



## D1.2 Data Management Plan (DMP)

26/01/2026

Andree Cappellari<sup>1</sup>, Andrea Battisti<sup>1</sup>

<sup>1</sup>University of Padova, Department of Agronomy, Food, Natural resources, Animals and Environment (Italy)



Funded by  
the European Union

**Prepared under contract from the European Research Executive Agency**

Grant agreement No. 101134200

EU Horizon Europe Research and Innovation Action

Project acronym:	<b>FORSAID</b>
Project full title:	<b>Forest surveillance with artificial intelligence and digital technologies</b>
Project duration:	01/09/2024 – 29/02/2028 (42 months)
Project coordinator:	Andrea Battisti, University of Padua (UNIPD)
Call:	HORIZON-CL6-2023-GOVERNANCE-01-16
Deliverable title:	Data Management Plan
Deliverable n°:	1.2
WP responsible:	WP1
Nature of the deliverable:	Document
Dissemination level:	Public
Lead partner:	UNIPD
Recommended citation:	Cappellari, A. & Battisti, A. (2026). <b>Data Management Plan</b> . FORSAID project deliverable D1.2.
Due date of deliverable:	Month 7
Actual submission date:	Month 7, updated at Month 17

Deliverable status:

Version	Status	Date	Author(s)	Reviewer(s)
1.0	First draft	14 January 2025	Andree Cappellari, Andrea Battisti (UNIPD)	Neda Modova, Peter Bozakov (PENSOFT), Manuela Branco (ISA)
2.0	Final (first uploaded version)	28 February 2025	Andree Cappellari, Andrea Battisti (UNIPD)	
3.0	Revised version	15 January 2026	Andree Cappellari, Andrea Battisti (UNIPD)	Maarten de Groot (SFI), Christelle Robinet (INRAE)
4.0	Final (second uploaded version)	26 January 2026	Andree Cappellari, Andrea Battisti (UNIPD)	

## Table of contents

Key takeaway messages .....	4
Summary .....	4
List of abbreviations .....	4
1 Data summary .....	5
1.1 Datasets generated.....	6
2 FAIR data.....	9
2.1 Making data findable, including provisions for metadata .....	9
2.1.1 Metadata structure .....	9
2.2 Making data accessible.....	10
2.2.1 Day-to-day data storing and sharing .....	11
2.2.2 DEFID2.....	11
2.3 Making data interoperable.....	11
2.4 Increase data reuse .....	13
2.5 Roles and responsibilities .....	13
2.5.1 FORSAID Zenodo Community responsibilities.....	14
3 Other research outputs .....	14
4 Allocation of resources.....	15
5 Data security .....	15
6 Ethics.....	16
7 Annex .....	17
7.1 Annex 1: Data and manuscript publication in FORSAID.....	17
7.1.1 Data publication .....	17
7.1.2 Manuscript publication .....	19

## Key takeaway messages

- The FORSAID project will generate a diverse range of data, encompassing various data types and formats.
- The Zenodo FORSAID Community is available at <https://zenodo.org/communities/forsaid/>.
- All final data and their metadata will be stored in the dedicated Zenodo Community and made available to everyone after the publication of the corresponding scientific paper.
- Final data that are not associated with a scientific publication will be stored in Zenodo and/or in institutional repositories, such as the one managed by the University of Padova.
- Raw data that are still being processed will be securely stored in institutional repositories, ensuring data protection and backup.

## Summary

Promoting Open Science is a central requirement for Horizon Europe projects, ensuring transparency, reproducibility, and long-term usability of scientific outputs. The FORSAID Data Management Plan outlines the overall framework of how research data will be managed within the project, with particular emphasis on the FAIR principles, *i.e.*, aiming to make research data Findable, Accessible, Interoperable, and Reusable. The project will produce a wide range of research data, including biological data, citizen science data, economic data, geospatial data, social science data, as well as other research outputs, such as scientific publications. All research data associated with a scientific publication will be stored and made available through the FORSAID Community on Zenodo, a widely used open-access repository. In addition, specific datasets produced within Work Packages 2 and 3 will be integrated into the Database of European Forest Insect and Disease Disturbances (DEFID2), coordinated and curated by the Joint Research Centre. Final research data that are not associated with a scientific publication will be stored in the FORSAID Community on Zenodo and/or in institutional repositories, such as the repository managed by the University of Padova. Before the publication of scientific papers, associated raw data that are still being processed will be securely stored in institutional repositories, ensuring data protection. Team Leaders of each participating institution will be responsible for overseeing data storage, maintaining backups, and ensuring compliance with data management protocols. The Data Management Plan will be updated as the project progresses to accommodate new findings and requirements.

## List of abbreviations

DEFID2: Database of European Forest Insect and Disease Disturbances  
DMP: Data Management Plan  
DOI: Digital Object Identifier  
EU: European Union  
FAIR: Findable, Accessible, Interoperable, Reusable  
GDPR: General Data Protection Regulation  
JRC: Joint Research Centre  
PI: Principal Investigator  
WP: Work Package

## 1 Data summary

The FORSAID project focuses on addressing the objectives of the European Green Deal, with particular emphasis on forest management. Forest trees are increasingly threatened by invasive pests, many of which are regulated within the EU territory. In this context, the overarching goal of FORSAID is to develop a comprehensive set of innovative digital technologies aimed at early detection of regulated forest pests, monitoring their occurrence across the territory, and providing critical information to support the adoption of phytosanitary measures to mitigate their spread and impact. The project involves seventeen partners from ten countries: Bulgaria, Denmark, France, Germany, Italy, Portugal, Slovenia, Sweden, Switzerland, and Ukraine (Table 1).

The present Data Management Plan (DMP) outlines the overall framework of how research data will be managed within the FORSAID project. The initial version of this document was developed at Month 6 of the project (February 2025). Following feedback from the external reviewers after the first periodic reporting, the DMP was updated in January 2026. The document may continue to be revised throughout the project lifecycle to accommodate new findings and requirements.

**Table 1: The FORSAID Consortium.**

Institution (short)	Institution	Role	Team Leader	Country
UNIPD	University of Padova	Coordinator	Andrea Battisti	Italy
CNR	National Research Council	Beneficiary	Alberto Santini	Italy
EFOS	Environment and food safety solutions and services	Beneficiary	Boštjan Božič	Slovenia
EPPO	European and Mediterranean Plant Protection Organization	Beneficiary	Olga Tikka	France
IEFC	European Institute of Planted Forest	Beneficiary	Christophe Orazio	France
INIAV	National Institute of Agrarian and Veterinary Research	Beneficiary	Helena Bragança	Portugal
INRAE	National Research Institute for Agriculture, Food and the Environment	Beneficiary	Hervé Jactel	France
ISA	University of Lisbon	Beneficiary	Manuela Branco	Portugal
KIT	Karlsruhe Institute of Technology	Beneficiary	Christian Pylatiuk	Germany
LNU	Linnaeus University Växjö	Beneficiary	Johanna Witzell	Sweden
MfN	Natural History Museum Berlin	Beneficiary	Rudolf Meier	Germany
PENSOFT	Pensoft Publishers	Beneficiary	Neda Modova	Bulgaria
SFI	Slovenian Forestry Institute	Beneficiary	Maarten De Groot	Slovenia
TPZF	Telespazio France SAS	Beneficiary	Jean-Charles Samalens	France
UCPH	University of Copenhagen	Beneficiary	Rasmus Fensholt	Denmark
UNFU	Ukrainian National Forestry University	Beneficiary	Iryna Matsiakh	Ukraine
WSL	Swiss Federal Institute for Forest, Snow, and Landscape	Associated Partner	Eckehard Brockerhoff	Switzerland

## 1.1 Datasets generated

The FORSAID project is expected to produce and use a significant amount of research data. A comprehensive overview of the final research data for each Work Package (WP) is provided in Table 2. These data are primarily associated with the corresponding Research Actions, which define the scientific and technical activities of the project. However, in some cases, they are generated in relation to specific deliverables or milestones, or they integrate outputs from multiple actions. These include observational and experimental data collected in the field and laboratory, citizen science-driven data from questionnaires, economic data, and geographical data obtained through remote sensing. Consequently, different types of datasets, including documents, spatial datasets, generic datasets, images, and software, will be produced in various formats. While file sizes will vary, no single research dataset is expected to exceed 50 GB, with the total estimated dataset size below 100 GB.

Most of these data will be generated within FORSAID. However, these will be supplemented by data originating from public databases, such as the Database of European Forest Insect and Disease Disturbances (DEFID2), coordinated and curated by the Joint Research Centre (JRC).

All data will be made openly accessible as soon as possible, typically upon the publication of the corresponding scientific paper. To ensure transparency, all research outputs will be licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license. Data will be valuable not only to the scientific community, but also to a wide range of stakeholders, including policymakers, forest managers, environmental agencies, and the general public.

**Table 2: Final research data that FORSAID is expected to produce or reuse, including information on the type, format, size, availability, and reusability licence.**

WP	Research data	Type	Generated / Reused	Format	Size	Availability	Reusability licence
WP2	Maps of defoliated areas by <i>Thaumetopoea pityocampa</i>	Spatial dataset	Generated	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Maps of defoliated areas by <i>Thaumetopoea pityocampa</i>	Spatial dataset	Reused, DEFID2	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Maps of tree mortality by <i>Ips typographus</i>	Spatial dataset	Generated	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Maps of tree mortality by <i>Ips typographus</i>	Spatial dataset	Reused, DEFID2	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Maps of tree damage by <i>Corythucha arcuata</i>	Spatial dataset	Generated	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Maps of tree damage by <i>Bursaphelenchus xylophilus</i>	Spatial dataset	Generated	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	HOMED project aerial survey in 2020 and 2024	Spatial dataset	Reused	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Maps of tree damage by <i>Ceratocystis platani</i>	Spatial dataset	Generated	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Database of damage by <i>Agrilus planipennis</i> using remote sensing spectroscopy	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP2	Database of damage by <i>Fusarium circinatum</i> using mid-infrared spectroscopy	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP2	Database of damage by <i>Cryphonectria parasitica</i> and <i>Bursaphelenchus xylophilus</i> using mid-infrared spectroscopy	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP2	Database of damage by <i>Diplodia sapinea</i> using mid-infrared spectroscopy	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP2	Database of plant damage by <i>Ceratocystis platani</i> using near-infrared and RGB images	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP3	Database of data collected by automated traps for <i>Ips typographus</i>	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP3	Database of data collected by automated traps for <i>Thaumetopoea</i> spp.	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP3	Images for AI identification of <i>Agrilus</i> spp.	Images	Generated	.tiff	> 10 GB	Open	CC BY-NC 4.0

WP3	Images for AI identification of <i>Monochamus</i> spp.	Images	Generated	.tiff	> 10 GB	Open	CC BY-NC 4.0
WP3	Images for AI identification of longhorn beetles associated with broadleaf trees	Images	Generated	.tiff	> 10 GB	Open	CC BY-NC 4.0
WP3	Images for AI identification of exotic and native bark and ambrosia beetles	Images	Generated	.tiff	> 10 GB	Open	CC BY-NC 4.0
WP3	Images for AI identification of beetles captured in generic surveillance traps near entry points	Images	Generated	.tiff	> 10 GB	Open	CC BY-NC 4.0
WP3	Protocol for the detection of the pine wood nematode, <i>Bursaphelenchus xylophilus</i> , in eDNA	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP3	Protocol for the detection of <i>Agrilus planipennis</i> in eDNA	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP3	Protocol for the detection of <i>Ceratocystis platani</i> in eDNA	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP3	Protocol for the detection of <i>Cryphonectria parasitica</i> in eDNA	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP4	Report on the accuracy of citizen science platforms in identifying quarantine pests	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP4	Report on patterns in spatial, temporal, and taxonomic biases in citizen science from citizen science platforms	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP4	Report on the reliability of citizen science data by comparing them with interception and establishment data	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP4	Report on the accuracy of phone apps and associated AI algorithms for pest detection	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP4	Report on the characterization of citizen scientists	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP4	Report on workshop with stakeholders	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP4	App for the identification of forest quarantine species	Software	Generated	software	< 1 GB	Open	CC BY-NC 4.0
WP5	Results of the questionnaire on needs, expectations, and priorities of European stakeholders regarding effective regulated forest pest surveillance	Dataset	Generated	.csv	50 M	Open	CC BY-NC 4.0
WP5	Database of cost-effectiveness of measures to contain and eradicate forest quarantine pests	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP5	Protocol for multicriteria analysis	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0

## 2 FAIR data

In order to implement FAIR (Findable, Accessible, Interoperable, Reusable) data management principles, all research data produced within the FORSAID project will be stored in secure and trusted repositories throughout the project lifecycle, both prior to and following publication. Published datasets will be made accessible through appropriate repositories, while unpublished or intermediate data will be securely stored in institutional repositories, in accordance with institutional policies and applicable legal and ethical requirements. The storage and management of unpublished data will be supported at the institutional level, with Team Leaders responsible for overseeing the proper handling and protection of data generated by their respective organizations. By upholding FAIR data principles, FORSAID supports forest health research, early pest detection, and policy decisions.

### 2.1 Making data findable, including provisions for metadata

All research data generated within FORSAID and associated with a scientific publication will be available in the Zenodo database, which guarantees long-term preservation and accessibility of the data and promotes data reuse. In this way, all participants will be able to easily share their data in any format, with file sizes of up to 50 GB. A dedicated Zenodo FORSAID Community, where all research data of the project will be uploaded, is available at <https://zenodo.org/communities/forsaid/>. A detailed description of the procedure to upload research data to Zenodo is available in the *Data publication* section of Annex 1.

The following actions will ensure that data and metadata produced within the FORSAID project adhere to the principles of findability:

- (Meta)Data are assigned a globally unique and persistent identifier (DOI, Digital Object Identifier).
- Data are described with rich metadata.
- Metadata clearly and explicitly include the DOI of the data it describes, as it is a top-level and mandatory field in the metadata of each record.
- (Meta)Data are registered or indexed in a searchable resource: metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing and is sent to DataCite servers during DOI registration and indexed there.

Research data not associated with a scientific publication will be stored in Zenodo and/or in institutional repositories. The repository provided by the University of Padova via Google Drive is available for that purpose (see paragraph 2.2.1 – Day-to-day data storing and sharing). For each dataset, a ReadMe file in .csv or .txt format will accompany the data, documenting all relevant metadata, processing steps, and the location of associated raw data within the partners' internal repository.

Raw data that are still being processed will be securely stored in institutional repositories, ensuring data protection.

#### 2.1.1 Metadata structure

A minimum set of metadata is mandatory for dataset registration. The Zenodo metadata scheme complies with DataCite Metadata Schema mandatory and recommended terms, with further enrichment fields available to meet specific field needs. These terms include:

- Resource type, e.g., dataset, image, model, publication, software.

- DOI; if a DOI is not yet available, Zenodo can generate one.
- Title and subtitle. Subtitles should include the name of the project, the number of the associated Research Action, and a short sentence describing the data, e.g., FORSAID\_RA3.2.1\_AI identification of *Agrilus*.
- Publication date in ISO 8601 format (YYYY-MM-DD).
- Creators, *i.e.*, the persons or organisations that have created the resource being uploaded (e.g., the authors in case of papers) and are listed in the academic citation.
- Licences and rights: the default license is Creative Commons Attribution 4.0 International (CC-BY), which allows re-distribution and reuse of a licensed work on the condition that the creator is appropriately credited.
- Descriptions (recommended by Zenodo, mandatory in FORSAID), including detailed information on data provenance, e.g., abstract, notes, and information on methods.
- Contributors (recommended), *i.e.*, persons or organisations that have contributed to the record, such as supervisors, contact persons, and sponsors.

FORSAID also requires the inclusion of the following mandatory information for all uploaded research data:

- The terms “European Union (EU)” and “Horizon Europe”.
- Action name, acronym, and grant agreement number: Forest surveillance with artificial intelligence and digital technologies – FORSAID, grant agreement 101134200.

## 2.2 Making data accessible

Access to FORSAID data and metadata after scientific paper publication (when relevant) will be open, with limited exceptions. At this stage, no significant embargo periods for making research data publicly accessible are anticipated. However, should specific datasets emerge during the project that are deemed to have potential for commercial exploitation or the development of marketable products, major embargo periods may be applied. Such measures would ensure adequate protection of Intellectual Property Rights and compliance with applicable regulations while balancing the principles of open access with the need to safeguard opportunities for innovation and commercialization. In addition, PhD theses produced during the project and made public by the Universities could undergo an embargo period of one-year maximum to ensure the data will be published.

FORSAID data accessibility policies will leverage the accessibility features provided by the Zenodo repositories. The following actions will ensure that data and metadata produced within the FORSAID project adhere to the principles of accessibility:

- (Meta)Data are retrievable by their identifier using a standardized communications protocol: metadata for individual records as well as record collections are harvestable using the OAI-PMH protocol by the record identifier and the collection name. Metadata is also retrievable through the public REST API.
- The protocol is open, free, and universally implementable.
- The protocol allows for an authentication and authorization procedure, where necessary: metadata are publicly accessible and licensed under the public domain. No authorization is ever necessary to retrieve it.
- Metadata are accessible, even when the data are no longer available: data and metadata will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least. In addition, metadata are stored in high-availability database servers at CERN, which are separate from the data itself.

Zenodo allows users to sign up via the creation of local accounts. Authentication is allowed via local username and password, ORCID, or GitHub credentials.

### 2.2.1 Day-to-day data storing and sharing

A secure, institutionally managed repository provided by the University of Padova via Google Drive is available to all partners for data storage and exchange. The existing contractual agreement with Google (Google Workspace for Education, accessible [here](#)) ensures that Google has no access to the uploaded data and that all information is stored in a protected environment. All partners may access this repository for the duration of the project and for at least two years following its completion through the restricted access section of the FORSAID website. Regular WP meetings and cross-WP exchanges, during which partners report on ongoing data generation and processing status, ensure project-wide awareness of data activities.

### 2.2.2 DEFID2

DEFID2 is a joint and voluntary effort among scientists to share and harmonize their geospatial observations of insect and disease disturbance. Data collected in WP2 and WP3 should be sent to [JRC-DEFID2@ec.europa.eu](mailto:JRC-DEFID2@ec.europa.eu) for upload to the DEFID2 database. The [DEFID2 protocol for data contribution](#) describes in detail the attributes needed for each record, which include:

- Information about the contributor and the data source.
- Key information about the disturbance.
- Complementary information mostly related to occurrences characterized by multiple agents or multiple hosts, climate-driven triggering factors, and silvicultural practices.
- Qualitative assessment of the damage.

In addition to contributing newly generated datasets, the project will make use of relevant data already available within the DEFID2 database to support specific Research Actions.

## 2.3 Making data interoperable

FORSAID data can be exchanged and reused within the project period among all the institutions involved in the project. The following actions will ensure that data and metadata produced within the FORSAID project adhere to the principles of interoperability:

- (Meta)Data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- (Meta)Data use vocabularies that follow FAIR principles.
- (Meta)Data include qualified references to other (meta)data.

Formats of research data will be universal, cross-platform, and open source, with open standards (Table 3). All details on formats of the expected data outcome are provided in Table 2.

**Table 3: Recommendations for file formats based on file type.**

File type	Format name	File extension(s)
Text	Microsoft PowerPoint XML	.pptx
	Microsoft Word XML	.docx
	OpenDocument Presentation	.fodp, .odp
	OpenDocument Text	.fodt, .odt
	Plain Text	.txt, .asc
Documentation	Adobe PDF/A	.pdf
	Microsoft Word XML	.docx
	OpenDocument Text	.fodt, .odt
Markup	CSS	.css
	HTML	.htm, .html
	SGML	.sgm, .sgml
	XML	.xml
Tabular	Comma-separated values (CSV)	.csv
	Microsoft Excel XML	.xlsx
	OpenDocument Spreadsheet	.fods, .ods
	Tab-separated values	.tsv, .tab
Image/graphics	GIF	.gif
	JPEG 2000	.jpxml, .jp3d, .jpf, .jpm, .jpx, .jp2
	PNG	.png
	Scalable Vector Graphics (SVG)	.svg
	TIFF	.tiff, .tif
Audio	WAV	.wav
	Motion JPEG2000	.mjp2, .mj2
Video	MPEG-4	.m4v, .m4r, .m4b, .m4p, .m4a, .mp4
Database	Minitab syntax and output	.lis, .tj
	R	.rds
	SAS syntax	.sas
	SPSS syntax	.sps
	Stata syntax	.do, .dct
	Structured Query Language	.sql
	CAD	.dwg
Geospatial	ESRI Shapefile	.shp, .shx, .dbf, .prj, .sbx, .sbn
	Geopackage	.gpkg
	Geo-referenced TIFF	.tif, .tfw

## 2.4 Increase data reuse

All datasets produced within FORSAID have the potential for reuse. The following actions will ensure that data and metadata produced within the FORSAID project adhere to the principles of reusability:

- (Meta)Data are richly described with a plurality of accurate and relevant attributes.
- (Meta)Data are released with a clear and accessible data usage license (Table 2).
- (Meta)Data are associated with detailed provenance.
- (Meta)Data meet domain-relevant community standards.

Metadata of deposited data will be made available under a Creative Commons Public Domain Dedication (CC 0) or equivalent.

For data structured in tables, dedicated metadata documentation must be provided to facilitate correct interpretation and reuse. For each table, a separate metadata file (in .csv or .txt format) will be provided, describing the structure and content of the dataset. This documentation will include clear definitions of all variables and columns, units of measurement, data types, and any relevant assumptions or data processing steps. Where applicable, references to methodologies, data collection protocols, and data quality considerations will also be included. This approach ensures transparency, supports reproducibility, and enhances the reusability of FORSAID datasets in line with the FAIR principles.

For images produced within Task 3.2 – Robot sorting coupled with advanced image analysis, a structured data format will be used, including the following information for each picture:

- Species name, *e.g.*, *Ips sexdentatus*.
- File name, which includes the species name, the institution abbreviation, the number of the picture, and the viewpoint, *e.g.*, *Ips\_sexdentatus\_UNIPD\_001\_dorsal*.
- Viewpoint, *i.e.*, dorsal, ventral, right lateral, left lateral.
- Stacking methodology, *e.g.*, B; 8; 4.
- Lens used, *e.g.*, TC5M-03-110/SL240409004B.

All metadata of geodata publicly shared in FORSAID will follow the INSPIRE requirements (<https://inspire-geoportal.ec.europa.eu>) and the open data strategies of the national countries.

For the remaining research data, there is no predefined structured metadata format at this stage of the project. The project will adopt adaptable metadata standards as needed, ensuring compatibility with emerging requirements and best practices.

## 2.5 Roles and responsibilities

The following responsibilities are assigned within the FORSAID project to support the implementation of the DMP:

- WP1 is responsible for overseeing the data management life cycle for all datasets to be collected, processed, or generated by the project.
- WP leaders are considered responsible for adhering to the specifications above in their respective WPs.
- The Team Leaders of each beneficiary organization (Table 1) are considered responsible (curators) for overseeing data storage, maintaining backups, and ensuring compliance with data management protocols. They will support WP1 in all issues related to research

data management. The Team Leader of each beneficiary should ensure that personnel working on the FORSAID project have read the DMP.

- Data collectors have the ultimate responsibility of complying with the specifics of the DMP, as well as with the related General Data Protection Regulation (GDPR) policies.

Additional details about result exploitation and Intellectual Property Rights are outlined in the FORSAID Consortium Agreement and Grant Agreement.

### 2.5.1 FORSAID Zenodo Community responsibilities

The FORSAID Community in Zenodo is available at <https://zenodo.org/communities/forsaid/>. Community roles are defined as follows:

- Community owner: Andrea Battisti (UNIPD), FORSAID project coordinator.
- Community manager: Andree Cappellari (UNIPD), FORSAID project manager.
- Community curators: 1 Team Leader for each beneficiary, appointed by procedures detailed in the Consortium Agreement.
- Community readers: all the FORSAID researchers belonging to the project beneficiary institutions.

Zenodo FORSAID Community data management policies are defined as follows:

- Data owners, via their institutional Zenodo profiles, trigger dataset upload requests.
- Community curators for the specific institution review the request and accept/reject the dataset submission.
- Data owners set the visibility level for the specific dataset.

## 3 Other research outputs

The project comprises 19 milestones and 25 deliverables across five WPs, including documents, reports, data, and other outputs, most of which are publicly accessible. Key deliverables and milestones contributing to the research community will be considered for further publication and DOI assignment.

FORSAID promotes adherence to open science best practices, ensuring that research outputs are openly available as early as possible. All scientific publications arising from FORSAID will be peer-reviewed and made available in open access, fulfilling the requirements for the Gold Open Access standard, whenever possible, or at least for the Green Open Access standard. In addition, all publications must be deposited in the FORSAID Community on Zenodo under a CC-BY 4.0 licence. Project members are also encouraged to share preprints on recognized platforms to facilitate early knowledge dissemination and feedback, especially for urgent discoveries, such as the detection of a new quarantine pest.

The project also emphasizes citizen science and public engagement, encouraging collaboration with the public, stakeholders, and policymakers by making research findings accessible through open reports, interactive platforms, and outreach activities. By implementing these practices, FORSAID enhances transparency and fosters interdisciplinary collaboration. Additional information on other research outputs is available in Annex 5 of the Grant Agreement.

## 4 Allocation of resources

The overall costs for data management are expected to be limited to Personnel Costs and are detailed as follows:

- 3 person-months are allocated to the Zenodo community manager.
- 1 person-month is allocated to each Zenodo community curator (Team Leaders).

At the time of writing this document, no costs for goods or services related to data management are foreseen.

## 5 Data security

Raw data that are still being processed will be securely stored in institutional repositories, ensuring data protection. The repository provided by the University of Padova is available for that purpose.

For the data published on Zenodo after scientific paper publication, the latter ensures robust security through various measures:

- **Physical security:** the data centres are located on CERN premises, and physical access is restricted to a limited number of people with appropriate training. For example, Zenodo staff do not have physical access to the CERN Data Centre.
- **Server and infrastructure security:** Zenodo servers are managed according to the CERN Security Baseline for Servers, meaning for example that the operating system and installed applications are kept updated with the latest security patches via an automatic configuration management system.
- **Network security and intrusion detection:** the CERN Security Team runs both host and network-based intrusion detection systems and monitors the traffic flow, pattern, and contents into and out of CERN networks to detect attacks. All access to zenodo.org happens over HTTPS, except for static documentation pages, which are hosted on GitHub Pages.
- **Password hashing and token encryption:** Zenodo stores user passwords using strong cryptographic password hashing algorithms (currently PBKDF2+SHA512). User access tokens to GitHub and ORCID are stored encrypted and can only be decrypted with the application's secret key.
- **Application and session security:** Zenodo employs techniques to protect user sessions from being stolen by attackers during login and runs vulnerability scans against the application.
- **Data access control and operational policies:** CERN staff with access to user data operate under CERN Operational Circular no. 5, meaning among other things that: *a)* staff should not exchange among themselves information acquired unless it is expressly required for the execution of their duties; *b)* access to user data must always be consistent with the professional duties and only permitted for resolution of problems, detection of security issues, monitoring of resources and similar; and *c)* staff are liable for damage resulting from any infringement and can have access withdrawn and be subject to disciplinary or legal proceedings depending on the seriousness of the infringement.

## 6 Ethics

FORSAID will ensure that all the data sharing activities comply with ethical principles and relevant national, international, and EU legislation. This will be achieved in collaboration with the ethics advisor, Maarten de Groot (SFI), who will provide guidance and address potential ethics issues, including the use of Artificial Intelligence.

WP4 and WP5 will involve human participants by focusing on citizen science and stakeholder engagement activities. The project will ensure the strict application of ethical requirements arising from surveys, interviews, and the collection of other data with privacy implications. All data that can encompass any personal data protection or privacy issues will not be publicly disclosed, and sensitive data will undergo a mandatory anonymization process:

- Only data needed to address the scientific purpose will be collected, *e.g.*, whenever possible, full personal names and other identifiable information will not be collected.
- Subject direct identifiers will be removed by eliminating or obscuring the specific part in the survey's hard/digital copies, *e.g.*, names, phone numbers, and email addresses.
- Anonymization will be performed during transcription or initial write-up, *e.g.*, names, surnames, and addresses will not be transferred while transcribing data from survey hard copies to electronic spreadsheets.

During interviews, consent forms and information sheets will be provided, which will include information about the project, how the data will be used and reused, and where and how data will be stored. All information will be stored in a confidential manner and in accordance with the EU Directive 95/46/EC and the EU GDPR regarding the use of personal data. Data like zip code and stakeholder or citizen category may be required, but it will be made sure that no answers can be traced back to a single participant. If the survey is performed online, it will be made sure that the encrypted connection via HTTPS is applied.

For all personal data, such as transcripts of interviews, the Team Leader and the Data Protection Officers of the corresponding beneficiary organizations will provide a declaration confirming compliance with Chapter 5 of the GDPR, as well as a declaration confirming compliance with the laws of the country where the data was collected.

Rules on research integrity are detailed in the *Manuscript publication* section of Annex 1.

## 7 Annex

### 7.1 Annex 1: Data and manuscript publication in FORSAID

This document was drafted in March 2025 and updated in January 2026.

#### 7.1.1 Data publication

All research data generated within the FORSAID project must adhere to the FAIR principles (Findability, Accessibility, Interoperability, and Reusability). For this reason, all research data (including manuscripts) must be published in an open repository. We have chosen Zenodo due to its ease of use and ability to accommodate a wide range of data types. Zenodo allows up to 100 files per upload, with a total storage limit of 50 GB per upload.

In addition, specific datasets produced within WPs 2 and 3 will be integrated into the Database of European Forest Insect and Disease Disturbances (DEFID2), coordinated and curated by the Joint Research Centre.

Raw data that are still being processed will be securely stored in institutional repositories, ensuring data protection. A secure, institutionally managed repository provided by the University of Padova via Google Drive is available to all partners for data storage and exchange. The existing contractual agreement with Google (Google Workspace for Education, accessible [here](#)) ensures that Google has no access to the uploaded data and that all information is stored in a protected environment. All partners may access this repository for the duration of the project and for two years following its completion through the restricted access section of the FORSAID website.

#### Publishing data on Zenodo

Publishing data on Zenodo is a quick and straightforward process. The Pensoft team has created a [short instructional video](#) outlining the steps. Although created for a different project, the steps remain largely the same.

In Zenodo, each publication record must include at least the mandatory DataCite terms, with additional recommended terms where applicable. These terms include:

- **Basic information (mandatory):**
  - DOI: if a DOI is not yet available, Zenodo can generate one.
  - Resource type: dataset, image, publication, etc.
  - Title and subtitle: to improve the discoverability of research data, we suggest using a standardized naming convention for subtitles, including the name of the project, the number of the associated Research Action, and a short sentence describing the data, e.g., FORSAID\_RA3.2.1\_AI identification of *Agrilus*.
  - Publication date: in ISO 8601 format (YYYY-MM-DD).
  - Creators: persons or organisations that have created the resource being uploaded (e.g., the authors in case of papers) and are listed in the academic citation.
  - Description: while recommended by Zenodo, this is mandatory for FORSAID and should include detailed information on data provenance.
  - License: the default license is Creative Commons Attribution 4.0 International, which allows redistribution and reuse of a licensed work on the condition that the creator is appropriately credited.
- **Recommended information (not mandatory):**

- Contributors: persons or organisations that have contributed to the record, e.g., supervisors, contact persons, and sponsors, and are not included in the academic citation.
- Keywords and subjects.
- Languages.
- Dates for acceptance, collection, creation, etc.
- Version: mostly relevant for software and dataset uploads
- Publisher: this is used to formulate the citation, so consider the prominence of the role. The default is Zenodo.
- **Funding (mandatory):** search for FORSAID, or Grant No. 101134200.
- **Additional information (not mandatory):**
  - Alternative identifiers.
  - Related works.
  - References.
  - For software: repository URL, programming language, and development status.
  - For manuscripts: journal or book title, ISSN or ISBN, volume, issue, page range or article number, etc.
  - For conferences: title, acronym, location, dates, website, and session.
  - Additional specific fields.
- **Visibility (mandatory):** we suggest specifying “Restricted” before manuscript publication and changing it to “Public” after manuscript publication (when relevant).
- **Community (mandatory):** all research data must be uploaded to the [FORSAID Community](#) on Zenodo. When uploading, search for the FORSAID community at the top of the submission page. If a dataset was uploaded without specifying the community, follow these steps: go to the record you want to submit to the community and click the cog-wheel icon to open the Communities menu, click Submit to community in the dropdown menu, then find the FORSAID community and click the Select button. For additional information, please visit [this web page](#).

### FORSAID Zenodo Community

The roles of the FORSAID Community in Zenodo are defined as follows:

- Community owner: Andrea Battisti (University of Padova), FORSAID Project Coordinator.
- Community manager: Andree Cappellari (University of Padova), FORSAID Project Manager.
- Community curators: the Team Leader of each beneficiary, appointed by procedures detailed in the FORSAID Consortium Agreement.
- Community readers: all FORSAID researchers.

### Data management and responsibilities

To ensure compliance with data management decisions, the following measures apply in FORSAID:

- The Project Coordinator and the Project Manager will make sure that the data will be correctly uploaded.
- The Team Leaders of each beneficiary organization are considered responsible (curators) for overseeing data storage, maintaining backups, and ensuring compliance with data management protocols. They will support WP1 in all issues related to research data management. The Team Leader of each beneficiary should ensure that personnel working

on the FORSAID project have read the FORSAID Data Management Plan and the present document.

- Data collectors have the ultimate responsibility of complying with the specifics of the FORSAID Data Management Plan, as well as with the related General Data Protection Regulation policies.

### **The database of European Forest Insect and Disease Disturbances - DEFID2**

DEFID2 is a joint and voluntary effort among scientists to share and harmonize their geospatial observations of insect and disease disturbance. Specific data collected in WP2 and WP3 should be sent to [JRC-DEFID2@ec.europa.eu](mailto:JRC-DEFID2@ec.europa.eu) for upload to the DEFID2 database. The [DEFID2 protocol for data contribution](#) describes in detail the attributes needed for each record, which include:

- Information about the contributor and the data source.
- Key information about the disturbance.
- Complementary information mostly related to occurrences characterized by multiple agents or multiple hosts, climate-driven triggering factors, and silvicultural practices.
- Qualitative assessment of the damage.

### **7.1.2 Manuscript publication**

The FORSAID project will produce multiple publications within and across its WPs. The success of the project strongly relies on the effective dissemination of research findings through publication in appropriate scientific journals. For this reason, preprint articles are encouraged, especially for urgent discoveries, such as the detection of a new quarantine pest.

FORSAID will make use of the Pensoft Open Access journal Neobiota, with the publication of a special issue toward the end of the project, and of the highly innovative Pensoft journal Research Ideas and Outcomes (RIO), where a special Open Science Collection for FORSAID will be published.

To ensure clarity and consistency regarding authorship and publication practices, the FORSAID project established the following publication rules.

#### **Publication rules**

1. Researchers are required to share the outcomes of their studies with the scientific community, avoiding unnecessary delays as well as artificial partitions of the research results for the sole purpose of obtaining a higher number of scientific products.
2. Researchers planning a publication should circulate a tentative title, list of authors, and additional preliminary information via the [Manuscript Proposal form](#) in the Publication and Outreach section of the internal communication platform of the FORSAID website as early as possible, and no later than one month before manuscript submission. This ensures proper acknowledgement, enhances awareness of ongoing activities, and allows improvements in synthesis work at an early stage.
3. All and only the researchers who have actively contributed to creating the research products should be included as authors of scientific publications. The planning of an experiment, the planning and execution of data collection, the analysis of the data, and the writing of a manuscript are all steps that can merit authorship. First authorship must be assigned to researchers who

made the most significant contributions to the results being published. PhD students and postdoctoral researchers are warmly invited to take the lead in scientific publications.

4. In collaborative papers, proper credit must be given to data owners. Options for acknowledgement include co-authorship, if the data contribute important information to the publication, or a mention in the acknowledgments section if the data are less relevant.

5. Researchers have the responsibility to ensure that the use of generative AI is consistent with the principles of scientific integrity.

6. Researchers must include the following statement in all scientific publications that involve FORSAID contribution: "This project has received funding from the European Union's Horizon Europe Research and Innovation programme under grant agreement No. 101134200 "FORSAID: Forest surveillance with artificial intelligence and digital technologies".

7. In choosing the publication venue, the researcher must assess its reputation in the scientific communities of reference, favouring, where possible, locations that can promote the widest circulation of ideas and research results. The researcher must make every effort to identify and avoid predatory publishing.

8. All scientific publications arising from FORSAID will be peer-reviewed and made available in open access, fulfilling the requirements for the Gold Open Access standard, whenever possible, or at least for the Green Open Access standard.

9. All scientific publications must also be stored in the Zenodo repository under a CC-BY 4.0 licence, as for datasets.