ECHO

ENGAGING CITIZENS IN SOIL SCIENCE:
THE ROAD TO HEALTHIER SOILS

Co-funded by the European Union

UK Research and Innovation

ECHOSOIL.EU

Project information

Project number	101112869
Project acronym	ECHO
Project name	Engaging Citizens in Soil Science: The Road to Healthier Soils
Call	HORIZON-MISS-2022-SOIL-01
Topic	HORIZON-MISS-2022-SOIL-01-09
Type of Action	HORIZON Research and Innovation Actions
Responsible Service	REA.B.2
Project staring date	01 June 2023
Project duration	48 months

Document Details

Deliverable	D7.2 – Data Management Plan (DMP)
Work Package	7 – Project Management and coordination
Task	7.3 - Open science and Data management
Deliverable Type	Document, Report
Dissemination Level	Public
Deliverable Lead	FC.ID/CIENCIAS
Date of publication	30/11/2023
DOI	10.5281/zenodo.10231350

Disclaimer

Funded by the European Union under GA no. 101112869 – ECHO and co-funded by UK Research and Innovation (UKRI).

Purpose of document

The European Commission has consistently acknowledged the significance of research data, placing it at the same level of publications and outcomes. Moreover, there is a growing awareness in the market, research communities, society, and industry regarding the tangible value of data for exploitation. This recognition highlights the necessity for a structured approach to the utilization of research data in both theoretical and applied research projects, striking a balance that meets the diverse needs of various stakeholders.
To seamlessly integrate data into a project, generate project-specific data, and share data outside the project, it's crucial to address how this data will be handled early in the project's life and consistently throughout its execution. The central focus revolves around managing, sharing, and creating data.
On these bases, the Echo Data Management Plan (DMP) describes the data management life cycle for the data to be collected, processed and/or generated by the ECHO project. The goal is to establish guidelines that offer straightforward and practical advice for the implementation and validation phases of the project. These guidelines aim to ensure compliance with legal obligations and encourage the adoption of optimal data management practices.
The DMP includes information on:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project).

Communication, dissemination and exploitation of ECHO data are detailed in the D6.1 Communication, Dissemination and Exploitation Plan.

Intellectual Property Rights (IPR) issues have been established in the Consortium Agreement and Grant Agreement.

Versioning and contribution history

Version	Date	Modified by	Notes
1.1	12/11/2023	Silvana Munzi (FC.ID/CIENCIAS)	Draft version
1.2	27/11/2023	Tanja Mimmo Raphael Tiziani (UNIBZ)	Draft version
1.3	27/11/2023	Oleg Osychenko (QUANTA)	Draft version
1.4	28/11/2023	Silvana Munzi (FC.ID/CIENCIAS)	Draft version
1.5	23/01/2024	Silvana Munzi (FC.ID/CIENCIAS) Tanja Mimmo (UNIBZ) Oleg Osychenko (QUANTA)	Final version

Foreword

Soil is a vital, yet often disregarded, resource that supports life on Earth by providing the foundation for agriculture, forests, and various other natural ecosystems. However, soil degradation is a growing concern around the world, and it can have severe consequences for our planet like reduced crop yields, increased greenhouse gas emissions, and decreased biodiversity. The ECHO project aims to prevent this by bringing together citizens and volunteer scientists from around Europe to work towards a common goal of protecting and preserving our soils, thus contributing to the transition towards healthy soils of the EU Mission: "A Soil Deal for Europe".

ECHO will generate new data on the health status of EU soils, complementing existing soil mapping and monitoring in EU Member States and Scotland, including the EU Soil Observatory (EUSO). The project will develop and deploy 28 tailor-made citizen science initiatives across EU Member States and Scotland, taking into account different land-uses, soil types, and biogeographical regions, as well as stakeholder needs. With 16 participants from all over Europe, including 10 leading universities and research centres, 4 SMEs, and 2 Foundations, under the coordination of the Free University of Bolzano-Bozen, ECHO will assess 16,500 sites in different climate and biogeographic regions to achieve its ambitious goals.

The project aims to engage citizens in protecting and restoring soils by building their capacities and enhancing their knowledge. Citizens will thereby not only actively contribute to the project's data collection but also promote soil stewardship and foster behavioural change across the EU. The ECHOREPO, a long-term open access repository with a direct link to the EUSO, will make the citizen science data available for exploitation not only by scientists but also by citizens, policy makers, farmers, landowners and other end-users, providing added value to existing data and other relevant soil monitoring initiatives. ECHOREPO will thus provide valuable information about the state of soil health in various regions, and help citizens make informed decisions about land use and conservation.

We believe that the ECHO project will have a significant impact on soil health and citizen engagement across Europe and become an important step towards protecting and preserving our soil for future generations. By working together, we can ensure that our soil remains healthy and productive, and that we continue to enjoy the many benefits it provides.

Co-funded by the European Union
UK Research and Innovation

Table of Contents

1. Data collection in relation to the project's objectives

Data collection plays a pivotal role in achieving ECHO's specified objectives:

Objective 1: To engage citizens and through increased knowledge and improved literacy stimulate their interest in soil health related issues and motivate them to protect and restore soils.

The primary objective of ECHO, as outlined in the EU Soil Mission, is to engage citizens by increasing knowledge and improving literacy, thereby stimulating their interest in soil health-related issues. ECHO adopts a citizen science approach to promote awareness of soil health issues and strengthen citizens' capacity to play an active role in soil health protection. The project involves identifying and mapping target citizen groups across Europe interested in environmental and soil health issues. ECHO actively involves citizens in the collection and assessment of soil data through user-friendly tools and citizen science initiatives. This engagement is expected to increase citizens' understanding of the relationship between soil properties, function, and management choices, promoting soil stewardship on a pan-European level. ECHO's approach, focusing on engaging, empowering, and enabling citizens, creates a positive feedback loop between citizens, end-users, and scientists.

Main quantitative targets/data collection tools/activities:

- A map of citizen groups (each group with 75 citizens) covering all 28 European countries (Member States plus Scotland)
- 28 Co-creation workshops and 28 citizen initiatives involving approx. 16500 citizens
- 30 one-to-one meetings with citizen groups to introduce and promote the project
- Production of dissemination material

Objective 2: To empower citizens to understand the functioning and value of soils and to be capable of taking an active role in soil science.

ECHO plans to develop a customized citizen science toolbox for soil health assessment, empowering citizens to actively engage in community development decision-making. Tailored citizen science initiatives will be implemented across European Member States, providing unique knowledge for both citizen scientists and researchers. This approach is expected to increase citizens' understanding of soil health, enabling them to comprehend the impact of their actions on soils and fostering behaviors that benefit soil health. Data generated by citizen scientists will contribute to the maintenance and restoration of soil health by generating comprehensive knowledge about soil health, including measurement and management for optimal ecosystem services.

Main quantitative targets/data collection tools/activities:

- toolbox including open access field guidelines translated in the 24 official European languages, assessment, monitoring protocols and a forum where users can ask questions and discuss with experts
- app to make the toolbox easily available
- 500 citizens involved in the testing of the toolbox to collect and analyse soil data
- 16500 sampling sites collecting measurements about 8 different soil parameters

Objective 3: to enable citizens to take an active role in directly participating in decision-making on soil issues based on acquired knowledge.

ECHO's European-wide citizen science initiatives, led by experts, aim to generate a vast dataset on chemical, physical, and biological parameters of European soils. To enhance the broader user community's understanding of soil health, ECHO focuses on creating user-friendly mechanisms for data discovery and access. This involves establishing a long-term repository adhering to FAIR, open data principles, and modern web-based interfaces. The repository aims to broaden soil data generation and interpretation capacities in Europe, providing reliable tools for soil end-users and the wider community. Beyond data collection, ECHO encourages the reuse of soil data and actively engages various end-user groups, ensuring the high added value of collected data. The project explores the potential of fully exploiting citizen science data on soil health within specific case studies, showcasing its utility. These efforts create opportunities for citizens to actively participate in local development, collaborative decision-making on soil health issues, and the formulation of related policies.

Main quantitative targets/data collection tools/activities:

- long-term digital repository for data collected from ECHO citizen science initiatives
- user testing of the software engaging five (5) experts and volunteers (200 contacted through a survey, 50 through interviews)
- One focus group per consortium country to identify and engage current and potential end-users of citizen science soil health data and identify case studies
- EU-level workshop to demonstrate to and inform policymakers and farmer communities

2. Data management plan

The ECHO DMP outlines the key components of the data management policy applicable to all project participants and their handling of various datasets. Each dataset is examined in terms of the following aspects:

- Overview of Policies: Considering relevant national/international initiatives and programs in the field.
- Roles: Identifying roles and associated responsibilities of the participants.

- Data Set Reference and Name: Assigning an identifier for the produced dataset.
- Data Set Description: Providing details on the data's origin, nature, scale, utility, and potential support for scientific publications. Also, addressing the existence of similar data, integration possibilities, and reusability.
- Standards and Metadata: Referring to existing standards on what and how metadata will be produced or outlining new suitable standards if necessary.
- Data Infrastructure: Describing assets for data production and provision.
- Data Sharing: Outlining the sharing process, including access procedures, possible embargo periods, technical dissemination mechanisms, required software/tools, and specifying open access or restricted access. Identifying the repository for data storage and reasons for non-sharing, if applicable.
- Archiving and Preservation: Describing procedures for long-term data preservation, specifying preservation duration, approximate data volume, associated costs, and planned coverage.

Each of the required points will be addressed on a dataset-by-dataset basis following Annex 1.

Sharing data brings various advantages, such as building a community centered around the dataset and becoming a catalyst for expediting future research. This collaborative environment promotes advancements in research, enhances visibility for project results, and amplifies their impact by referencing the dataset.

Therefore, Findable, Accessible, Interoperable and Re-usable data (FAIR) and open data sharing are the default for the results of ECHO. In compliance with the EU's Open Science policy, data will be shared “as early as possible in the research process with all relevant actors to help diffuse the latest knowledge”.

This deliverable is a working document, and it is designed to adapt throughout the project's lifespan. Updates will encompass changes in data management details (such as dataset specifics) as well as adjustments related to notifications to relevant Data Protection Agencies, shifts in EU Data Protection Legislation, national regulations, emerging international treaties, and any aspect of evolving legislation. A minimum of three updates during the project duration (M6, M24, M48) is anticipated to ensure relevance and compliance.

This deliverable also appoints the partner responsible for data management. The DMP coordinator, FC.ID/CIENCIAS, as leader of the Task 7.3 Open Science and Data Management, is responsible for implementing the DMP, and ensuring it is reviewed and revised.

3. FAIR data

The data and/or other research results generated will be managed using the FAIR principles (FP):

(FP1) Making it findable: unique and persistent identifiers will be used for the project's outputs. The consortium will openly publish all the code developed (mainly in WP5), metadata, user manuals and instructions to replicate the platform on GitHub under European Public Union Licence (EUPL). The use of the European Open Science Cloud (EOSC, https://eosc-portal.eu) will ensure that all digital objects resulting from the project, like publications, databases and reports, are FAIR and accessible to institutional, national and international stakeholders, initiatives and infrastructures. All datasets will include comprehensive metadata to ensure discoverability and reusability. The use of EOSC, open-source software, and the Zenodo (https://zenodo.org/) repository will ensure long-term data curation after the life of the project. Scientific publications will be published under gold or green open access regimes while protecting author copyright. All relevant knowledge actors participating in the co-creation of results will be properly acknowledged. We will advertise its availability on the project website and document as well as public presentations. All the datasets created by citizen science initiatives during the project will be clearly identified.

(FP2) Making it openly accessible: The consortium will endeavor, whenever possible, to render data and other research results as open data or through open services. Project deliverables marked as public will be made openly available through the project website and related platforms such as Zenodo, OpenAIRE, in conformity with the Grant Agreement and the Consortium Agreement, with important features such as DOI support, GitHub integration, and that will guarantee the data's preservation and longevity.

The ECHO project is set to establish a state-of-the-art cloud-based data infrastructure, specifically designed for the efficient storage, management, and dissemination of scientific data. This cyberinfrastructure is characterized by several advanced components delivered by ECHOREPO (see Section 5), including a scalable cloud-oriented database for scientific data, an API for machine-to-machine communication, and a user interface for data discovery and accessibility.
ECHOREPO's cyberinfrastructure will ensure accessibility via metadata interfacing with EUSO by using established open-source data management systems, to expose scientific data to external entities. This integration will facilitate the availability of data to the EU Soil Observatory's metadata harvester in a standardized format, promoting data sharing with the global scientific community. The data exposure in the ECHO project's infrastructure will adhere to established standards, conforming to specific protocols and formats mandated by the EU Soil Observatory. This approach will ensure that the data is compatible and can be effectively harvested, understood, and utilized by external systems and users. This component also supports (FP3).

In public deliverables, all personal data will be anonymised. All the datasets created by citizen science initiatives will be accessible through the platform created in WP5.

(FP3) Making it interoperable: Actions will be implemented to ensure adherence to standards, especially those referred to SOIL data provided by citizen science initiatives. When possible, we will follow the standards suggested in the European Soil Data Center (ESDAC). We will follow standardized formats for variables, code, code list and metadata. We will consider the following guidelines: non-proprietary and non-tied to specific software, open and documented standards, common format used by the scientific community, standard representation (UTF-8, ASCII, etc.) unencrypted, uncompressed, etc.
(FP4) Making it reusable: Research data will be licensed under an open access licence whenever possible (EUPL for code, CC BY-NC-SA for documents and datasets, etc.). This will be conditional on the nature and level of privacy and IPR involved in the datasets (in any case we will recognize the ownership of the citizen science community for datasets created in the citizen science initiatives). This will be ensured by delivering the data not only in the project platform, but also in repositories such as Zenodo and linked platforms such as EUSO. From the technological standpoint, reusability can be assured by assigning to datasets in question their unique PIDs.
The data quality will be taken into account on two different levels. First, the quality of the data generated by the project will be assessed by the EUROREPO-CI software in accordance with the algorithms tailored specifically for all types of data. The second aspect of the data quality that will be guaranteed is the standardization of naming conventions, formats, and units to comply with FAIR principles, and to facilitate data harvesting.
The majority of the envisaged project results will constitute knowledge that will be publicly disseminated through presentations and scientific/technical publications, and therefore not to be subject to intellectual property rights protection. As the project progresses and the data are identified and collected, further information on increasing data reuse will be outlined in subsequent versions of the DMP (D7.2, M6, M24, M48).

4. ECHO website

The ECHO's website (echosoil.eu) was created as Deliverable 6.2, part of WP6 Dissemination and Communication. WP6 leader Plantpress is responsible for the content management and editorial responsibility of the project website.

ECHO´s website will host the public dissemination deliverables, the Citizen Science platform including the toolbox and mobile app (WP2) and will be linked to the ECHOREPO digital repository (WP5). The website will stay updated and available up to six years from the start of the project, exceeding the lifetime of the project. The main content of the website will be in English, with summary information, project abstract, main outcomes translated in the languages covered by the project consortium (Italian, German, French, Spanish, Portuguese, Greek, Finnish, Polish, Romanian).

ECHO will also develop open-source, easy-to-understand training materials, which will be displayed on the project's website. Training materials will allow easy adoption and use of the Citizen Science toolbox and will include video guides on how to use the different components of the toolbox; instructions on how to package and send soil samples to an ECHO lab for offsite assessments; instructions and protocols on how to perform on-site assessments. Furthermore, the comprehensive materials will include a territorial guide on the main soil uses across Europe, as well as a description of each of the eight soil health indicators used within ECHO.

All materials developed in this subtask will be translated into the official languages of the EU.

5. ECHO datasets

ECHO recognizes the following datasets as relevant to be processed throughout the project activities: (i) Research Data/Outputs - Online/Offline questionnaires and surveys; outputs (raw data) from focus groups; surveys and interviews; data for characterisation of CS initiatives, including pre-existing datasets from several sources; images, audio and video recordings; documents and spreadsheet for all WPs and finally, data provided by citizen scientist in the CS initiatives. (ii) Personal Data - Personal details of human participants in relevant activities, minimizing it to the strictly necessary for successful project implementation. This data might include email, age, contact details. (iii) Source code: The consortium will develop software components whose code and related documents will be written in the tasks described in WP5. All code developed during the project will be released under EUPL licences. For ECHO specific datasets, the consortium can guarantee preservation for more than 20 years at very low cost. Cost for data care and preservation are covered by the requested EU contribution and by in-kind contributions from the beneficiaries with project's IT physical infrastructure, commercial and EOSC cloud-based cold storage (in the order of tens of thousands of EUR). For the longer term, additional expenditures may ensue if a cloud storage (EOSC) is used, or to cover the cost of preservation of physical copies.

Main types of research data collected by ECHO are:

- **Derived or compiled data:** Resulting from data mining or statistical analysis, this type of data can be reproduced if the analysis process is thoroughly documented.
- **Observational data:** Captured in real time, this type of data is often unique and cannot be precisely reproduced.
- **Experimental data:** Derived from laboratory and equipment experiments, this data is usually reproducible, although the cost may be a limiting factor.

The adopted naming convention will succinctly detail the content, the data-collecting institution, and the month of publication. Version control will be invoked in instances where a participant wishes to withdraw their data; here, a version number will be incorporated into the file name.

Reuse of pre-existing data (from papers, datasets, technical reports, grey and white literature among others) will involve a clear and informative description of the data's origin, encompassing its collection methods, source, and any other pertinent details that contribute to a thorough understanding of the dataset.

The expected volume of data, based on the following presumed factors: a) number of participants, b) data types, c) frequency of submissions, d) data retention policy is of the order of tens of Gigabytes, in a lifespan of the project.

As stated in Article 8 of the Consortium Agreement, the results are owned by the Party that generates them. This researcher/entity will be also the “controller” of the dataset in question. The Data Controller is responsible for:

- assuring that the dataset always contains i) Contact Name (telephone and email contact details), ii) Date of First Version (Date the first version was completed); iii) Date of Last Update (Date it was last changed).
- data sharing – specifying the details of data sharing, including access procedures, potential embargo periods, delineation of technical mechanisms for dissemination, and the essential software and tools facilitating reuse.
- determining whether access will be broadly open or limited to specific groups. Likewise, they will pinpoint the repository for data storage, specifying its type (institutional, standard repository for the field, etc.), if already identified.
- establishing data archive and preservation procedures, including storage and backup measures, for business continuity and the long-term preservation of the data.

6. ECHOREPO

ECHO will also deliver significant technological innovations with the development of ECHOREPO, the ECHO digital data repository for soil data collected from citizen science and linked to EUSO (European Union Soil Observatory) as described in Section 2. Most of the current citizen observatories are highly biased towards reporting observations in areas such as biodiversity and not soil. Therefore, they do not provide a robust cyber-infrastructure delivering a digital repository and interfaces specifically oriented to reporting and consulting soil-related information.

ECHOREPO-CI (digital repository cyber-infrastructure) contributes to the project by developing a scalable, highly available, and comprehensive cyberinfrastructure delivering a long-term repository to facilitate the reporting, sharing, and consulting of soil information. Central to the ECHO project's infrastructure will be a cloud database, dedicated to storing scientific data. This database will be optimized for managing large-scale datasets typical in scientific research, offering enhanced capabilities for data storage, retrieval, and management. This would enable the consolidation of a larger community of scientists, the general public, and end-users for activities to support the assessment of European Soil health. Furthermore, one of the key

aspects of ensuring the long-term engagement of volunteers reporting soil data is developing linked web applications based on the user approach. Currently, there are no dedicated web applications specifically designed for soil monitoring.

ECHO will stick to the industrial standards in programming interfaces (e.g. REST API), data repository technologies, and the support for the open protocols and frameworks (RDF, OAI-MHP, and similar). The API will facilitate machine-to-machine communication and enable automated access and handling of data, allowing for seamless integration with a variety of software systems, data analysis tools, and automated workflows.

ECHOREPO-UI (digital repository user interfaces) contributes to the project with the development of interfaces for both the human user and machine-to-machine, facilitating the development of tools and connection with existing soil monitoring systems, using the proposed co-design methodology, and considering user needs from the initial stages of the process.

ECHOREPO-UI will leverage existing open front-end citizen science solutions, making use of the developed technologies and approaches.

EU Specific deliverables integrated in ECHOREPO in line with FAIR principles are:

- D5.1 Specification for interoperability and integration with existing soil monitoring systems: Report describing the interfaces and methods for connecting the digital repository with existing soil monitoring systems and databases.
- D5.2 Digital repository cyberinfrastructure code: ECHOREPO-CI source code delivered through a git-based repository using a permissive open-source license.
- D5.3 Digital repository user interfaces code: ECHOREPO-UI source code delivered through a git-based repository using a permissive open-source license.

7. Results

7.1 Ownership of results

Results are owned by the Party that generates them.

7.2 Joint ownership

Joint ownership is governed by Grant Agreement Article 16.4 and its Annex 5, Section Ownership of results, with the following additions:
Unless otherwise agreed:

- each of the joint owners shall be entitled to use their jointly owned Results for non-commercial research and teaching activities on a royalty-free basis, and without requiring the prior consent of the other joint owner(s).
- each of the joint owners shall be entitled to otherwise Exploit the jointly owned Results and to grant non-exclusive licenses to third parties (without any right to sub-license), if the other joint owners are given: (a) at least 45 calendar days advance notice; and (b) fair and reasonable compensation.

The joint owners shall agree on all protection measures and the division of related cost in advance.

7.3 Transfer of results

Each Party may transfer ownership of its own Results, including its share in jointly owned Results, following the procedures of the Grant Agreement Article 16.4 and its Annex 5, Section Transfer and licensing of results, sub-section "Transfer of ownership".

- Each Party may identify specific third parties it intends to transfer the ownership of its Results to in Attachment (3) of this Consortium Agreement. The other Parties hereby waive their right to prior notice and their right to object to such a transfer to listed third parties according to the Grant Agreement Article 16.4 and its Annex 5, Section Transfer of licensing of results, sub-section "Transfer of ownership", 3rd paragraph.
- The transferring Party shall, however, at the time of the transfer, inform the other Parties of such transfer and shall ensure that the rights of the other Parties under the Consortium Agreement and the Grant Agreement will not be affected by such transfer. Any addition to Attachment (3) after signature of this Consortium Agreement requires a decision of the General Assembly.
- The Parties recognise that in the framework of a merger or an acquisition of an important part of its assets, it may be impossible under applicable EU and national laws on mergers and acquisitions for a Party to give at least 45 calendar days prior notice for the transfer as foreseen in the Grant Agreement. In this case, notice of the transfer will be given as soon as possible and, in any case, not later than 30 (thirty) days after the transfer.
- The obligations above apply only for as long as other Parties still have - or still may request - Access Rights to the Results.

8. Personal data collection, processing, and reuse

In the ECHO project, there will be a limited collection and processing of personal data, involving the tracking or observation of participation in citizen science activities. In doing so the consortium will take into account the general obligation to maintain confidentiality in Article 21 and the rules concerning personal data according to Article 23 of the EU-GA. Participant Informed Consent Form is presented in Annex 2.

All project partners are bound by national and European data protection laws. The legislation in each participating partner's country upholds key principles:

- Data must be processed fairly, lawfully, and solely for the purpose for which it was collected and further processed.
- Data cannot be disclosed without authorization, except in cases of overriding legal acts or legitimate grounds.
- Individuals, with certain exemptions, have the right to access information about them and request correction of inaccurate data.
- Information cannot be transferred beyond the EEA boundaries without consent or the adoption of other adequate protection measures.
- Organizations are typically required to register or notify the processing of personal data unless the processing is straightforward, or a data protection officer has been appointed.
- Organizations must implement adequate security measures.

Measures for protection of personal data will be adopted, especially during those project activities requiring the collection of personal data (e.g., field activities and training courses). Data management activities will involve data and information that the users voluntarily share, and data and information that the ECHO consortium generates and shares. Declarations of consent will be required whenever necessary. Any personal data collected will be managed according to the provisions of the Regulation (EU) 2016/679 (General Data Protection Regulation) and will solely be used for the project activities.

As stated in section '4.5 Specific responsibilities regarding data protection' of the project Consortium Agreement "*Where necessary, the Parties shall cooperate in order to enable one another to fulfil legal obligations arising under applicable data protection laws (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 relevant national data protection law applicable to said Party) within the scope of the performance and administration of the Project and of this Consortium Agreement. In particular, the Parties shall, where necessary, conclude a separate data processing, data sharing and/or joint controller agreement before any data processing or data sharing takes place*."

Ethical aspects are considered more in detail in Task '7.4 - Ethics issues'.

9. Synergies with other projects and platforms

One of the main objectives of the data management solution is the integration of the staging ECHO data repository with the EU Soil Observatory (EUSO) data platform.

The main mission of the EUSO is to make EU soil-related data available to the maximum extent possible, and this aligns with the objectives of the ECHO project. In particular, EUSO will:

- provide an integrated and actionable access point to scattered and heterogeneous soil data;
- follow FAIR principles;
- improve trust, willingness, and ability to share and reuse soil data and knowledge;
- work with EU countries to identify and present relevant national soil data, using advanced web service technologies;
- incorporate the European Soil Data Centre (ESDAC). This is the thematic centre for soil-related data in Europe since 2006;
- support ESDAC as the single reference point for all relevant soil data at European level: 1) the European Soil Database, 2) the series of LUCAS Soil data.

The exact technology to integrate with the EU Soil Observatory data platform is currently under development. In the current approach the interaction with the EUSO will be performed through SoilWise (https://soilwise-he.eu/).

SoilWise is an open access knowledge and data repository to safeguard soils. It provides an integrated and actionable access point to scattered and heterogeneous soil data and knowledge in Europe in accordance with FAIR principles. It is envisaged that the platform will comply with the following requirements:

- Data interfaces
- Search capability
- Access (ease of access, equal access, etc.)
- Persistency (long-term availability)
- Data integrity
- Redundancy
- Fail-over
- Business continuity, disaster recovery
- Building quality products
- Report back to data “providers” (e.g., citizens)
- Operation & maintenance / sustainability plan
- Continuous integration
- Cybersecurity plan
- FAIR assessment (e.g., FIP profile)

A special emphasis is placed on a possible use of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) as an open industry standard mechanism for repository interoperability. The data providers exposing their metadata according to the protocol can be

linked to the service requestors (or harvesters, i. e. a client application that issues OAI-PMH requests) that harvest these metadata. The latest specification of OAI-PMH is deemed powerful enough to cover all the potential metadata complexity that we expect for ECHO.

Annex 1- Dataset description

1 Brief description of the described research output
1.2 Is it physical or digital?
1.3 Are you generating or re-using it?
1.4 What types and formats of data will the project generate/collect?
1.5 What is its expected size?
1.6 What is its origin/provenance?
1.7 To whom might it be useful ('data utility')?
1.8 What is the purpose of the data collection/generation and its relation to the objectives of the project?

2 Making data findable
2.1 What type(s) of persistent identifier(s) are used for the described dataset/output?
2.2 Will you provide metadata for the described dataset/output?
2.3 What type(s) of metadata?
2.4 What naming conventions and standardized vocabularies are used?
2.5 Please provide URL/description of used vocabularies
2.6 Are the metadata searchable?
2.7 How are searchable metadata provided?
2.8 Are keywords provided in the metadata?
2.9 Are metadata harvestable?

3 Making data openly accessible
3.1 Will this dataset/output made openly available? If this dataset cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.
3.2 How will the data be made accessible (e.g. by deposition in a repository)?
3.3 What methods or software tools are needed to access the data?
3.4 Is documentation about the software needed to access the data included?
3.5 Is it possible to include the relevant software (e.g. in open source code)?
3.6 Have you explored appropriate arrangements with the identified repository?
3.7 Is there a need for a data access committee?

4 Repositories and publications
4.1 In which repository will the dataset/output be deposited?
4.2 Is the selected repository a trusted source?
4.3 Does the repository(ies) assign datasets/outputs with persistent identifiers?

4.4 Does the repository(ies) resolve the identifiers to a digital object?
4.5 Does the repository support versioning?
4.6 What level of access to data will be provided (and why)?
4.7 Is there a data availability statement provided along with the publication?
4.8 Are the publications derived by this dataset open access?
4.9 Are all the contributors acknowledged in the publications?

5 Data, including data security
5.1 What is the described dataset/output title?
5.2 How is the dataset/output shared?
5.3 Are there any methods or tools required to access the dataset/output?
5.4 Is the described dataset/output supported by a data access committee?
5.5 Please specify how the dataset/output will be accessed during and after the project ends
5.6 Please specify how long after the project has ended the dataset/output will be made accessible
5.7 Do you make use of other national/funder/sectorial/departmental procedures for data management?
5.8 Documentation of other procedures
5.9 What security measures are followed?
5.10 What conditions do the security measures meet?
5.11 How will you preserve the described dataset/output in the long term?

6 Metadata
6.1 Will you provide metadata even if the described dataset/output cannot be openly shared?
6.2 Under which license will metadata be provided?
6.3 Do metadata provide information about how to access the described dataset/output?
6.4 Will metadata remain available after the dataset/output is no longer available?

7 Making data and other outputs interoperable
7.1 Does your (meta)data use a controlled vocabulary?
7.2 Have you applied a standard schema for your (meta)data?
7.3 What is the methodology followed?
7.4 Does the described dataset/output provide qualified references with other outputs?
7.5 In case you used uncommon or generated project specific ontologies or vocabularies, are you providing mappings to more commonly used ontologies?

8 Increasing data and other outputs reuse
8.1 What internationally recognised licence will you use for your dataset/output?

8.2 When will the data be made available for reuse? If an embargo is sought, specify why and how long this will apply
8.3 What reusability and/or reproducibility methods are followed?
8.4 Will you provide the described dataset/output in the public domain?
8.5 Do you intend to ensure (re)use by third parties after your project finishes? If the reuse of some data is restricted, explain why.
8.6 How long is it intended that the data remains re-usable?
8.7 Is provenance well documented?
8.8 What documented procedures for quality assurance do you have in place?

9 Allocation of resources
9.1 What will be the cost of making the described dataset/output FAIR?
9.2 How will this cost be covered?
9.3 Identify the people who will be responsible and their role(s) in the management of the described dataset/output
9.4 Are the resources for long term preservation established (costs and potential value, who decides and how what data will be kept and for how long)?

10 Ethical aspects
10.1 Are there any ethical or legal issues that can have an impact on data sharing?
10.2 Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

Annex 2- Participant informed consent form

You are invited to take part in the citizen science activities of the project "ECHO - Engaging Citizens in soil science: the road to Healthier Soils" an initiative aimed at increasing knowledge and awareness of soil health among EU citizens. This research study is conducted in collaboration with [mention the institution organizing the activity and partners], one of the partners of the project. Please carefully read this form and sign it. In case of doubts, you can ask any questions to [responsible for the activity] before deciding to participate. A copy of this form will be provided for your records.

Purpose: The objective of this study is to [briefly outline the nature and purpose of the research.

Procedures: If you agree to participate, you will be requested to [explain the participant's involvement] concerning [specify the topic]. This is expected to take approximately [mention the estimated time commitment].

Risks and Benefits: We do not anticipate any specific risks associated with this study [or acknowledge unpredictable risks if applicable]. While there are no direct benefits for you [or specify any benefits or incentives], your involvement will contribute to our understanding of [highlight potential benefits for the researcher].

Voluntary Nature of Participation: Your decision to participate or not will not impact your current or future relations with [the institution administering the study]. You have the freedom to withdraw at any point without affecting these relationships. You may also choose not to answer any questions or perform any activity you find uncomfortable.

Confidentiality: This information will not be disclosed to third parties without your consent, except to meet government, legal or other regulatory authority requirements. Your anonymity will always be safeguarded in this research. All data collected through [describe the data collection procedure, e.g., survey, interview, etc.] will be securely stored in a protected file accessible only to the researchers. This consent form will be stored separately and destroyed at the conclusion of the study.

Contacts and Questions: The researcher overseeing this study is [researcher's name]. If you have any questions later, you can reach out at [researcher's contact information].

For questions about your rights as a participant, you can contact the [institution's name] at [contact information].

Statement of Consent: I have received information about this research study, including its risks and benefits, and have had the opportunity to ask and have my questions answered. I willingly consent to participate in this study. I agree to allow my de-identified data collected as part of this study to be utilized in a future study.

Signature _____ **Date:** _____