



# icosHELLs

## D8.2 DATA MANAGEMENT PLAN

**Cheryl Cordeiro** – Senior Scientist

**Erik Sindhøj** – Senior Scientist

**RISE Research Institutes of Sweden**



Funded by  
the European Union

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 the granting authority can be held responsible for them.

## Project Information

<b>Project name</b>	iCOSHELLS
<b>Full project name</b>	INNOVATIVE CO-CREATION SOIL HEALTH LIVING LABs
<b>Grant number</b>	101157394
<b>Project coordinator</b>	RISE Research Institutes of Sweden
<b>Project duration</b>	1 September 2024 - 31 August 2028

## Document Information

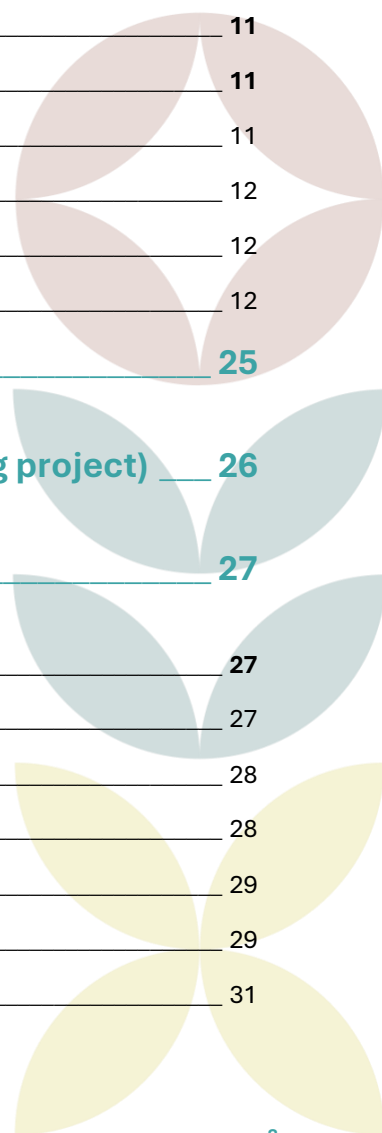
<b>Deliverable name and number</b>	<b>D8.2 Data Management Plan</b>
<b>Due date</b>	14/03/2025
<b>Actual submission date</b>	06/03/2025
<b>Contributing partners</b>	RISE, (input from Zabala, CSCP, CETENMA, IFAU, ATB, G!E, GAIA, ISINNOVA, Clube, AUP)
<b>Reviewer partners</b>	UPM
	<b>Deliverable type</b>
<b>R</b>	Document, report
	<b>Dissemination level</b>
<b>PU</b>	Public

Version	Date	Author	Comments
1	20/02/2025	RISE / Cheryl Cordeiro	
2	07/03/2025	UPM/Ernesto Sanz	

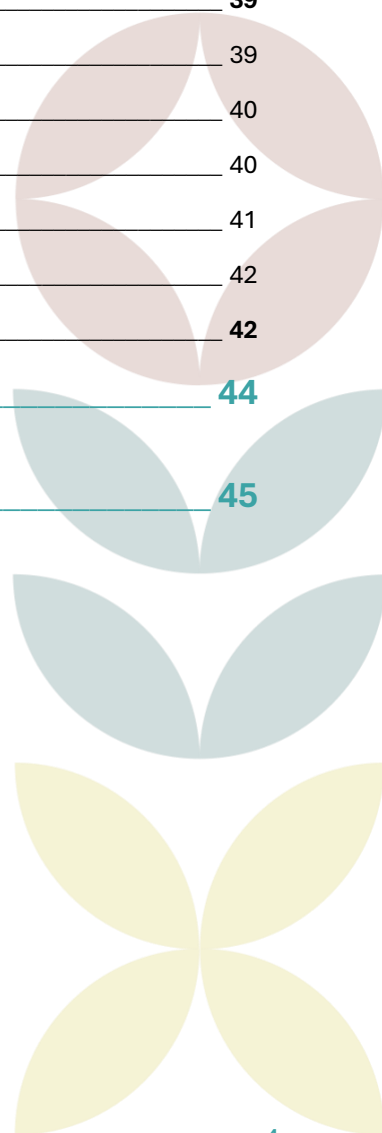


# Table of contents

<b>About iCOSHELLS</b>	<b>6</b>
<b>List of tables</b>	<b>7</b>
<b>List of abbreviations</b>	<b>8</b>
<b>1. Introduction</b>	<b>10</b>
<b>Purpose</b>	<b>10</b>
<b>2. Data Summary</b>	<b>11</b>
<b>Purpose of data collection/generation</b>	<b>11</b>
<b>Data identification and description</b>	<b>11</b>
Types of formats of data	11
Size of data and expected volume	12
Data utility	12
Origin of data	12
<b>3. Data managers</b>	<b>25</b>
<b>4. Data organization and exchange (within consortium, during project)</b>	<b>26</b>
<b>5. FAIR data management</b>	<b>27</b>
<b>Making data findable, including provisions for metadata</b>	<b>27</b>
Dataset reference, name and versions	27
Naming conventions	28
Standards and metadata	28
DMP review process and data inventory	29
Archiving, preservation and data quality assurance	29
Publications and keywords	31



Digital Object Identifiers _____	31
<b>Making data accessible _____</b>	<b>32</b>
Open Access _____	32
Overview of access to data _____	33
<b>Making data interoperable _____</b>	<b>34</b>
<b>Increase data re-use _____</b>	<b>35</b>
<b>Other research outputs _____</b>	<b>36</b>
<b>6. Allocation of resources _____</b>	<b>37</b>
<b>Costs _____</b>	<b>37</b>
<b>Data management responsibilities _____</b>	<b>37</b>
<b>7. Data security _____</b>	<b>39</b>
<b>Data storage and back-up _____</b>	<b>39</b>
<b>Data security within Consortium _____</b>	<b>39</b>
Deployment of a threefold security protection strategy _____	39
Authentication, authorization and encryption _____	40
Focus on data aggregation and pseudonymization techniques _____	40
Internal threats and human errors _____	41
Internal threat prevention _____	42
<b>Data security as specified for Zenodo _____</b>	<b>42</b>
<b>8. Ethics _____</b>	<b>44</b>
<b>9. Other issues _____</b>	<b>45</b>



## Disclaimer

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. Neither the European Union nor the granting authority can be held responsible for them.



# About iCOSHELLS

Soil is essential for life on Earth, yet 60%-70% of EU soils are unhealthy due to pollution, urbanisation, and intensive agriculture – issues made worse by climate change. This soil degradation leads to significant economic, social, and environmental challenges, including reduced land productivity and biodiversity loss.

The iCOSHELLS project supports the EU Mission '[A Soil Deal for Europe](#),' aiming to restore healthy soils by 2030. Specifically, iCOSHELLS focuses on three key objectives: **reducing soil pollution and promoting restoration, improving soil structure and biodiversity, and increasing soil literacy among society.**

To achieve these goals, iCOSHELLS leverages **six Living Labs** located in the **Basque Country, Bulgaria, Greece, Italy, Spain, and Sweden**. These LLs bring together diverse local stakeholders to co-design and test practical strategies for soil health improvement.

The project employs a systematic approach that strengthens stakeholder capacities, bridges scientific research with practical solutions, enhances understanding of soil indicators, and replicates effective recovery methods. Its ultimate purpose is to develop and validate scalable solutions that can be applied across Europe.



# List of tables

Table 1: Data identification and description. .... 13



# List of abbreviations

Acronym	Meaning
<b>ANOVA</b>	Analysis of Variance
<b>API</b>	Application Programming Interface
<b>CA</b>	Consortium Agreement
<b>CC-BY-4.0</b>	Creative Commons Attribution 4.0 License
<b>COOT</b>	Coordination Team
<b>CSV</b>	Comma-Separated Values (File Format)
<b>D&amp;C</b>	Dissemination and Communication
<b>DMRT</b>	Duncan Multiple Range Test
<b>DMP</b>	Data Management Plan
<b>DMT</b>	Data Management Team
<b>DOI</b>	Digital Object Identifier
<b>EC</b>	European Commission
<b>ESDAC</b>	European Soil Data Centre
<b>ETL</b>	Extract, Transform, Load (Data Processing)
<b>EU</b>	European Union
<b>EUSBSR</b>	EU Strategy for the Baltic Sea Region
<b>EUSO</b>	EU Soil Observatory
<b>FAIR</b>	Findable, Accessible, Interoperable, and Reusable
<b>FTP</b>	Funding and Tenders Portal
<b>GA</b>	Grant Agreement
<b>GDPR</b>	General Data Protection Regulation
<b>GMS</b>	Grant Management Services
<b>HE</b>	Horizon Europe
<b>IAM</b>	Identity and Access Management
<b>IP</b>	Intellectual Property
<b>IPR</b>	Intellectual Property Rights
<b>JSON</b>	JavaScript Object Notation
<b>LL</b>	Living Lab
<b>MongoDB</b>	NoSQL Database System
<b>NCA</b>	Necessary Condition Analysis
<b>NoSQL</b>	Non-Structured Query Language (Database)
<b>OA</b>	Open Access
<b>ORDP</b>	Open Research Data Pilot
<b>PC</b>	Project Coordinator
<b>PTC</b>	Project Technical Committee
<b>QCA</b>	Qualitative Comparative Analysis

<b>REST</b>	Representational State Transfer (API architecture)
<b>RP</b>	Reporting Period
<b>WP</b>	Work Package
<b>Zenodo</b>	Open-Access Repository (used in Horizon Europe)



# 1. Introduction

## 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 iCOSHells project consortium will follow. The DMP guidelines for iCOSHells 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 iCOSHells project, the 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*<sup>1</sup>. The development of the DMP will enable iCOSHells' partners to address all data protection issues, including ethical concerns and security strategies. iCOSHells participates in the Open Research Data Pilot (ORDP), which aims to enhance and maximize access to, and reuse of research data generated by HE projects, such as the data produced by the iCOSHells platform during its deployment and validation. 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 iCOSHells 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 iCOSHells project progresses. This report outlines the procedures for data collection, storage, and processing, and provides an overview of iCOSHells' 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, including ZABALA, CSCP, CETENMA, UPM, IFAU, ATB, and GIE. 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 iCOSHells project.

---

<sup>1</sup> [Reference Documents | EU Funding & Tenders Portal](#)

## 2. Data Summary

### Purpose of data collection/generation

The main objective of iCOSHELLS is to enhance soil health across the EU by utilising six Living Labs situated in different European regions (Sweden, Southeast Spain, Basque Country, Italy, Greece, Bulgaria), integrating a wide array of local stakeholders to co-design and test effective strategies to improve soil health. In this perspective, iCOSHELLS will generate technical data, after collecting 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 examination, mapping, and strategic planning for the different iCOSHELLS' LLs in their co-creation processes.
- Perform an extensive analysis of existing soil health indicators and the promotion of sustainable soil management practices.
- Definition of the Open Call, evaluation and selection of projects (and iCOSHELLS' 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 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 on the impact and success of LL activities.
- 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.

### Data identification and description

#### Types of formats of data

Data will be organized into datasets based on the category of the data and the site of collection. 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.

iCOSHELLS is set to incorporate external data like soil indicators, biodiversity metrics, and management practices. Additionally, data from stakeholder engagement will be meticulously stored to ensure participant confidentiality.

All project data will be treated with meticulous 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' data. By default, in compliance with GDPR (General Data Protection Regulation, EU GDPR 2016/679), all project data will be handled with the utmost care.

iCOSHELLS 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 iCOSHELLS data are outlined in Table 1.

### 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 iCOSHELLS 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 Table 1.

### Data utility

iCOSHELLS' collected data will be suitable for use by:

- iCOSHELLS consortium, with more details of linkages between WPs in Table 1
- Stakeholders involved in the project (academia, researchers, policy makers, end-users such as farmers, education sector, citizens and other actors)
- Scientific community
- European Commission services and European Agencies
- Related initiatives, such as LIVINGSOILL, SOILCRATES, LILAS4SOILS or GOV4ALL.
- Living Labs platforms, such as SOILL
- General audience

### Origin of data

iCOSHELLS 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 Table 1. 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 iCOSHELLS partners, upon request
- Stakeholders network (interviews with groups and individuals, surveys, workshops, etc.)

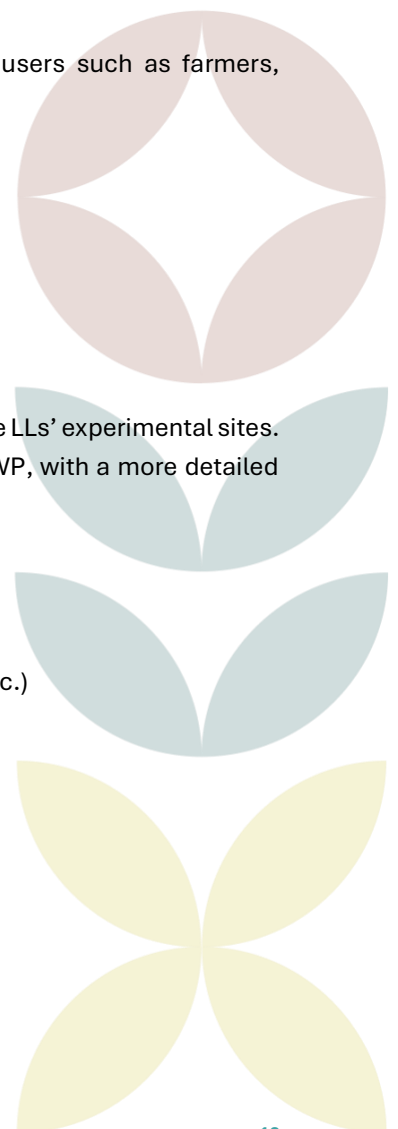


Table 1: Data identification and description.

<p><b>WP1 (CSCP): iCOSHELLs Living Labs Co-creation activities Dataset</b></p>	<b>Contributor(s)</b>	ATB, IFAU, RISE, GAIA, LLs leaders
	<b>Ownership</b>	ATB, IFAU, RISE, GAIA, LLs leaders
	<b>Description</b>	Descriptive dataset that integrates quantitative metrics and qualitative insights to support stakeholder analysis, market evaluation, and strategic planning for the different iCOSHELLs' LLs
	<b>Type</b>	Observational
	<b>Format</b>	Structured (e.g., *.csv, *.xlsx) and unstructured (e.g., *.txt) formats
	<b>Size</b>	Less than 50 GB
	<b>Provenance</b>	New data
	<b>Methodology for collection/generation, process and analysis</b>	<p><b>Quantitative data</b> will be obtained through structured surveys, databases, and reports to gather metrics such as stakeholder counts, relationship strength, power and influence scores, engagement levels, and indicators related to economics, market dynamics, technological adoption, and environmental impacts.</p> <p><b>Qualitative data</b> will be collected through interviews, focus groups, and surveys to provide contextual insights into stakeholder profiles, operational practices, challenges, opportunities, motivations, relationship dynamics, policy frameworks, trends, regulatory environments, and social factors.</p> <p><b>Analytical methods</b> such as statistical analysis, network mapping, and thematic coding will be applied to integrate the datasets and enable a comprehensive understanding of stakeholder networks and market conditions, ensuring both quantitative rigor and qualitative depth for informed decision-making.</p>
	<b>Storage</b>	Identity revealing information that has been collected in a digital manner will be stored on the local partners' internal cloud servers in folders that are accessible to limited project team members and protected by high security passwords.
	<b>Link with other WPs</b>	The data generated will support the design and operations of the iCOSHELLs (Living Lab), stakeholder engagement strategies, and tailored approaches for different regional contexts while refining project aims. It will identify and recruit key stakeholders, map synergies, and foster collaborations, providing insights into baseline regional conditions, local challenges, policy frameworks, and market dynamics. This data will inform activities across work packages (WP2, WP4, WP5-7), guide further engagement strategies, and support decision-making. Long-term, it will support to measure iCOSHELLs solutions' impacts, scale

		successful approaches, inform policy recommendations, and guide sustainable business model development.
	<b>Applicability</b>	It can be valuable to soil stakeholders, including policymakers and governance bodies, who can leverage it to inform decision-making and regulatory frameworks. Researchers and academics may use the data to advance scientific understanding and support further studies, while civil society organizations can apply the insights to advocate for sustainable practices and engage communities in soil-related initiatives.
	<b>Metadata</b>	Analytical and procedural information; title of dataset, author of dataset, license of dataset

<b>WP2 (RISE): LLs implementation dataset</b>	<b>Contributor(s)</b>	LLs leaders, ATB, IMIDA, SLU, UPM, CAJAMAR, EUSKAMPUS, FORUA, UTH
	<b>Ownership</b>	RISE and LLs partners
	<b>Description</b>	<p>WP2 is responsible for the management and coordination of operationalization processes across LLs, consolidating and structuring data collected by the LLs for broader project integration:</p> <ul style="list-style-type: none"> <li>- Operational &amp; Implementation Data: Documentation of LL setup, governance structures, and execution across regions.</li> <li>- Stakeholder &amp; Process Documentation: Information on partnerships, governance models, and strategies for local engagement.</li> <li>- Aggregated Reports: Compilation of monitoring, validation, and iterative learning insights from LLs into structured deliverables supporting WP3 (evaluation), WP4 (funding), and WP5 (scaling).</li> </ul>
	<b>Type</b>	<p>Compiled and administrative data, categorized as:</p> <ul style="list-style-type: none"> <li>- Derived/compiled data: Aggregated reports from LL monitoring and implementation processes.</li> <li>- Administrative/managerial data: Project governance documents, LL partnership agreements, and execution strategies.</li> <li>- Survey and stakeholder data: Co-creation session outputs, policy engagement records, and knowledge exchange reports</li> </ul>
	<b>Format</b>	Text-based reports & analysis (*.pdf, *.docx); Survey and stakeholder records (*.xlsx, *.csv); Environmental & observational data (in accordance with formats described by soil LL leads)



<b>Size</b>	Project documentation and reports: 50–200 GB; Stakeholder and governance data: 10–50 GB; Compiled LL monitoring data: Dependent on contributions from LL leads
<b>Provenance</b>	<p><b>New Data:</b> from operationalization frameworks, stakeholder engagement records, and project-level implementation reports.</p> <p><b>Re-used Data:</b> Consolidation and structurization of data generated by the LLs but does not independently collect scientific field data.</p>
<b>Methodology for collection/generation, process and analysis</b>	<p>WP2 functions as a data aggregator rather than a primary collector of environmental data.</p> <ul style="list-style-type: none"> <li>- LL leads provide data: Each LL generates site-specific datasets based on field activities and monitoring.</li> <li>- WP2 aggregates and structures data: WP2 compiles progress updates, governance models, and implementation reports to support other WPs.</li> <li>- Stakeholder engagement reports: WP2 documents interactions with policymakers, industry representatives, and regional stakeholders.</li> </ul>
<b>Storage</b>	<p>Primary Storage: WP2’s outputs, including implementation reports and governance models, will be stored in secure project repositories.</p> <p>Access Control:</p> <ul style="list-style-type: none"> <li>- Public Data: Aggregated reports and policy recommendations.</li> <li>- Restricted Data: Stakeholder engagement records, governance documentation with sensitive project details</li> </ul>
<b>Link with other WPs</b>	<p>WP3 (Monitoring and Evaluation): Provides operational data for assessing LL performance; WP4 (LL Expansion and Funding Calls): Contributes to funding assessment by documenting LL readiness; WP5 (Scaling and Replication): Informs replication strategies by sharing best practices from LL implementation.</p>
<b>Applicability</b>	<p>Policymakers developing regional or EU-wide soil health strategies, Research institutions studying Living Lab methodologies and Industry stakeholders involved in sustainable soil management innovations.</p>
<b>Metadata</b>	<p>Since WP2 primarily consolidates and manages data from Living Labs (LLs), its metadata will focus on describing implementation reports, governance documents, and stakeholder engagement records. The metadata structure aligns with FAIR data principles to ensure findability, accessibility, interoperability, and reusability.</p>



<b>WP3 (CETENMA): Soil monitoring database</b>	<b>Contributor(s)</b>	CETENMA, UPM, LLs leaders
	<b>Ownership</b>	CETENMA, UPM, LLs leaders
	<b>Description</b>	This dataset will contain all the information gathered from the experimental sites across all the LLs: values of the monitoring indicators, edaphoclimatic information, satellite images, etc.
	<b>Type</b>	Observational and experimental
	<b>Format</b>	Depending on the indicator to measure: *.jpeg, *.png for images; *.xlsx for laboratories results; *.txt for field reports; Geolocated point for soil data and algorithms.
	<b>Size</b>	~1 GB per LL
	<b>Provenance</b>	Both new (i.e. that provided by the measured indicators) and re-used (i.e. soil classification, previous management...)
	<b>Methodology for collection/generation, process and analysis</b>	The methodology for data collection is provided in D 3.1. Analysis of the collected data will be carried out using appropriate statistical methods and modelling techniques following a data science approach such as: Levene's test, Shapiro-Wilk test, ANOVA and Duncan Multiple Range Test (DMRT).
	<b>Storage</b>	Data, once obtained, will be uploaded in the web application which is going to be developed specifically for this project. This application will be developed with a visual display tool of the data to provide information at-a-glance (dashboards). The NoSQL (Non-Structured Query Language) database MongoDB (NoSQL Database System) will be used in the storage layer, which allows BigData capabilities and can be scaled to larger areas if necessary. In addition, an ETL (Extract, Transform, Load) 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.
	<b>Link with other WPs</b>	The purposes of the data generated is to assess the best performing solution and feed the database for prediction purposes (T3.3). This database will then be the starting point to model the relationship between the site-specific soil management and the soil quality parameters (predictors) and to assess the predictive efficiency of the developed models.
	<b>Applicability</b>	The EC (European Commission), other projects, solution testers, farmers, etc., might be interested in using this data.
<b>Metadata</b>	Defined in D3.3.	



<b>WP3 (UPM): Predictive tool</b>	<b>Contributor(s)</b>	CETENMA, UPM, LLs leaders
	<b>Ownership</b>	CETENMA, UPM, LLs leaders
	<b>Description</b>	This dataset features an online predictive tool using all the data collected from Living Lab that will allow potential replicators to predict the improvement that a specific solution could have under their specific conditions.
	<b>Type</b>	Simulation/derived/compiled.
	<b>Format</b>	Multivariate model-type format (e.g. multiple linear regression, random forest...).
	<b>Size</b>	It is difficult to estimate the volume of data at this stage. What we can say at this stage is that it is anticipated to be large, basically because models are expected to be multivariate and complex. This means that the aggregate volume of data requirement might be in the scale of TB.
	<b>Provenance</b>	Both new and re-used data.
	<b>Methodology for collection/generation, process and analysis</b>	<p>Data collected will be used together with, potentially, new data from outside the LLs (e.g. other living labs and EUSO, European soil observatory) to improve the predictive model. For modelling purposes, Multiple linear regression (MLR) and random forest (RF) analysis will be employed to predict soil health indicators of chrono-sequence sites in terms of the relationship between soil health and its controlling factors in different study sites. Additionally, these models will also be fed by site-generated data from other LLs or replicators.</p> <p>To ensure the robustness of machine learning models, researchers and practitioners will apply different strategies of data preprocessing, regularization through elastic net (combination of L1 and L2 techniques), cross validation, ensemble method and data augmentation. This will ensure that models perform well in diverse, real-world scenarios and can be trusted to make reliable predictions. In addition, we will conduct a rigorous feature selection step to understand the importance of each factor and reject noisy features. In this respect, we will use permutation importance techniques and Shapley values.</p>
	<b>Storage</b>	As described for the previous database.
	<b>Link with other WPs</b>	The main objective of the data generated is for replication purposes in order to evaluate the performance of solutions to improve soil health (WP2, WP6)
<b>Applicability</b>	The EC, other projects, Living Labs, solution testers, farmers...	

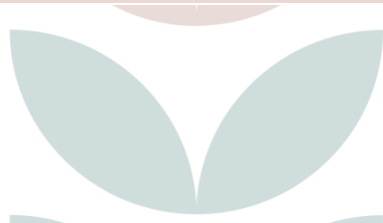


	<b>Metadata</b>	Those describing each of the models (e.g. samples used, multivariate model chosen, etc.)
<b>WP4 (ZABALA): Open Call dataset</b>	<b>Contributor(s)</b>	ZABALA, WPs leaders, LLs leaders
	<b>Ownership</b>	ZABALA and RISE
	<b>Description</b>	<ul style="list-style-type: none"> <li>- Call documentation (guidelines for applicants, template for sub-grantee agreement, application form, Q&amp;A section...)</li> <li>- Analytics on the submitted proposals</li> <li>- Selection of nominated new stakeholders in the LLs</li> </ul> <p>The contents of such dataset will be the result of aggregating (taking measures to prevent identification of personal data) participants' attributes, e.g. country of origin, challenge targeted and so on, for statistical purposes and analysis of the performance of the Open Call. A spreadsheet with the information gathered as result of proposals' submissions will be generated. The contents will be generated exclusively in English.</p>
	<b>Type</b>	<p>Observational data: The database of the applicants to the Open Call will collect:</p> <ul style="list-style-type: none"> <li>- Administrative data: Applicants information (Legal name, VAT number, address of the organization or the coordinator, webpage, contact person's name and surname, email address, phone number and company figures if the case).</li> <li>- Technical approach: Description of soil restoration solution, innovation and technical risks assessment.</li> <li>- Team: Details about profiles and dedication to the project of each member.</li> <li>- Project Planning and Value for Money: The activities and steps to achieve the expected results and the timing and resources for completion during the programme.</li> </ul>
	<b>Format</b>	Mainly survey responses and will be mainly stored in CSV format.
	<b>Size</b>	1-2 MB
	<b>Provenance</b>	New Data. There will be two points in time when information about participants in the iCOSHELLS Open Calls will be collected, namely: a) proposal submission form; b) tracking of sub-grantees progress within the implementation process.
	<b>Methodology for collection/generation, process and analysis</b>	All data will be aggregated or anonymized to not compromise personal details of applicants nor any other confidential information about their projects.



	<b>Storage</b>	All the information related to Open Calls will be stored in an encrypted file hosted in a SharePoint space to which only iCOSHELLS consortium members and external experts, as Data Processors, will have limited access. Besides, the private key to be able to read the contents will be shared by secure channels among consortium members.
	<b>Link with other WPs</b>	Data collected will be useful for consortium members (WP2,3) and External Evaluators in order to perform the evaluation of the submitted proposals and decide whether they should be selected in WP4 to participate in the iCOSHELLS programme or not. If they are selected, they will receive technical and financial support from other WP participants. This data will also serve to generate statistics about the popularity of the different services offered by the project, origin of the applicants, modalities of submissions, etc.
	<b>Applicability</b>	Only anonymized or aggregated data from the Open Call participation statistics will be made public through iCOSHELLS website. The entire metadata will only be used for research purposes: in the case of the releasing some scientific publication from the running of Open Calls, information will be completely anonymised and/or aggregated first.
	<b>Metadata</b>	<p><b>Descriptive:</b> submission title, applicant name(s), organization or affiliation, relevant keywords and tags, abstract or summary of the proposal.</p> <p><b>Structural:</b> submission components (documents, images, or datasets, versioning data to track drafts and revisions).</p> <p><b>Administrative:</b> submission ID, date and time, status (e.g., pending, under review, accepted, rejected), reviewer assignments, and the decision date.</p> <p><b>Technical:</b> file format and size, software used (if relevant), access logs detailing, and submission platform details.</p> <p><b>Provenance:</b> review feedback and scores from evaluators, decision rationale outlining why a submission was accepted or rejected, and communication logs recording emails or notifications sent to applicants</p>
<b>WP4 (ZABALA): Sub-grantees dataset</b>	<b>Contributor(s)</b>	ZABALA, WPs leaders, LLs leaders
	<b>Ownership</b>	ZABALA and RISE
	<b>Description</b>	<p>This dataset will be composed of:</p> <ul style="list-style-type: none"> <li>- iCOSHELLS’ participants portfolio: list of applicants selected, and information reported to justify the subgrantees.</li> <li>- Payments: information of the distributed funding to third parties</li> </ul>

	<p>The contents of such dataset will be the result of aggregating sub-grantees' attributes, e.g. country of origin, stages in which they participate and so on, for statistical purposes and analysis of the performance of the participation process.</p> <p>A database with the information gathered as result of the execution of the solutions proposed by each awarded third party will be generated. The contents will be generated exclusively in English.</p>
<b>Type</b>	Derived / Compiled from the sub-call dataset
<b>Format</b>	Reports mainly stored in CSV format.
<b>Size</b>	1-2 MB
<b>Provenance</b>	New data collected from the awarded applicants of the Open Call
<b>Methodology for collection/generation, process and analysis</b>	<p>After the conclusion of the Open Call, the shared spreadsheet will be filled in only with information about those participants that qualify to start the following phases of the program. Only for those teams progressing to the following stages, their details in the sub-grantees' dataset will be updated. Double checking of the fields describing a sub-grantee will be carried out to ensure the accuracy of the information gathered.</p> <p>As regards data capture methods, there is one single way of creating content about sub-grantees, i.e. to extract metadata from the signed sub-grant agreements, according with template, and metadata from evaluation forms together with letters for communication of results.</p>
<b>Storage</b>	<p>All the information related to sub-grantees is stored in the sub-grantees internal database. RISE and ZABALA, as controllers of this information, ensures to handle security and privacy issues. This information will be stored in an encrypted spreadsheet file stored in a SharePoint space to which only iCOSHells consortium members, other than RISE and ZABALA as Data Processors, will have limited access. Besides, the private key to be able to read the contents will be shared by secure channels among consortium members. Upon request, this information will be also made available to the EC.</p> <p>RISE and ZABALA considers retaining generated data about sub-grantees during the length of the project. Anonymised statistical data can be retained longer, after the end of the project lifespan, for research purposes. If so, RISE and ZABALA estimates no additional cost for this. If anonymised collected metadata and statistical data will be retained after the length of the project, both entities have the infrastructure to retain data safely.</p>



	<b>Link with other WPs</b>	The information from the selected proposals will be shared with the leaders of WP2 (RISE), 3 (CETENMA), 5 (IFAU) and 6 (ATB) as they will work with the selected applicants of the Open Call to implement their proposed soil restoration solutions and track the results to integrate the outcomes of the third parties into the overall conclusions of the LLs. Data collected will be useful for consortium members in order to perform follow-up and monitoring of the progress of the selected sub-grantees during the Open Call. This dataset will also allow the consortium to publish reports about the solutions addressing soil restoration challenges developed thanks to the iCOSHells programme.
	<b>Applicability</b>	This dataset will be used mainly to disseminate metadata about the selected participants and associated soil restoration solutions that have been developed by iCOSHells.
	<b>Metadata</b>	<b>Descriptive:</b> dataset title, name of the sub-grantee, grant associated with the dataset; <b>Administrative:</b> dataset ID, creation date, modification date, and access rights, including any restrictions or licensing terms; <b>Technical:</b> file format, size, structure; <b>Provenance:</b> tracking access and usage logs, along with version history
<b>WP4 (ZABALA): External experts dataset</b>	<b>Contributor(s)</b>	ZABALA
	<b>Ownership</b>	ZABALA and RISE
	<b>Description</b>	External experts datasets: - For the Open Call for Experts: General information (name, surname, email, LinkedIn profile, phone number, gender, country of residence, nationality, CV) - For the selected Experts: - <i>Personal details for self-employees (name, surname, phone, email, ID, address country) or in case of experts working for a company (name of organization, VAT number, address, country, name, surname and ID of employee)</i> - <i>IBAN account</i> - <i>Profile (expertise area)</i> - <i>Description of activities carried out for iCOSHells</i> - <i>Payments issued for their work for iCOSHells.</i>
	<b>Type</b>	Observational from survey results and compiled from existing datasets
	<b>Format</b>	It will be mainly stored in CSV format.
	<b>Size</b>	1-2 MB



	<b>Provenance</b>	New Data: collected from the selection procedure of external experts that will evaluate the proposals submitted by third parties to the Open Call.
	<b>Methodology for collection/generation, process and analysis</b>	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 iCOSHELLS where they will be explicitly informed that ZABALA will keep their personal data whilst they work for iCOSHELLS and to be notified about iCOSHELLS 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.
	<b>Storage</b>	All the information related to external experts and stakeholders is stored in the external experts' and stakeholder's internal database. RISE and ZABALA, as controller of this information, ensures to handle security and privacy issues. This information will be stored in an encrypted spreadsheet file stored in a SharePoint space to which only iCOSHELLS consortium members will have limited access. Besides, the private key to be able to read the contents will be shared by secure channels among consortium members.  RISE and ZABALA consider retaining generated data about external experts and stakeholders during the length of the project. Anonymised statistical data can be retained longer, after the end of the project lifespan, for research purposes. If so, RISE and ZABALA estimates no additional cost for this. If anonymised collected metadata and statistical data will be retained after the length of the project, both entities have the infrastructure to retain data safely.
	<b>Link with other WPs</b>	Data collected will be useful exclusively for iCOSHELLS consortium members in order to perform follow-up and monitoring of the progress of the program, as well as for dissemination and communication issues. Particularly, the gathered data will be useful to organize evaluation in WP4, of the applicants in iCOSHELLS programme.
	<b>Applicability</b>	No (this dataset will not be publicly available).
	<b>Metadata</b>	<b>Descriptive:</b> title of the dataset or report, name of the external expert or organization; <b>Administrative:</b> dataset ID, creation date, modification date, and access rights, including any restrictions or licensing terms; <b>Technical:</b> file format, size, structure; <b>Provenance:</b> tracking access and usage logs, along with version history



<b>WP5 (IFAU): Replication potential dataset</b>	<b>Contributor(s)</b>	IFAU, ATB, CSCP, LL participants, ESK
	<b>Ownership</b>	IFAU, ATB, LLs, solution owners
	<b>Description</b>	Qualitative and quantitative data about the replication and scaling up potential for the solutions developed by the LLs and Open Call solution owners.
	<b>Type</b>	Observational and Compiled.
	<b>Format</b>	Formats of observational data include answers to interviews, minutes from workshops. Raw data will be captured in questionnaires or reporting templates. Mainly .txt, .docx, .pdf and .csv.
	<b>Size</b>	1-2 MB
	<b>Provenance</b>	New data from interviews and workshop; Re-used data stemming from desk research.
	<b>Methodology for collection/generation, process and analysis</b>	Data for determining the replication potential of the solutions will be gathered from replication workshops with the LLs and the solution owners including the Open Call solution providers. Data will be generated by using a facilitated discussion approach at the workshops and interviews with solutions owners and key stakeholders based on a set of open-ended questions designed to identify and assess the replication potential in qualitative, quantitative and spatial terms and, for collecting data about best practices for using a solution. Collected data will be submitted to an explorative analysis to identify replication potential of the solutions
	<b>Storage</b>	Data will be stored as word and excel files at the iCOSHELL Teams shared drive.
	<b>Link with other WPs</b>	Data generated in WP1 provides information about local ecosystems that constitute the environment for taking up solutions. Data from WP2 defines the soil health solutions to be replicated as well as the requirements for using the solutions. WP3 provides information about the best performing soil health solutions in environmental terms. WP4 captures the best solutions from the Open Call. Therefore, WP1-WP4 provides all relevant information to delineate the soil health solutions and the requirements for using these solutions, hence guidance for replication work.
	<b>Applicability</b>	Minutes from replication workshops and interview summaries are for internal use only. Processed data will be made available to the public through deliverables D5.1 and D5.5.
<b>Metadata</b>	N/A	



<b>WP6 (ATB): Lessons learned</b>	<b>Contributor(s)</b>	All iCOSHELLS partner organizations.
	<b>Ownership</b>	ATB will manage the data collected within WP6. The data belongs to the project Consortium iCOSHELLS. If partners require any data derived within WP6, this can be communicated and coordinated internally. No conflicts of interests are foreseen.
	<b>Description</b>	The data generated in this WP will come from extensive literature review and captured in real time through survey results, workshops, interviews, policy forums, roundtable discussions and other dedicated meetings. The data will be captured in real time.
	<b>Type</b>	Observational through interviews, survey results, scientific literature, project results and participatory methods such as workshops, forums, roundtable discussions and dedicated meetings. We will also compile data from existing reports and scientific results.
	<b>Format</b>	Mainly .txt, .docx, .pdf and .csv.
	<b>Size</b>	~ 20 MB
	<b>Provenance</b>	New data generated through research within the project ´s lifetime. Re used data from other WPs and project results.
	<b>Methodology for collection/generation, process and analysis</b>	Literature review, surveys and participatory methods such as workshops, discussions, interviews. Analysis through descriptive statistics, qualitative comparative analysis (QCA), necessary condition analysis (NCA) and archetype analysis.
	<b>Storage</b>	Plan to store all data in our internal ATB cloud and in the iCOSHELLS folders.
	<b>Link with other WPs</b>	We will gather data from all WPs in order to synthetize and integrate lessons learned from the innovation processes of the LLs.
	<b>Applicability</b>	The wider soil health community could benefit from the data and lessons learned generated in WP6. This refers to stakeholders investing in soil health globally, including European Soil Mission partners as well as land managers, decision makers, technology developers, research institutes, industry and civil society.
	<b>Metadata</b>	Around 5% of data volume associated with the data files.



### 3. Data managers

The iCOSHELLS DMP will be stored as a project deliverable document, with regular updates and successive versions. The first version of the iCOSHELLS DMP integrates the discussed procedures of the work packages (WPs) and their interfaces. RISE is responsible for producing the iCOSHELLS DMP and coordinating data management. All WPs contribute to the iCOSHELLS DMP by collecting relevant project data, providing data for use in other WPs, and delivering information about the datasets for recording in the iCOSHELLS 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: CSCP
- WP2: RISE
- WP3: CETENMA, UPM
- WP4: ZABALA
- WP5: IFAU
- WP6: ATB
- WP7: G!E
- WP8: RISE, ZABALA



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

The development, implementation, oversight, and updating of iCOSHELLS' Data Management Plan (DMP) are integrated into project management (WP8, T8.4) throughout the project's duration. The responsible team member (see Section 3) regularly receives contributions from all work packages (WPs). During the project, iCOSHELLS 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 iCOSHELLS, 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 iCOSHELLS 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 iCOSHELLS 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 iCOSHELLS' 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 iCOSHELLS members must be accessible to other members during the project period. Members interested in accessing this data must also comply with iCOSHELLS' 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.

## 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](#), European soil data centre) and other repositories (e.g., OpenAgrar) following the FAIR principles.

### Making data findable, including provisions for metadata

#### 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:

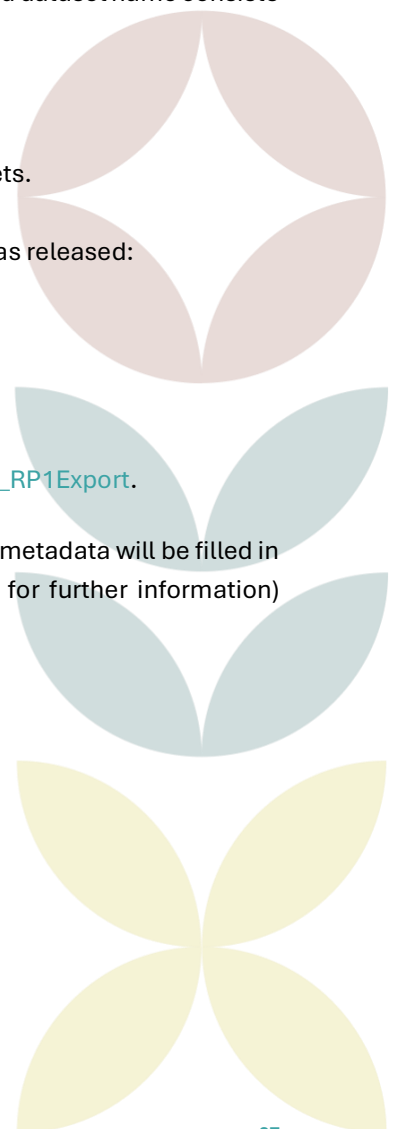
[ProjectName\\_DatasetName\\_Version](#), where:

1. The ProjectName is iCOSHells, in order 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:
4. DB the live database during project lifetime.
5. RP1Export export of the database at M15.
6. RP2Export export of the database at M27.
7. 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: [iCOSHells\\_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 (see 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.

### 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 follows: [YYYYMMDD\\_iCOSHells\\_DatasetName\\_Vzz.docx](#).

### 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 the iCOSHells project is essential for facilitating open access to the data. iCOSHells records metadata for the specific produced datasets described in Table 1, 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 Data Management Plan, 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<sup>2</sup>, based on DataCite's metadata schema<sup>3</sup>, 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 iCOSHells, 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 (<https://zenodo.org/communities/clarity/>))
- grants (List of European Commission FP7 grants which have funded the research for this deposition (730355). (Needed to establish the relationship to iCOSHells in OpenAIRE)

To improve metadata and the obtained results, iCOSHells participants will consider mentioning controlled vocabulary that significantly enhances metadata and increases the findability of the data. For many disciplines,

---

<sup>2</sup> [Developers | Zenodo](#)

<sup>3</sup> [DataCite Schema](#)

there are already specialized classifications and thesauri (see [BARTOC.org](http://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 UPM. Partners from each Living Lab solution will be in charge of measuring and feeding the database. To obtain a harmonized database, CETENMA and UPM 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.

A set of general keywords that should apply to all public datasets, scientific publications, and public deliverables will be defined and updated in the next version of the DMP.

### 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 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.

### Archiving, preservation and data quality assurance

#### Archiving and preservation

The iCOSHells partners have agreed on the procedures to ensure the long-term preservation of the datasets. Datasets will be stored on [Zenodo](http://Zenodo), a generic repository for EC-funded research developed by CERN and launched in May 2013. To be an effective catch-all that eliminates 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 iCOSHells programme will be informed of this requirement and requested to act accordingly. This may require anonymizing the uploaded contents of survey respondents, among other measures.

Alongside Zenodo, iCOSHells will contribute to the European Soil Observatory ([EUSO](#)), an initiative by the European Commission's Joint Research Centre (JRC) that serves as a centralized platform for soil data, policy monitoring, and knowledge sharing across Europe. EUSO is a policy-driven initiative designed to collect, integrate, and harmonize high-quality soil data across Europe, monitor soil health indicators such as carbon content, biodiversity, erosion, and contamination, support EU Soil Mission objectives related to sustainable land use and regenerative agricultural practices, and provide an open-access infrastructure for research and policymaking.

iCOSHells will leverage EUSO to ensure dataset harmonization and interoperability by aligning its soil health data with EUSO's standardized formats and methodologies, ensuring consistency and comparability across EU-funded projects. It will enhance soil health monitoring by contributing real-time and field-based data, improving EUSO's predictive models and soil health tracking systems. The project's research findings will inform EU policy frameworks on sustainable soil management, aligning with the European Green Deal and Soil Health Law proposals. Additionally, integrating iCOSHells datasets with EUSO's open-access repositories will facilitate scientific collaboration by providing researchers and policymakers with access to real-world soil health data, fostering cross-project analysis. While Zenodo will serve as a primary archiving platform, EUSO will support long-term data preservation and accessibility, ensuring that datasets remain available and valuable beyond the project's duration.

### Data quality assurance

iCOSHells 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.

## Publications and keywords

### Publications

The publications issued during the project include 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. 101157394”.**

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 iCOSHELLS 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 (see 5.3). The list of keywords will be compiled during the project and included in the second versions of the DMP.

#### Appropriate Keywords for iCOSHELLS Data and Publications

1. Living Labs
2. Soil Health
3. Indicators
4. Co-creation
5. Sustainable Soil Management
6. Soil Monitoring
7. Soil Restoration
8. Soil Quality
9. Soil Biodiversity
10. Soil Contamination
11. Regenerative Agriculture
12. Circular Economy
13. Agroecology
14. Carbon Sequestration
15. Nitrogen Flux
16. Nutrient Recycling
17. Remote Sensing
18. Proximal Soil Sensing (PSS)
19. Data Interoperability
20. Precision Agriculture

#### Digital Object Identifiers

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



Open iCOSHells results deposited in the iCOSHells 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 iCOSHells 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.

## Making data accessible

### 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)

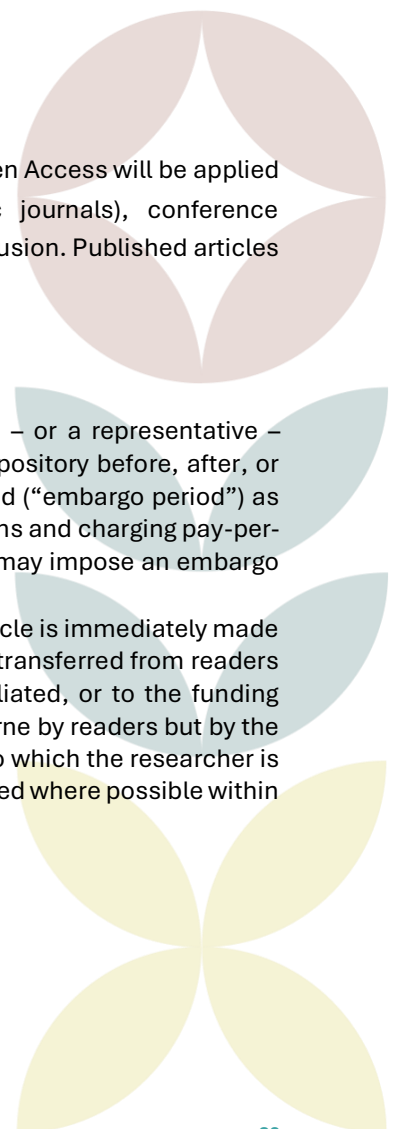
In accordance with the H2020 Guidelines on Open Access to Scientific Publications<sup>4</sup>, Open Access will be applied to peer-reviewed publications (scientific research articles published in academic journals), conference proceedings, and workshop presentations conducted during and after the project's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

There are two primary routes for open access to scientific peer-reviewed publications<sup>5</sup>:

- 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 iCOSHells project.

<sup>4</sup> [h2020-hi-oa-pilot-guide\\_en.pdf](#)

<sup>5</sup> [Open-Access-to-Publications-and-Data-in-Horizon-2020-Frequently-Asked-Questions-\(FAQ\).pdf](#)



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<sup>5</sup>.

### Overview of access to data

Open Access <span style="color: #008080;">●</span>		Confidential Access <span style="color: #FFD700;">●</span>	
Data type	Access	Main exploitation guidelines	
<b>Stakeholders / Personal data</b>	<span style="color: #FFD700;">●</span>	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 <sup>6</sup> 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 Directive 95/46/EC (General Data Protection Regulation) <sup>7</sup> .	
	<span style="color: #008080;">●</span>	Open to public in final reports, project website and in press releases.	
<b>Data from the LLs and management practices</b>	<span style="color: #FFD700;">●</span>	Confidential access for contractual terms, financial Information and management plans.	
	<span style="color: #FFD700;">●</span>	Confidential access for personal information in compliance with GDPR.	
<b>Data collected for databases</b>	<span style="color: #008080;">●</span>	Open to public in final reports, web application project website and press releases. Selected datasets will be transferred to EUSO-ESDAC for long-term accessibility and integration with EU-wide soil health monitoring frameworks.	
	<span style="color: #FFD700;">●</span>	Confidential access for contractual terms, Open Call metadata, Information and business plans.	
	<span style="color: #FFD700;">●</span>	Confidential access for personal information in compliance with GDPR.	

<sup>6</sup> 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](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046-EN)

<sup>7</sup> 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>

<b>Data exchanged among partners</b>	●	Contact details, file transfer protocol and secured collaborative space (Teams channel).
<b>Conference presentations, Workshops, Events, Deliverables</b>	●	Open access in the project website.
<b>Peer-reviewed publications</b>	●	Open access in the project website, Zenodo repository and partners' own repositories.
<b>Soil monitoring database</b>	●	Data will be transferred to a public web application and transferred to a public web application and integrated with EUSO-ESDAC for interoperability with EU soil data platforms.
	●	Datasets and the report on best performing solutions will be confidential (D 3.4. and D 3.5.).
<b>Open Call dataset</b>	●	Public but only the following fields will be published in the dataset, since this dataset has to 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 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.

## 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 NoSQL database MongoDB 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.

## 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.

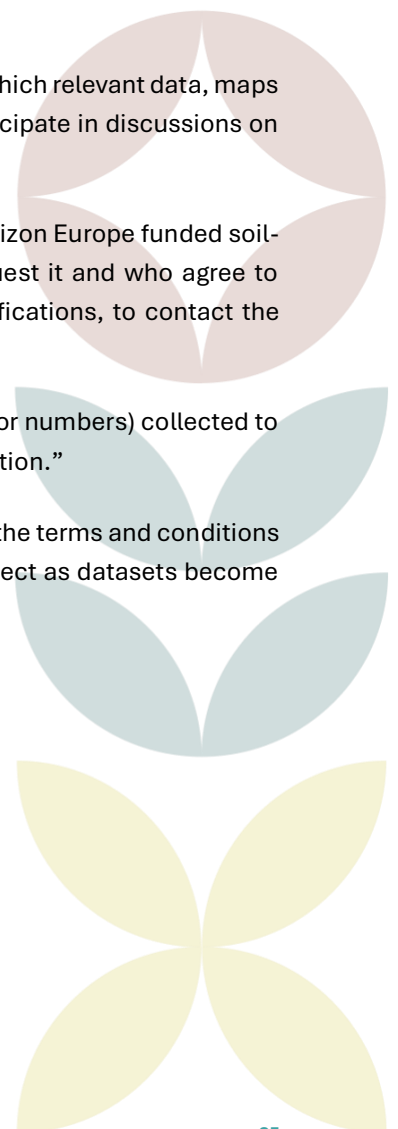
To ensure an effective contribution to the EUSO, the consortium will identify in the DMPs which relevant data, maps and information might be of relevance to the EUSO and appoint a contact person to participate in discussions on data management with the JRC's EUSO and the project SoilWISE.

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 iCOSHELLS (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 iCOSHELLS 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 iCOSHELLS 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 iCOSHELLS 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.

## Other research outputs

All known outputs have been included in the main sections of the DMP.



## 6. Allocation of resources

### 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.

### 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 (coordination team) will regularly update the access list for personnel within partner teams to files and sensitive information.

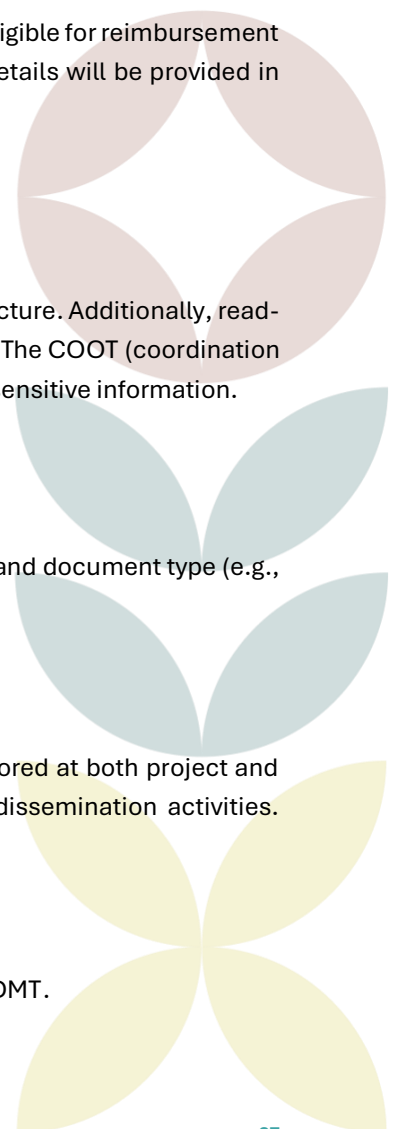
ZABALA will manage the data repository in Teams, with responsibilities including:

- 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 DMT.



- 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](#) 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



# 7. Data security

## Data storage and back-up

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 iCOSHells servers using state-of-the-art data protection software and procedures.

Regarding the **predictive tool**, the backend will include a secure REST (representational state transfer) service that serves data to be consumed by the front-end. The REST backend services will be developed using Python 3 and FastAPI (a high-performance web framework for building HTTP-based service APIs). The REST API will have several endpoints to serve all the indicators and models developed at that time. The NoSQL database MongoDB will be used in the storage layer, which allows BigData capabilities and can be scaled to larger areas if necessary. In addition, an ETL processes will be generated to update the database information regularly deploying the processes using Apache Airflow, which is an open-source platform for batch-oriented workflows. The interactive visualisations will be built using Apache eCharts (a free and open-source data visualization library). AGD will create the web application and dataset architecture. (For more information, consult D 3.3).

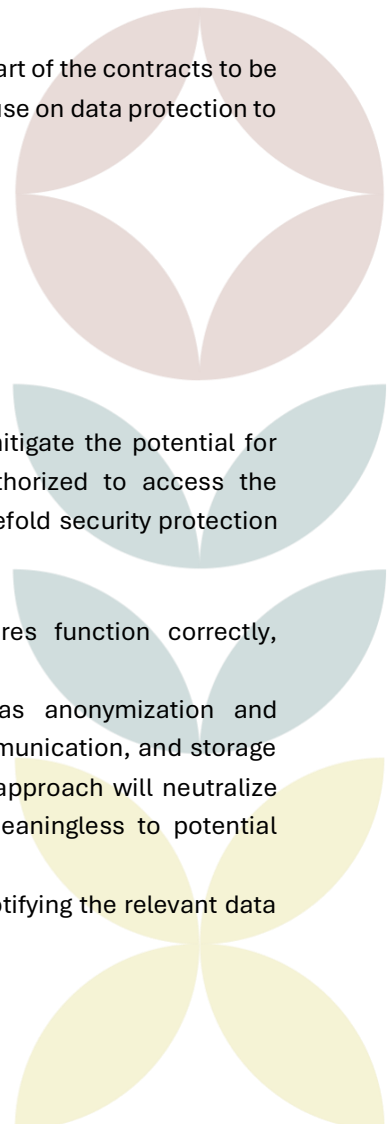
In **Open Call** management, WP4 will collect and process sensitive and personal data as part of the contracts to be signed with the sub-grant beneficiaries and evaluators. These contracts will include a clause on data protection to ensure compliance with relevant data privacy regulations.

## Data security within Consortium

### 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 iCOSHells 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 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.



### 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 iCOSHells. For identity management and data protection mechanisms, iCOSHells 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 internetted 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 iCOSHells 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 iCOSHells 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 iCOSHells 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.

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

### Internal threats and human errors

Most organizations focus on managing data risks from external threats, but the majority of breaches actually 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.

### Internal threat prevention

The implementation of a robust and adaptable security policy is crucial for iCOSHells. Such a policy can establish clear rules and permissions that are easily understood by both employees of iCOSHells' partner organizations and those responsible for enforcement, ensuring that personal data remains within the office. iCOSHells' 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:

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 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

## 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.



## 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 General Data Protection Regulation (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 iCOSHELLS 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.



## 9. Other issues

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





**icosHELLs**

[www.icoshells.eu](http://www.icoshells.eu)



**Funded by  
the European Union**

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 the granting authority can be held responsible for them.