D1.2 Data Management Plan (DMP)

28/02/2025

Andree Cappellari[1], Andrea Battisti[1]

[1]University of Padova, Department of Agronomy, Food, Natural resources, Animals and Environment (Italy)

Funded by the European Union

FORSAID receives funding from the European Union's Horizon Europe Research and Innovation Programme under grant agreement 101134200. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the EU nor the REA can be held responsible for them.

Prepared under contract from the European Research Executive Agency
Grant agreement No. 101134200
EU Horizon Europe Research and Innovation Action

Project acronym:	**FORSAID**
Project full title:	**Forest surveillance with artificial intelligence and digital technologies**
Project duration:	01/09/2024 – 28/02/2028 (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., & M., & Battisti, A. (2025). **Data Management Plan**. FORSAID project deliverable D1.2.
Due date of deliverable:	Month n° 7
Actual submission date:	Month n° 7

Deliverable status:

Version	Status	Date	Author(s)	Reviewer(s)
1.0	Draft	14 January 2025	Andree Cappellari, Andrea Battisti (UNIPD)	Neda Modova, Peter Bozakov (PENSOFT)
2.0	Draft	28 February 2025	Andree Cappellari, Andrea Battisti (UNIPD)	Manuela Branco (ISA), Neda Modova (PENSOFT)

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 10
2.1.2 DEFID2 10
2.2 Making data accessible 11
2.2.1 Zenodo sign-up and sign-in 11
2.2.2 Data and metadata accessibility 11
2.2.3 Data management responsibilities 12
2.2.4 FORSAID Community on Zenodo 12
2.3 Making data interoperable 12
2.4 Increase data re-use 14
3 Other research outputs 14
4 Allocation of resources 14
5 Data security 14
6 Ethics 15

Key takeaway messages

- The FORSAID project will generate a diverse range of research assets, encompassing various data typologies and formats.
- All project research assets and their metadata will be stored in a dedicated Zenodo Community, after the publication of the corresponding scientific paper.
- Raw data that are still being processed, *i.e.*, not yet associated with a published article, will be securely stored in institutional repositories, ensuring data protection and backup.

Summary

Promoting Open Science is a central requirement for Horizon Europe projects, promoting 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 - ensuring research data are Findable, Accessible, Interoperable, and Reusable. The project will produce a wide range of research assets, including biological data, citizen science data, economic data, geospatial data, and social science data. All research assets 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. Before the publication of scientific papers, associated raw data that are still being processed will be securely stored in institutional repositories, such as the repository provided by the University of Padova, ensuring data protection. All scientific publications produced within the project will be peer-reviewed and made available in open access, as well as uploaded to the FORSAID Community on Zenodo. 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 at month 6 of the project (February 2025). The DMP could be updated 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 Lavrentjeva	France
IEFC	European Institute of Planted Forest	Beneficiary	Christophe Orazio	France
INIAV	National Institute for Agrarian and Veterinarian 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 number of research assets. A comprehensive overview of the final research assets for each Work Package (WP) is provided in Table 2. These research assets include a vast range of datasets, including 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 typologies of research assets, including documents, spatial datasets, generic datasets, images, and software will be produced in various formats. While file sizes will vary, no single research asset is expected to exceed 50 GB, with the total estimated dataset size below 100 GB.

Most of the research assets used in the project 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.

All research assets 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.

For each research asset, we anticipate producing at least one scientific paper. In addition, preprint articles will be published for urgent discoveries, such as the detection of a new quarantine pest. Data associated with preprints will be made available once the final scientific paper is published.

These 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: All Data Management Plan-relevant final research assets FORSAID is expected to produce, including information on the type, format, size, availability and reusability licence

Work Package	Research Asset	Type	Generated / Reused	Format	Size	Availability	Reusability licence
WP1	Project Quality Handbook	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP1	Data Management Plan	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP2	Maps of defoliated areas by *Thaumetopoea pityocampa*	Spatial dataset	Generated	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Maps of defoliated areas by *Thaumetopoea pityocampa*	Spatial dataset	Reused, DEFID2	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Maps of tree mortality by *Ips typographus*	Spatial dataset	Generated	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Maps of tree mortality by *Ips typographus*	Spatial dataset	Reused, DEFID2	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Maps of tree damage by *Corythucha arcuata*	Spatial dataset	Generated	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Maps of tree damage by *Bursaphelenchus xylophilus*	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 *Ceratocystis platani*	Spatial dataset	Generated	.gpkg	> 5 GB	Open	CC BY-NC 4.0
WP2	Database of damage by Agrilus *planipennis* using remote sensing spectroscopy	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP2	Database of damage by *Fusarium circinatum* using mid-infrared spectroscopy	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP2	Database of damage by *Cryphonectria parasitica* and *Bursaphelenchus xylophilus* using mid-infrared spectroscopy	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP2	Database of damage by *Diplodia sapinea* using mid-infrared spectroscopy	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP2	Database of plant damage by *Ceratoxystis platani* 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 *Ips typographus*	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP3	Database of data collected by automated traps for *Thaumetopoea* spp.	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP3	Images for AI identification of *Agrilus* spp.	Images	Generated	.tiff	> 10 GB	Open	CC BY-NC 4.0

WP3	Images for AI identification of *Monochamus* 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 the pine wood nematode, *Bursaphelenchus xylophilus*, in eDNA	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP3	Protocol for the detection of *Agrilus planipennis* in eDNA	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP3	Protocol for the detection of *Ceratocystis platani* in eDNA	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP3	Protocol for the detection of *Cryphonectria parasitica* 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	Report on needs, expectations and priorities of European stakeholders regarding effective regulated forest pest surveillance	Document	Generated	.pdf	< 1 M	Open	CC BY-NC 4.0
WP5	Database of data on the economic impact of forest quarantine pests	Dataset	Generated	.csv	100 M	Open	CC BY-NC 4.0
WP5	Database of cost-benefits 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 assets produced within the FORSAID project will be stored and made accessible via the FORSAID Community data space on Zenodo. Specific datasets produced within Work Packages 2 and 3 could also be shared with the Database of European Forest Insect and Disease Disturbances (DEFID2), which includes harmonized georeferenced records of European forests that were disturbed by insects and diseases. Raw data that are still being processed, *i.e.*, not yet associated with a published article, will be securely stored in institutional repositories, ensuring data protection. The repositories provided by the University of Padova are available for that purpose. Team Leaders of each participating institution will be responsible for overseeing data storage, maintaining backups, and ensuring compliance with data management protocols. 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 assets generated 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 size (with a limit of 50 GB) and format. The dedicated FORSAID Community on Zenodo is available at https://zenodo.org/communities/forsaid/. To ensure unique identification, a DOI (Digital Object Identifier) will be automatically assigned by the repository at the time of data record creation. When uploading research assets to the Zenodo repository, it will be mandatory to specify the funding grant, *i.e.*, European Commission (EU), Forest surveillance with artificial intelligence and digital technologies – FORSAID, grant agreement 101134200.

The following actions will ensure 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 a 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.

To improve the discoverability of research assets within the Zenodo community, a standardized naming convention will be implemented. Each research asset name will include:

- The name of the project, FORSAID.
- Underscore.
- The number of the associated Research Action, *e.g.*, RA2.1.2.
- Underscore.
- A short sentence describing the research asset, e.g., *Ips typographus* attacks on Norway spruce with remote sensing.

In this example, the complete name of the research asset would be FORSAID_ RA2.1.2_*Ips typographus* attacks on Norway spruce with remote sensing.

2.1.1 Metadata structure

All metadata will be available in the .csv format for easy accessibility.

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.*, image, model, publication, software.
- A DOI.
- Title and additional title.
- Publication date in ISO 8601 format (YYYY-MM-DD).
- Creators, *i.e.*, the persons or organisations that have created the resource being uploaded.
- Licences and rights: Zenodo defaults to the Creative Commons Attribution 4.0 International (CC-BY) license.
- Descriptions (recommended), *e.g.*, abstract, notes, or information on methods. Although description is recommended, it is mandatory to include a section on data provenance.
- Contributors (recommended), *i.e.*, persons or organisations that have contributed to the record such as supervisors, contact persons, and sponsors.

In addition to Zenodo minimal mandatory metadata, FORSAID will require the inclusion of the following mandatory information for all research assets:

- 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.

For WP3 images produced for AI identification, 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 assets, 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.

2.1.2 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 could be sent to JRC-DEFID2@ec.europa.eu for upload in 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 occurrence characterized by multiple agents or multiple hosts, climate-driven triggering factors and silvicultural practices.
- Qualitative assessment of the damage.

2.2 Making data accessible

FORSAID data accessibility policies will leverage the accessibility features provided by the Zenodo repositories.

The following actions will ensure 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 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.

2.2.1 Zenodo sign-up and sign-in

Zenodo allows users to sign up via the creation of local accounts. Login is allowed via local username and password, ORCID, or GitHub credentials. In all cases, Zenodo has a local account and links external identities to this local account. This allows a user to log into the same account from both GitHub, ORCID or with a local password.

2.2.2 Data and metadata accessibility

Access to FORSAID raw and final data and metadata after scientific paper publication (when relevant) will be open, with limited exceptions.

At this stage, no significant major 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 thesis 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.

2.2.3 Data management responsibilities

To ensure compliance with data management decisions as they relate to the DMP, the following measures apply in FORSAID:

- 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 FORSAID DMP, as well as with the related GPDR policies.

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

2.2.4 FORSAID Community on Zenodo

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 FORSAID consortium agreement.
- Community readers: all the FORSAID researchers belonging to the project beneficiary institutions.

Zenodo FORSAID Community data management policies are defined as follows:

- FORSAID researchers belonging to the project beneficiary institution, 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.

2.3 Making data interoperable

FORSAID data can be exchanged and re-used within the project period among all the institutions involved in the project. The following actions will ensure 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 assets will be universal, cross-platform, and open source, with open standards. For example, tables could be stored in simple text-based tabular format as .csv

supported by any open-source software, geospatial data as .shp or .tif, documents as .pdf, images as .gif or .png (Table 3). All details on formats of 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 re-use

All datasets produced within FORSAID have the potential for re-use. The following actions will ensure 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 Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles.

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 of 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.

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. Project members are encouraged not only to publish findings in open-access journals, but also to share preprints on recognized platforms to facilitate early knowledge dissemination and feedback. 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 the Annex 5 of the FORSAID 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.

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, *i.e.*, not yet associated with a published article, will be securely stored in institutional repositories, ensuring data protection. The repositories provided by the University of Padova are available for that purpose.

For the raw and final 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 run 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, also 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 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 re-used, and where and how data will be stored. All information will be stored in a confidential manner and accordance with the EU Directive 95/46/EC and EU General Data Protection Regulation (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 single participants. 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 of confirming compliance with Chapter V of the GDPR, as well as a declaration of confirming compliance with the laws of the country where the data was collected.