ForestPaths

Co-designing holistic forest-based policy pathways for climate change mitigation

D7.5 Data Management Plan (DMP)

01/03/2023

Authors: Nikol Yovcheva, Anna Sapundzhieva, Pavel Stoev, Hans Verkerk, Jerker Brolén, Katriina Pajari, Thomas Pugh

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Research and Innovation Programme (ID No 101056755), as well as from the United Kingdom Research and Innovation Council (UKRI). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the EU nor the EC can be held responsible for them.

Prepared under contract from the European Commission and the United Kingdom Research and Innovation Council.
Grant agreement No. 101056755
EU Horizon Europe Research and Innovation Action

Project acronym:	**ForestPaths**
Project full title:	**Co-designing Holistic Forest-based Policy Pathways for Climate Change Mitigation**
Project duration:	01.09.2022 – 28.02.2027 (54 months)
Project coordinator:	Dr. Hans Verkerk, European Forest Institute (EFI)
Call:	HORIZON-CL5-2021-D1-01
Deliverable title:	D7.5 Data Management Plan (DMP)
Deliverable n°:	D7.5
WP responsible:	WP7
Nature of the deliverable:	Data Management Plan
Dissemination level:	Public
Lead partner:	PENSOFT
Recommended citation:	Yovcheva, N., Sapundzhieva, A., Stoev, P., Verkerk, P.J., Brolén, J., Pajari, K. & Pugh, T. (2023). ***Data Management Plan (DMP)***. ForestPaths project deliverable D7.5.
Due date of deliverable:	Month 6
Actual submission date:	Month 6

Deliverable status:

Version	Status	Date	Author(s)
1.0	Draft	08 February 2023	Yovcheva, N., Sapundzhieva, A., Stoev, P., Pensoft Publishers
2.0	Review	10 February 2023	Brolén, J., Pajari, K., Verkerk, H., European Forest Institute
3.0	Revised draft	15 February 2023	Yovcheva, N., Sapundzhieva, A., Stoev, P., Pensoft Publishers
4.0	Second review	21 February 2023	Brolén, J., Pajari, K., Verkerk, H., European Forest Institute
5.0	WP leaders review	27 February 2023	TUM, ULUND, VITO, LUKE, EFI
6.0	Final	28 February 2023	Yovcheva, N., Sapundzhieva, A., Stoev, P., Pensoft Publishers
7.0	DMP Update 1	10 November 2023	TEES, TUM, ULUND, VITO, LUKE, CMCC, PENSOFT, EFI

Table of contents

Key takeaway messages

- Deliverable D7.5 Data Management Plan (DMP) provides a comprehensive overview of ForestPaths' data management practices.
- The DMP documents the use of data in ForestPaths' scientific work and describes the data types, licences, formats and data repositories planned to be utilised by the project.
- The DMP defines data management recommendations for all partners when generating, collecting, or using research data in the framework of ForestPaths in order to safeguard the FAIR data principles, while considering the ethical aspects of data management, such as sensitive or personal data.

Summary

In order to warrant the findability, accessibility, interoperability and reusability of its research data and other outputs, ForestPaths has developed a tailored Data Management Plan (DMP). It is based on the open access principles established in the project's description of action, as well as on information collected from a consortium-wide survey. The plan first provides a summary of which project partners will generate and reuse data, for what purposes, under what types, formats and sizes, as well as when will the data become available and for whom would it be useful (Chapter 2). The DMP then goes over the FAIR data principles, identifying how ForestPaths will implement these in its data management and listing a number of recommendations that partners are advised to follow in order to ensure FAIR data (Chapter 3).

Since ForestPaths will produce other research outputs besides data (software, models, protocols), the DMP also considers their management and how they can be made available to stakeholders (Chapter 4). The plan then goes over the administrative aspect of data management, such as the allocation of resources (Chapter 5), data security (Chapter 6), intellectual property rights (Chapter 7) and ethics (Chapter 8). The DMP is a living document which will be updated to provide a finer level of granularity as the project progresses and when significant changes occur. The plan will have at least one update, planned for M48. The changes from updates are highlighted in green, with corresponding dates provided in footnotes[1].

List of abbreviations

APC	Article Processing Charges
API	Application Programming Interfaces
DMP	Data Management Plan
DOI	Digital Object Identifier
DPO	Data Protection Officer
EML	Ecological Metadata Language
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation

[1] Added in DMP Update 1, 10 November 2023.

WP	Work Package
UK	United Kingdom

1 Introduction

As an EU Horizon Europe project, ForestPaths' approach to the scientific process follows the principles of open science, i.e., it is based on open cooperative work, tools and diffusing of knowledge. Adhering to these principles warrants that ForestPaths' research results and important discoveries extend past the project and accelerate research progress and its societal impact. The project is therefore to ensure that partners have at their fingertips a series of procedures and guidelines which enable them to provide open access to their research outputs free of access charge or other barriers to all potential end-users.

In light of that, ForestPaths work package 7 is tasked with developing Deliverable D7.5 Data Management Plan (DMP) in month 6. The DMP aims to 1) document the use of data in ForestPaths' scientific work; 2) enable subsequent use of the research data or its derivatives for further research with machine-readable metadata; and 3) facilitate data security, protection and documentation. It consolidates the project's data management practices and describes the data types, licences and formats to be used by ForestPaths in order to ensure the implementation of the FAIR data principles. In addition, D7.5 contains a series of recommendations regarding the procedures which project members are encouraged to follow when generating, collecting, or using research data, while also considering the ethical aspects of data management, such as sensitive or personal data which has to be properly protected. The DMP determines the datasets which the project will publish for open use and identifies trusted data repositories where these could be published. To add value to the DMP, an additional one-pager with Data Management Guidelines will be produced and shared with partners, summarising the current recommendations and serving as a guiding tool.

The DMP has been elaborated based on the initial open access principles established in the project's description of action, as well as on information collected from a consortium-wide consultation process. Project partners were approached with a detailed survey containing 13 questions (available in Annex 1), covering findability, accessibility, interoperability and reusability of the project's research results and data. The survey was filled in by the 12 project organisations that will store, generate or reuse data in relation to ForestPaths' research: EFI, ULUND, TUM, KIT, LUKE, WR, VITO, PBL, PI, UTBV, UEDIN and TEES. Each of these organisations submitted a comprehensive answer to the survey (*Milestone 16 Data management survey filled out by partners*), identifying the datasets that the organisation will generate and/or reuse, as well as detailing their institutional data management practices and preferences.

2 Data summary

ForestPaths will collect, generate and reuse data in relation to the following project objectives:

- **Improve the understanding of factors that shape decision-making behaviour** by forest practitioners across Europe **on adopting and locally adapting climate and biodiversity-smart forest management** to inform policy pathways [WP1&3].

- **Develop advanced high-resolution monitoring methods of climate change related risks from forest disturbances** to advance the modelling of future impacts on forest composition, structure, and carbon pools in Europe [WP2&3].
- **Assemble a next generation integrated assessment framework** incorporating improved methods, models, tools and data to assess holistically the contribution of European forests, forest management and the forest-based sector in achieving climate, energy and biodiversity targets under future climate conditions [WP3-5].
- **Co-design, quantify and evaluate holistic forest-based policy pathways** for climate change mitigation that consider climate change related risks, impacts and feedbacks of the forest socio-ecological system, as well as adaptive management practices, biodiversity goals, and the provision of ecosystem services [WP5&6].
- **Maximise ForestPaths' impacts by fostering collaboration** with universities, research institutes and intergovernmental bodies, and liaising with relevant networks and initiatives to promote synergies, integration and cooperation and facilitate knowledge exchange with targeted stakeholders [WP7].

ForestPaths' research will rely on a significant amount of self-generated data collected through field work, remote sensing, modelling, interviews, surveys and literature reviews. At this initial stage, the preliminary details of the datasets the project plans to generate are described in Table 1. On the whole, the vast majority of the datasets will be open access, with WP1 and WP6 anticipating 2 datasets with restricted access in order to comply with GDPR regulation ((EU) 2016/67)* and out of concerns that based on previous experience, stakeholders are more willing to participate if their data is not shared.

Overall, the project's datasets will contain both quantitative and qualitative data, concerning mainly demographic (from interviews and surveys), geographic (from modelling and remote sensing), and market (from modelling) information. The majority of the generated datasets are expected to be fairly small, i.e., under 3 GB in size each. However, there is a likelihood that datasets from EFISCEN-space, RCA, LPJ-GUESS and Landsat composites will exceed this size. Among the envisioned formats of the generated data are .docx, .xlsx, .csv, .tsv, .py, .mp4, .netcdf and .tif.

Next to the above-mentioned formats, ForestPaths will also reuse data under the .xml, .pdf, ecospold2, .shp, .hdf5 and .las formats. For own analyses the project will solicit datasets from the Food and Agriculture Organization of the United Nations (FAO), the GenTree Project, Web of Science, National Forest Inventories institutes (NFI), national institutes and governmental agencies, climate modelling centres, Natural History Museum (London), United States Geological Survey (USGS), National Aeronautics and Space Administration (NASA), European Statistical Office (EUROSTAT), ecoinvent, and FCBA technological institute.

Besides contributing to the objectives of ForestPaths and the specific internal needs of its work packages, the project's generated data will also be of use to the following stakeholder groups (more information on them in *Deliverable D7.1 Communication Plan (CP) and Plan for the Exploitation and Dissemination of Results [PEDR]*):

- Policy and governance (P)
- Forestry and related practitioners (FP)
- Forest value-chain and economic actors (FA)

- Research and academia (A)
- Civil society (CS)
- General public (GP)

An initial data summary for ForestPaths can be found in Table 1 and Table 2, reflecting partners' responses to the DMP survey. Table 1 contains information about the data ForestPaths expect to generate and Table 2 contains the details of the data that the project intends to reuse. Since not all the information is available this early in the implementation of the project, the DMP is perceived as a living document which will gradually be updated until it reflects the full picture of ForestPaths' research data.

Table 1: Summary of the data ForestPaths anticipates generating*

No	Name of the dataset	Name of the generator	Relevant task	Generated via	Size	Format	Type of data	Sensitive Personal data	Personal data	Delivery	Users	Access
1	Climate and biodiversity smart forest management practices	EFI	T1.1.3	Literature review	TBD	.docx	Qualitative data	No	No	M13	A	Open
2	Acceptability of climate-smart forestry practices (I)	EFI	T1.3	Interviews	TBD	.docx	Qualitative data	Yes	Income Name Email Gender	M8-M12	WP1	Closed due to ethical concerns and open sharing affecting willingness to participate
3	Acceptability of climate-smart forestry practices (II)	EFI	T1.3	Survey	TBD	.xlsx	Quantitative data	Yes	Income Name Email Gender	M13-M18	WP1	Anonymised data will be published after the embargo period
4	EFI-GTM scenario projections	EFI	T5.3	Forest sector scenario analysis with EFI-GTM	TBD	.csv	Quantitative data	No	No	M46	A	Open
5	Stakeholder mapping	PI	T6.1 T1.2	Publicly available data online Contact data provided by project partners	TBD	.xlsx	Contact data Demographic data	No	Name Email	M8	WP6	Closed as per GDPR regulation ((EU) 2016/67)*
6	Factors influencing forest management	TEES	T1.2	Literature analysis	TBD	.docx	Qualitative data	No	No	M16	A, P	Open

7	Forest management practices across Europe	TEES	T1.1	Literature analysis	TBD	.docx	Qualitative data	No	No	M16	A, P	Open
8	Substitution and biodiversity Displacement factors	VITO	T4.3.2	Data processing	2 MB	.xlsx	Quantitative data	No	No	TBD	A, FP	Open
9	EU wood flow	VITO	T4.1.2	Data processing	5 MB	.xlsx	Quantitative data	No	No	TBD	A, FP	Open
10	GWPbioEU impact assessment method	VITO	T4.4.1	Data processing	2 MB	.xlsx	Quantitative data	No	No	TBD	A, FA	Open
11	EFISCEN-Space outputs	WR	T5.2	Modelling	100 Gb	.xlsx .tsv	Quantitative data	No	No	TBD	A, P	Only on aggregated level (maps, gridded)
12	Policy tool and predicted outputs (statistical/MCA tool for analysing simulated impacts)	LUKE	T5.5	Data analysis	TBD	R script .xlsx, jpg	Qualitative data	No	No	M50	A, CS, P	Open
13	Scenario framework and definitions	LUKE	T5.1	Review Policy labs in WP6	TBD	.xlsx .csv .tsv	Qualitative data	No	No	M40	A	Open
14	Demo case validation results for simulated scenarios (TBD)	LUKE	T5.5	Data analysis Review	TBD	.xlsx .csv .tsv	Quantitative Qualitative data	No	No	M50	Forest Paths	Open
15	Evaluation simulations for LPJ-GUESS	ULUND	T3.1.3 T3.3.3	Modelling	1 Gb	.netcdf	Quantitative data	No	No	M27 (6–12-month embargo)	A, P	Open
16	Exploratory scenarios from LPJ-GUESS	ULUND	T5.2.1	Modelling	10 Gb	.netcdf	Quantitative data	No	No	M34 (6–12-month embargo)	A, P	Open
17	Policy pathway scenarios from LPJ-GUESS	ULUND	T5.2.1	Modelling	10 Gb	.netcdf	Quantitative data	No	No	M44 (6–12-month embargo)	A, P	Open

18	Policy pathway scenarios from RCA-GUESS	ULUND	T5.4.1	Modelling	100 Gb	.netcdf	Quantitative data	No	No	M44 (6–12-month embargo)	A, P	Open
19	Biodiversity responses to management	PBL	T3.3.1	Literature review	2-3 Mb	.xlsx .csv	Quantitative data	No	No	M18	A, FA	Open
20	Biodiversity impact factors	PBL	T4.3.3	Data processing	2-3 Mb	.xlsx .csv	Quantitative data	No	No	M36	A, FA	Open
21	GLOBIO output	PBL	T5.4.2	Modelling	TBD	.netcdf	GIS data	No	No	M46	A	Open
22	IMAGE scenario data per region	PBL	T5.2 T5.3 T5.4	Modelling	TBD	.xlsx .csv	Quantitative data	No	No	M46	A	Open
23	IMAGE scenario data Gridded map	PBL	T5.2 T5.3 T5.4	Modelling	TBD	.netcdf	GIS data	No	No	M46	A	Open
24	Landsat composites	TUM	T2.1	Remote sensing	>100 GB	GeoTIFF	GIS data	No	No	M13	A	Open
25	Pan-European forest Disturbance maps	TUM	T2.1	Remote sensing	2-3 GB	GeoTIFF	GIS data	No	No	M13	A, FA, FP	Open
26	Pan-European forest composition map	VITO	T2.2	Remote sensing	1-10 Gb	Cloud Optimized GeoTIFF	.R (GIS data)	No	No	M24	A, CS, FP, GP, P	Open
27	Pan-European forest structure (maps)	VITO	T2.2	Remote sensing	1-10 Gb	Cloud Optimized GeoTIFF	.R (GIS data)	No	No	M24	A, CS, FP, GP, P	Open
28	Land use change scenarios (CRAFTY output)	KIT	WP5	Modelling	TBD	GeoTIFF .netcdf	Spatial	No	No	TBD	A, CS, FA, FP, GP, P	Open
29	PLUM results	UEDIN	T3.4 T5.2	Modelling	TBD	.csv and others	Maps use and market information	No	No	TBD	A	Open
30	Romanian forests projections	UTBV	T3.1	Modelling	1 GB	.csv	Quantitative data	No	No	M16	TBD	Open & shared on request

Table 2: Summary of the data ForestPaths expects to reuse*

No	Name of the dataset	Relevant task	Size	Format	Sensitive Personal data	Personal data	Access	Origin	Ownership	Licence
1	FAOSTAT	T1.2 WP5	TBD	.xml	No	No	Open	FAO	FAO	N/A
2	Forest Management Atlas	T1.1.3	TBD	TBD	No	No	Open	GenTree Project	EFI	N/A
3	Factors influencing forest management	T1.2	TBD	.pdf	No	No	Open	Web of Science	Authors, Journal	N/A
4	Forest management practices across Europe	T1.1	TBD	.pdf	No	No	Open	Web of Science	Authors, Journal	N/A
5	Ecoinvent	T4.2.1	400 MB	ecospold 2	Yes	No	Closed	Mixed	Ecoinvent association	Commercial license
6	ForestPaths_WP1_Task 1.2.3_Forest practitioner survey _FI_Forest Centre data	T1.2.3	TBD	.xlsx	No	Yes	Restricted	Other organisation	Finnish Forest Center	N/A
7	Published data on forest structure	T5.5	TBD	gis	No	No	Open	Luke	Luke	CC-BY
8	National forest inventory data	T2.2, T3.1, WP5	TBD	.csv	No	No	Restricted	National Forest Inventories	Data producers (NFI)	Various, many of them restrictive
9	CMIP6 climate data	WP3 WP5	100s Gb	.netcdf	No	No	Open	Climate modelling centres	N/A	Open
10	PREDICTS	T3.3.1	TBD	.xlsx	No	No	Open	Natural History Museum, London	NHM	Not required
11	Disturbance reference data	T2.1	~100 MB	.shp	No	No	Open	Manually interpreted reference data on forest disturbances across Europe	No owner	CC4.0
12	Landsat	T2.1	~100 TB	GeoTIFF	No	No	Open	USGS/NASA	USGS/NASA	N/A

13	IceSat-2	T2.2	TBD	.hdf5	No	No	Open	NASA	N/A	CC BY 4.0
14	GEDI	T2.2	TBD	.hdf5	No	No	Open	NASA	N/A	CC BY 4.0
15	LUCAS	T2.2	TBD	.csv	No	No	Open	Eurostat	N/A	CC BY 4.0
16	Airborne LiDAR	T2.2	TBD	.las	No	No	Open	Different providers	N/A	Varies
17	HILDA+	WP3, WP5	TBD	TBD	No	No	Open	Multiple	KIT	None
18	FCBA technological institute	T4.2.1	5MB	ecospold 2	No	No	Restricted	FORMIT, GESFOR, other projects	FCBA	N/A

*The information provided in Table 1 and Table 2 is based on an initial data mapping and is subject to change in order to reflect the project's development and arising data needs. Updated information, as well as adjustments to the current information, will be provided on an as-needed basis and at the latest when updating the DMP in M48.

3 FAIR data

Transparency and accessibility of research results, as well as of the sources and methods used to create them are essential for scientific development which depends on the broadest possible level of data availability. To support such data transparency and availability, ForestPaths' DMP is designed to implement the FAIR data principles (Findable, Accessible, Interoperable, Reusable). These four elements are independent and separable but work in synergy to ensure that the project's data is as available to stakeholders as possible. In this context, research data refers to all the data used in the course of scientific work, which includes both primary data (raw data/input data) and processed data (output data), forming the basis of published results.

3.1 Making data findable

ForestPaths will ensure its finalised datasets which are made publicly available are findable by assigning them globally unique and persistent identifiers (e.g., Digital Object Identifiers). These identifiers will be one of the elements included in the project datasets' rich metadata. The term metadata refers to the standardised and structured dataset characteristics describing, among others, its origin, purpose, time, geographic location, creator, terms of access and terms of use. The availability of metadata helps locate resources and provides searchable information which enables users to easily find and cite existing data. ForestPaths will ensure its data is thoroughly described with metadata that clearly and explicitly includes the identifier of the data it describes, offers descriptive and structural information and provides a rich representation of the dataset.

In generating metadata, the project would aim to follow a unified metadata description standard which supports the findable, accessible, interoperable and reusable nature of data. It is important to select a standard which suits the data's type, ensuring its interoperability with other datasets in the field and enhancing its discovery. Since ForestPaths will produce several types of data, the project has identified a number of suitable standards which can be used depending on the data's nature.

Recommendation

ForestPaths partners are encouraged to apply one of the following metadata standards:

- ISO 19115: provides information about the identification, extent, quality, spatial and temporal aspects, content, spatial reference, portrayal, distribution, and other properties of digital geographic data and services;
- ISO/IEC 19506: suitable for representing existing software assets, their associations, and operational environments;
- Ecological Metadata Language (EML): includes modules for describing the spatial, temporal, taxonomic, and thematic extent of data, as well as research methods and protocols;
- INSPIRE metadata: requirements for the creation and maintenance of metadata for spatial data sets, spatial data set series and spatial data services set by the EC.

Recommendation

If no metadata standard is used, ForestPaths' generated data should at the very least be accompanied by metadata including (if relevant):

- Author(s)
- Year
- Dataset title
- Dataset description
- Data repository/archive
- Date of deposit
- Embargo period
- Global persistent identifier
- Version or subset
- Language
- Metadata language
- Licence of use
- Date of metadata creation
- Hierarchy level
- Character encoding
- Format version
- Keywords
- Horizon Europe funding: grant project name, acronym and number

In addition, ForestPaths aims to support the sharing of information within the consortium, which is why it will follow a specific naming convention for its data. Naming documents in a standardised and intuitive way enables partners to collaborate efficiently and easily discover project datasets when necessary.

Recommendation

ForestPaths datasets should follow the following unified naming convention: [ForestPaths_dataset-name_version_creation-date], whereby data format should be DDMMYYYY, number style version should be 01, 02, 03.

Example: ForestPaths_DisplacementFactors_v01_06022023.xlsx

Lastly, in order to ensure project (meta)data is registered, indexed and can be harvested in a searchable resource, ForestPaths will upload it to trusted repositories that offer search engines and indexing, such as Zenodo.

3.2 Making data accessible

ForestPaths builds on an open cooperative work approach and systematic sharing of knowledge and data as early and widely as possible. Therefore, the project will strive to provide open access to peer-reviewed scientific publications and the underlying datasets relating to its results and funded or co-funded by the project, with the exception of the datasets identified as closed in Table 1.

To ensure access to scientific publications, partners will deposit in a trusted repository a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication at the latest at the time of publication. Immediate open access will be provided to the deposited publication via the repository. Where relevant, information will be provided via the repository about any research output or tool needed to validate the conclusions of the scientific publications.

Recommendation

ForestPaths partners can choose to deposit their scientific publications in:

- a trusted certified repository;
- a trusted community-recognised repository;
- a trusted institutional repository.

The publicly shared underlying research data for scientific publications[2] generated during ForestPaths will be made available in open access through trusted repositories as soon as possible after the paper has been published and no later than the end of the reporting period during which the paper was published (respecting relevant embargo periods). Partners will also provide information via the repository about potential research outputs or tools needed to reuse or validate the data. The research data underlying ForestPaths' deliverables will be deposited in trusted open access repositories which support uploads with embargo status. The maximum embargo period for this type of data will be 12 months after the deliverable's submission date[3].

[2] Added in DMP Update 1, 10 November 2023.
[3] Added in DMP Update 1, 10 November 2023.

Recommendation

ForestPaths members can choose between two ways of making their underlying data open access:

- They can upload the data to an open access research data repository under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence with equivalent rights. They could use a generic repository such as Zenodo, a thematic one like GBIF or a trusted institutional repository.
- They can publish datasets as open access data papers in an academic journal and then deposit them in an open access repository.

The majority of the partners who expressed a preference for a trusted repository in which to deposit their datasets identified Zenodo as a suitable option, which they are familiar with and have previously explored appropriate arrangements regarding deposition.

Recommendation

ForestPaths partners are encouraged to deposit their datasets to Zenodo as the project can verify that this trusted repository covers all the requirements of FAIR data, most importantly:

- A DOI can be issued to every published record on Zenodo.
- Metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing.
- Metadata of each record is sent to DataCite servers during DOI registration and indexed there.
- Metadata for individual records as well as record collections is harvestable using the standard, open, free and universal OAI-PMH protocol by the record identifier and the collection name.
- Metadata is publicly accessible and licensed under public domain. No authorisation is ever necessary to retrieve it.
- Data and metadata will be retained for the lifetime of the repository, ensuring that the metadata will be accessible, even when the data is no longer available.
- Metadata is stored in high-availability database servers at CERN, which are separate from the data itself.

The project will follow the principle of data being “as open as possible as closed as necessary” and if deemed appropriate, datasets will have restricted access. An overview of the envisioned restricted/closed datasets and the respective reasons can be found in Chapter 2 and Table 1. In some cases, when the datasets are complete, partners will review them to determine whether

they can be made available in an aggregated form in order not to disclose sensitive information. Table 1 also contains the envisioned embargo periods for certain datasets.

Lastly, in line with the FAIR principles and if it does not go against legitimate interests, the project's metadata will be deposited under a Creative Common Public Domain Dedication (CC 0) or equivalent.

3.3 Making data interoperable

To ensure the automatic findability and interoperability of datasets, ForestPaths will follow a series of community-endorsed interoperability best practices. It will use common metadata vocabularies, standards, formats and ontologies that are often used for knowledge representation. In case it is necessary for partners to use uncommon ontologies and vocabularies, mappings will be provided to more commonly used ones. Furthermore, partners will assess on a case-by-case basis, whether the newly generated ontologies or vocabularies (if any) would be openly published to allow reusing, refining or extending.

Recommendation

ForestPaths partners are encouraged to deposit their publications to Zenodo which:

- uses JSON Schema as internal representation of metadata;
- offers export to other popular formats such as Dublin Core or MARCXML;
- for certain terms, uses open and external vocabularies, such as license (Open Definition), funders (FundRef) and grants (OpenAIRE).

ForestPaths will not produce large databases that need to be made available through Application Programming Interfaces (API). The project will ensure interoperability by using common file formats as identified in Table 1. Using such file formats, linked to rich, standard and machine-readable metadata (for example, EML) and deposited in open access trusted repositories that provide programmatic access (for example, Zenodo) ensures that ForestPaths' data can be queried and read by any programing language and without the use of proprietary software. Lastly, project data will include qualified references to other data where relevant to further enhance interoperability.

Recommendation

ForestPaths partners should accompany their data with a separate document containing descriptive folder structure, names, headings, units of variables, etc.

3.4 Increasing data reuse

ForestPaths' data will be described with a plurality of accurate and relevant attributes. Furthermore, it will be released with a clear and accessible data usage license and will include detailed documentation of the data's provenance using the relevant metadata standards, as described in Chapter 3.1.

Recommendation

ForestPaths partners are encouraged to make their datasets available under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence with equivalent rights. If they want to publish data associated with a journal article under a different license, authors should explicitly inform ForestPaths' coordinator.

3.5 Monitoring

Recommendation

To ensure the quality of their data, ForestPaths partners should perform regular data quality assurance, determining and screening anomalies in their data by means of data profiling, removing obsolete information and data cleaning.

To ensure the smooth implementation of the FAIR data principles, ForestPaths will develop and distribute a "FAIR data checklist". This document will have a dual purpose: to remind partners of the steps necessary to ensure FAIR data and to provide a monitoring mechanism for WP7. The checklist will be created in the form of an Excel sheet which partners have to fill in a month prior to their datasets' delivery. Based on the delivery months identified in Table 1, WP7 will monitor this document to ensure that partners are on the right track with their data management. The checklist will also regularly be shared with partners via ForestPaths' internal newsletter.

Recommendation

To ensure the smooth management of data in the project, partners are strongly encouraged to fill in the "FAIR data checklist" at the latest a month prior to their datasets' delivery date. WP7 will provide the checklist on the project's Teams channel and will share it with the consortium.

4 Other research outputs

Besides ensuring FAIR data management, with its DMP, the project also aims to address the management of other research outputs generated and/or reused by ForestPaths, such as software, models, policy briefs and training materials. Where possible, the project will strive to adhere to the FAIR principles detailed in Chapter 3.

However, ForestPaths will, among others, improve multiple existing models in which case it will respect access and use policies for each of these models. Accordingly, they will be accessible upon registration or request, based on scientific collaboration, and in accordance with the use policies for each model. In addition, the LCA tools used and developed in ForestPaths are based on BRIGHTWAY2 and available under BSD software license conditions. Further guidelines for the management of project models and tools will be provided when updating the DMP and when these have reached a more mature development stage.

Recommendation

ForestPaths members must manage the digital research data generated in the action in line with the FAIR principles (as described in Chapter 3 of the DMP). When applied to other research outputs, the FAIR data guidelines from Chapter 3 can be summarised as:

- Findable: strive to deposit your research outputs described with rich metadata in trusted repositories which assign them globally unique and persistent identifiers and offer search engines and indexing;
- Accessible: aim to publish your outputs in open access journals which accept the publication of less traditional research outputs or deposit them in trusted open access repositories;
- Interoperable: use standard formats, vocabularies and ontologies and accompany your outputs with a separate human-readable description of the output;
- Reusable: strive to make your outputs accessible under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence with equivalent rights.

ForestPaths' CANOPY platform, policy briefs and training materials will be uploaded on the project's website which will be maintained for at least five years after the project's end. Additionally, further actions will be explored to ensure their availability to stakeholders in line with the above-mentioned recommendation. The project will explore their publication in relevant journals and platforms which support the publication of less traditional outputs, such as the Research Ideas and Outcomes (RIO) Journal, Open Research Europe, One Ecosystem, Knowledge4Policy, the EOSC training marketplace, EFI's Policy Briefs series, etc.

Recommendation

To enhance the availability and openness of research outputs, project members should consider the opportunities offered by current open science practices, such as:

- early and open sharing of research through preregistration, registered reports, preprints, etc.;
- providing open access to less traditional research outputs through trusted repositories or publication in journals;
- participation in open peer review;
- use of 'Open Research Europe (ORE) Platform' as an open access platform for scientific publications and add data availability statements to publications where relevant;
- use of the Research Ideas and Outcomes (RIO) Journal, Open Research Europe, etc. as a means to publish less traditional research outputs, such as project reports, software descriptions, data papers, etc.

5 Allocation of resources

FAIR data management can be related to several types of costs, which could be grouped under two main categories:

1. article processing charges (APC) for publishing data in open access journals
2. fees for depositing data in global data repositories.

Sufficient resources have been allocated to ForestPaths members, and particularly to WP7 leaders who are responsible for the project's Data Management Plan. The DMP will be maintained up to date, reflecting the current data needs and practices of ForestPaths and ensuring the data's long-term preservation.

6 Data security

As established in Chapter 3, once ForestPaths open access datasets are compiled, they will be deposited in trusted repositories, which ensure long-term preservation and curation. Furthermore, project partners will safeguard non-finalised datasets on their institutional servers, which have in place regular and automatic backup procedures. Access to the servers is monitored via two-factor authentication processes and/or is protected by firewalls.

Recommendation

To ensure the security of ForestPaths data, partners are encouraged to:

- perform daily backup procedures;
- enquire with full accuracy who the official Data Protection Officer (DPO) of their institution is and maintain contact with them.

Furthermore, ForestPaths is dedicated to protecting personal data in relation to the project by complying with applicable data protection rules.

Recommendation

To protect personal data, ForestPaths partners should:

- inform respondents and participants of interviews, surveys and meetings about what will happen with their personal data and what are their rights in this respect;
- give respondents the opportunity to contact relevant project members with a request to have their personal data deleted;
- include informed consent for data sharing and long-term preservation in questionnaires dealing with personal data;
- use personal information of respondents and participants of interviews, surveys and meetings only if they have given their consent for its use;
- anonymise/delete the personal information when it is no longer needed;
- give access to personal information only on a need basis to fulfil tasks.

7 Intellectual property rights

Management of intellectual property will be conducted in accordance with the Consortium Agreement (CA) and the Grant Agreement. Results will be owned by the Party that generates them. Whenever results have been produced jointly by two or more participants, the ownership of the results will be shared among them. The terms of this joint ownership, protection, and sharing, and costs for possible protection will be agreed upon in writing via a joint ownership agreement. Each participant will be responsible for examining possibilities to protect results that may be commercially or industrially exploited, such as, for example, WP4 LCA data and tools.

8 Ethics

As established in the project's Grant Agreement and in the DMP's Chapter 6, ForestPaths will adhere to data security and ethical principles. Only data that is relevant and limited to the purposes of the project will be collected and processed. When personal information is collected as part of the research activities, participants will be informed that their personal information is used for the purposes of the project only. Personal data will be processed in accordance with applicable data protection rules. Consent will be asked from participants in interviews, surveys, and workshops, for the processing of personal data and the data will not be used for any other purpose without further consent.

If research data can be reasonably analysed without direct identifiers and there are no research grounds for storing identifiers, only data from which identifiers have been removed will be produced for research purposes and stored for secondary research. Where that is not possible, the digital subject identifiable data will be protected by a username and a password and will be

available only to the researchers directly working with that specific data. Results will be published in such a way that individual respondents cannot be identified from them.

Each project participant conducting surveys in WP1 will be a data controller and responsible for handling their own data in accordance with applicable data protection rules and ethical guidelines and procedures. Personal data collected under WP1 will be anonymised, securely stored in the repositories of these organisations.

Personal data, including public contact information such as names, email addresses, phone numbers, organisations, professional roles, and languages spoken, will also be collected for the Policy Engagement Forum and policy labs (WP6). This data will be collected and stored securely, with restricted and password-protected access by the partner who is arranging the activity. A privacy notice will be attached to event invitations. Data will be collected also for dissemination and communication purposes, for example, newsletter subscription, website statistics and cookies (WP7). The privacy policy and regulations have been established in accordance with applicable data protection rules and are available on the project website. Any photographs and videos produced for any DEC activities will require explicit consent for their use, which will be requested when any materials may identify individuals. The use of visual materials will be limited only to the dissemination of the project and will not be used otherwise.

9 Annex: Data Management Questionnaire

ForestPaths

Resume later

0%

ForestPaths Data Management Survey

1. First and last name

2. Organisation/institution

Questions 3 and 4 concern the data you will **generate**.

3. Please provide the following provisional information for your **generated data**:

1. **Name of the dataset**

2. **Name of the generator**: name of the person who will generate this data.

3. **Relevant task**

4. **Generated via**: for example, field work, modelling, data processing (LCA, MFA), remote sensing, literature review, policy review, interview, surveys.

5. **Size**: a rough estimate only if you know.

6. **Format**: for example, .docx; .xlsx; .pdf; .mp4; .xml; .csv.

7. **Type of data**: for example, qualitative data; semi-quantitative data; quantitative data; analogue data; digital data; GIS data.

8. **Sensitive data**: Yes/no. If yes, please specify, for example, racial, political, ethical, health, and more here.

9. **Personal data**: Yes/no. If yes, please specify, for example, name, surname, address, email, IP address, location data.

10. **Delivery**: a rough estimate of a timeline. If there is an embargo period, specify why and how long it will apply.

11. **Metadata**: the metadata accompanying your datasets.

12. **Users**: to whom they might be useful.

13. **Access**: will they be open access? If not, please indicate the reasons, for example, ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related, contract.

14. **Re-use**: potential documents or tools needed to re-use or validate the data.

 This question concerns the data you will **generate**.

 Please fill in at least one answer

	Dataset 1	**Dataset 2**	**Dataset 3**	**Dataset 4**	**Dataset 5**
Name of the dataset					
Name of the generator					
Relevant task					
Generated via					
Size					
Format					
Type of data					
Sensitive data					
Personal data					
Delivery					

Metadata					
Users					
Access					
Re-use					

4. Would you consider anonymising your **generated datasets** and publishing them in an aggregated form so as to not disclose private information?

If yes, please specify to which dataset(s) that would apply. If not, please explain why.

This question concerns the data you will **generate**.

Choose one of the following answers

- Yes
- No
- Not relevant

Please enter your comment here:

Question 5 concerns the data you will obtain from elsewhere and **reuse**.

5. Please provide the following information for the **existing data you will reuse**:

1. **Name of the dataset**
2. **Relevant task**
3. **Size**
4. **Format**: for example, .docx; .xlsx; .pdf; .mp4; .xml; .csv.
5. **Sensitive data**: Yes/no. If yes, please specify, for example, racial, political, ethical, health, and more here.
6. **Personal data**: Yes/no. If yes, please specify, for example, name, surname, address, email, IP address, location data.
7. **Metadata**: the metadata accompanying your datasets.
8. **Access**: open/restricted/closed access.

9. **Origin**: what is the origin of the data?

10. **Ownership**: who owns the data you will reuse?

11. **Licence**: under what licence can you use the data?

12. **Re-use**: potential documents or tools needed to re-use or validate the data.

☐ This question concerns the data you will obtain from elsewhere and **reuse**.

	Dataset 1	**Dataset 2**	**Dataset 3**	**Dataset 4**	**Dataset 5**
Name of the dataset					
Relevant task					
Size					
Format					
Sensitive data					
Personal data					
Metadata					
Access					
Origin					
Ownership					
Licence					
Re-use					

Questions 6-13 concern your **data management practices**.

6. Please provide a brief summary of your **institutional data management practices**, specifying:

1. **Data location**: where and how data will be stored, for example, institutional server or web hosting.

2. **Server location**: EU or non-EU; compliant or not-compliant with applicable data protection rules (for example, GDPR).

3. **Backup procedures**: type of backup procedures and their frequency.

4. **Protection**: how data security is ensured, for example, password or two-factor authentication.

5. **Responsible**: name the person from your team who will bear primary responsibility for data management and serve as a contact person if questions arise.

This question concerns your **institutional data management**.

Data location	
Server location	
Backup procedures	
Protection	
Responsible	

7. Do you follow a specific naming convention?

If yes, please specify.

This question concerns your **data management**.

Choose one of the following answers

- ○ Yes
- ○ No

Please enter your comment here:

8. Do you use any standard metadata vocabulary, standards or methodologies when creating your datasets?

If yes, please specify.

This question concerns your **data management**.

Choose one of the following answers

- Yes
- No

Please enter your comment here:

9. Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

If yes, please specify.

This question concerns your **data management**.

Choose one of the following answers

- Yes
- No

Please enter your comment here:

10. How will you licence your data?

If other, please provide a justification.

This question concerns your **data management**.

Comment only when you choose an answer.

- Creative Commons Attribution International Public License (CC BY) (or equivalent)
- Creative Commons Public Domain Dedication (CC 0) (or

equivalent)

☐ Other

11. Would you be interested in publishing your data in the form of data papers?

If yes, please give an example of a suitable dataset. If not, please explain why.

This question concerns your **data management**.

Choose one of the following answers

○ Yes

○ No

Please enter your comment here:

12. Do you have a preference for a trusted repository where to store your research data?

If yes, please specify.

This question concerns your **data management**.

Choose one of the following answers

○ Yes

○ No

Please enter your comment here:

13. Can you identify potential obstacles (e.g., technical, social, policies) that would prevent delivering FAIR data during ForestPaths' lifetime and beyond? Information on FAIR data here.

If yes, please specify.

This question concerns your **data management**.

Choose one of the following answers

Yes

No

Please enter your comment here:

Any additional comments?

moveprev

movesubmit