URSOiLL

D8.2 DATA MANAGEMENT PLAN

under review

28th February 2026

Nargish Parvin and Tora Råberg

Funded by the European Union

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor REA can be held responsible for them.

Technical References

Project acronym	URSOILL
Project full title	Innovative solutions for sustainable urban soil restoration through Living Labs
Call	HORIZON-MISS-2024-SOIL-01
Grant number	101219012
Project website	https://ursoill.eu
Coordinator	RISE - Research Institute of Sweden
Project duration	1st September 2025 – 31st March 2030

Deliverable No.	8.2
Deliverable nature	R
Work Package (WP)	WP 8
Task	D8.2 Data Management Plan
Dissemination level*	PU
Number of pages	51
Keywords	FAIR principles, Data Management, Open Science, GDPR, Data Governance, Data Ownership.
Authors	RISE
Contributors	ZABALA
Reviewer partners	ZABALA
Due date of deliverable	28th of February 2026
Actual submission date	27th of February 2026

Under review

R = Document, report

DEC = Websites, patent filings, video, etc.

DATA = data sets, microdata, etc.

DMP = Data Management Plan

*The labels used:

SEN = Sensitive, limited under the conditions of the Grant Agreement

PU = Public, fully open, e.g., web (Deliverables flagged as public will be automatically published in CORDIS project's page)

EU classified — RESTREINT-UE/EU-RESTRICTED, CONFIDENTIEL-UE/EU-CONFIDENTIAL, SECRET-UE/EU-SECRET under Decision 2015/444. For items classified under other rules (e.g. national or international organization), please select the equivalent EU classification level.

Document History

V	Date	Beneficiary	Author
V0.1	15/12/2025	Zabala	Marina Ruilope and Damián Muruzábal
V0.2	18/02/2026	RISE	Nargish Parvin, Tora Råberg
V0.3	23/02/2026	Zabala	Marina Ruilope, Damián Muruzábal
V1	25/02/2026	RISE	Nargish Parvin, Tora Råberg

Disclaimer

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 European Union nor REA can be held responsible for them.

Under review

1. Introduction 8
1.1. URSOILL project in a nutshell 8
1.2. Purpose 8
2. Data Summary 10
2.1. Purpose of data collection/generation 10
2.2. Data identification and description 10
2.2.1. Types of formats of data 10
2.2.2. Size of data and expected volume 11
2.2.3. Data utility 11
2.2.4. Origin of data 11
2.2.5. Re-use of existing data 11
3. Data managers 28
4. Data organization and exchange (within consortium, during project) 29
5. FAIR data management 30
5.1. Making data findable, including provisions for metadata 30
5.1.1. Dataset reference, name and versions 30
5.1.2. Naming conventions 31
5.1.3. Standards and metadata 31
5.1.4. DMP review process and data inventory 32
5.1.5. Archiving, preservation and data quality assurance 32
5.1.6. Publications and keywords 33
5.1.7. Digital Object Identifiers 34
5.2. Making data accessible 34
5.2.1. Open Access 34
5.2.2. Overview of access to data 35
5.3. Making data interoperable 37
5.4. Increase data re-use 37
5.5. Other research outputs 39
6. Allocation of resources 40
6.1. Costs 40
6.2. Data management responsibilities 40
7. Data security 42
7.1. Data storage and back-up 42
7.2. Data security within Consortium 42
7.2.1. Deployment of a threefold security protection strategy 42
7.2.2. Authentication, authorization and encryption 43
7.2.3. Focus on data aggregation and pseudonymization techniques 43

under review

7.2.4. Internal threats and human errors 45
7.2.5. Internal threat prevention 45
7.3. Data security as specified for Zenodo 46
8. Ethics 47
9. Other issues 48

under review

List of Tables

Table 1: Data identification and description.....12
Table 2. Internal threats prevention methodology 45

under review

Funded by the European Union

List of Abbreviations

Acronym	Meaning
AE	Affiliated Entity
CA	Consortium Agreement
CFS	Certificate on the Financial Statements
COOT	Coordination Team
D&C	Dissemination and Communication
EC	European Commission
FS	Financial Statement
FTP	Funding and Tenders Portal
GA	Grant Agreement
GenA	General Assembly
GMS	Grant Management Services
HE	Horizon Europe
IPR	Intellectual Property Rights
LL	Living Labs
PC	Project Coordinator
PTC	Project Technical Committee
RP	Reporting Period
WP	Work Package

under review

1. Introduction

1.1. URSOILL project in a nutshell

Urban soils are the foundation of a city's healthy ecosystem, supporting carbon storage and water filtration, and providing habitat for diverse species. Yet, these soils are so delicate that sealing, compaction or contamination can reduce their ability to protect cities from floods, heat and environmental stress. In this context, the EU-funded URSOILL project will establish a network of five Living Labs where scientists, communities and authorities can work on practical solutions to protect urban soils. Specifically, the project aims to restore soil health, promote ecosystem services and sustain greener and more resilient cities through a combination of nature-based solutions and smart technologies.

In the face of rapid urbanization, with 70% of the global population expected to live in cities by 2050, urban soils have become critical in providing essential ecosystem services (ESs) supporting healthy urban ecosystems. Among the ESs urban soils are essential for regulating the carbon and water cycles, mitigating flooding risk and heat islands, provide provisioning services, such as supporting biodiversity and urban food production, and enhancing urban green spaces that improve mental and physical well-being, and landscape aesthetics.

Despite their importance, urban soils are often degraded by anthropogenic activities, including sealing, compaction, contamination with heavy metals and other pollutants which hinders their ability to provide ESs. This degradation poses risks to human and environmental health and worsens the effects of natural disasters like floods and heatwaves.

URSOILL aims to address these challenges by establishing a network of five Living Labs (LLs) across EU cities to serve as interactive platforms where local stakeholders collaborate to co-create, test, and implement innovative solutions for urban soil restoration. Each LL will focus on improving soil health, enhancing ecosystem services, and promoting sustainable land management using innovative technologies and nature-based solutions (NBSs). URSOILL will adopt a multi-actor, interdisciplinary approach and engage different stakeholders. By placing urban communities at the core, the project ensures that the solutions developed are not only scientifically validated but also socially and economically viable. The co-creation of solutions in real-world settings will allow for effective knowledge transfer, fostering long-term cooperation and facilitating the replication of successful practices in other urban areas. By restoring urban soils and enhancing their capacity to deliver essential ESs, URSOILL will help cities become more climate-resilient, sustainable, and liveable.

1.2. Purpose

Projects funded by the EU's Horizon Europe (HE) program must outline the types of information and results they will generate or collect, as well as the methods and timelines for storing and publishing this data. The primary goal of the Data Management Plan (DMP) presented in this deliverable, report D8.2, is to offer guidelines for incorporating the key elements of the data management policy that URSOILL project consortium will follow. The DMP guidelines for URSOILL project will ensure that research data adheres to the FAIR principles (Findable, Accessible, Interoperable, and Reusable), facilitating the integration and reuse of knowledge in future research and projects. According to the Grant Agreement (GA) for the URSOILL project, the

Under review

first version of the DMP is due by the M6 (Deliverable D8.2), with subsequent updates scheduled M30 (D8.3) and M48 (D8.4).

The Data Management Plan (DMP) has been structured according to the Guidelines on FAIR Data Management in Horizon Europe (HE) provided by the European Commission in May 2021. The DMP outlines the types of data that will be generated and/or collected during the project, the standards to be used, the methods for data exploitation and sharing (for verification or reuse), and the preservation strategies. This DMP has been prepared using the *Data Management Plan Template HE*[1]. The development of the DMP will enable URSOILL' partners to address all data protection issues, including ethical concerns and security strategies. Additionally, under HE, each beneficiary must ensure open access (OA) to all peer-reviewed scientific publications related to its results, which will also be made available through the public section of the URSOILL website. All these aspects have been considered in the preparation of the DMP.

The Data Management Plan (DMP) will be a living document throughout the project, with this initial version evolving as the URSOILL project progresses. This report outlines the procedures for data collection, storage, and processing, and provides an overview of URSOILL' security protection strategy, addressing general ethical and data protection concerns.

RISE is the primary partner responsible for developing the DMP. This first version of the DMP was prepared in collaboration with partners from the DMP group (see Section 3), including ZABALA, CETENMA, TUM, ATB, and G!E. However, the implementation of the DMP is the responsibility of all project partners, particularly those involved in the collection, storage, organization, maintenance, use, and publication of data, information, and results within the URSOILL project.

under review

[1] Reference Documents | EU Funding & Tenders Portal

2. Data Summary

2.1. Purpose of data collection/generation

The main objective of URSOILL is to enhance urban soil health across the EU by establishing a network of five Living Labs (LLs) across EU cities to serve as interactive platforms where local stakeholders collaborate to co-create, test, and implement innovative solutions for urban soil restoration. In this perspective, URSOILL will generate technical, social, economic and environmental data at different levels, in compliance with all national and EU ethics and legal requirements within the scope of the following activities:

Support stakeholder mapping, engagement and strategic planning for the different URSOILL' LLs in their co-creation processes.

- Perform an extensive analysis of soil health indicators and promote sustainable soil management practices.

Define the Open Call, evaluate and select projects (and URSOILL's new partners), monitoring of the project execution and management of sub-grantees.

- Contribution to the development of promising techniques that may contribute to further advancements in urban soil health monitoring and management.
- Development of a dataset to create the online predictive tool using all the data collected from LLs, allowing potential replicators to predict the improvement that a specific solution could have under their specific conditions.
- Development of an interactive online application featuring predictive models capable of assessing the effectiveness of soil recovery solutions under various conditions, utilizing the LLs data.
- Engage relevant stakeholders in shaping and refining the experiments conducted in the LLs.
- Where applicable, collect stakeholder input and feedback on the impact and success of LL activities.
- Support the Decision Support Tool for replication.
- When possible, contribute to EUSO (EU Soil Observatory) and integrate urban soil data into EU-wide monitoring frameworks
- Enhance soil literacy among citizens and stakeholders.

Only the data necessary to execute project activities will be collected. Participants will not be requested to provide personal information unless it is essential.

2.2. Data identification and description

2.2.1. Types of formats of data

Data will be organized into datasets based on the category of the data and the site of soil sampling. Dataset units will be gathered, analysed, and produced throughout the project's duration. Each dataset will include, at a minimum, the work package it belongs to and the data typology, along with a brief description. URSOILL is set to incorporate external data like soil indicators, biodiversity metrics, and management practices. Additionally, data from stakeholder engagement will be stored to ensure participant confidentiality. All project data will be treated

Under review

with care and precision, ensuring the utmost respect for participants' rights. The external datasets, along with those resulting from personal surveys and interviews, will be stored with strict confidentiality to protect participants' personal data. By default, in compliance with GDPR (EU GDPR 2016/679), all project data will be handled with rigorous care.

URSOILL will compile and produce project datasets of long-term value within the context of regional implementations and assessments. A dataset may include various types of formats. The types and formats of URSOILL data are outlined in **Error! Reference source not found.**. The type and format of the output data shown in **Error! Reference source not found.** relate to successive and interrelated WPs/tasks. The output data formats will be coordinated between the work-packages and tasks to ensure the use of output from one WP/task as input for another WP/task. Collected data will be qualitative and quantitative data, with text formats (PDF/A, .txt, .doc/.docx), Spreadsheets (.xls/.xslx), Pictures (jpg, png), Shapefile (.shp), Databases (.cvs; accdb), AutoCAD (.dwg), WEBM File and other.

2.2.2. Size of data and expected volume

Dataset sizes will be continuously assessed throughout the project and will vary based on the scope and nature of the data provided. The storage used in URSOILL will cover this size with no additional costs, and exchange of data with different project members will be secure through the use of available Teams Channel Cloud storage. More details of data expected sizes can be found in **Error! Reference source not found.**.

2.2.3. Data utility

URSOILL' collected data will be suitable for use by:

- URSOILL consortium, with more details of linkages between WPs in **Error! Reference source not found.**
- Stakeholders involved in the project (academia, researchers, policy makers, end-users such as education sector, citizens and other actors)
- Scientific community
- European Commission services and European Agencies
- Related initiatives, such as iCOSHELLs, LIVINGSOILL, SOILCRATES, LILAS4SOILS, BENCHMARKS, SPADES, ARAGON, HUMUS, CURIOSOIL and GOV4ALL.
- Living Labs platforms, such as SOILL
- General audience

2.2.4. Origin of data

URSOILL will collect data from participants' facilities, research laboratories and from the LLs' experimental sites. Depending on the type of data, there will be various methods and data origins for each WP, with a more detailed description in **Error! Reference source not found.**. In general, main data origins will be:

- New generated data in experimental sites (soil indicators, etc.)
- Literature study/review and open-sourced data (re-use of existing data)
- Personal data from URSOILL partners, upon request
- Stakeholders network (interviews with groups and individuals, surveys, workshops, etc.).

Under review

2.2.5. Re-use of existing data

Data will be sourced from several already existing datasets and/or reports and will be updated making use of the knowledge of project partners.

The re-use of existing data available from research projects and from other European projects is being encouraged. References from Horizon projects would be considered as advantageous re-use data source. Table 1 includes also the detail per WP in case any data is re-used from existing sources.

Table 1: Data identification and description.

WP1: Stakeholder Mapping	
Contributor(s)	POLIMI, iUE, LL LEADERS
Data set description	Database with site mapping and stakeholder mapping for each Living Lab. The initial version (M3) will serve as a baseline and will be updated throughout the project, notably to reflect stakeholder participation in co-creation sessions.
Data type	Observational
Data format	.xls
Provenance of data: sources	New data
Methodology for data collection/generation, process and analysis.	Structured templates used to collect inputs from Living Lab Leaders.
Storage	URSOILL Teams server, for the duration of the project.
Purpose for data collection and link with other WPs	Establish the baseline for the development of each Living Lab. The data primarily supports WP1 and is also relevant for WP2, WP3, WP5, WP6 and WP7.
Data applicability	Not envisaged.
Expected delivery	M3 as part of D1.1; further updates throughout the project
Data ownership	POLIMI
Public or confidential database?	Confidential; the file must not be shared beyond the consortium.
Size (estimation)	< 1 MB.
Standards (if any)	N/A
Metadata (if any)	N/A

WP1: Co-creation data	
Contributor(s)	POLIMI, iUE, LL LEADERS
Data set description	Database collecting outputs from the co-creation sessions (8 per Living Lab), including anonymous qualitative information provided by participants.
Data type	Observational
Data format	Text-based material collected to support the co-creation session reports (D1.3–D1.7).
Provenance of data: sources	New data

under review

Methodology for data collection/generation, process and analysis.	Structured templates used to collect inputs from Living Lab Leaders.
Storage	URSOILL Teams server. For the duration of the project
Purpose for data collection and link with other WPs	Support and steer the development of the Living Labs based on stakeholder feedback; relevant for all URSOILL WPs.
Data applicability	Not in its raw form.
Expected delivery	Aggregated outputs in: D1.3 (M10), D1.4 (M22), D1.5 (M34), D1.6 (M48), D1.7 (M54).
Data ownership	Living Lab leader responsible for the co-creation session
Public or confidential database?	Confidential; the file must not be shared beyond the consortium.
Size (estimation)	Less than 10 MB.
Standards (if any)	N/A
Metadata (if any)	Target population, Data collection method (online, in-person, etc.), Dates and geographic scope of event, Question wording and response scales

WP1: Soil literacy survey results	
Contributor(s)	SLU, iUE, POLIMI, LL LEADERS
Data set description	Database with survey results on soil literacy levels in the Living Labs (at least two survey rounds).
Data type	Observational
Data format	Survey responses in .xls (or .csv).
Provenance of data: sources	New data; may also include qualitative data from re-used sources (e.g., existing studies on soil literacy levels).
Methodology for data collection/generation, process and analysis.	Structured templates used to collect inputs from Living Lab Leaders; processed by SLU for analysis purposes.
Storage	URSOILL Teams server for the duration of the project. If non-confidential, the dataset will be shared via a selected trusted open repository.
Purpose for data collection and link with other WPs	Establish the baseline of soil literacy levels in the Living Labs and track evolution towards project end. Primarily guides WP1 and is also relevant for WP7.
Data applicability	May inform soil literacy research beyond the project.
Expected delivery	M18 (baseline survey), M32 (second survey), with further updates at project end.
Data ownership	SLU
Public or confidential database?	Public; no sensitive data.
Size (estimation)	< 1 MB
Standards (if any)	N/A
Metadata (if any)	Survey title and purpose, Target population, Sampling method, Data collection method (online, in-person, etc.), Dates and geographic scope, Question wording and response scales, Variable definitions and codes, Data license and access conditions.

Under review

Funded by the European Union

WP1: Soil literacy activity database	
Contributor(s)	CLUBE, SLU, iUE, POLIMI, LL LEADERS
Data set description	Database with information on soil literacy activities organised in the Living Labs.
Data type	Observational
Data format	.xls (or .csv).
Provenance of data: sources	New data.
Methodology for data collection/generation, process and analysis.	Structured templates used to collect inputs from Living Lab Leaders; processed by CLUBE and SLU for analysis purposes.
Storage	URSOILL Teams server for the duration of the project. If non-confidential, the dataset will be shared via a selected trusted open repository.
Purpose for data collection and link with other WPs	Analyse soil literacy activities in the Living Labs. Primarily guides WP1 and is also relevant for WP5 and WP7.
Data applicability	May inform soil literacy research beyond the project.
Expected delivery	Analysis will feed into D1.9 (M54).
Data ownership	CLUBE
Public or confidential database?	Public.
Size (estimation)	< 1 MB
Standards (if any)	N/A
Metadata (if any)	Target population, Description of activity, Key learnings, Dates and geographic scope.

WP2: Living Labs Operational & Implementation	
Contributor(s)	RISE (lead), LL leaders, IUE, POLIMI, ATB, site owners/municipal partners (as applicable).
Data set description	Operational documentation and implementation records required to establish and run the Living Labs, including: site selection notes and implementation plans; governance/roles; intervention designs; implementation & maintenance logs (what/when/where/who); operational risks, safety notes and fieldwork logistics; stakeholder engagement session metadata (dates, counts, topics) — metadata only (no raw personal data). These materials underpin D2.1–D2.5 and D2.9–D2.10.
Data type	Derived/compiled & operational (administrative/operational records; not scientific measurements).
Data format	.docx, .pdf/A, .xlsx/.csv (tabular logs), .pptx; selected GIS references as .geojson/.gpkg for site boundaries.
Provenance of data: sources	New data curated in WP2 from Living Lab activities (drawing on WP1 inputs and local partner materials; no direct reuse of external protected datasets).
Methodology for data collection/generation, process and analysis.	WP2 structured templates; KoM-defined governance; monthly Living Lab reporting cadence; change-controlled updates at key deliverables.

under review

Storage	SharePoint with role-based access for project members. Preservation extracts: selected annexes embedded within public deliverables (e.g., redacted plans) where appropriate.
Purpose for data collection and link with other WPs	Supports WP3 (monitoring design context), WP4 (FSTP readiness evidence), WP5 (replication playbooks), WP6 (sustainability/governance synthesis).
Data applicability	Limited (project management and LL operations). Aggregated, non-sensitive content may be reused by municipalities or other Living Labs as good practice.
Expected delivery	Continuous; summarised at M12, M24, M39, and M48–50 in D2.1–D2.5 and D2.9–D2.10.
Data ownership	RISE and the producing beneficiary for their contributions. Some content may be co-owned with municipalities (as per CA).
Public or confidential database?	Confidential, with redacted/public extracts in deliverables where feasible.
Size (estimation)	0.5–1.5 GB.
Standards (if any)	EC reporting structure; PDF/A for archiving; Dublin Core fields for document-level metadata.
Metadata (if any)	Title, author, date, LL, version, location, WP/task link, confidentiality tag.

WP2: — Experimental & Soil Analytical Dataset	
Contributor(s)	LL leaders; CETENMA, TUM;
Data set description	Field and laboratory measurements (soil physical/chemical/biological indicators; baseline and follow-ups), plot registers and sampling events; monitoring observations (sensors/time series, microclimate, vegetation indices, geotagged photos); geospatial layers (plots, transects, sampling points). Supports D2.3–D2.8 and validation reports.
Data type	Experimental & Observational (primary scientific data).
Data format	Tabular: CSV/Parquet (primary), .xlsx for field sheets; Geospatial: GeoPackage/GeoJSON, rasters (GeoTIFF); Temporal series: CSV (ISO-8601 timestamps). Images: JPEG/PNG with EXIF (geotags where appropriate).
Provenance of data: sources	New data collected by Living Lab partners under common sampling/analysis guidance validated by WP3. External data (e.g., Copernicus) is referenced but not republished unless licensing allows.
Methodology for data collection/generation, process and analysis.	Harmonised sampling protocols (WP3 approval pre-sampling), lab SOPs, indicator lists and QA/QC routines; calibration/blank/duplicate samples; geospatial referencing with declared local CRS where relevant; versioned processing scripts (if used) and method sheets.
Storage	Internal: SharePoint project space with restricted subfolders for raw data. Public deposition: processed/cleaned datasets and methodology in Zenodo/OpenAIRE; dataset DOIs referenced in D2.x deliverables and linked to WP3 outputs.
Purpose for data collection and link with other WPs	Feeds WP3 (evaluation and indicator harmonisation), WP5 (replication evidence) and WP6 (sustainability lessons).

Under Review

Data applicability	May support research, policy, replication and EUSO integration (subject to licensing and ethics).
Expected delivery	Baseline in the first Living Lab season; updates at key milestones M13, M24, M26, M36–39 and M50 (per D2.3–D2.10).
Data ownership	Collected by the producing beneficiary; shared under project CA terms.
Public or confidential database?	Public by default after embargo (processed/aggregated); raw data may be restricted (IPR/ethical reasons).
Size (estimation)	2–8 GB per LL, depending on raster/imaging volume.
Standards (if any)	File: CSV/Parquet, GeoPackage/GeoJSON, GeoTIFF; Metadata: OpenAIRE + INSPIRE-like geospatial fields; Pictures (jpg, png), Units & timestamps harmonised; controlled vocabularies for indicators.
Metadata (if any)	Rich dataset-level metadata + data dictionary/codebook.

WP2: — Prototype Development Dataset	
Contributor(s)	RISE, LL leaders, CETENMA, TUM, IUE; technology providers where engaged.
Data set description	Prototype design specifications, bill of materials and operating parameters; commissioning/field test logs and performance metrics; comparative results across iterations (linked to D2.6–D2.8).
Data type	Derived/Compiled + Experimental results (mostly derived from experimental runs and design iterations).
Data format	pdf/A (design notes/specs), .csv/.xlsx (KPIs), selected CAD/diagram exports as .svg/.pdf; photos/videos (if any) as JPEG/MP4.
Provenance of data: sources	New data generated in WP2 and built on prior publications (properly cited).
Methodology for data collection/generation, process and analysis.	Design–build–test cycles documented using versioned templates; KPI definitions linked to the WP3 indicator framework to ensure comparability.
Storage	Internal: restricted SharePoint (IPR/safety).Public deposition: metadata and redacted results summaries in Zenodo; full specifications may be restricted or delayed due to IPR.
Purpose for data collection and link with other WPs	Supports replication/scaling (WP5) and policy/sustainability contexts (WP6).
Data applicability	Public summaries may support adopters, while full details may remain restricted.
Expected delivery	Iterations aligned with D2.6–D2.8 at M15, M26 and M38.
Data ownership	Producing beneficiary; IPR per CA.
Public or confidential database?	Mixed: public summaries; restricted technical detail.
Size (estimation)	0.5–2 GB.
Standards (if any)	KPI glossary aligned with WP3; QMS-style versioning.
Metadata (if any)	Version, status (draft/prototype vX), LL, safety/IPR flags.

WP3: — Soil Monitoring Datasets	
Contributor(s)	CETENMA, UPCT and Living Lab leaders

under review

Data set description	Dataset consolidating information gathered from experimental sites across all Living Labs, including monitoring indicator values, edaphoclimatic information, satellite imagery, etc.
Data type	Observational, experimental, derived
Data format	Multiple formats depending on indicators, including images, .xlsx laboratory results, .txt field reports, geolocated point soil data, algorithms and shapefiles.
Provenance of data: sources	Combination of new data (measured indicators) and re-used data (e.g., soil classification, previous management).
Methodology for data collection/generation, process and analysis.	Data collection methodology is defined in D3.1. Analysis will use appropriate statistical methods and modelling techniques (data science approach), including Levene's test, Shapiro–Wilk test, ANOVA and Duncan Multiple Range Test (DMRT).
Storage	Stored in the project SharePoint repository, structured by Living Lab and document type. Each partner uploads data to designated folders and notifies by email to enable validation. An Excel file consolidates Living Lab data for later integration into the project web application. Data will be stored for four years.
Purpose for data collection and link with other WPs	Assess best-performing solutions and populate the prediction database. The database supports modelling relationships between site-specific soil management and soil quality parameters (predictors), and assessing predictive efficiency of developed models.
Data applicability	Yes; potentially relevant for the EC, other projects, solution testers and farmers.
Expected delivery	M10-M52 (T3.2-T3.3)
Data ownership	CETENMA and LL leaders
Public or confidential database?	Data will feed a public web application, but the datasets and the report on best-performing solutions will be confidential (D3.5 and D3.6).
Size (estimation)	>1 GB size for each Living Lab.
Standards (if any)	Continuous feed into the general database (web application) developed by UPCT; partners measure and upload data, with CETENMA and UPCT supervising data structure and integration to ensure harmonisation.
Metadata (if any)	As defined in D3.3.

WP4: Open Call dataset	
Contributor(s)	ZABALA, WPs leaders, LLs leaders
Data set description	Dataset including call documentation (guidelines, sub-grantee agreement template, application form, Q&A); analytics on submitted proposals; and selection of nominated new stakeholders in Living Labs. Contents will be generated in English and will include aggregated participation attributes (e.g., country of origin, challenge targeted) for statistical analysis, applying measures to prevent identification of personal data. A spreadsheet will be generated from proposal submissions.

Under review

Data type	Observational. The applicants' database collects: administrative data (legal name, VAT, address, webpage, contact details, and company figures if applicable); technical approach (solution description, innovation and technical risk assessment); team (profiles and dedication); and planning/value for money (activities, steps, timing and resources).
Data format	Mainly survey responses, primarily stored as CSV.
Provenance of data: sources	New data collected at two points: (a) proposal submission form; (b) tracking of sub-grantees' progress during implementation.
Methodology for data collection/generation, process and analysis.	Data will be aggregated and/or anonymised to avoid compromising applicants' personal details and other confidential project information.
Storage	Stored in an encrypted file in SharePoint with limited access for consortium members and external experts acting as Data Processors. The private key will be shared via secure channels among consortium members.
Purpose for data collection and link with other WPs	Supports evaluation of proposals by consortium members (notably WP2, WP3) and External Evaluators for selection decisions in WP4. Also used to produce statistics on services uptake, applicants' origin, submission modalities, etc.
Data applicability	Only anonymised/aggregated Open Call participation statistics will be made public via the URSOILL website. Metadata may be used for research purposes, but any publication will rely on fully anonymised/aggregated information.
Expected delivery	M27
Data ownership	ZABALA and RISE
Public or confidential database?	Mixed. call data is public; call analytic data is confidential.
Size (estimation)	1-2 MB
Standards (if any)	N/A
Metadata (if any)	Descriptive (title, applicant name(s), affiliation, keywords/tags, abstract); structural (submission components, versioning); administrative (ID, timestamps, status, reviewer assignments, decision date); technical (format/size, software if relevant, access logs, platform details); provenance (review feedback/scores, decision rationale, communication logs).

WP4: Sub-grantees dataset	
Contributor(s)	ZABALA, WPs leaders, LLs leaders
Data set description	Dataset including: URSOILL participants portfolio (selected applicants and justification information) and payments (funding distributed to third parties). It will also include aggregated sub-grantee attributes (e.g., country of origin, stages participated) for statistical analysis of the participation process. A database

under review

	documenting execution of solutions by each awarded third party will be generated, in English.
Data type	Derived / Compiled from the sub-call dataset
Data format	Reports mainly stored in CSV format.
Provenance of data: sources	New data collected from awarded applicants.
Methodology for data collection/generation, process and analysis.	After the Open Call, the shared spreadsheet will be populated only for participants qualifying for subsequent phases; records are updated only for teams progressing. Fields will be double-checked to ensure accuracy. Content is created by extracting metadata from signed sub-grant agreements (per template), evaluation forms and result communication letters.
Storage	Stored in the internal sub-grantees database. RISE and ZABALA (controllers) ensure security and privacy. Data will be stored in an encrypted spreadsheet in SharePoint with limited access for consortium members (as Data Processors). The private key will be shared via secure channels. Information will be made available to the EC upon request. Data is retained for the project duration; anonymised statistical data may be retained longer for research purposes with no additional cost expected.
Purpose for data collection and link with other WPs	Shared with WP2 (RISE), WP3 (CETENMA), WP5 (TUM) and WP6 (ATB) to support implementation, monitoring and integration of third-party outcomes into Living Lab conclusions. Also supports follow-up and reporting on solutions developed through URSOILL.
Data applicability	Mainly for dissemination of metadata about selected participants and associated solutions developed by URSOILL.
Expected delivery	M28, M50
Data ownership	ZABALA and RISE
Public or confidential database?	Mixed. Call awardees name-reference is public; personal data of awardees is confidential.
Size (estimation)	1-2 MB
Standards (if any)	N/A
Metadata (if any)	Descriptive (title, sub-grantee name, grant); administrative (ID, creation/modification dates, access rights/restrictions/licensing); technical (format, size, structure); provenance (access/usage logs, version history).

WP4: External experts dataset	
Contributor(s)	ZABALA
Data set description	Dataset covering: (i) Open Call for Experts (name, surname, email, LinkedIn, phone, gender, country of residence, nationality, CV); and (ii) selected experts (personal details for self-employed or company-employed experts; IBAN; profile/expertise area; description of URSOILL activities; and payments issued).

under review

Data type	Observational (survey results) and compiled from existing datasets.
Data format	Primarily CSV.
Provenance of data: sources	New data collected through the selection procedure of external experts evaluating third-party proposals.
Methodology for data collection/generation, process and analysis.	External experts will issue invoices for their carried-out work. They will be asked to fill in a form together with the contract the first time that they carry out work for URSOILL where they will be explicitly informed that ZABALA will keep their personal data whilst they work for URSOILL and to be notified about URSOILL related facts. This information notice must include the fact that non-confidential information about external experts might be shared with the European Commission and sub-grantees for transparency purposes.
Storage	All information related to external experts is stored in an internal database. RISE and ZABALA, as data controllers, ensure compliance with security and privacy requirements. Data are stored in an encrypted spreadsheet within a SharePoint space with restricted access limited to URSOILL consortium members. The private key required to access the contents is shared through secure channels. Data are retained for the duration of the project. Anonymised statistical data may be retained beyond the project lifetime for research purposes. If retained after project completion, anonymised metadata and statistical data will be stored securely within the existing infrastructure.
Purpose for data collection and link with other WPs	The dataset supports project monitoring and follow-up activities, as well as communication and dissemination where relevant. In particular, it enables the organisation and implementation of proposal evaluations under WP4.
Data applicability	No.
Expected delivery	M28
Data ownership	ZABALA and RISE
Public or confidential database?	Confidential
Size (estimation)	1-2 MB
Standards (if any)	N/A
Metadata (if any)	Descriptive (title, sub-grantee name, grant); administrative (ID, creation/modification dates, access rights/restrictions/licensing); technical (format, size, structure); provenance (access/usage logs, version history).

WP5: — TEF questionnaire	
Contributor(s)	LL Leaders / Solution providers
Data set description	The dataset includes information on the definition of the market and end-users of the solution; expected costs and revenues (including estimated variability/sensitivity or ecosystem gains); required skills for implementation and operation; technological

under review

	aspects; connection points and infrastructure; required investments; certifications; and non-marketed outputs.
Data type	Observational (online survey on costs and benefits).
Data format	.pdf
Provenance of data: sources	New data
Methodology for data collection/generation, process and analysis.	Online survey questionnaire distributed as an editable file (e.g. Word format).
Storage	TUM PuR drive (institutional servers) with backup and preservation strategies, and URSOILL SharePoint.
Purpose for data collection and link with other WPs	The dataset supports the identification of the most promising soil health solutions for potential beacons. The collected data are used to perform a cost-benefit analysis, providing quantitative and qualitative evidence on each solution's potential and identifying possible needs for intervention in current practices.
Data applicability	Not expected.
Expected delivery	M52
Data ownership	TUM
Public or confidential database?	Costs and other sensitive data are treated as strictly confidential and are not disclosed to third parties. Only anonymised information may be published.
Size (estimation)	1 MB
Standards (if any)	N/A
Metadata (if any)	N/A

WP5: — Social acceptance study	
Contributor(s)	LL Leaders / Solution providers
Data set description	The dataset addresses potential barriers and enablers for social acceptance, based on interviews with solution owners, findings from existing research, and stakeholder surveys. It includes statistical modelling to elicit stakeholder preferences (willingness to accept or pay), using stated preference methods (e.g. choice modelling) or revealed-preference methods, where applicable.
Data type	Observational (survey on social acceptance, including willingness to pay or accept).
Data format	pdf / csv
Provenance of data: sources	New data and re-used data from the literature.
Methodology for data collection/generation, process and analysis.	Online survey questionnaire distributed as an editable file (e.g. Word format) or implemented through survey software (e.g. Unipark, Qualtrics).
Storage	TUM PuR drive (institutional servers) with backup and preservation strategies, and URSOILL SharePoint.
Purpose for data collection and link with other WPs	The dataset supports the identification of the most promising soil health solutions by analysing potential barriers and drivers of social acceptance. The collected data feed into the techno-economic feasibility assessment, providing quantitative and

under review

	qualitative evidence on each solution's potential and possible needs for intervention in current practices.
Data applicability	Not expected.
Expected delivery	M52
Data ownership	TUM
Public or confidential database?	Costs and other sensitive data are treated as strictly confidential and are not disclosed to third parties. Only anonymised information may be published.
Size (estimation)	1 MB
Standards (if any)	N/A
Metadata (if any)	N/A

WP5: — Lighthouse feedback	
Contributor(s)	LL Leaders / Solution providers
Data set description	The dataset contains information on potential Lighthouses identified by Living Lab leaders, including details on site characteristics, capacities, technologies, foreseen soil health benefits, and solution descriptions.
Data type	Derived/compiled (Lighthouse data provided by Living Lab leaders).
Data format	pdf / xls
Provenance of data: sources	New data
Methodology for data collection/generation, process and analysis.	Lighthouse feedback report distributed as an editable file (e.g. Word or Excel format).
Storage	TUM PuR drive (institutional servers) with backup and preservation strategies, and URSOILL SharePoint.
Purpose for data collection and link with other WPs	The dataset supports the identification of sites qualifying as potential Lighthouses, characterised by innovative approaches that enhance solution effectiveness and adaptability across contexts, enabling upscaling and improved soil health outcomes.
Data applicability	Not expected.
Expected delivery	M54
Data ownership	TUM
Public or confidential database?	Costs and other sensitive data are treated as strictly confidential and are not disclosed to third parties. Only anonymised information may be published.
Size (estimation)	1 MB
Standards (if any)	N/A
Metadata (if any)	N/A

WP5: — Living Labs workshop	
Contributor(s)	LL Leaders / Solution providers
Data set description	The dataset documents replication workshops organised by the Living Labs to transfer good practices from experimental sites and support replication activities.
Data type	Derived/compiled (Lighthouse-related data provided by Living Lab leaders).

under review

Data format	.pdf
Provenance of data: sources	New data
Methodology for data collection/generation, process and analysis.	Workshop reports prepared and shared as editable files (e.g. Word format).
Storage	TUM PuR drive (institutional servers) with backup and preservation strategies, and URSOILL SharePoint.
Purpose for data collection and link with other WPs	The workshops aim to transfer good practices from Living Labs experimental sites, and to identify challenges and opportunities for engaging and supporting replicators within and beyond the Living Labs.
Data applicability	Not to be expected.
Expected delivery	M54
Data ownership	TUM
Public or confidential database?	Costs and other sensitive data are treated as strictly confidential and are not disclosed to third parties. Only anonymised information may be published.
Size (estimation)	1 MB
Standards (if any)	N/A
Metadata (if any)	N/A

WP5: — Roadmap Evidence Base and Drafting File	
Contributor(s)	ESK, TUM, ATB, Living Lab leaders/partners; inputs from other WPs as needed.
Data set description	Curated evidence base supporting the drafting of D5.5, including synthesis notes, structured matrices of drivers and barriers, stakeholder engagement inputs, references to WP5 outputs (T5.1–T5.3), and successive roadmap draft versions.
Data type	Derived/compiled and observational.
Data format	Working files in .docx/.xlsx/.csv; final report in .pdf; optional meeting notes in .txt/.docx.
Provenance of data: sources	New data compiled from URSOILL internal results (T5.1–T5.3 and other WPs), roadmap drafting workshops and interviews, and re-used research sources.
Methodology for data collection/generation, process and analysis.	Desk-based synthesis of URSOILL outputs using structured extraction templates (drivers, barriers, strategies). Drafts are consolidated into a roadmap structure with version control and traceability to source deliverables and meeting minutes.
Storage	Working files stored in the consortium's access-controlled shared workspace (SharePoint). Final publishable outputs deposited in a trusted repository (Zenodo/OpenAIRE) once cleared for dissemination. Data retained up to five years after project end.
Purpose for data collection and link with other WPs	Supports the preparation of D5.5 and informs dissemination, replication, and exploitation activities by consolidating scaling-up pathways.
Data applicability	Yes. Aggregated and non-confidential roadmap insights are relevant for cities, Living Labs, policymakers, solution providers, and Mission "Soil Deal for Europe" stakeholders.

under review

Expected delivery	M52 (Task 5.4 runs M36–M52; D5.5 delivered within this period).
Data ownership	Originating partners retain ownership of their inputs. The compiled evidence base is managed by the Task 5.4 leader (ESK) under the Consortium Agreement. Personal data are handled in compliance with GDPR and project ethics requirements. Sensitive data are mitigated through anonymisation and aggregation.
Public or confidential database?	Mixed. Anonymised/aggregated datasets and the final roadmap report are public; raw notes containing personal or confidential information remain restricted.
Size (estimation)	4 GB
Standards (if any)	Open document standards where feasible; consistent file naming and versioning; controlled vocabularies where applicable.
Metadata (if any)	Dataset-level metadata (title, authors, description, dates, keywords, funding, grant information) following repository schemas.

WP5: — Exploitation and KER Management Dataset	
Contributor(s)	All partners
Data set description	Structured inventory of Key Exploitable Results (KERs), results ownership list, IPR analysis notes, exploitation workshop materials, business model canvases, commercialisation plans, and stakeholder/contact lists for exploitation (where permitted).
Data type	Derived/compiled and observational.
Data format	.xlsx/.csv (KER inventories and ownership lists); .docx/.pdf (analyses and plans); optional .pptx (workshop materials).
Provenance of data: sources	New data generated during the project (KER updates, IPR analyses, exploitation roadmaps), with references to technical results from other research where relevant.
Methodology for data collection/generation, process and analysis.	Partner self-reporting through structured templates, consolidation by Task 5.5 leader, IPR screening and classification (commercial vs non-commercial), and iterative updates aligned with reporting periods.
Storage	Access-controlled consortium SharePoint during the project; confidential exploitation/IPR files retained by KER owners and the Task leader.
Purpose for data collection and link with other WPs	Ensures long-term uptake of URSOILL results, supports dissemination and sustainability planning, and clarifies exploitation pathways and ownership.
Data applicability	Partially applicable. Public-facing summaries may be released; sensitive IPR and negotiation details remain confidential.
Expected delivery	M54 (Task 5.5 runs M6–M54; D5.6 delivered at project conclusion).
Data ownership	Result ownership and background IP are governed by the Consortium Agreement. Sensitive data include IPR details, ownership notes, and business model elements.There will be sensitive data: IPR details, ownership/negotiation notes, and some business model elements may be confidential.

Under Review

Public or confidential database?	Predominantly confidential; selected summaries may be public.
Size (estimation)	2 GB
Standards (if any)	Structured KER templates; consistent classification scheme.
Metadata (if any)	Dataset-level metadata for openly deposited items (authors, project, grant, version/date, keywords).

WP6: — List of Best Practices to improve Urban Living Lab performance	
Contributor(s)	CLUBE, LL Participants, LL Leaders, POLIMI, IUE, FORUA
Data set description	Dataset detailing best practices for the continuation and improvement of Urban Living Labs.
Data type	Derived/compiled from literature on Soil Health Living Labs and observational data from the five Living Labs.
Data format	Report in .pdf; supporting data in .xml.
Provenance of data: sources	Re-used data from ongoing Mission Soil projects and new data from the five Living Labs.
Methodology for data collection/generation, process and analysis.	Structured templates used to collect inputs from Living Lab leaders and participants.
Storage	ATB institutional server and URSOILL SharePoint Teams server for the duration of the project; potential publication in a data journal.
Purpose for data collection and link with other WPs	Best practices identified through literature review and Living Lab experience inform the development of roadmaps (short-, medium- and long-term strategies) and support deliverables D6.1–D6.5.
Data applicability	Yes. Relevant for the scientific community working on Living Labs, Soil Health Living Labs, and soil governance.
Expected delivery	M18 (D6.1), with updates at later stages.
Data ownership	Data will be owned by URSOILL authors.
Public or confidential database?	Public
Size (estimation)	>100 Mb
Standards (if any)	N/A
Metadata (if any)	Scientific standards

WP7: — Communication and Dissemination dataset	
Contributor(s)	All partners
Data set description	Dataset used for communication and dissemination purposes, including internal and external contact data (e.g. names, email addresses, images), collected with consent.
Data type	Observational (not research-related).
Data format	.txt, .pdf, .csv, .xlsx, .jpeg, .svg, .png, video formats.
Provenance of data: sources	Re-used data and newly collected contact information.
Methodology for data collection/generation, process and analysis.	GDPR-compliant procedures for collecting and processing personal data (e.g. surveys, emails, templates).
Storage	Stored internally on G!E servers with restricted access and established backup procedures.

Under review

Purpose for data collection and link with other WPs	Supports communication and dissemination activities in compliance with GDPR.
Data applicability	No
Expected delivery	Continuous
Data ownership	G!E
Public or confidential database?	Confidential.
Size (estimation)	From a few MB (text files) up to several GB (e.g. video files).
Standards (if any)	N/A
Metadata (if any)	N/A

WP8 — Project Management Records	
Contributor(s)	All partners
Data set description	Administrative and coordination metadata related to deliverable preparation and review cycles, meeting metadata, risk and quality monitoring, amendment and reporting logs (metadata only), and workflow tracking for GA/CA signing (metadata only).
Data type	Derived/compiled internal project metadata.
Data format	.xlsx, .csv, .pdf/A, .docx, .pptx, .txt.
Provenance of data: sources	New data generated continuously from internal project activities.
Methodology for data collection/generation, process and analysis.	Structured templates (Management Handbook, RMP, QARP); meeting minutes; reporting tools (SYGMA, SharePoint logs); periodic EC reporting procedures.
Storage	SharePoint workspaces (ZABALA and RISE) with institutional backups. Selected metadata may be transferred to Zenodo annexes when appropriate.
Purpose for data collection and link with other WPs	Ensures project governance, EC reporting, delivery tracking, and risk management.
Data applicability	Limited to EU reviewers and auditors.
Expected delivery	Continuous; compiled at M6, M18, M36, M54.
Data ownership	RISE
Public or confidential database?	Confidential (only aggregated metadata may become public).
Size (estimation)	50–150 MB.
Standards (if any)	Horizon Europe reporting standards; PDF/A archival format.
Metadata (if any)	Minimal metadata (date, creator, version, WP link, status).

WP8 — URSOILL Data Management Plan Dataset (DMP Metadata Registry)	
Contributor(s)	All partners
Data set description	Central registry compiling all WP-specific dataset descriptions submitted through DMP input forms, including dataset names, data types and sources, formats, licences, standards, repository targets, and FAIR compliance statements.
Data type	Derived / Compiled
Data format	.xlsx (registry), .docx (WP inputs), .pdf (final annex in D8.2/D8.3/D8.4).

under review

Provenance of data: sources	New data generated from WP input forms and updated at M34 and M54.
Methodology for data collection/generation, process and analysis.	WP leaders submit DMP input forms; WP8 validates completeness and FAIR alignment; ZABALA performs quality control; RISE compiles the final registry.
Storage	Internal SharePoint during drafting; final DMP deposited in Zenodo under CC-BY / CC0.
Purpose for data collection and link with other WPs	Provides an audit trail for FAIR compliance and supports EUSO compatibility.
Data applicability	Yes. The DMP (excluding confidential elements) is publicly accessible.
Expected delivery	M6, M34, M54.
Data ownership	RISE, with co-ownership of WP leads for their respective contributions.
Public or confidential database?	Public (final DMP); internal drafts confidential.
Size (estimation)	5–20 MB per version.
Standards (if any)	FAIR principles; EOSC alignment; Horizon Europe DMP structure.
Metadata (if any)	Dataset-level metadata per WP.

WP8 — URSOILL Ethics & Gender Action Plan (GAP) Dataset	
Contributor(s)	All partners
Data set description	Dataset including ethics-related metadata: status tracking of ethics deliverables (D8.5 and WP9 deliverables), GAP monitoring indicators, compliance checklists, ethics review records (metadata only), and consent template versions (metadata only).
Data type	Derived / Compiled
Data format	.xlsx, .docx, .pdf/A
Provenance of data: sources	Newly created data, with partial reuse of EC ethics templates.
Methodology for data collection/generation, process and analysis.	Collected from partner ethics focal points and Ethics Advisor inputs; validated by RISE and ZABALA.
Storage	Secure SharePoint with restricted folders for ethics documentation.
Purpose for data collection and link with other WPs	Ensures legal, ethical, and GDPR-compliant implementation of project activities.
Data applicability	Partially applicable: GAP content may be public; internal ethics logs remain confidential.
Expected delivery	M10, with updates during reporting periods.
Data ownership	RISE; Ethics Advisor retains authorship of assessments.
Public or confidential database?	Mixed: GAP public; ethics compliance logs confidential.
Size (estimation)	10–30 MB total
Standards (if any)	EC Ethics Appraisal Procedure (EAP); GDPR; Horizon Europe ethics requirements.

under review

Metadata (if any)	Version control, status tracking, and ethical issue categorisation.

3. Data managers

The URSOILL DMP will be stored as a project deliverable document, with regular updates and successive versions. The first version of the URSOILL DMP integrates the discussed procedures of the work packages (WPs) and their interfaces. RISE is responsible for producing the URSOILL DMP and coordinating data management. All WPs contribute to the URSOILL DMP by collecting relevant project data, providing data for use in other WPs, and delivering information about the datasets for recording in the URSOILL DMP. Each WP may have more than one data manager, depending on the number of different tasks that require data management efforts. The identified data manager(s) for WP tasks are as follows:

- WP1: IUE
- WP2: RISE, CLUBE, LIST, ESKILARA, INNOVHUB
- WP3: CETENMA
- WP4: ZABALA
- WP5: TUM
- WP6: ATB
- WP7: G!E
- WP8: RISE, ZABALA
- WP9: RISE

under review

4. Data organization and exchange (within consortium, during project)

The development, implementation, oversight, and updating of URSOILL's DMP is integrated into project management (WP8, T8.3) throughout the project's duration. The responsible team member (see Section 3) regularly receives contributions from all work packages. During the project, URSOILL will utilize ZABALA's central unit's IT infrastructure to facilitate exchanges between WPs and safeguard the value of long-term data.

A Teams Channel repository for URSOILL, hosted on ZABALA's server, has been established to efficiently manage project documentation. This tool is designed for collaborative document work and sharing final documents of common interest. It serves as the project's online collaboration platform, providing user-friendly and secure access to data through a web interface, allowing users to view, sync, and share files across devices easily. The system uses an open architecture, extensible via API for applications and plugins, and is compatible with any storage.

The Teams Channel is accessible only to URSOILL WP members, includes daily backups, and maintains high standards of data security. All consortium partners will have access to the project SharePoint within the Teams Channel and will be able to upload and download datasets. The availability of datasets, which will be delivered from one WP as input for another, will be organized by project management and is pre-defined in the project schedule. Partners of the project consortium are responsible for data consistency, quality, and storage during the project. To ensure the exchange and storage of project data, especially long-term data, the uploaded data on the Teams Channel will be backed up using ZABALA's IT infrastructure resources.

In the Teams Channel repository, the metadata of datasets will be publicly accessible to all consortium partners. This aims to enhance transparency for URSOILL research, attract more partner attention, facilitate cross-checking among members, and encourage further research collaboration. According to the policies outlined in Section 7.2 of URSOILL's data security protocol (which includes a threefold security protection strategy: authentication, authorization, and encryption; focusing on data aggregation and pseudonymization techniques; and preventing internal threats and human errors), any observation data or primary data generated or collected by URSOILL members must be accessible to other members during the project period. Members interested in accessing this data must also comply with URSOILL's data security policy. However, certain partners may request to withhold generated data for a limited period for publishing purposes. In such cases, each data controller must provide a detailed explanation regarding the purpose and duration of the data withholding.

Under review

5. FAIR data management

Data access and sharing activities will be conducted in compliance with national and EU privacy and data collection regulations, as well as H2020 and HE rules. All project deliverables and data will be stored and shared in Teams folders restricted to the project consortium. Initially, only consortium partners will have access to the project cloud storage where datasets and metadata are filed. Subsequently, scientific publications, articles, dataset deliverables, and final research results will be stored and published through Zenodo, EUSO (via ESDAC) and other repositories (e.g., OpenAgrar) following the FAIR principles.

5.1. Making data findable, including provisions for metadata

5.1.1. Dataset reference, name and versions

Each file is given a unique name to distinguish and easily identify datasets. This name also serves as the identifier for the datasets. We follow a specific practice to create these dataset names. Each designed dataset name consists of three different parts, separated by an underscore ('_') character:

ProjectName_DatasetName_Version, where:

1. The ProjectName is URSOILL, to clearly identify the origin for all datasets.
2. The DatasetName represents the full name of the dataset.
3. The Version of the dataset represents in which phase of the project the dataset was released:
    a. DB the live database during project lifetime.
    b. RP1Export export of the database at M15.
    c. RP2Export export of the database at M27.
    d. FinalExport export of the database at the end of the project, i.e. M48.

An example of a dataset's name could be the following:

URSOILL_SoilIndicatorsSWELL_RP1Export.

In order to catalogue data in the repository as well as to facilitate their search and re-use, metadata will be filled in when uploading datasets in Zenodo (or ESDAC, Arxiv, GitHub, etc.) (see Section 5.1.5 for further information) encompassing at least the following elements:

- Title
- Author/s
- Publication Date
- Description
- Access right
- License
- Funding

Additionally, all open data, publications and open-source software deposited in the Zenodo repository (se 5.1.5) will use DOI versioning. DOI versioning allows for updating a dataset after it has been published and to cite either a specific version of a dataset or all versions of a dataset.

under review

5.1.2. Naming conventions

File names will include at least a version number and/or a time stamp. Files and folders at data repositories will be versioned and structured by using a name convention consisting as follow: YYYYMMDD_URSOILL_DatasetName_Vzz.docx.

5.1.3. Standards and metadata

Providing as much accurate and rich metadata as possible will enhance the data's findability and reusability. Metadata for describing the data collected and generated by URSOILL project is essential for facilitating open access to the data.

URSOILL records metadata for the specific produced datasets described in **Error! Reference source not found.**, which will use a standard format for metadata and will specify the data and metadata standards used. The governance of metadata is a crucial part of the DMP, as even metadata without obvious identifiers could qualify as personal data if rich metadata could lead to reidentifying a data subject when combined with other data sources.

Several metadata standards exist (e.g. DCC, RDA), and it may not be possible to find one that fits all purposes. Therefore, a practical approach is to agree on a common and minimal catalogue metadata schema for datasets published in public catalogues and data repositories, using data-type specific schema extensions if necessary.

The Zenodo deposition metadata domain model[2], based on DataCite's metadata schema[3], minimum and recommended terms will be used for open data generated by the project and deposited in an appropriate repository. These standards will be used for releasing data outside the project.

For URSOILL, the following deposition metadata fields are mandatory:

- title (of the deposition)
- description (abstract or description for deposition)
- files (Deposition files identifiers, filenames, size of the files in bytes)
- upload_type (Type of the deposition from a controlled vocabulary (publication, dataset, software, etc.))
- publication_date (Date of publication in ISO8601 format (YYYY-MM-DD))
- creators (The creators/authors of the deposition)
- license (Open license from controlled vocabulary "Open Definition Licenses Service")
- DOI (Digital Object Identifier assigned by the DOI registrant)
- keywords (descriptive of the dataset content)
- related_identifiers (Persistent identifiers of related publications, datasets, and software)
- communities (List of communities the deposition appears in (Zenodo)
- grants (List of European Commission Horizon Europe grants which have funded the research for this deposition (730355). Needed to establish the relationship to URSOILL in OpenAIRE.

[2] Developers | Zenodo

[3] DataCite Schema

Under review

To improve metadata and the obtained results, URSOILL participants will consider mentioning controlled vocabulary that significantly enhances metadata and increases the findability of the data. For many disciplines, there are already specialized classifications and thesauri (see BARTOC.org) as well as standard data for persons, institutions, or research funding agencies.

For soil monitoring database and development of the predictive tool, data from each Living Lab solution will continuously feed the general database (web application) developed by CETENMA. Partners from each Living Lab solution will oversee measuring and feeding the database. To obtain a harmonized database, CETENMA will supervise data structure and integration into the application.

For Open Call's datasets, some metadata standard to be considered are:

- Dublin Core Metadata Initiative (DCMI), This is one of the most widely used standards and provides a basic set of elements to describe a wide variety of resources. It is flexible and can be adapted for different types of data.
- ISO/IEC 11179 Metadata Registry (MDR): This standard focuses on the management of metadata registries and is useful for ensuring data consistency and quality over time.
- Data Documentation Initiative (DDI): Ideal for social science data, this standard provides a detailed framework for data documentation, including technical and administrative aspects.
- Statistical Data and Metadata Exchange (SDMX): This standard is excellent for the exchange of statistical data and metadata between organizations, ensuring interoperability and consistency.
- ISO 27001: Although not a metadata standard per se, ISO 27001 provides a framework for information security management, which is crucial when handling confidential data.

5.1.4. DMP review process and data inventory

The internal process of quality evaluation and reporting is active throughout the entire project duration to assess both project data/products and processes. All research generated data and publications will be analysed and described using the Data Inventory Table (to be included in the next update of this deliverable), which WP leaders and publication authors are required to fill in periodically. A data inventory is a centralized metadata collection that indicates all the datasets the project collects and maintains.

The Data Controllers at each LL will be responsible for completing the metadata, uploading the public datasets they have generated, and assigning specific keywords relevant to these datasets. These keywords must be descriptive of the dataset content.

Further updates to the DMP will include the eventual updating of the online research data repository where data are collected and shared, as well as the description of the datasets and research data gradually generated and collected.

5.1.5. Archiving, preservation and data quality assurance

Archiving and preservation

URSOILL partners have agreed on the procedures to ensure the long-term preservation of the datasets. Datasets will be stored on Zenodo, a generic repository for EC-funded research developed by CERN and launched in May 2013. To be an effective catch-all that eliminates

Under review

barriers to adopting data-sharing practices, Zenodo does not impose any requirements on format, size (currently accepts up to 50GB per dataset), access restrictions, or license.

Additionally, datasets stored on Zenodo are automatically part of OpenAIRE, the EC-funded initiative that supports the Open Access policy of the European Commission via a technical infrastructure, thus integrating them into existing reporting lines to funding agencies like the European Commission.

Archiving on Zenodo is free, thus eliminating costs. According to Zenodo's Terms and Conditions, the uploader is responsible for ensuring the content is suitable for open dissemination, complying with all applicable regulations, including data protection, privacy, and intellectual property laws. All relevant stakeholders in the URSOILL project will be informed of this requirement and requested to act accordingly. This may require anonymizing the uploaded contents of survey respondents, among other measures.

Data quality assurance

URSOILL is committed to delivering quality data and adopts data quality assurance procedures to achieve this goal. Quality control of each dataset is the responsibility of the relevant WP leader, supported by the Project Coordinator. Depending on the context, 'quality' may have different meanings, based on the utility and re-usage scenarios of the dataset. Data quality assurance may involve editing and moderation, cleaning, pre-processing, adding metadata, transforming to a more convenient format, or providing easier access. Information about the consortium's efforts to address data quality issues is provided for each type of dataset.

All stakeholders acknowledge and agree to comply with the traditional principles relating to data quality, which have been updated by the GDPR as principles relating to the processing of personal data. These principles include, among others, that data shall be processed lawfully, fairly, and transparently, collected for specified purposes, and be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed. Lastly, data quality shall also refer to the dataset as a whole, ensuring that population subgroups are not discriminated against or otherwise harmed due to misrepresentation in the dataset.

Under review

5.1.6. Publications and keywords

Publications

The publications issued during the project include and acknowledge the Grant Number, acronym and a reference to the Horizon Europe Programme funding, including the following sentence:

"THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON EUROPE RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO. 101219012".

When displayed together with another logo, the EU logo will have appropriate prominence.

Search keywords

When uploading data and publications to Zenodo, data owners will enter keywords to facilitate searches. All open URSOILL results deposited in a repository will include search keywords along with their metadata (see 5.1.3). Keywords for open data will be selected from controlled vocabularies appropriate for the specific type of data. The list of keywords will be compiled during the project and included in the second versions of the DMP.

Keywords used so far: Living Labs, soil, soil health, soil indicators, co-creation, urban soils, ecosystem services provided by soils, soil management, soil pollution, soil remediation, soil

compaction, soil sealing, spatial development and architecture, land use, regional planning, remote sensing, urban agriculture, urban water management, ecosystem services, soil monitoring, stakeholder engagement, real world testing, nature based solution, Copernicus, soil reuse.

5.1.7. Digital Object Identifiers

All open data, publications, and open-source software produced in URSOILL (open URSOILL results) will be identifiable and locatable by means of a persistent Uniform Resource Locator (URL). If possible, open URSOILL results will be assigned a Digital Object Identifier (DOI) to make content easily and uniquely citable. URSOILL relies on external services for this, as DOIs can only be assigned by DOI registrants through a DOI registration agency (see DOI® Handbook).

Open URSOILL results deposited in the URSOILL default Open Access repository (Zenodo, see 5.1.4) will be automatically assigned a DOI and will benefit from Zenodo's DOI versioning support. Open URSOILL results deposited in institutional repositories, repositories of scientific publishers, or other data and research repositories will be at least identifiable by a persistent URI. If the institution is a DOI registrant with an agreement with a DOI registration agency, a DOI will also be assigned.

Whether scientific publications will be assigned a unique identifier like DOI, Publisher Item Identifier (PII), International Standard Serial Number (ISSN), etc., depends on the open access strategy (green or gold) chosen by the editors and thus also on the respective scientific publisher and the chosen research repository. Zenodo is, for example, one of the open data repositories that can generate DOIs for research results.

5.2. Making data accessible

5.2.1. Open Access

The H2020 Open Access Mandate aims to ensure that research data generated by HE projects is accessible with minimal restrictions, while also safeguarding personal or sensitive data due to privacy, commercial, or security concerns. Open Access in Research and Development projects typically pertains to "scientific information", which includes two main categories:

- Peer-reviewed scientific research articles (published in academic journals)
- Scientific research data (data underlying publications and/or raw data)

In accordance with the H2020 Guidelines on Open Access to Scientific Publications[4], Open Access will be applied to peer-reviewed publications (scientific research articles published in academic journals), conference proceedings, and workshop presentations conducted during and after the project's conclusion. Open-access published articles (accepted author version) will be made available on the URSOILL website.

There are two primary routes for open access to scientific peer-reviewed publications[5]:

[4] h2020-hi-oa-pilot-guide_en.pdf

[5] Open-Access-to-Publications-and-Data-in-Horizon-2020-Frequently-Asked-Questions-(FAQ).pdf

under review

- Self-archiving (also known as “Green” Open Access): This involves the researcher – or a representative – archiving the published article or the final peer-reviewed manuscript in an online repository before, after, or alongside its publication. Access to the article is often – but not necessarily – delayed (“embargo period”) as some scientific publishers may seek to recoup their investment by selling subscriptions and charging pay-per-download fees during an exclusivity period. Depending on the journal, the publisher may impose an embargo period of 6 to 12 months.
- Open Access publishing (also known as “Gold” Open Access): This means that an article is immediately made available in Open Access mode by the scientific publisher. The associated costs are transferred from readers to (for example) the university or research institute to which the researcher is affiliated, or to the funding institutions supporting the research. In this model, the costs of publishing are not borne by readers but by the authors, meaning these costs will be covered by the university or research institute to which the researcher is affiliated, or by the funding agency supporting the research. Gold Access will be pursued where possible within the URSOILL project.

Open Access will not impact the intellectual property generated by research results, as the decision to publish Open Access documents will follow the procedure outlined in Article 16 of the GA to first seek protection for intellectual property rights[5].

5.2.2. Overview of access to data

Open Access 🟢	**Confidential Access** 🔴	
Data type	Access	Main exploitation guidelines
Stakeholders / Personal data	🔴	User data will be treated with confidentiality. Disclosure is conditioned by the systematic authorization of end users provided in the "informed consent form" and "information sheets". The collection, storage, protection, retention will be done in compliance with EU Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data[6] and in compliance with the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing

[6] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of

personal data and on the free movement of such data: EUR-Lex - 31995L0046 - EN

Under review

		Directive 95/46/EC (General Data Protection Regulation)[7].
Data from the LLs and management practices	🟢	Open to public in final reports, project website and in press releases.
	🔴	Confidential access for contractual terms, financial Information and management plans.
	🔴	Confidential access for personal information in compliance with GDPR.
Data collected for databases	🟢	Open to public in final reports, web application project website and press releases.
	🔴	Confidential access for contractual terms, Open Call metadata, information and business plans.
	🔴	Confidential access for personal information in compliance with GDPR.
Data exchanged among partners	🔴	Contact details, file transfer protocol and secured collaborative space (Teams channel).
Conference presentations, Workshops, Events, Deliverables	🟢	Open access in the project website.
Peer-reviewed publications	🟢	Open access in the project website, Zenodo repository and partners' own repositories.
Soil monitoring database	🟢	Data will be transferred to a public web application.
	🔴	Datasets and the report on best performing solutions will be confidential (D 3.4. and D 3.5.).
Open Call dataset	🟢	Public but only the following fields will be published in the dataset, since this dataset must be made publicly available and informed through the Funding and Tenders Portal: Legal name, Organisation type, Country, Funding awarded, and Funding paid.
	🔴	More detailed dataset will be kept internally, e.g. external experts' data, link to the evaluation template filled in for those applicants who met the criteria to be

under review

[7] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the

processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504

		evaluated during the Open Call, the IBAN of the sub-grantees where payments are issued.

Accessible datasets will be licensed with a Creative Commons (2023) License.

5.3. Making data interoperable

This section can only be partially addressed at present, as data collection is ongoing. However, most datasets from LLs and WPs are stored in a shared location within a private Teams® Channel, created and managed by ZABALA. The Teams software system is open-source and compatible with a wide range of computing devices and operating systems, providing easy access for reading or uploading datasets.

The consortium will endeavour to collect and document data in a standardized manner to ensure datasets are correctly understood, interpreted, and reused. In addition to metadata, the data files will include relevant templates, surveys, and codebooks used to generate the data. Partners will be asked to provide documentation describing the main variables in the datasets, where necessary, to support interpretation and reuse. Standard vocabulary will be used for all data types in the dataset to facilitate interdisciplinary interoperability.

It is also recommended the use of non-proprietary, open formats, avoiding vendor-specific software. Even with commonly used software packages like Microsoft Office, backward compatibility and long-term readability cannot be guaranteed. Therefore, files should be saved in open formats whenever possible. Archived files should be unencrypted, uncompressed, patent-free, and created in an open, documented standard. This aligns with the FAIR Data principles for interoperability and data reuse.

Regarding the database for soil monitoring, the project cloud will be used in the storage layer, which allows Big Data capabilities and can be scaled to larger areas if necessary. In addition, a ETL processes will be generated to update the database information regularly deploying the processes using the open-source platform Apache Airflow.

Regarding the Open Call, the subgrantees datasets will be linked to the open call datasets as those selected applicants listed in the open call datasets will be part of the subgrantees datasets.

Different repositories will be used by the partners, such as Arxiv or Zenodo for publications, and GitHub or SourceForge for models and algorithms. Zenodo uses JSON schema as the internal representation of metadata and offers export to other formats such as Dublin Core, MARCXML, BibTeX, CSL, DataCite and export to Mendeley. The data record metadata will utilise the vocabularies applied by Zenodo. For certain terms, these refer to open, external vocabularies, e.g.: license (Open Definition), funders (FundRef) and grants (OpenAIRE). Reference to any external metadata is done with a resolvable URL.

5.4. Increase data re-use

Public data will be made available for re-use. To avoid any potential doubt, the consortium will attach specific licenses to the deposited data to define all conditions under which the work is provided under either open or restricted access.

To ensure an effective contribution to the EUSO, the consortium will identify in the DMPs which relevant data, maps and information might be of relevance to the EUSO and appoint a contact

Under review

person to participate in discussions on data management with the JRC's EUSO and the project SoilWISE.

Solicited stakeholders will be asked if they agree to their data being shared with other Horizon Europe funded soil-related projects. However, this sharing will only be done with project leaders who request it and who agree to respect the data management conditions proposed by URSOILL (or, in case of modifications, to contact the stakeholders for agreement).

According to the European Commission, "research data is information (particularly facts or numbers) collected to be examined and considered, and to serve as a basis for reasoning, discussion, or calculation."

Open access to research data is the right to access and reuse digital research data under the terms and conditions set out in Article 16 of the Grant Agreement. This section will be compiled during the project as datasets become available for URSOILL and will include information on:

- Licensing of data
- Availability of data and embargo period
- Re-use of data by third parties
- Data quality assurance processes
- Duration of data for re-use

Open Data and Open-Source Software

Data and software are owned by the beneficiary that generates them. However, owners of open results arising from the URSOILL project are encouraged to release their work under a Creative Commons license, preferably Creative Commons Attribution 4.0 (CC-BY-4.0).

Scientific Publications

Authors of scientific publications arising from the URSOILL project are encouraged to seek an agreement with the scientific publisher that allows the authors to:

- Retain the ownership of the copyright for their work
- Deposit the publication in an Open Access repository

Creative Commons Licensing will be used to protect the ownership of the datasets. An embargo period may be applied if the data (or parts of data) are used in published articles in "Green" open access journals. The recommended maximum embargo period length by the European Commission is 6 months. For datasets deposited in a public data repository (Zenodo), access is unlimited. Restrictions on re-use policy are applied for all protected data, whose re-use will be limited within the project partners.

Other restrictions could include:

- The "embargo" period imposed by journals' publication policy (Green Open Access)
- Some or all of the following restrictions may be applied with Creative Commons licensing of the dataset:
    - Attribution: Requires users of the dataset to give appropriate credit, provide a link to the license, and indicate if changes were made.
    - Non-commercial: Prohibits the use of the dataset for commercial purposes by others.

Under review

5.5. Other research outputs

In addition to the management of research data, URSOILL will ensure that all other research outputs generated or re-used throughout the project are handled in accordance with Horizon Europe expectations on responsible research output management. These outputs may include digital products (e.g., software, computational workflows, protocols, soil assessment models, monitoring scripts) and physical outputs (e.g., soil samples, substrates, amendments, prototypes, materials used for nature-based solutions, or other experimental artefacts). For each category of output, the project partners will assess which elements of the FAIR principles apply and will implement appropriate measures to maximise their value and long-term usability. This includes the application of structured metadata, persistent identifiers where relevant, and adequate documentation to support reproducibility and reuse.

Outputs will be stored, catalogued, and preserved in appropriate infrastructures, such as institutional repositories, OpenAIRE-compatible repositories, or physical storage facilities adhering to recognised curation standards. Access conditions will be defined based on ethical, legal, safety, and intellectual property considerations. By adopting these practices, URSOILL ensures that all research outputs—beyond datasets alone—are shared or made available in ways that support transparency, reproducibility, and sustained impact, fully aligned with Horizon Europe guidance on research output management.

under review

Funded by the European Union

6. Allocation of resources

6.1. Costs

There is no charge for using the Teams infrastructure provided by the central service. The project budget covers: (i) maintaining the repository, (ii) personnel costs, and (iii) developing, implementing, overseeing, and updating the DMP.

The costs associated with making data FAIR include:

- Fees for publishing scientific articles containing the project's research data in 'Gold' Open Access journals. Cost-sharing among multiple authors will be decided on a case-by-case basis.
- Project website operation: to be determined.
- Data archiving at Zenodo and other online databases (i. e. GitHub): free of charge.
- Copyright licensing with Creative Commons: free of charge.

Each partner is responsible for the data they produce. Any fees incurred for Open Access through the scientific publication of the data will be the responsibility of the data owner (authors) or partners.

Any unforeseen costs related to open access to research data in Horizon Europe (HE) are eligible for reimbursement during the project's duration, as defined in the Grant Agreement (Article 6). Additional details will be provided in future versions of the DMP as needed.

6.2. Data management responsibilities

All project datasets will be shared among Consortium partners via a Teams repository structure. Additionally, read-only access will be provided to the EC and external reviewers via direct link upon request. The COOT will regularly update the access list for personnel within partner teams to files and sensitive information.

ZABALA will manage the data repository in Teams, with responsibilities including:

- Defining, creating, and updating the repository structure
- Collaborating with project partners to create folders/sub-folders for each user group and document type (e.g., data, metadata, templates)
- Timely uploading of data/publications to the repository following validation by RISE
- Ensuring appropriate versioning, metadata, access, and dissemination levels
- Providing final approval for content uploads to the repository

Data management activities span the entire project and must be coordinated and monitored at both project and WP levels. Data management is also linked to the publication of project results and dissemination activities. Therefore, the following roles and responsibilities are identified:

RISE, as the Project Data Manager (WP8 leader), is responsible for:

- Developing the data management plan and policy in cooperation with the COOT and the data management team (see section 3).

Under review

- Coordinating the technical implementation in WPs (data survey, data repositories, metadata catalogues, etc.)
- Monitoring data management activities (both collection and publication) and deadlines, and sending reminders to WP data managers
- Providing support to WP data managers
- Creating the DMP and its periodic updates (D8.2, D8.3 and D8.4)
- Providing solutions for specific issues in accordance with project management
- Coordinating with the Dissemination Manager to assist in choosing the appropriate publication path (Green or Gold open access), offering customized help and guidance for publishing scientific publications, ensuring that the journal's open access policy complies with HE open data requirements before manuscript submission, monitoring that Green access (self-archiving) publications are deposited in repositories and sending reminders to partners, ensuring that research data related to a publication is made available in repositories and linked to the respective publication, monitoring possible embargo periods and sending reminders to partners, and ensuring that publications are available.

The Work Package Data Managers (see section 3) are responsible for:

- Implementing the data management policy in their respective WPs/tasks
- Monitoring data management activities and deadlines, and sending reminders to partners
- Offering customized help and guidance for filling out the WP data surveys
- Requesting missing information or clarifications from partners
- Providing input to the DMP by analysing and summarizing WP-specific data surveys
- Offering customized help and guidance for publishing open data and open-source software
- Monitoring that open results (data and software) are deposited in the default repository and sending reminders to partners
- Contacting the project manager and Ethics Advisor in case of questions and ethical or privacy issues that may prevent data publication

under review

7. Data security

7.1. Data storage and back-up

The Teams Channel service is administered by the ZABALA IT group, which follows the data security standards of the National Cybersecurity Institute of Spain (INCIBE), including maintaining regular backups of data, with copies stored in multiple locations. To ensure the storage of project data, especially of long-term data, the uploaded data on the Sharepoint in Teams will be back-up at ZABALA IT-infrastructure resources. The ZABALA's resources have daily weekly, and monthly back-up schedules and IT group at ZABALA, which has experiences in data issues, will be responsible for data storage and back-up. Team ZABALA IT member, Gaizka Reta, will be responsible for backup and storage of URSOILL data in ZABALA's infrastructure.

Individual datasets will be simultaneously stored in partners' local IT- infrastructure and updating information of local storages will be performed by responsible manager(s) in each WP/task. Partners are responsible for ensuring data security and backing up datasets during data collection and processing. Typically, a local backup system will be maintained throughout the project's duration. Once data is uploaded to SharePoint, managed by ZABALA, it will be backed up and stored on internal URSOILL servers using state-of-the-art data protection software and procedures.

Regarding the predictive tool, the backend will include a secure REST service that serves data to be consumed by the front-end. The REST backend services will be developed using Python 3 and FastAPI. The REST API will have several endpoints to serve all the indicators and models developed at that time. The project cloud will be used in the storage layer, which allows BigData capabilities and ca be scaled to larger areas if necessary. In addition, a ETL processes will be generated to update the database information regularly deploying the processes using Apache Airflow. The interactive visualisations will be built using Apache eCharts. AGD will create the web application and dataset architecture.

In Open Call management, WP4 will collect and process sensitive and personal data as part of the contracts to be signed with the sub-grant beneficiaries and evaluators. These contracts will include a clause on data protection to ensure compliance with relevant data privacy regulations.

7.2. Data security within Consortium

7.2.1. Deployment of a threefold security protection strategy

The process of collecting, processing, and storing data may present some risks. To mitigate the potential for malevolent, criminal, and/or terrorist abuse, including by malicious individuals authorized to access the information, the URSOILL Consortium is regularly examining the deployment of a threefold security protection strategy throughout the project lifespan:

1. Ensuring that the employed security layers and privacy-preserving measures function correctly, maintaining access logs, and following best practices for system administration.
2. Employing techniques to prevent information leakage 'on-the-fly,' such as anonymization and pseudonymization of personal and sensitive information during collection, communication, and storage (e.g., through encryption schemes, hash functions, and/or tokenization). This

Under review

approach will neutralize eavesdropping and similar hacking attempts, making the data completely meaningless to potential attackers in the unlikely event of successful retrieval.

3. Implementing good practices in the unlikely event of a data breach, including notifying the relevant data protection authorities and proceeding according to the GDPR.

7.2.2. Authentication, authorization and encryption

According to the GDPR, the implementation of both computerized authentication and procedures for managing authorization credentials is required. To ensure the security and trust in the system, and to properly protect the rights and freedoms of natural persons, it is essential to provide technical solutions to ensure data security regarding the services offered by URSOILL. For identity management and data protection mechanisms, URSOILL will follow standard practices in the security research community.

Identity management involves identifying individuals (authentication) and controlling access (authorization) to resources in a system. All Privacy Enhancing Technologies associated with identity management aim at identity verification with minimal identity disclosure and protection against identity theft. Due to internetworked services and Cloud technology, the need for secure identity management has grown increasingly. Identity and access management (IAM) is the security and business discipline that "enables the right individuals to access the right resources at the right times and for the right reasons." It addresses the need to ensure appropriate access to resources across increasingly heterogeneous technology environments and to meet increasingly rigorous compliance requirements. Technologies, services, and terms related to identity management will be analysed by the consortium and applied, if applicable, for the external services and platforms used during project implementation.

More specifically, following the "Privacy by default and by design" principles included in the GDPR, the URSOILL website will adopt an integrated and multilevel approach to protect user information from fraudulent access and consumption.

To ensure the privacy and security of datasets generated by the URSOILL project that will remain private, GPG (GNU Privacy Guard) will be used. GPG allows encrypting files using public-key cryptography. GPG uses the user's private key for signing and encrypting files that can later be decrypted using the user's public key. In the URSOILL project, GPG's symmetric encryption algorithm will be used, allowing files to be encrypted using a password known by all consortium members authorized to access those files. Additionally, these encrypted files will be stored in a private section of the Teams Channel, granting access only to the appropriate consortium members.

7.2.3. Focus on data aggregation and pseudonymization techniques

Personal and sensitive data will be made publicly available only after informed consent has been obtained and suitable anonymization techniques have been applied. Before starting project activities, a thorough investigation of privacy and security issues has been and will be conducted, particularly focusing on Swedish, Spanish, Greek, Italian and Luxembourgish privacy laws, as these are the geographical locations for obtaining data from LLs.

Regarding anonymization techniques, data confidentiality, integrity, and privacy will be ensured when collecting and processing data. The information for each person in the release cannot be distinguished from that of a certain number of other individuals whose information also appears in the release. Additionally, pseudonymization of data is another method of ensuring

Under Review

confidentiality, according to the Article 29 Working Party Opinion on Anonymization Techniques and the EU General Data Protection Regulation. When data are particularly sensitive (e.g., detailed personal narratives), the risks to confidentiality increase. In such cases, participants will be carefully informed of the potential risks. This does not absolve the applicant from ensuring that maximal anonymization procedures are implemented. A detailed description of the measures to prevent improper use, improper data disclosure scenarios, and 'mission creep' (i.e., unforeseen usage of data by any third party) within the aforementioned security protection strategy will be provided as an update to this deliverable.

The optimal solution will be determined using a combination of different techniques, considering the practical recommendations developed in the Article 29 Working Party Opinion on Anonymization Techniques. It should be noted that, although pseudonymization reduces the likability of a dataset with the original identity of a data subject and is a useful security measure, it does not qualify as an anonymization technique according to both the Article 29 Working Party Opinion on Anonymization Techniques and the GDPR, as it enables data to re-identify the data subject to which it refers. Therefore, pseudonymized data is considered personal data, and its processing must fully comply with the GDPR.

These techniques should adhere to certain requirements to comply with data protection and privacy-related legislation in the EU. The following set of requirements (among others) has been extracted from the GDPR and the Article 29 Working Party Opinion on Anonymization Techniques and will guide the security protection strategy drafting:

- User authentication: The system must provide adequate mechanisms for user authentication.
- Limited access: The system must ensure that data is only provided to authenticated and authorized persons. The list of authorized persons for each dataset and their user privileges shall be restricted to the minimum necessary.
- Protection against unauthorized and authorized access: The records of an individual must be protected against unauthorized access.
- Notice about the use of data: Users should be informed about any access to their records.
- Access and copy users' own data: The system must provide mechanisms to access and copy users' own data.
- Modification of the database: If an attacker breaches the system, the system must detect modifications and inform the system administrator about the attack.
- Data protection by design and by default: Privacy and data protection standards shall be considered from the outset of the project. Therefore, taking into account the state of the art, the cost of implementation, and the nature, scope, context, and purposes of processing, as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of determining the means for processing and at the time of processing itself, implement appropriate technical and organizational measures designed to implement data-protection principles effectively and integrate the necessary safeguards into the processing to meet the requirements of the GDPR. Additionally, the controller shall implement appropriate technical and organizational measures to ensure that, by

Under review

default, only personal data necessary for each specific purpose of the processing are processed.

7.2.4. Internal threats and human errors

Most organizations focus on managing data risks from external threats, but the majority of breaches stem from internal vulnerabilities. These vulnerabilities can be seen as part of the same risk spectrum. This section addresses internal vulnerabilities and strategies to mitigate them. There are two primary types of internal threats:

- Human Error: Security can be compromised by mistakes, such as an employee copying information from an entire database table into an email for troubleshooting and accidentally including external email addresses in the recipient list.
- Internal Attacks: While accidental internal breaches are common, deliberate attacks by insiders also account for a significant portion of database breaches. These are often disgruntled employees who use their privileged access to cause harm.

Many of these attacks exploit the various data outlets on modern PCs, including USB and Firewire ports, CD and DVD recorders, and built-in storage media slots. With the rapid increase in storage capacity on portable devices, business professionals can now use personal storage devices, such as USB memory sticks, digital cameras, and smartphones, to remove or copy sensitive information for malicious purposes or personal gain.

7.2.5. Internal threat prevention

The implementation of a robust and adaptable security policy is crucial for URSOILL. Such a policy can establish clear rules and permissions that are easily understood by both employees of URSOILL' partner organizations and those responsible for enforcement, ensuring that personal data remains within the office. URSOILL' policy is grounded in the EU's security regulations, which, when properly enforced, are typically sufficient to prevent breaches. This policy is summarized in the following five-point methodology:

Table 2. Internal threats prevention methodology

No.	Prevention Method	Implementation	Beneficiaries, Timeline
1	Data protection policies	Use national or local legal guidelines for data protection and privacy policies (DP)	Consortium and disseminated electronic documents (Teams channel). Within first 6 months
2	Internal data protection policies	Written policies and procedures for all staff to sign in and agreed	Consortium and new coming employees. Within first 6 months
3	Clear staff roles definition and responsibilities	Staff training, awareness and clear roles and staff responsibilities on data for access to data with checklists	Consortium. Regular monthly organization trainings or integrated with internal meetings
4	Access control	Manage change in staff and have leave processes in place	Project coordinator, Organizations

Under review

			leaders. When there is a change
5	Sanctions and audits	Disciplinary actions for breach of DP and process guidelines by staff and threat of audits	Consortium. When a violation occurs

7.3. Data security as specified for Zenodo

Open results deposited in the Zenodo repository are safely stored for long-term preservation. The following list describes the security settings for Zenodo:

- Versions: Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and records are preserved.
- Replicas: All data files are stored in the CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- Retention period: Items will be retained for the lifetime of the repository. The host laboratory of Zenodo, CERN, has defined a lifetime for the repository of the next 20 years minimum.
- Functional preservation: Zenodo makes no promises of usability and understandability of deposited objects over time.
- File preservation: Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- Fixity and authenticity: All data files are stored along with an MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.
- Succession plans: In case of closure of the repository, a guarantee has been made from Zenodo to migrate all content to suitable alternative institutional and/or subject-based repositories.

under review

8. Ethics

The project consortium unanimously agrees that protecting personal data is a top priority. All partnership activities, especially those involving stakeholder engagement, dissemination, and communication, will adhere to the EU's GDPR. Additionally, when publishing data, all scientific ethical principles will be observed. Partners will be guided on anonymizing personal research data before making it openly available, thus complying with both open research data and data protection rules. They will also follow the dissemination rules outlined in the Consortium Agreement (Sec. 8.4). However, complete anonymization is not always feasible, particularly with raw data like recorded interviews. Partners will obtain informed consent to disseminate data in public reports, communications, etc., and for long-term storage. Nevertheless, informed consent cannot legitimize the use of data in an open access environment when future data use purposes are unknown. In such cases, data will remain confidential. The appointed URSOILL Ethics Advisor will review the main data collections, including personal data from the project, and will be responsible for reviewing and approving the methods used to anonymize or de-identify this data.

under review

9. Other issues

URSOILL will not be making use of other national, sectorial or departmental procedures for data management.

under review

Funded by the European Union