the Biodiversity Data Centre (BDC) held by European Environmental Agency or emerging platforms as part of the European Pollinators Initiative to guarantee long term preservation and accessibility as well as promoting data re-use.

List of abbreviations

EU	European Union
FAIR	Findable, Accessible, Inter-operable, Re-usable
GDPR	General Data Protection Regulation
BDC	Biodiversity Data Centre
DM	Data Manager
DMP	Data Management Plan
PCC	Project Coordination Committee
CA	Consortium Agreement

1. Data Overview

Significant progress in both research on pollinators and public awareness of their importance has been made during the last 20 years, substantially driven by EU-funded projects. However, major knowledge gaps remain, existing data is fragmented, often not publicly accessible, and an integrated framework for preserving pollinators and their functions is lacking. Safeguard aims to accelerate the rate of scientific and societal progress made so far to finally halt and reverse pollinator declines and maintain Natural Capital in the EU.

Safeguard will implement the concepts and methods of FAIR and intensive data use, developing a database that supports the entire data life cycle, including not only data collection, analyses and preservation, but also open-access sharing, integrative analysis, citation and re-use. Safeguard will create a broad range of data types, including observational and experimental primary data, geographical data, economic data, social science data, citizen science-driven data, and modelling results. Further, existing data on multiple pollinator taxa, ecosystem services and environmental and socioeconomic global change drivers will be compiled, documented and archived. To foster the attractiveness and value of the database for all partners and multiple stakeholders, web-based scientific workflow functions will be provided for data quality check, aggregation, analysis, visualisation, and new frequently requested procedures. Such functionalities will promote cross-disciplinary analysis between partners and strongly support the overall data synthesis and integrative modelling approaches in Safeguard. To ensure long-term accessibility, the database will be transferred to the Biodiversity Data Centre (BDC) held by European Environmental Agency or emerging platforms as part of the European Pollinator Initiative. The final decision depends on the development of the Biodiversity Information System for Europe (BISE) and recommendations of involved DGs and EEA.

Relevant Safeguard datasets will be published as data papers to make them citable ensuring credits for data generators, in open access data journals or in public data repositories with Digital Object Identifier (DOI) assignment such as GFBio or Dryad. All geodata publicly shared in Safeguard will be in accordance with the INSPIRE requirements (<https://inspire-geoportal.ec.europa.eu/>) and the open data strategies of the national countries. The project will ensure strict application of ethical requirements arising from surveys, interviews, and collection of other data with privacy implications.

1.1. Data collection

Safeguard will compile empirical pollinator distributional data and trait data from past and ongoing European and national research, institutional and citizen-science driven monitoring schemes and digitalisation of museum collections. The original distributional and trait data will be stored in a tailored database hosted at Mons whilst aggregated datasets for data synthesis will be uploaded to the Safeguard database.

Safeguard will collect field data from countries across Europe (Belgium, Estonia, France, Germany, Hungary, Italy, Romania, Serbia, Spain, Sweden, Switzerland, UK) to investigate spatial distributions and temporal trends of pollinator biodiversity, pollination services and responses to multiple pressures in (semi-)natural habitats, agricultural landscapes, and

urban areas. All field data will be documented and stored in the Safeguard database, made accessible through our Safeguard database and will be transferred to the Biodiversity Data Centre (BDC) held by European Environmental Agency or emerging platforms as part of the European Pollinator Initiative at the end of project, unless legal constraints prevent us from doing so.

1.2. Data types and formats

Research data in Safeguard can be categorized due to different curation methods into structured data and unstructured data. Structured data is preserved and indexed in database tables, which provides further manipulation possibilities such as filtering, aggregation, statistical analysis and visualisation directly from the server. Structured data uses ASCII format for uploading/downloading to guarantee maximum interoperability. Unstructured data will be stored as Binary large object (BLOB) in the database management system. It includes spatial data such as shapefiles, satellite images, videos, audios, and images. We use ESRI shapefiles for spatial vector data, Geo-tiff format for spatial raster data, MP4 format for videos, MP3 format for audios, JPEG format for images, in order to be consistent and compliant with the most common software applications across different operating platforms.

Our metadata will be in XML format, which is standardized throughout the project and is machine-readable. This metadata standard is based on Darwin Core and Ecological Metadata Language, because a large part of the data generated in our project is from earth and environmental science. We will further simplify its syntax and append it with new semantic vocabularies from socioeconomic science and other disciplines involved. We will use this metadata standard for all the data types presented in Safeguard to allow inter-disciplinary interoperability. If ontologies are proven necessary during the data collection and documentation, we will map our vocabularies to ontologies that is compatible with ontologies offered by open data centres such as INSPIRE.

1.3. Data size

The initial estimation of total size for research data is 4TB. Collected and generated data will be curated and preserved in the Safeguard database that will follow data quality, security, and privacy standards and relevant EU legislation (i.e., GDPR). Each WP as well as local partners will use the Safeguard database for research data preservation and exchange.

2. FAIR data management

The overall approach of Safeguard aims to make data and results visible and freely accessible and to ensure long term data preservation. For Safeguard research data, in accordance with FAIR data principles, we will provide a multi-user online database – Safeguard Database - that provides open access to metadata, secured authentication for data users, intuitive procedure for data preservation and versioning, a request-grant procedure complying with Safeguard data sharing and publishing policy for data exchange, as well as toolset for data analysis on the fly. Safeguard database is accessible via the Safeguard homepage (www.safeguard.biozentrum.uni-wuerzburg.de).

2.1. Making data findable, including provision for meta data

Data of the Safeguard project will be preserved and exchanged between project members in the Safeguard database. This database will be developed based on BExIS (Biodiversity Exploratory Information System, www.bexis.uni-jena.de). It will adapt the specific requirements of the Safeguard project and be further improved and developed with new functionalities by UWUE. We chose our own database solutions instead of other data centres to guarantee that 1) full control of research data access 2) own designed meta data structure to meet the specific need of Safeguard project and is compatible with international standard 3) datasets are deeply integrated via the shared study sites, merging and aggregation tools are offered by the database for data synthesis 4) trainings and real time consultant are offered throughout the project phase. We follow the “as open as possible, as closed as needed” principle to restrict and regulate data access using the following approaches: 1) metadata is openly accessible while research data is restricted with user log in; 2) only project members are allowed to create a log-in account to the database; 3) only authors can view and download his/her datasets, others have to go through the online automatic request-grant procedure where authors can decide when to grant access to request; 4) every data download is recorded and traceable; 5) users must agree to the Safeguard data sharing and publishing policy before downloading.

All the research data preserved under the Safeguard database will follow the FAIR principle with the following details: 1) data identifier: Every dataset will have a persistent and unique identifier called Dataset ID throughout the entire project; 2) metadata: Metadata input is mandatory for all research data and is recorded at the time of dataset creation. Users will input their metadata manually via the online metadata creation wizard of the Safeguard database. Safeguard metadata uses the Extensible Markup Language (XML) standard. It will be based on Darwin Core (<https://dwc.tdwg.org/>) and Ecological Metadata Language EML (<https://eml.ecoinformatics.org/>), further simplified and extended upon specific requirements of interdisciplinarity in the Safeguard project and be made compatible with metadata standard from data archive centres (e.g. INSPIRE, Pangaea). The metadata is openly accessible for all audiences and does not require user log in. It will help to attract international interests and guarantee transparent research; 3) search function: Users can search for various types of keywords (author, year, research topic, or other free text) in the metadata, thus making

datasets easily findable; 4) versioning: Versioning is included in all structured datasets and is identifiable via Version ID. One dataset can be linked with multiple versions. When downloading a dataset, a specific version ID need to be selected in order to download the desired version.

2.2. Making data accessible

During the funding phase, research data is made accessible for project members via the Safeguard database, which is hosted by UWUE. The database is made available via a website (Safeguard homepage) encrypted by Hypertext transfer protocol secure (HTTPS) and is accessible internationally in all types of electronic devices and software platforms.

In Safeguard database the metadata of datasets are made publicly accessible for all audiences beyond the scope of the Safeguard project. This aims at better transparency for Safeguard research as well as attracting more public attention and possible further research collaboration. Observation data, or primary data collected/generated by Safeguard members is closed during the project period where only authors are allowed to access and edit. Other members of Safeguard, who are interested in accessing these data, should follow the Safeguard data sharing and publishing policy and a request-grant procedure (see Figure 1) provided by the Safeguard database.

After the funding phase of Safeguard, datasets will be reviewed by the Safeguard data manager and data owners and must be approved before becoming candidates for contribution to the open research pilot. Following in principle approval, Safeguard will make the datasets available through Biodiversity Data Centre (BDC) held by European Environmental Agency or emerging platforms as part of the European Pollinator Initiative. After migration a unique DOI will be assigned to the safeguard dataset and be updated in the metadata. Safeguard database is expected to be running (though might without maintenance) several years beyond the funding period.

2.3. Making data inter-operable

Safeguard data can be exchanged and re-used within the project period among all the institutions involved in the Safeguard project. Safeguard uses data formats that are universal, cross platform, open source, with open standard, and not lock-in formats. For example, structured data will be exchanged in ASCII format such as txt or csv, spatial vector data stored as shapefiles, raster data as geotiff format, documents as PDF, images as JPG. Details of data types and formats are specified in Chapter 1.2.

Data exchange is based on agreement of Safeguard data sharing and pushing policy (see Appendix) and follows a request-grant procedure as shown in Figure 1: 1) the requester writes a small proposal text on the database website about why and how he/she plans to use these data; 2) the database will automatically send an email to all authors with the proposed request and the requester, but only the first author will receive a link to grant or deny access; 3) when authors read this email and make a decision on this request, the first author would click on the link to either grant or deny this access request; 4) the requester will get an automatic email

notification after his/her request is granted or denied. If the requester is granted with access, he/she is able to download this dataset.

This procedure allows authors to have full control of who can access their data, how and when their data should be implemented in synthetic research. Safeguard data sharing and data publishing policy (see Appendix) is available as a public page in the Safeguard homepage and is included as attachment in the downloaded dataset. Any Safeguard member who wants to request others' data needs to agree to this policy before he/she can continue with the downloading procedure. Documentation and tutorial of the database will be presented in text format as well as video format.

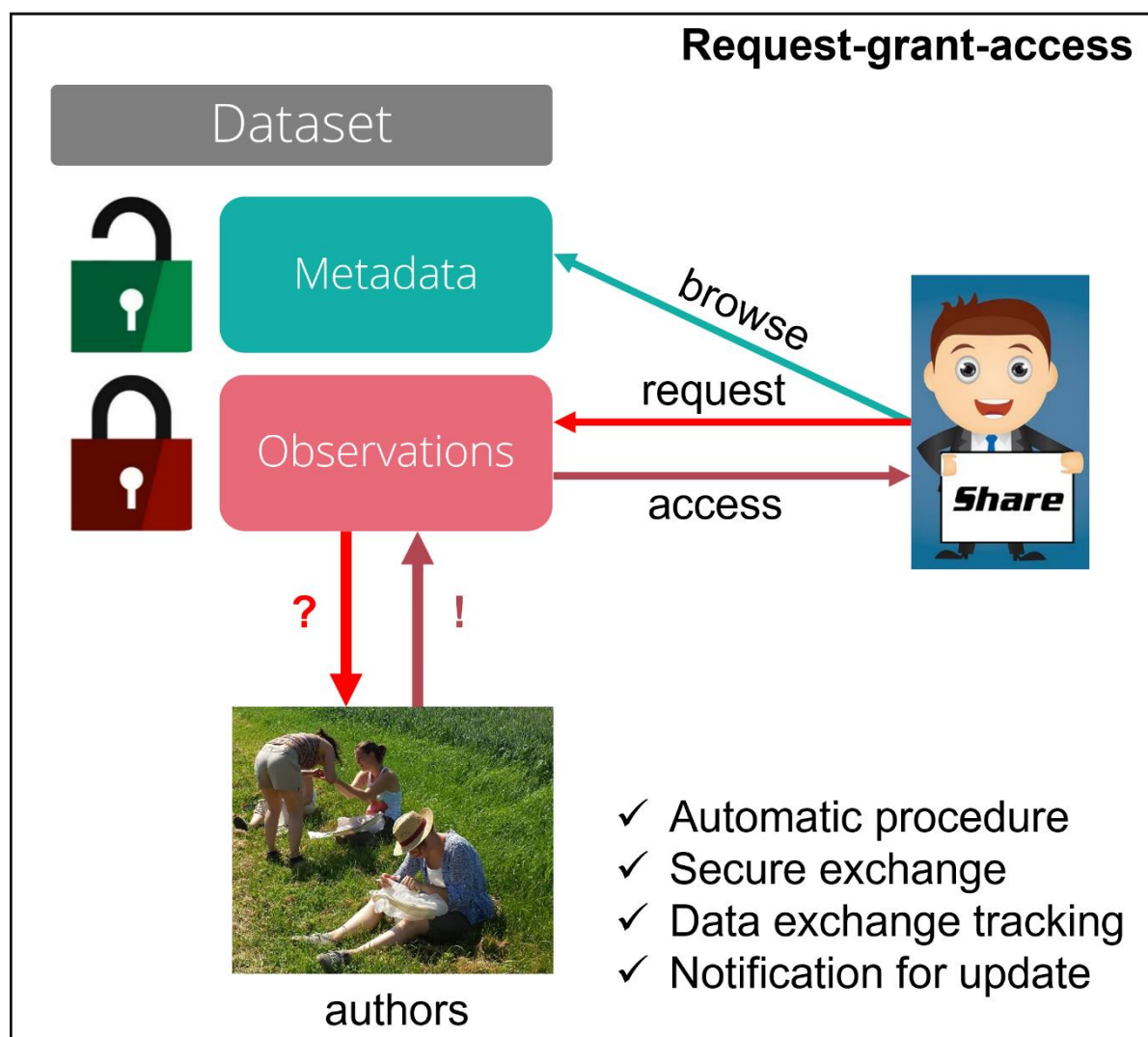


Figure 1: Request-grant procedure for data exchange in Safeguard

2.4. Data re-use

Safeguard will compile and re-use data from past and ongoing European and national research, institutional and citizen-science driven monitoring schemes and digitalisation of

museum collections. Building upon strong involvement of partners in major European pollinator projects (Table 1) and an extensive national, European and global network of collaborations with researchers, public institutions and NGOs, Safeguard will significantly improve the knowledge base on pollinators, their spatial distribution, temporal trends, biotic interactions and ecosystem services. Safeguard will assemble European-scale data for an extensive set of pressures utilising high-resolution data from the Copernicus Land Monitoring and Climate Change Services and multiple other sources.

Table 1: Examples of international research, innovation and policy activities feeding into Safeguard.

Acronym	Project details	Relevant outputs	Safeguard partners
STEP	Status and Trends of European Pollinators (FP7, 2010-2015)	Status, trends and drivers of pollinator populations; potential drivers; first EU Red List of bees	UREAD (coord), + 9 partners
PoshBee	Pan-European assessment, monitoring, and mitigation of Stressors on the health of Bees (H2020, 2018-2023)	Pesticide, pathogen and nutritional stressors on bees and tools to mitigate these	RHUL (coord), + 8 partners
EPI	European Pollinators Initiative (2018-2020 and then to 2030)	Strategic objectives and actions for EU and Member States to address pollinator declines	IUCN, IEEP, UREAD and many partners
EUPMS	European Pollinator Monitoring Scheme expert group	Design of EU monitoring scheme and pollinator indicators for CAP	UREAD (chair), + 5 partners
IPBES	Intergovernmental Platform on Biodiversity and Ecosystem Services (2014-2019), 'Pollination' and 'Global' assessments	Assessment of drivers, state, impacts and responses for pollinators, wider biodiversity and ecosystem services	UREAD ('Pollination' chair), UFZ ('Global' chair)
Red Lists	Bee, butterfly, hoverfly EU and national Red Lists	Conservation status, data and maps of EU pollinators	IUCN (lead), + 3 partners
Orbit	ORBIT: Taxonomic resources for European bees (DG ENV, 2021-2024)	A centralised taxonomic facility groundwork for the identification of European wild bee	UMons (coord), + 2 partners
eBMS & ABLE	European Butterfly Monitoring Scheme, & Assessing Butterflies in Europe	Maps, indicators, models and methodologies for EU butterflies	UFZ (co-lead)
Promote Pollinators	Promote Pollinators (PP; formerly Coalition of the Willing on Pollinators)	Links to >30 national pollinator strategies	PP subcontracted to UREAD
LIBER-ATION	Linking farmland Biodiversity to Ecosystem services for ecological intensification (FP7, 2013-2017)	Effectiveness of interventions on pollination and other services at multiple spatial scales	WU (coord), + 6 partners

EKLIPSE	Knowledge and Learning Mechanism on Biodiversity & Ecosystem Services	Knowledge synthesis processes; key societal & policy needs	UFZ (new coord)
MAES	Mapping and Assessment of Ecosystems and their Services	Framework and indicators to map and assess ecosystem conditions	JRC (lead),
VOODOO	Viral Eco-evolutionary Dynamics of Wild and Domestic Pollinators under Global Change (BiodivERsA 2020-2023)	Links between bee biodiversity, drivers, disease dynamics and human wellbeing	INRAE (coord) + 2 partners
Open-NESS	Operationalisation of natural capital and ecosystem services (FP7, 2012-2017)	Natural capital and ecosystem services concepts and applications	UFZ, WU, JRC, MTA-OK
OBServ	Open Library of Pollinator Biodiversity and Ecosystem Services Scenarios (BiodivERsA)	Cost-benefit analysis on use of models to develop scenarios of pollinators and services	CSIC (coord), WU
SUPER-B	Sustainable Pollination in Europe (COST Action)	State-of-the-art knowledge on pollination; pollinator loss	Most Safeguard partners
Showcase	Showcasing synergies between agriculture, biodiversity and ecosystem services to help farmers capitalising on native biodiversity (H2020, 2020-2025)	Develops, tests and scales-up biodiversity indicators and farmland interventions	WU (coord, + 8 partners)
Scales	Securing the conservation of biodiversity across administrative level and spatial, temporal, and ecological scales (FP7, 2009-2014)	Co-benefits of pollinators for pest control, spill-over effects among protected areas and agricultural land	UFZ (coord), + 5 partners
ALARM	ALARM: Assessing large-scale environmental risks with tested methods (FP 6, 2004-2009)	Methods for pollinator monitoring, environmental pressures, pollination services	UFZ (coord), + 5 partners

Safeguard will participate in the Open Data Pilot of the EU Horizon 2020. Open access to research data will be provided whenever there is no conflict with the protection of results and GDPR and under the supervision of the consortium. During the project phase, collected and generated data will be curated and preserved in the Safeguard database where project members are able to exchange their data. At the end of the project, validated datasets will be transferred to the Biodiversity Data Centre (BDC) held by European Environmental Agency or emerging platforms as part of the European Pollinators Initiative to promote data re-use. Embargo times for opening access to datasets are defined by each dataset owners in the metadata, but in theory not later than 3 years after data collection. In practice if one needs longer embargo time, he/she shall inform the Safeguard data security manager the reason for

delay and the estimated time for data publishing. All data that can encompass any personal data protection or privacy and IPR issues will not be publicly disclosed, and datasets will be stripped of these issues (e.g., anonymized) before publication following the procedures described in chapter 5 (Ethical aspect).

We will adhere to the Open Access policy in H2020 by depositing published versions or final peer-reviewed, accepted manuscripts in a repository for scientific publications using the Open Access Infrastructure for Research in Europe (OpenAIRE) as an entry point and publish articles in open access journals thereby fulfilling the requirements for the “gold Open Access (OA)” standard. Where it is not suitable to publish in fully open access publications, we will ensure a minimum Green Open Access standard. Further, we plan to use open access journals like Research Ideas and Outcomes (RIO) to publish entire research cycles including research plans, study designs, methodologies, statistics and results. Relevant Safeguard datasets will be published as data papers to make them citable ensuring credits for data generators, in open access data journals or in public data repositories with Digital Object Identifier (DOI) assignment such as GFBio or Dryad. All geodata publicly shared in Safeguard will be in accordance with the INSPIRE requirements and the open data strategies of the national countries. The project will ensure strict application of ethical requirements arising from surveys, interviews, and collection of other data with privacy implications (see Chapter 5 Ethical aspect).

3. Allocation of resource

WP8 is responsible for the data management life cycle monitoring for all datasets to be collected, processed or generated by the project. The overall cost for data management is ca. 17 personal month. To ensure compliance with data management decisions as they relate to the DMP, the following measures apply in Safeguard:

- WP leaders are considered responsible for adhering to the specifications above in their respective work packages.
- The principal investigators and the Data Protection Officer (see deliverable D9.1) of each beneficiary organization are considered responsible for the DMP actions. They will support WP8 in all issues related to research data management. The principal investigators of each beneficiary should ensure that personnel working on the project have read the Data Management Plan and apply/exercise all the principles as described in the Safeguard DMP (this document).
- Data collectors have the ultimate responsibility of complying with the specifics of the Safeguard Data Management Plan, as well as with the related GDPR policies.
- For the overall Safeguard project activities, the Data Manager (Jie Zhang, UWUE) is the main contact point to assist compliance of all project partners with the DMP guidelines to ensure secure and correct data handling in Safeguard, including technical support for the database. The DM regularly reviews all technical components of secure data handling and storage and communicates whenever the need arises to ensure safe and correct handling of data exploitation issues.

UWUE is in charge of research data management in Safeguard. This includes costs for a server (server certificate, tuning, maintenance, backup), software, and the DM for developing the database, providing counsel and support, organising data workshops and other training. To achieve long term data preservation, we will transfer our data to the Biodiversity Data Centre (BDC) held by European Environmental Agency or emerging platforms as part of the European Pollinator Initiative. The data migration will be carried out by the DM and data owners. The goal is to keep Safeguard data accessible and re-usable for at least 10 years after the funding phase.

4. Data security

Safeguard database will be hosted in a Windows server provided by the computer centre of UWUE. The server is backed up automatically on a daily basis, secured by firewall of UWUE and the firewall of the server itself. Safeguard database uses PostgreSQL as database server, which offers encryption at several levels and provides flexibility in protecting data from disclosure due to database server theft, unscrupulous administrators, and insecure networks. Safeguard database website uses encrypted connection via HTTPS and applies encrypted Forms-Based Authentication. Each user account in the Safeguard database is verified and validated by the Data Manager (DM). In theory this database only accepts accounts from Safeguard partners. For external users, e.g. project reviewers, their account will be discussed within the PCC about its necessity and be approved individually. PIs will be responsible for the data security in their local facilities before data uploading to the Safeguard database.

5. Ethical aspect

Ethical aspects regarding data management are described in detail in Deliverables D9.1 (Ethics requirements). Safeguard will ensure that Article 34 of the Horizon 2020 Framework Programme (Grant Agreement- Ethics) is properly implemented and, in particular, that all the data sharing and re-use activities for this Horizon 2020 project comply with ethics principles and relevant national, EU and international legislation.

Ethical aspects in using material from surveys and interviews: The project will ensure strict application of ethical requirements when preparing, conducting, and evaluating interviews. Procedures of using material from surveys and interviews will be documented, and informed consent from all participants will be ensured before starting the interviews. We will confirm the content to be generated, processed, and analysed, and products to be created in the scope of the project that may include anonymised data. This especially includes explicit agreement of all participants on how to document, store and use sensitive data. The consent form will include information about the project, how the data will be used (only for research purposes), and a checklist indicating the types and/or intended uses of consented material including but not limited to (1) audio recording; (2) video recording; (3) photographs; (4) transcriptions and annotations; (5) translations. Anonymity will be ensured.

This applies to: (a) data retrieval: consent forms and information sheets will be provided, (b) data processing (oral biographies / stories): Interview partners will sign written agreements in which they indicate which data they allow to use, and under which conditions (e.g., modification of data to avoid any risk of correlation) the data may be re-used, (c) data storage: Interview partners will give their consent to location and (restricted) accessibility of data, and to potential further re-use of such data.

Protection of Personal Data: The collection and use of individual data & more generally private information will be reduced to a minimum on a “need to use basis” maintaining data confidentiality. All participating partners need to ensure legal requirements related to the use of private data according to EU rules and standards. It will be made sure that also possible

subcontractors will adhere to these rules. For the planned socio-economic surveys data like zip-code and function (e.g., owner of a business, consumer etc.) may be required but it will be made sure that no answers can be traced back to single participants or single groups of participants. If the survey is performed online, it will be made sure that the encrypted connection via HTTPS is applied, data will be stored in institute computer and the Safeguard database instead of cloud storage offered by third party, IP-addresses of the participants are also be treated according to national and European law. Storage time of anonymized data is not limited (no longer than the specified maximum conservation time), but again according to prevalent laws, the informed consent and the time needed to reach the goals for this project.

Furthermore, the people involved will handle the data accurately, securely, confidentially, anonymously, and ensure that data are processed in accordance with the data subject's rights. All information will be stored in a confidential manner and in accordance with the EU Directive 95/46/EC and GDPR regarding use of personal data. It will be ensured that the following requirements are met: (a) Confirmation that data will be collected on a need to know basis only (b) The guarantee of withdrawal rights and oblivion rights as made compulsory by the European court of justice in 2014 (c) Avoidance of merging data sets in order to prevent any unforeseen personal information disclosure (d) Provision of detailed information on the procedures that will be implemented for data collection, processing, storage, protection, retention and destruction and confirmation that they comply with national and EU legislation.

Non-EU participants in Safeguard include partners from Switzerland and Serbia. Regarding the ethical issues in Safeguard, especially data protection, it will be made sure by the principal investigators and the Data Protection Officer (see Deliverable D9.1) of each beneficiary organization that all measures undertaken in the study comply with ethics principles and relevant national, EU and international legislation. For all personal data (e.g. transcripts of interviews), the project, more specifically the Principal Investigators and the Data Protection Officers of the corresponding beneficiary organizations will provide a declaration of confirming compliance with Chapter V of the EU General Data Protection Regulation (GDPR), as well as a declaration of confirming compliance with the laws of the country where the data was collected.

Personal data will be anonymized before it can be uploaded to the database. All the data stored in the database will be exchanged among partners regardless of EU or non-EU participants following the request- grant procedure (Figure 1) and agreement to the data sharing and publishing policy (Appendix).

Safeguard uses an external service SendInBlue for newsletter sent to project members. It has access to the email information of project members in Safeguard. SendInBlue is committed to processing this data in compliance with the applicable laws and regulations, and in particular Law No. 78-17 of 6 January 1978 relating to Information Technology, Data Files and Civil Liberties (the Data Protection Act), and Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data (GDPR).

Geographic coordinates: Geographic coordinates of the study sites as well as their landscape and climate characteristics will be stored as one dataset in Safeguard database. The identifier of study sites (the primary key of this dataset) will be called Plot ID. Other datasets which use the study sites will use Plot ID as reference but not including the actual coordinates. When we

migrate our data to open access data archives, the Plot IDs will be made publicly accessible while the study sites dataset will apply another license to make sure the sensitivity of the geographical locations (e.g. coordinates of a farm) are protected.

Appendix: Safeguard data sharing and publishing policy

Safeguard Data Sharing and Publishing Policy

Within the Safeguard project, many publications will arise within and across Work Packages (WPs). In addition, the Safeguard project offers exciting opportunities for collaborative data synthesis and publications. This offers the advantages of synergy and increased scientific progress. The Safeguard project specifically requires agreements on the documentation, quality, sharing, and publication of data. These agreements are the subject of this policy.

A. DATA POLICY OF THE SAFEGUARD PROJECT

§1 Coverage and definitions

- a. This policy applies to all members and staff
- b. “Project data” in the sense of this agreement are conventional measurements (data from measuring instruments, field surveys, generally in tabular form), images (digital photos, satellite data, maps, etc.), GIS-Data (spatial data, raster maps), model scripts, genetic data as well as resulting (digital) publications, thesis, proceedings and posters generated within the project.
- c. “Metadata” shall mean any data describing and documenting the project data.

§2 Data Management and Publication

- a. The Safeguard Project Coordination Committee (PCC) sets the data standards and also serves to adjudicate possible disputes relating to this policy. In case of disputes, a majority of votes is necessary to make a decision. The PCC is headed by the PI of WP8.

§3 Rights and responsibilities

- a. Project members (PIs, Postdocs, PhD students, other staff) have a right of access to project data according to the following provisions. The right of access to project data is limited to scientific use.
- b. Furthermore, project members have a right, that their intellectual input and their academic interest is respected with regard to the project data in accordance to §5.
- c. Each member and staff agrees to have her/his data with metadata stored in the central database of the Safeguard project at the University of Würzburg.

d. The Data Manager (DM) and the respective PIs are responsible for the management of this database.

e. In cases of a serious violation of obligations under this agreement the PCC of the Safeguard project may impose adequate sanctions.

§4 Documentation of datasets

All datasets must be documented with meta-information. All data sets will be stored with information on who collected them, version of the data and date of the latest update.

a. Details on deadlines and data formats will be given by WP8.

§5 Access to data in the Safeguard project database

All members and staff have free access to the meta-information.

a. After request and approval by the respective data owner, each member or scientific staff will get access to the original data in due time. The scientists originally obtaining the data and the PIs of their projects will be informed when data are used by other partners and give their consent prior to data publication. Of course, scientists making use of data supplied by other scientists for scientific publication must acknowledge the use of the data appropriately (see below).

b. Release of data to non-members of the Safeguard project will be an exception for which explicit permission has to be sought of the PCC.

c. All project data will be made publicly available considering privacy issues and in accordance with standards set by the Safeguard CA EU legislation, as well as in accordance with the EU Data Protection Directive 95/46/EC and GDPR 2016/679.

§6 Use of data collected by Safeguard project

a. Data use must always be based on an agreement between the original data supplier and the data user. Original data suppliers are the scientists originally obtaining the data and the PIs of the respective projects. The latter are responsible for supplying the data obtained by all staff of their project to the database.

b. Data accessed by a scientist must only be used for purposes necessary to carry out his/her own work in the Safeguard project. Data accessed must only be used for scientific purposes, i.e., commercial use of data is not allowed. It is prohibited to distribute other scientist's data to a third party without the written consent of the scientist.

§7 Delivery of data and quality control

a. Data needs to be deposited as fast as possible, normally within half a year, and at the latest one year after the field sampling or laboratory analysis has been completed.

Data quality is controlled by careful review of the different components saved to the database.

B. PUBLICATION POLICY OF THE SAFEGUARD PROJECT

The success of the Safeguard project depends on successful publication in appropriate journals. Many collaborative papers will have several authors. The people most important in the derivation of a particular result are acknowledged by first authorship (usually the PhD students or Postdoc). To avoid uncertainties about access to data or authorship, Safeguard project agrees on the following publication policy.

§1 General publication rules

a. The Safeguard project adheres to the guidelines on good scientific practice in European countries including

France (<https://comite-ethique.cnrs.fr/wp-content/uploads/2020/09/COMETS-GUIDE-EN.pdf>),

Germany (https://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/),

Netherlands (https://eneri.eu/wp-content/uploads/2018/10/Netherlands2BCode2Bof2BConduct2Bfor2BResearch2BIntegrity_2018_UK.pdf),

Spain (http://enacti2ng-itn.cbm.uam.es/images/Documentation/CBP_CSIC_en.pdf), Sweden (https://www.vr.se/download/18.5639980c162791bbfe697882/1555334908942/Good-Research-Practice_VR_2017.pdf) and UK (<https://ukrio.org/publications/code-of-practice-for-research/>).

b. The planning of an experiment or a field campaign, the planning and execution of data collection, the analysis of the data, and the writing of a manuscript all are necessary steps that can merit authorship.

c. Of course, usually the person who originally obtained the data (often PhD students) will publish the results of his/her disciplinary study. To also promote swift publication of collaborative papers in appropriate journals, the following procedure is adopted. Before writing a collaborative paper, credit to the data suppliers is discussed. Often, this credit will consist of co-authorship. Because collaborative synthesis does complement rather than forestall disciplinary publications, the access to data needed for collaborative synthesis shall be provided by the data supplier.

d. Members and staff planning a publication distribute a tentative title of the paper, a preliminary list of authors and the anticipated journal to the internal Safeguard project mailing list as early as possible, latest one month before the manuscript is finished. This serves for claiming acknowledgement as under §2, rapid information on ongoing activities and allows improvements of synthesis work at an early stage.

Members and staff planning to submit a paper to a journal distribute the final version to the internal Safeguard project mailing list as early as possible, latest 2 weeks before submission. This serves for claiming acknowledgement as under §2, rapid information on ongoing activities and allows final improvements of synthesis work.

§2 Acknowledgement of original data suppliers and co-authorship

a. In the case of providing unpublished data to a paper, the credit of the original data supplier can be in the form of co-authorship, mentioning in the acknowledgements or in the figure or table legend or as “personal communication”. Co-authorship is appropriate if the data contribute important information to the publication. The other options are appropriate if the data are only used as additional side information. The original data suppliers should decide what form they find appropriate.

b. PhD students and PIs are encouraged to co-author where appropriate. For example, if a PhD project receives intellectual input into the design from a PhD student or a PI from another project, then co-authorship may well be appropriate.