D7.2 Open Science and Data Management Plan, v1

M06/FEB 2024

Funded by
the European Union

This project has received funding from the Horizon Europe research and innovation programme under Grant Agreement No 101112838.

Acronym	SoilWise
Project Full Title	**An open access knowledge and data repository to safeguard soils**
GA number	**101112838**
Topic	**HORIZON-MISS-2022-SOIL-01-01**
Type of Action	**HORIZON Innovation Actions**
Project Duration	**48 months**
Project Start Date	**1-Sep-23**
Project Website	**www.soilwise-he.eu**
Deliverable Title	**Open Science and Data Management Plan, v1**
Delivery Time (DOA)	**M06**
Deliverable Submission Date	**29/02/2024**
Status	**draft**
Dissemination Level	**PU - Public**
Deliverable Lead	**Tomáš Řezník (MU)**
Author(s)/Organisation(s)	**Tomáš Řezník (MU), Tomáš Pavelka (MU), Paul van Genuchten (ISRIC), Rob Lockers (WR), Isidora Stojacic (BIOS), Nick Berkvens (EV ILVO), Thorsten Reitz (WE), Somakanthan Somalingam (WE), Radu Giurgiu (EV ILVO), Dajana Snopková (MU), Lukáš Herman (MU)**
Contributor(s)	**MU, EV ILVO, WE, WR, WU, ISRIC, BIOS**
Peer-Reviewers	**Fenny van Egmond (ISRIC), Anna Fensel (WU)**
Contact	**tomas.reznik@sci.muni.cz**
Work Package	**WP7**
Dissemination level	**Public**
Keywords	**Data management plan, metadata, knowledge, mission soil projects, FAIR, data security, data storage, SoilWise repository**
Abstract	**This is the first version out of three deliverables describing the foreseen open science and data management practices of the SoilWise Repository (SWR). The SWR is being developed as a part of the EU Mission: A Soil Deal for Europe (Mission Soil). The SWR is designed as a system to access reliable, interoperable existing and new data and knowledge collected at local, national and EU levels to allow informed decision-making at all scales to support the proposed EU Soil Monitoring, Resilience Directive and the EU Soil Strategy.**

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.

List of Abbreviations

AI	Artificial Intelligence
BIOS	BIOSENSE INSTITUTE - RESEARCH AND DEVELOPMENT INSTITUTE FOR INFORMATION TECHNOLOGIES IN BIOSYSTEM
DMP	Data Management Plan
EEA	European Economic Area
ESDAC	European Soil Data Centre
ESP	European Soil Partnership
EUSO	European Soil Observatory
EV ILVO	EIGEN VERMOGEN VAN HET INSTITUUT VOOR LANDBOUW- EN VISSERIJONDERZOEK
GDPR	General Data Protection Regulation
GSP	FAO Global Soil Partnership
IPR	Intellectual Property Right
IACS	Integrated Administration and Control System
ICT	Information and communications technology
IETF	Internet Engineering Task Force
IDSA	International Data Spaces Association
ISRIC	STICHTING INTERNATIONAL SOIL REFERENCE AND INFORMATION CENTRE
IUSS	International Union of Soil Sciences

JRC	European Commission's Joint Research Centre
ML	Machine Learning
MU	Masarykova univerzita
OGC	Open Geospatial Consortium
PU	Public
R&I	Research and Innovation
REA	European Research Executive Agency
SWR	SoilWise Repository
UNESCO	United Nations Educational, Scientific and Cultural Organization
W3C	World Wide Web Consortium
WE	WETRANSFORM GMBH
WP	Work Package
WR	STICHTING WAGENINGEN RESEARCH

Table of Contents

List of Tables and Figures

1 Introduction

1.1 Project summary

Now more than ever, soil health is an issue that needs to be addressed urgently, as recent assessments state that 60-70% of European soils can be considered unhealthy (Bouman, 2022). The EU Mission 'A Soil Deal for Europe', the EU Soil Strategy and the proposal for an EU Soil Monitoring and Resilience Directive (5 July 2023), aim to have 75% of EU soils healthy or significantly improved by 2030 and all soils healthy in 2050 (Panagos et al. 2022). Reaching such an ambition requires, among others, access to reliable, harmonised existing and new data and knowledge collected at local, national and EU levels to allow **informed decision-making at all scales to support the proposed EU Soil Monitoring and Resilience Directive and the EU Soil Strategy**.

The SoilWise project will provide an integrated and actionable access point to scattered and heterogeneous soil data and knowledge in Europe, making them FAIR (Findable, Accessible, Interoperable and Reusable) (Wilkinson, et al. 2016) and improve trust, willingness, and ability to share and re-use soil data and knowledge. In three project development cycles, **co-creation and co-validation by multi-stakeholder groups are at the centre of project activities**. SoilWise recognises existing workflows and repositories for specific user needs and aims to work with them to enhance their discoverability, approachability, and interconnection. An overview of existing repositories foreseen for data integration with SoilWise is schematised in Figure 1. We also foresee visualisation by multiple coordinated views approach, as discussed, e.g. by Langner et al. (2018). An open, modular, scalable, and extensible knowledge and data repository building on existing and new technologies will be provided while respecting data ownership, access policies and privacy. AI- and ML- techniques will be employed to interlink scattered data and knowledge, automatise the processes, infer new knowledge and increase FAIRness. **SoilWise applies infrastructure thinking instead of project thinking to design a repository for at least a decade to support EUSO evolvement accordingly**. The SoilWise repository and community are designed to be a joint starting point and common ground for countries, the European Commission, and other stakeholders to jointly guide soil and related spatial policy and informed decision-making towards the 2030 goals of the Green Deal (Montaldo, S. 2022), achieve healthy soils in 2050 and ensure broad uptake and implementation by land managers, policy, research, and industry.

All personal data acquired through SoilWise is processed in strict accordance with the relevant EU privacy regulations, highlighting our dedication to uphold to the highest standards of data privacy and security for our users.

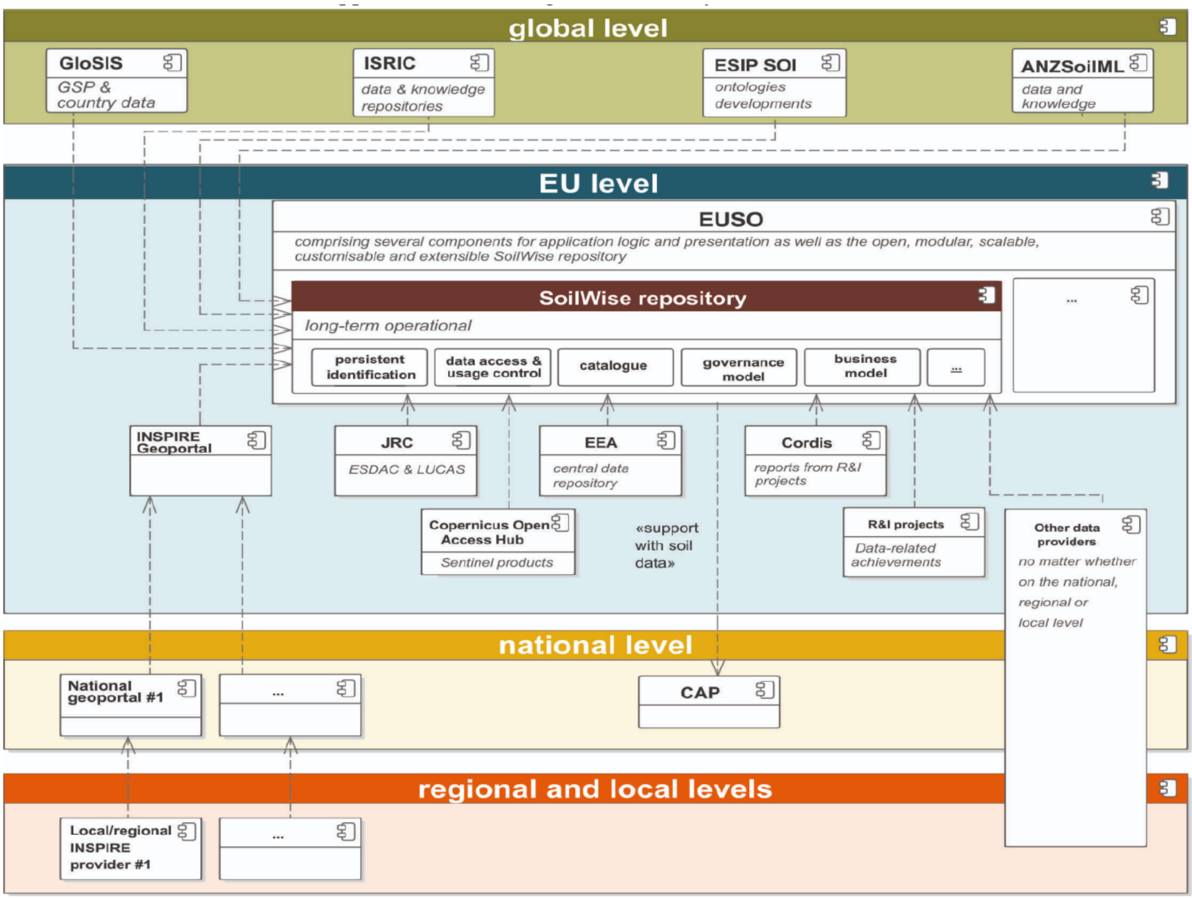


Figure 1 A schematized overview of data and knowledge flows between the SoilWise repository and other repositories.

1.2 Document scope

The scope of this document is to describe the initial and, later on, updated plans for data and knowledge management and open access publications of the SoilWise project, including our primary foreseen achievement: the SoilWise repository (SWR). The primary goal of SoilWise data and knowledge integration and management is to facilitate the analysis, reasoning, querying, and visualization of data and knowledge from different sources for different purposes. We aim to create coherence in the currently dispersed data and knowledge ecosystems on soil and soil health to facilitate improved decision-making by various actors towards improved soil health.

SoilWise wholeheartedly supports the Open Science Policy under Horizon Europe; thus, appropriate open science actions are being implemented as an integral part of the project.

Moreover, the presented data management plan (DMP) stands in the centre of the SoilWise project as well as in the centre of attention of external stakeholders, including the Joint Research Centre of the European Commission and other research and innovation projects funded under the mission 'A Soil Deal for Europe' (or Mission Soil Projects). In line with the discussions and needs of these stakeholders so far, we introduce a document structure different to a common DMP structure. The need for SoilWise to provide guidance on the SWR inputs, processing, and outputs was repeatedly mentioned by all the relevant stakeholders as a priority. As such, this deliverable primarily presents open science and data management practices/plans of the SWR. The open science and data management practices/plans of the SoilWise project itself are intentionally introduced as secondary, in section 6. Nevertheless, this DMP describes the data and knowledge management aspects of the following project activities as well:

- Support of academic articles,
- generating new knowledge,
- data processes facilitation,
- support of data to prove that a given solution works as expected,
- evidence that the project has been carried out effectively.

The presented document is the first iteration out of the three foreseen. The second iteration, 'Deliverable D7.3 – Open Science and Data Management Plan, v2' is scheduled for M27 (November 2025). Such a second iteration will primarily update the sections on i) knowledge handling possibilities, ii) information gathered from stakeholders and its implementation, iii) functionality aligned with SWR architecture and proposed components, iv) EUSO support, and v) project-related data and knowledge management. The third and last iteration, entitled 'Open Science and Data Management Plan, v3', is scheduled for the very end of the project, due in M48 (August 2027). The final version of the DMP will, among others, describe all the lessons learned on data management and open access publications as an information source, guide and inspiration on data handling to other Mission Soil (research and innovation) projects and beyond.

1.3 Document structure

This document is comprised of the following chapters:

- **Chapter 1** presents an introduction to the document and its relation to the project;
- **Chapter 2** explains a strategy towards provisioning of external data/knowledge/metadata inputs to SWR;
- **Chapter 3** describes the SWR storage strategy, its data and knowledge handling and processing, including dealing with sensitive assets;
- **Chapter 4** explains the SWR strategy of making the data and knowledge 'FAIR', that is making it findable, openly accessible, interoperable and re-usable;
- **Chapter 5** presents the means of SWR support towards EUSO;
- **Chapter 6** describes the handling of data produced by the project related to the project management, documentation and dissemination;
- **Chapter 7** describes the resources that SoilWise will allocate in order to follow the FAIR principles and to guarantee data and knowledge preservation;
- **Chapter 8** deals with potential critical issues and technical risks, and describes the actions for their mitigation;
- **Chapter 9** defines how SoilWise will deal with the security aspect of the data and knowledge management during the project;
- **Chapter 10** is about the potential ethical issues associated to the datasets and knowledge handled in the SoilWise project, including compliance with the GDPR.

1.4 Relationship to other project deliverables

This document is related to and complements the following deliverables:

WP1:

- D1.3 – D1.4 Repository architecture [M8, M42]
- D1.5 – D1.6 Repository GM [M21, M42]

WP2:

- D2.1 – D2.4 Developed & Integrated DM components [M13, M18, M31, M46]
- D2.5 – D2.6 Report on strategy for FAIRness on soil data [M27, M42]

WP3:

- D3.1 – D3.4 Developed & Integrated KM components [M13, M18, M31, M46]
- D3.5 – D3.6 Report on strategy for effective soil KM [M27, M42]

WP4:

- D4.1 – D4.4 Repository infrastructure, components and APIs [M13, M18, M31, M46]
- D1.5 – D1.7 Repository Data and Knowledge Resources [M21, M34]

WP6:

- D6.5 – D6.6 IPR, Business Model Report and Policy Brief [M24, M47]

WP7:

- D7.1 Project Management handbook [M1]
- D7.3 – D7.4 Open Science and Data Management Plan [M27, M48]

1.5 SoilWise utility

SoilWise identifies a variety of target groups benefiting from the project results. Within the project, target groups are represented by 5 Use cases:

1. Soil health performance indicators for **Land Managers**;
2. Leveraging a network of Soil R&I Knowledge and Data to facilitate **scientists**;
3. Facilitate **policy makers** in policy Making & Evaluation to safeguard soil;
4. Enhanced capacities of **Public Authorities** and **Living Labs** actors;
5. New products, technologies, and services for **business**.

WP5 on User Cases planning, demonstration and assessment is dedicated to demonstrating how different target groups can capitalize on the (re)use of the SWR.

An exhaustive description on Data Utility is provided by Deliverable D1.1 – Usage Scenarios, Requirements, v1 [M6], and will be further elaborated in next version (D1.2 [M36]).

1.6 Data and knowledge

The SWR aims to improve the state-of-art on soil data and knowledge findability, accessibility, interoperability and reusability (FAIR). While there is generally a clear idea of what we define as data, there are different views on how knowledge is to be defined.

Within SoilWise, we define knowledge as follows: First, knowledge can be considered as the explicit knowledge, relevant for SWR users, which is “hidden” in unstructured content, like documents, web pages etc. SWR users would benefit from making such currently fragmented explicit knowledge sources better findable, accessible and reusable. Second, there is knowledge that is required to “understand” how both data and knowledge assets can be interpreted or reasoned over in a way that supports answering an end user’s question. This semantic knowledge could for instance be in the form of taxonomies, knowledge graphs or large language models used to support such reasoning. Third, there is knowledge that structures the relationships and interlinking between data and knowledge assets (e.g. linking a dataset with the project it was generated in and/or with the standard operating procedure used). While the core challenge for SoilWise is to make the knowledge expressed within “unstructured content” usable for its end users, this cannot go without the use (and development) of the other types of knowledge mentioned. At the same time, the latter types of knowledge are mostly hidden under the hood of the SWR to “run the engine”. From the perspective of a SWR end user, the reusable parts will be served as “just another knowledge asset”.

For an end user, both data and knowledge are assets that are made FAIR through the SWR. There are however good reasons to make the separation, since the ways of storage and processing, but also e.g. handling sensitivity, licensing etc., will be different. Due to the specific knowledge definition, the differences will be examined and described in more detail in following iterations of this document. As such, this DMP distinguishes between data and knowledge where needed, regarding the requirements, procedures and other aspects.

The variety of data and knowledge within the Mission Soil is enormous. Some efforts, typically Mission Soil (EU Horizon) projects, are at a level of plain text files, some structured data in a proprietary form like Microsoft Excel, some use traditional GIS relational databases, some deal with data on a semantic level, being well-based on ontologies and thesauri. We are witnesses of various data and metadata standardisation levels, similar to the high variety of data.

2 Provision of data and knowledge to the SWR

2.1 Data and knowledge origin

The SoilWise project will interact closely with key EU research projects that have the explicit aim of building soil stakeholder communities towards the Mission Soil Implementation and beyond, in Europe and globally for different user groups and purposes, such as SOIL O-LIVE, BENCHMARKS, ISLANDR, EJP SOIL, ORCaSa, and initiatives (EUSO, ESP, GSP, IUSS, CODATA). Moreover, SoilWise partners are involved in soil knowledge and data (infrastructure, analysis and interpretation) projects and initiatives (EJP SOIL, GS Soil, SIEUSOIL, ORCaSa, BENCHMARKS, MARVIC, EUSO, GSP, ESP, IUSS and more). ISRIC has an explicit (UNESCO) mandate to foster such communities and soil information infrastructure worldwide and be the World Data Centre for Soils.

The SoilWise consortium will re-use its existing networking threads with major R&I and other initiatives and consider their development outlook. Metadata will be populated through automatic extraction from common metadata models in various communities as much as feasible (ISO 19115, INSPIRE, DCAT, DataCite, STAC, AGROVOC) through harvesting capabilities of OGC/INSPIRE catalogues and DCAT harvesters.

SoilWise will establish data flows that will connect existing repositories and workflows with the SWR, and in a later stage, also with the EUSO, see section 5 SWR support to EUSO. A simplified overview of existing repositories foreseen for data integration with SoilWise is schematised in Figure 1. The data, metadata and knowledge will be handled as described in section 3.1 Storage strategy.

The data and knowledge assets foreseen for SoilWise integration are as follows:

- ESDAC (European Soil Data Centre);
- The INSPIRE Geoportal soil domain related high-value datasets (soil, ortho-imagery, elevation, geology, natural risk zones, environmental monitoring facilities, agricultural and aquacultural facilities, area management / restriction / regulation zones & reporting units, land cover, land use, hydrography), as identified by EJP Soil project;
- EEA central data repository;
- national (soil) data and knowledge repositories that may not be accessible in the INSPIRE GeoPortal yet;
- soil-related data & knowledge achievements of R&I projects including their interlinking with Cordis reports and OpenAire;
- scientific knowledge and data repositories with R&I results such as DataVerse, EJP SOIL, BonaRes and ORCaSa;
- Sentinel products in Copernicus Open Access Hub;
- EC (European Commission) DestinE (Destination Earth);
- IACS data;
- Other data resources discovered thanks to the multi-stakeholder approach will be integrated on-the-fly.

Note that the SWR linkage to data spaces, i.e. primarily the Green Deal Data Space and Agriculture data spaces, are subject of ongoing discussions. The existing mechanisms, such as Dataspace protocol, will most likely be used. Description of the interaction of SWR and data spaces will be a part of this deliverable in its future iterations.

2.2 Data and knowledge types and formats

SoilWise will generate a relatively small amount of new soil data as data and knowledge are primarily foreseen to remain at the data provider and linked to SWR through metadata records. The following **types** of new and derived data and knowledge are foreseen to be ingested within SWR. The criteria on which data should be ingested will be defined in deliverable Repository Governance Model (D1.5, D1.6 [M21, M42]):

- metadata repositories,
- metadata records,
- standardized data and knowledge,
- derived soil data and knowledge products (such as soil health basic data, indicators and thresholds, MRV input data and knowledge, sustainable soil management practices suitable for a specific soilscape, etc.),
- EO and land-related data knowledge graphs,
- farm data,
- other relevant data types,

SWR will log data, knowledge and metadata in many standardized **formats**. The specific formats, and their possible limitations, will be defined during the soil-related R&I projects screening and the stakeholder elicitation process (T1.1).

Note that SoilWise will not consider multimedia (images, videos) as separate knowledge assets. They can however be part of a knowledge construct, in which case textual context and/or tags should be sufficient to elicit the required information for indexing and deriving metadata of the construct.

2.3 Expected size of data and knowledge assets

Our assumption on the expected size of potentially relevant data and knowledge assets is based on the discussions of Mission Soil Horizon Europe projects at the Mission Soil Week (EUSO Stakeholders Forum, meetings of the EUSO Working Group on Data Harmonization & European Mission Soil Week in Madrid November 2023). There is a rough estimation of hundreds of terabytes of soil data assets and knowledge assets generated within the Mission Soil, i.e. a full scope of potentially applicable assets. As far as we are aware, several projects mentioned e.g. that it was not feasible to upload their datasets to Zenodo as their size is bigger than 50 GB. Further discussion on this topic continues in section 3.1 Storage strategy.

Expected size of relevant data and knowledge generated by SWR will be updated in the following iterations, after the range of project processing functionalities will be defined (Repository architecture D1.3 [M8], D1.4 [M42]; Developed & Integrated DM components D2.1 [M13], D2.2 [M18], D2.3 [M31], D2.4 [M42]; Developed & Integrated KM components D3.1 [M13], D3.2 [M18], D3.3 [M31], D3.4 [M42]; Repository Data and Knowledge Resources D4.5 [M21], D4.6 [M34], D4.7 [M46]).

Expected size of external data collected and stored within SWR will be updated after the soil-related R&I projects screening and stakeholder feedback (Usage Scenarios, Requirements D1.1 [M6], D1.2 [M36]).

2.4 Metadata

Metadata is descriptive data about resources such as data(sets) and knowledge items. It facilitates the discovery of these resources in an infrastructure. Metadata, if available online, will reference at its original online location or be generated as part of discovery optimisation. The collected set of metadata (references) is a dataset itself, produced by the SoilWise project and will be archived at a yearly interval in a persistent repository, under a CC-BY 4.0 license.

For remote metadata to be included in the central catalogue index, the adoption of standardised metadata models by the relevant repositories is of interest, to prevent the central catalogue needing to introduce dedicated pipelines for each remote repository.

Depending on the choices of the data holder, there might or might not be sufficient metadata available for a specific source. In the case that sets of unstructured content are stored and published in a structured manner (e.g. CORDIS), the structuring might provide relevant metadata elements. These can be translated to an interoperable metadata format used in the SWR. For now, no specific metadata standards or formats are required to allow the linkage of knowledge sources to SWR.

SWR will define a common recommended metadata format for data and knowledge on soil, based on broadly accepted vocabularies such as Dublin Core. Dublin Core is a relatively simple standard, where most elements should be relatively easily mappable to unstructured content resources. Depending on specific (user) requirement from the soil community, and provided that they can be acquired from common resources, additional metadata elements could be added.

SWR semantics and knowledge graphs will facilitate the data to metadata linking by connecting to and employing relevant "external" ontologies, e.g. to EuroSciVoc for linkage to CORDIS knowledge assets.

2.5 Ownership and data rights

Intellectual Property Right (IPR) of the author always remains intact. The description of the IPR, as part of the metadata (defined for example with a machine-readable license or contract information), will be inextricably linked to the data during metadata submission. In this way, IPRs will be always visible to the user in case of reuse.

The GDPR management is exhaustively described in Chapter 13 of D7.1 Project Management handbook. Moreover, the deliverable IPR, Business Model Report and Policy Brief (D6.5, D6.6 [M24, M47]) are dedicated, among others, to report on the IPR.

Data holders sharing their data and knowledge with SWR will have a control of their assets and metadata.

2.6 Sensitive data and knowledge

There are different types of sensitive data that are potentially processed by SoilWise as well as by connected upstream or downstream services. Such sensitive data typically falls into one of the following categories:

- Personal data identifying a natural person by name, address, location or other means. This group can include farmers or other landowners, or a researcher or technician who worked on a data set or knowledge (governed by the GDPR);
- Other data linked or linkable to a natural person, such as financial information, transaction records or communication (governed by the GDPR);
- Data collected by personal devices, such as mobile devices and private weather stations, such as health data or meteorological parameters (governed by the Data Act and the Data Governance Act);
- Economically sensitive data that discloses proprietary information to third parties such as competitors, such as the detailed value of crops sold from any given plot of land or other financial information;
- Ecologically sensitive data such as the exact position of protected species of flora or fauna;
- Data where existing licenses/access restrictions only permit specific usage.

For each data asset catalogued in the SWR, the metadata shall indicate which category (categories) of sensitive data this asset belongs to, if any, with a short rationale which parts of it fall under which specific category. Available information on licensing and IPR will be harvested where applicable and possible and will be integrated into the metadata.

For the protection of knowledge resources, the SWR will rely on the mechanisms implemented by the holder of the knowledge assets to protect and/or describe the usage conditions of public, non-public or otherwise sensitive content. Content publicly accessible on the Internet will be considered accessible.

3 SWR storage and processing strategy

3.1 Storage strategy

In general, the SWR storage strategy is to store a **minimal amount of data and knowledge** within its own repository, but a duplicate of **all relevant metadata** harvested from external resources will be stored and changes logged. Such an approach will ensure efficient identification of duplicates and versioning, to assess the dataset and knowledge persistence through time. It allows tracking when the original asset is deleted or uploaded again in a new version at the original repository.

SWR will in principle not duplicate and store data and knowledge assets. We distinguish between data and knowledge from persistent sources and non-persistent assets. Exceptions could be made in specific cases, such as when the persistence of high value data/knowledge assets could otherwise not be guaranteed. This issue will be described in more detail in the following iterations of this deliverable. Some limited **derived data products** (e.g., interoperable data generated by SWP processing capabilities) are going to be temporarily stored in the SWR, or uploaded to Zenodo on demand.

Data and knowledge will be handled equally in terms of supporting their **persistence**. Persistent content is considered to be stored in a trustworthy, persistent repository. We expect those storages to store the asset compliant with the applicable legally and scientifically required terms and periods for storage of the content, and to use a DOI or other persistent URI for persistent identification. These can be safely referred to from the SoilWise catalogue. For long-term preservation and availability of data and knowledge assets, SWR relies on the repository holders and their responsibility to keep it available. In this sense, the Mission Soil projects will be encouraged to archive their data and generated knowledge primarily to Zenodo, or alternatively to other compliant storage infrastructure (e.g. the official repository of European Data, trusted repositories registered at Directory of Open Access Repositories or Re3Data), to preserve their accessibility after the project ends. SoilWise will provide cookbooks on data handling, i.e. manual on how to upload data to Zenodo, to support findability, accessibility and reusability, and proper linkage to SWR. Annex I shows an example of such manual.

Non-persistent data and knowledge are the ones that are not guaranteed to persist by the repository or data and knowledge holder and/or does not guarantee a persistent URI for reference for at least 10 years. In practice many non-persistent knowledge sources and assets exist that could be relevant for SWR, e.g. on project websites, in online databases, at computers of researchers, etc. Due to their heterogeneity in structure and underlying implementing technologies etc., it is not possible nor desirable to store those in the SWR, with exception of high value data/knowledge assets.

From a sovereignty and control perspective, the ideal situation is that all data and knowledge reside at their respective sources and are linked to the SoilWise catalogue by metadata. These sources might be 3rd party project storages with long-term data preservation, or existing Open Access repositories such as OpenAIRE and Zenodo.

3.2 Data and knowledge logging

The SWR will systematically search for available data and knowledge, including their updates, in the identified resource repositories. As such, SWR will be capable of e.g. keeping an overview of the portfolio of available soil data and knowledge, their availability and updates. SWR will log to the system all relevant soil-related data and knowledge in the form of metadata. Whenever possible, metadata will be harvested from source repositories automatically. A link persistence validator component will determine the lifespan of the linked data and knowledge, i.e. if the link is still active or not.

The following events related to both data and knowledge management processes will be logged:

- Date of first / last harvesting/generation of the metadata
- Persistency (Y/N)
- Date of (permanent) breakage of the link to a non-persistent asset
- Data quality and completion status
- Source of data/knowledge origin
- The version and standard of the semantics used to index the knowledge asset
- Metadata record sensitivity and licensing

The possibilities and need of unique identification of versions of metadata for external resources by SoilWise will be examined and described in following iterations of this document. However, it is expected that versions of metadata are stored and can be retrieved by their date of harvest/generation.

3.3 Persistent identification strategy

Resources included in the SWR are referenced by their external persistent identification. If a remote resource uses a remote identification framework, such as DOI or GDI-DE registry, then preferably the identification provided by this service is used.

For resources produced by the SWR or external resources which do not include a unique identification, the SWR suggests assigning an identification based on the DOI service.

If data or knowledge is processed within the SWR, for which the results need to be persisted, an identifier for that resulting resource will be generated using the DOI service. References to the data and knowledge used as a source are added to the resource metadata using their persistent identification. An approach towards assets without identification (for example because their source is dynamic) will be examined and described in following iterations.

3.4 Storage of sensitive data and knowledge

As outlined in section 2.6, there are different types of sensitive data or knowledge that may be processed or stored by the SWR. Generally, a Protection Level Analysis needs to be performed per asset to fully determine the specific measures that have to be applied. However, the following rules can be applied widely.

SWR will follow the minimalization of information approach, i.e. will only process and store sensitive data and knowledge – both of assets holders and of users of the SWR – where it is necessary, and only for the shortest useful duration. All sensitive data or knowledge that does need to be stored is protected through the following technical and organisational measures:

- Encryption in transfer: State of the art encryption is used to ensure the data or knowledge remains safe in transfer.
- Encryption at rest: State of the art encryption is used to ensure the data or knowledge remains safe while stored.
- Encryption of backups: Backups must also be encrypted.
- Safekeeping of keys: Keys to decrypt data or knowledge shall be kept in safe storage, such as GoPass, and shall only be accessed through the respective processes, not manually.
- Limited logging: Especially activities of anonymous users are not logged unless absolutely necessary.
- No direct access: No members of the project have direct access to sensitive assets, e.g. through denial of specific rights for their cloud accounts.
- No full export: There are no interfaces or other means that make it easy to fully export sensitive data, such as user lists.
- No acceptance of data or knowledge without clear sensitivity metadata: external data or knowledge that has no information on its degree of sensitivity is not stored or processed.
- No acceptance of personal data that is highly sensitive: generally, data assets that are to be processed and stored in SWR should be stripped at the source from personal data as far as possible or shall use location encryption or aggregation.

3.5 Data and knowledge processing

We foresee that the SWR will enable to manually ingest data and knowledge for their processing within the SWR. Such a functionality is challenging when considering the significantly limited storage capacity of the SWR. As such, on-the-fly processing is the preferred option, i.e. without allowing a user to store the results within the SWR.

The primarily focus will be on the processing capabilities towards increasing interoperability. The detailed technical specification remains a subject of ongoing discussions with all the stakeholders. Based on our pre-study, we see two major candidates as a content data processing/transformation standard: processing to increase interoperability towards conformance to (1) INSPIRE and (2) GloSIS. Both initiatives went through intensive developments, both are interoperable to ISO 28258 and both have their conceptual models expressed for relational and semantic equivalents (note, INSPIRE semantics as well as a data model for grids remain under development, i.e. very recently a GeoPackage implementation has been submitted as Good Practice). Moreover, data processing should in the future comprise unit conversions, as also defined by requirements from stakeholder groups (Usage Scenarios, Requirements D1.1).

The SWR processing of knowledge focusses on 1) the crawling of identified knowledge resources to derive the metadata and index the content, through temporarily caching the unstructured content; 2) support of interoperability and reusability of knowledge through metadata schema transformation (for the time being Dublin Core, potentially extended with additional metadata elements), and 3) the indexation of the knowledge asset, using a feasible knowledge graph.

3.6 Back up strategy

Given that we assume the SWR to be built on Kubernetes and Docker containers, we expect to apply the following backup strategy:

1. Create volume snapshots frequently (30- or 60-minutes interval, depending on the system component) and store these for at least 30 days,
2. Regularly perform recovery exercises from these snapshots,
3. Create asset and database backups and store these in a different location/platform for at least 90 days.

4 Provision of data and knowledge by the SWR

This chapter describes mainly the principles on which SWR will promote FAIRification of logged assets, i.e. external data and knowledge, and SWR derived assets.

Note that SoilWise catalogue described in this chapter is part of the wider SWR system. SoilWise architecture deliverable (D1.3, D1.4 [M8, M42]) will delineate role of SoilWise catalogue and its relation to other SWR components.

4.1 Data and knowledge types and formats

Data and knowledge types and formats will be dependent on the external inputs to SWR (section 2.2 Data and knowledge types and formats), and on SWR Data and knowledge processing (section 3.6 Data and knowledge processing). The specific formats from external sources, as well as formats needed by stakeholder groups (and therefore considered during processing components development), will be defined during the soil-related R&I projects screening and the stakeholder elicitation process (T1.1).

The SWR data and knowledge processing will support the interoperability principles described in section 4.4 Interoperability, with emphasis on, among others, 5* Open data standards.

4.2 Findability

The findability of soil data and knowledge will be supported through the coherent use of persistent identifiers (see section 3.3 Persistent identification strategy) to unambiguously identify the data and knowledge, and facilitate citation, through the SoilWise catalogue and rich enhanced metadata (e.g., Dublin Core/RDF, DataCite, ISO19100 series incl. INSPIRE modifications, etc.). The SoilWise catalogue will be based on a set of open-source libraries and a web application for the management and discovery of geospatial (meta)data, supporting, among others, user-friendly data preview, data download and validation log. From a user perspective, it represents a tool for searching and finding relevant resources, such as geospatial and non-geospatial datasets, Web services, sensor measurements, map compositions, models, documents, Web pages, reports, legislation, or an e-shop. As for knowledge, we will explore the possibilities for persistent identification strategy and other functionality on both sides, on the side of the knowledge source and the side of the SoilWise catalogue. There is some documentation on repository knowledge handling, such as OpenAIRE, however, we need to examine the applicability of solutions on knowledge as defined in SoilWise. SWR will provide an important context for the interpretation of the data and knowledge and make it easier for machines to conduct automated analysis, powered by AI- and ML-means.

The datasets published in SoilWise catalogue will be semantically interlinked to relevant datasets, knowledge, and other sources, such as related R&I projects (thanks to the connection to the CORDIS), derivate or parent datasets, methodology documents, websites, etc. A map preview of a resource (dataset/knowledge/service/…) will be available from the source graphic, if applicable.

Finally, as for dataset and knowledge versioning, it is possible to specify the version of the dataset or knowledge as part of the metadata included with the data. Therefore, any published asset will include this information. If we want to keep different versions of the asset, it is possible to add different files to the asset itself, indicating the version of each one.

4.3 Accessibility

A lot of data and knowledge that will be linked to the SWR is owned by either public or private organisations. Access rights and usage licensing will be aligned with the policies of these stakeholders. The main provisioning point of SWR will be the SoilWise catalogue, as described in section 4.2 Findability. The SoilWise catalogue will provide the links which will enable the access to the data and knowledge itself, including links for direct download.

Once the metadata is published in the SoilWise catalogue, users will be able to navigate through the linked information and access or download data or knowledge from the source. The SoilWise catalogue will also provide APIs accessible by other systems. The foreseen APIs for the SoilWise catalogue are OGC Catalogue Service for Web and OGC API – Records. As for the data and knowledge gathered from external sources, before making it available, the SWR will support automatised filtering and transformations required for being compliant with GDPR and any related legislation.

Data and knowledge generated by SoilWise will be primarily stored at external open repositories created for such a purpose, e.g. Zenodo/OpenAIRE. Metadata in the SoilWise catalogue will explicitly contain links to external open repositories where the relevant data can be found.

SoilWise components will be primarily based on open-source products to increase accessibility. We will use GitHub to document progress of SoilWise developments, publicly available at: github.com/soilwise-he.

4.4 Interoperability

To make data FAIR, interoperability is essential. There are several aspects to achieving interoperability:

Shared Semantics

Especially in a varied and complex domain such as soil, shared vocabularies, conceptual models and their implementations into data (exchange) models are essential for achieving any kind of data interoperability and data re-usability. Such conceptual models and vocabularies are available from INSPIRE Soil, GloSIS, SoilML, SOTER, national soil data standards and many others, as listed in ISRIC's 'International' soil standards list. These often include the data representations of soil description and classification standards, such as the FAO Guidelines for Soil description, the World Reference Base (WRB), the USDA soil taxonomy, national soil description and classification systems etc. And many of them use OGC Observations and Measurements or SOSA to model some of the concepts in the model in a similar way. Within the project and the repository, we will use only such models and vocabularies that are well-established and published in such a way that they are machine-readable and have a persistent, versioned, ideally resolvable, identifier to each term or concept. Likely candidates are the INSPIRE Soil and GloSIS soil data exchange models, both are based on the ISO28258 model and can therefore relatively easily be mapped or translated to each other.

Open Formats and Encodings

Open Formats enable a wide range of systems and software to read or write the data. Ideally, existing support for these formats is widespread and complete, and the formats support the features of the conceptual model well. In SoilWise, we will select appropriate open formats for the delivery and ingestion of data, such as those specified by the IETF, W3C or OGC. These standards will have at least three stars according to the 5-Star Open Data model.

Open, standardised APIs

APIs enable software and systems to access the data. When using existing standards, existing software can directly access the data resources, with no special integration needed, SoilWise will use standardised APIs as specified by the OGC and other relevant standardisation bodies, such as OGC WFS.

Metadata

Metadata describes data assets, clarifies their content, lineage, quality, and helps making the asset discoverable. In SoilWise, we will fully build on well-established metadata standards, such as:

- Dublin Core structure in XML and JSON
- INSPIRE (ISO 19115 and 19119) in XML and JSON
- DCAT-AP in XML
- IDSA Asset and Endpoint Metadata as RDF

Interoperability is also not absolute – there is a spectrum ranging from a very lightweight Minimum Interoperability Mechanism to full interoperability that starts at data capture and then continues through the full lifecycle. In the case of data and metadata provided through the repository, we aim for full interoperability for metadata and APIs, and aim for an extended Minimum Interoperability Mechanism (MIM) for data.

In this soil data MIM, we require that any data sets described by metadata in the SWR also provide, for each object, a machine-readable annotation to the data types/tables in the dataset, with the following information:

Data holders have to publish machine-readable annotations of as-is data assets in terms of the commonly agreed Vocabularies of the SWR that their data fits with. These annotations have to cover the relevant data classes defined in the dictionary, the mandatory attributes of these and the value mapping to the defined controlled vocabularies. Such mandatory attributes are usually:

- the persistent unique key property
- the classifying properties defined in the shared vocabulary, if present
- the main measurement, if present

The goal of this Soil Data MIM is to ensure that a minimum amount of information and knowledge, e.g. for soil health monitoring, can be done in an automated, low effort manner. The majority of detailed data will not be affected by this mechanism but can be transformed on demand. This on-demand transformation will mostly work between standardised data models (GloSIS, INSPIRE, national models, ...) where a mapping or transformation is made once and when the source and target model is indicated can be executed automatically, but data holders can also bring their own data harmonisation processes, e.g. as hale»studio transformation projects.

4.5 Reusability

The interoperability actions, SWR data processing, and open science approach will set suitable conditions for reusability of collected and generated data and knowledge beyond project partners, for other users and beyond project life.

The data and knowledge owner states his or her IPR for research data and knowledge while selecting a proper Creative Commons license model in the metadata. IPRs of the author remains intact. The CC-BY is recommended as a standard license for all research data and knowledge stored in the SWR repository.

External data or knowledge without explicit information on license will not be made publicly available within the SWR.

The SoilWise-generated data and knowledge will be well-documented and will have clear licence and provenance information. By default, open access data and knowledge will be made available to the public for reuse following appropriate licensing schemes (incl. Open Data Commons licenses). The licensing of derived data and knowledge will be in line with licenses of original data, defined by the providers. The restrictions on reusability of any data or knowledge will be as defined by the license.

To support the reusability of data and knowledge, the consortium will also exploit and create new networks to maximise the visibility. WP6 is dedicated to, among others, set up an interactive stakeholder network, considering national and subnational initiatives across Europe.

4.6 Updated and faulty research data or knowledge

Regarding research data and knowledge that have been allocated by SoilWise, a DOI will not be altered.

Handling of errors occurring in the data or knowledge, or other cases of faulty assets, including the responsibility issues, is foreseen to be described in deliverable Innovative governance model for open and accessible knowledge (D1.5, D1.6 [M21, M42]).

5 SWR support to EUSO

The connection between the EUSO (EU Soil Observatory) and SWR is described in a standalone section due to the specific position of the EUSO. The SWR will be part of EUSO 2.0, and therefore, coordination activities are needed to support the cooperation with ESDAC (European Soil Data Centre) and EUSO. Note that more information on the SoilWise – EUSO – ESDAC cooperation is available within the section '6 User groups JRC and EUSO' of the SoilWise Usage Scenarios, Requirements (D1.1) document.

The SWR is foreseen as a key data and knowledge hub to support further growth of the EUSO in terms of extended functionality and a rising number of users. The SWR is considered a backend hub of the EUSO. As such, the EUSO users will use the SWR functionality and search through the SWR metadata holdings and enhanced knowledge and data sources. Moreover, the SWR acts as a broker in terms of data findability, accessibility, interoperability and download. Note that governance models are beyond this deliverable's scope, as they are foreseen within the SoilWise D1.5&D1.6 documents [M21, M42]. The SWR will benefit from the EUSO frontend, including its authentication layer, enabling a single sign-on approach.

To understand the JRC requirements well from the beginning of the project, a meeting took place before the kick-off of the SoilWise project in ISPRA (February 2023) between the JRC and WP leads of the SoilWise project. This resulted in documented requirements from JRC for the repository. These requirements cover the expected source data and knowledge, desired results of the Mission Soil projects, data and metadata, knowledge, storage, discovery and publishing services, AI and ML services and overall ESDAC, EUSO, SoilWise cooperation. Discussing the JRC requirements early on allowed them to be used well in preparing and implementing the co-design process (Usage Scenarios, Requirements D1.1, Repository architecture D1.3).

The desired SWR support to EUSO and specifically its consequences to data and knowledge management will be the subject of the second iteration of this deliverable, as the EUSO 2.0 developments are still in progress.

6 SoilWise project-related data and knowledge management

6.1 Project's data storage strategy

Project documentation is stored on Microsoft Sharepoint at ILVO, which only project partners can access. Deliverables are also uploaded to SyGMa (System for Grant Management) and stored for on demand access (public documents) by REA and other EC stakeholders. Three months after project finalisation the archivable content (deliverables) is archived to repositories (open/restricted, depending on the origin of data) and to ILVO's servers for archiving, with a due description of the content in the metadata. Some partners may have dedicated storage for draft documents and meeting notes related to the SoilWise project. This case is beyond the scope of this document, a dedicated data management plan is expected to be available for such a scenario.

Some work items are stored on Github. It contains configuration of service deployment and container images, source code of applications, documentation, user feedback and task definition and planning. Ownership to the SoilWise-HE repositories may be handed over to JRC or project partners after project finalisation. A snapshot of the Github contents will be archived to Zenodo at three months after the project finalisation.

Content of the project website is stored in a mysql database. At project finalisation the content of the website will be archived to Zenodo. The soilwise-he.eu domain is expected to stay operational for the next 10 years. The option to acquire ownership of the domain during or after 10 years will first be offered to the former project partners, REA and JRC, before handing it back to the market.

The SoilWise project is active on social media (LinkedIn, Facebook, X). After project finalisation the activity on LinkedIn, X and Facebook will be archived to ILVO servers and available on demand for partners, REA and JRC, and the accounts are foreseen to be closed. The specific timespan is matter of ongoing discussion. If a project partner, JRC or REA expresses interest to continue the account, they receive ownership.

Usage statistics of the project website are stored on Google (Analytics), these statistics will be archived to Zenodo and removed from Google cloud at project finalisation.

Usage statistics of the SWR services are stored in plain form for up to 30 days. Aggregated summaries of the usage statistics are stored in a relational database and will be archived to Zenodo at project finalisation.

Availability statistics are collected in a cloud solution offered by Uptimerobot.com to optimise the operationalisation of the SWR and project website. These statistics are archived to Zenodo at project finalisation.

Note that all section of chapter 6 will be developed in following iterations of this document, after the launch of the first SWR prototype (M13 – September 2024).

6.2 The expected size of SoilWise data and knowledge

The expected size of data and knowledge within SoilWise project is listed in the Table 1.

Table 1 Expected size of data and knowledge within SoilWise

Type of data	Size	Lifecycle
Project documentation (GA, (draft)deliverables, meeting-notes, timesheets)	1Gb	Permanent
Ingested data from partners (metadata, data, knowledge)	20Gb	Temporary
Results after (meta)data processing	10Gb	Permanent
Community contributions (github/feedback/social media)	10Gb	Permanent
Plain usage statistics	1Gb	Temporary
Aggregated usage & availability statistics	250mb	Permanent
Project website	100mb	Permanent

6.3 Implementation of open-science practices

SoilWise is committed to fostering Responsible Research and Responsible Innovation, aiming to help instigate systemic changes in R&I systems by encouraging inclusiveness, openness, and knowledge sharing. The objective is to facilitate the co-development of responsible innovations, encompassing a broader spectrum of innovation processes beyond research-driven ones. Responsible Innovation aligns with economic, social, ethical, and environmental objectives, incorporating material, organizational, and discursive dimensions. In line with the European Open Science Cloud's ambition, SoilWise supports the principle of being 'as open as possible' and aligns with the Open Science Policy under Horizon Europe.

Open science actions will be integral to SoilWise's methodology, implementing a **co-creation approach** that actively involves stakeholders from the project's inception. This ensures that developed solutions meet end-users' needs. Drawing inspiration from Responsible Research and Innovation (RRI) and Open Innovation, SoilWise employs a co-creation approach to define customized engagement strategies and agile methodologies, maximizing sustained social acceptance. Sensitive data gathered from involved end-users will be pseudo-anonymized for GDPR compliance where needed, processed, and stored for a defined period with prior ethical and legal approval (including informed consent).

SoilWise aims for **early and open sharing** of research results and data using the Open Research Europe scholarly publishing service for Horizon Europe. All SoilWise publications will be accessible through various platforms, including the project's website, EU open access publishing services (e.g., European Open Science Cloud, OpenAIRE), and research community services (e.g., ResearchGate or Academia).

To **promote science literacy**, SoilWise actively engages stakeholders to enhance their access to and understanding of scientific information. Summaries and results from co-creation and Responsible Innovation labs will be publicly available through the project's website, complying with GDPR provisions and licensed via a Creative Commons Licence.

Open access to research outputs and participation in open peer-review are emphasized, with machine-readable copies of manuscripts available on the website within six months of publication. The Dissemination leader will leverage EC open access portals and tools to make project knowledge widely accessible.

7 Allocation of resources

7.1 The cost to be FAIR

SoilWise is a Horizon Europe funded Soil Mission project. Therefore, it has an agreed and fixed budget, including personnel, traveling and other costs. The successful implementation of the SoilWise repository, particularly adhering to the FAIR principles – Findable, Accessible, Interoperable, and Reusable – must be achieved within the confines of a pre-determined and fixed budget provided by the European Commission. The estimated cost associated with various components of data and knowledge management (WP2, WP3), repository infrastructure (WP4), personnel (consortium partners) and technological tools will be transparently outlined at the reporting periods and will respect the set of financial boundaries.

In this context, SoilWise aims to best adhere at the FAIR principles, meaning the project partners will prioritize on the technologies and strategies that best fit this approach. The SWR will be integrated into EUSO under the management of JRC which are following closely the project development and giving input in the process. The costs and resources needed to sustain and further develop the platform past the SoilWise project cycle will be taken in consideration during the four years of iterative development.

The development of the platform will use GitHub platform for issue tracking and project management which will incur no additional cost. These platforms will also allow seemingly transition to EUSO, as JRC is utilizing the same software.

7.2 Responsibilities for Data Management

The overall responsibility for data management in the SoilWise project will be on the Project Coordinator (EV ILVO), assisted at several levels by the Technical Manager (EV ILVO) and the Scientific Coordinator (MU), with input from the Work Package leaders.

As SoilWise project interacts with the Mission Soil projects, it has the objective to establish a direct liaison with REA and JRC to facilitate communication, collaboration, and alignment of the repository with evolving programme standards and objectives. Therefore, the EC officials have a certain responsibility for the SoilWise DMP. As the SWR will be integrated to EUSO at the end of the project lifecycle, JRC will have evolving responsibility in steering, managing and completely owning the final product. Regular meetings are being set between SoilWise and JRC where relevant topics of the DMP will be discussed.

SoilWise will designate a Data Manager who will be the contact person with the Mission Soil projects and will have the responsibility to update the DMP continuously, releasing the updated version (public documents) at the set timeline of the deliverables. The Data Manager will also communicate and organize the work package partners in the development of the platform, adhering to the FAIR principles and using the DMP as an alignment document between the parties.

WP2 and WP3 will develop the required data and knowledge components as defined in WP1 that will be used for populating and validating the repository infrastructure (WP4) and finally demonstrate its functionality with the users (WP5). Therefore, the DMP will be the foundation encompassing several of the strategies that will guide prioritisation of certain technical components for each iteration. The development will be done in GitHub where clear roles and responsibilities, together with timelines and objectives will be highlighted. All participants will have access to raise issues and give feedback, while the Data Manager will continuously update the plan in this dynamic, agile approach. For more information see the section '7 Validation Framework' of the SoilWise Usage Scenarios, Requirements (D1.1) document.

7.3 Cost and potential value of long-term preservation

This section follows the layout of section 7.1 The cost to be FAIR, acting under the fixed budget allocation agreed in the grant agreement. The persistence of data and knowledge is an important topic in the SoilWise project, and it leads to selecting certain technical components and strategies that efficiently ensure the long-term accessibility of the data and knowledge. This section will be further expanded in the next versions of the DMP (D7.3, D7.4 [M27, M48]). Besides this top priority in the project, the SWR will be integrated in EUSO, meaning the cost of long-term preservation will be fully taken up by JRC. In that sense, the development of the platform will be in close collaboration with JRC so that the effect of the after the project cycle integration will be seamless and with clear expectations in terms of costs.

8 Issues, risks and mitigation actions

Implementing Open Science and a robust DMP entails addressing various issues and risks to ensure the effective sharing and preservation of research data and of course means of mitigation. Recognizing and addressing risks requires a holistic approach that combines technological solutions, capacity building, and adherence to best practices and regulatory requirements.

The Risks, impact and mitigation plans are summarised in the table below and further described in the following sub-chapters.

Table 2 Risks and mitigation

Risk description	WP (1-7)	Likelihood (Low-Medium-High)	Impact (Low-Medium-High)	Severity of occurrence (Low-Medium-High)	Mitigation plan
The deviation from a common Data Management Plan (DMP) **implicating stakeholder confusion**	7	Low	Medium	High	To provide a detailed and transparent explanation for the chosen DMP structure. To offer examples or visual aids to help stakeholders grasp the benefits of the selected structure in facilitating the project's goals.
Delays in project timelines and outcomes due to SWR linkage to external repositories	2, 3, 4, 6, 7	Medium	Medium	Medium	To identify key stakeholders involved in the discussions about SWR linkage. Engage in proactive communication with these stakeholders to ensure their perspectives are considered. To anticipate potential technical challenges associated with the integration of SWR with external repositories.
Ineffective data integration and interoperability, due to diverse (meta)data models/	2, 3, 4, 6	High	High	Medium	To establish a comprehensive (meta)data mapping strategy to identify commonalities and differences between different standards.

lack of data holder cooperation					
Stakeholders not willing or able to standardise and expose their data and knowledge bring to **data security and privacy considerations**	2, 3, 4, 6, 7	High	Medium	High	To provide information on SoilWise/SWR in an open manner, including good practices introduction. To implement advanced data anonymization techniques to protect the identity of individuals associated with sensitive data. Implement useful function for strict access controls and authorization mechanisms to limit access to sensitive data only to authorized personnel and safeguard the users.
The expected size of soil data assets can lead to **challenges related to storage and processing capabilities**	2, 7	Low	Low	Low	To re-use open and trustful external repositories, where Zenodo seems the first choice. To regularly assess the project's storage and processing requirements in response to evolving data volumes. Implement a dynamic infrastructure strategy that scales alongside data growth.
Lack of awareness and training among researchers regarding Open Science principles and proper data management practices.	1, 2, 6	Medium	Low	Medium	Re-use already well-established established dedicated resource centers or online portals that serve as a centralized repository for Open Science guidelines, toolkits, and case studies.
Lack of user engagement poses a significant challenge to the **viability of a repository**,	1, 6	Medium	Medium	High	To develop the SWR hand-in-hand with the EUSO that already has hundreds of thousand users. To develop the repository with a user-centric design. Solicit feedback from potential users during the development phase to incorporate

					their preferences and requirements.
IPR conflict between partners and non-compliance with established policies could bring to **(legal) vulnerability of a project and SWR to data loss.**	2, 3, 4, 7	Low	Medium	High	To establish clear Intellectual Property Rights (IPR) policies at the outset of the project. Clearly define ownership, usage rights, and responsibilities for each partner. Ensure that these policies are communicated and agreed upon by all partners. To ensure that there is a valid lawful basis (consent, contract, legal obligation, vital interests, public task, legitimate interests) for processing personal data.
SWR's functionalities not being appropriate for the needs of the target groups.	1, 6	Medium	High	Medium	To follow iterative development, in which each development cycle also comprises stakeholders's feedback. To conduct thorough assessments and surveys to understand the specific needs and preferences of the target user groups. Regularly gather feedback to inform continuous improvement and updates to SWR functionalities.
Chosen technologies and tools for the SoilWise Repository fall short of expectations due to unforeseen difficulties in development	2, 3, 4	Medium	High	Medium	To adopt agile development methodologies to facilitate flexibility and responsiveness to changing requirements. Frequent iterations and feedback loops allow for quick adjustments, ensuring that issues are addressed promptly without causing significant disruptions.
Incomplete or inaccurate metadata could hinder the usefulness of data	2, 3, 4	High	Medium	Low	To establish and adhere to standardized metadata formats and guidelines for documenting data. Ensure that all contributors are aware of and follow these good

archived in Zenodo, and retrieval may become challenging if not well-documented					practices and standards to maintain consistency across the archived datasets. Utilize widely accepted metadata schemas to enhance interoperability.
Shortfall in data availability	2, 3, 4, 6	Low	High	Medium	To prioritize data quality over quantity. To use data replication techniques to duplicate datasets across geographically dispersed locations.
Budget constrains related to the funds already allocated, while technical risks could bring additional efforts which require funds	2, 3, 4, 7	Low	Medium	High	To implement efficient resource allocation strategies to maximize the impact of available funds. Prioritize initiatives that deliver significant value and align with the primary goals of the SoilWise project.

8.1 Critical issues

Firstly, the deviation from a common DMP structure introduces potential challenges in terms of understanding, implementation, and collaboration. Common DMP structures typically follow established guidelines and standards, ensuring a uniform approach across projects. The departure from this norm may create **confusion among stakeholders** who are accustomed to conventional formats.

The SWR linkage to data spaces introduce a level of uncertainty that could potentially lead to **delays in project timelines and outcomes.** Data spaces, such as the Green Deal Data Space and Agriculture data spaces, play a crucial role in the integration and utilization of data. Any ambiguity or indecision regarding the SWR's connection to these spaces may hinder the project's overall progress.

Additionally, **data security and privacy considerations** pose a risk, especially when dealing with sensitive information. Standardization becomes imperative to provide a uniform structure, facilitating easier integration and interpretation of data across diverse platforms. For instance, in the context of the SWR, the emphasis on standardization aims to interconnect diverse entities within the soil science domain. By providing consistent metadata, the repository will facilitate efficient collaboration and will ensure that data is not only shared but also comprehensible and applicable across different research initiatives and platforms. In the case that, for instance in SoilWise, countries and other stakeholders not willing or able to standardise and expose their data and knowledge, this would be a significant issue for the SWR. The acknowledgment of various categories of sensitive data within the SoilWise project highlights the need for careful handling to comply with regulations and ethical standards. Failure to adequately address the sensitivity of certain data types may result in legal consequences, ethical dilemmas, and potential harm to individuals or entities associated with the data.

The expected size of soil data assets reaching the scale of hundreds of terabytes introduces potential **challenges related to storage and processing capabilities**. Traditional infrastructure may struggle to efficiently handle and manage such large volumes of data, leading to potential bottlenecks and performance issues.

On the other hand, there is a significant threat, that a lack of user engagement will make the repository less viable. A lack of user engagement poses a significant challenge to the **viability of a repository**, hindering its collaborative potential and diminishing its overall effectiveness. To address this issue, it is imperative to establish clear communication channels, ensuring that the repository's purpose, guidelines, and usage instructions are well-documented and easily accessible. Building a sense of community around the repository through open communication platforms, such as forums or discussion boards, fosters collaboration and encourages users to share ideas. Implementing effective feedback mechanisms, maintaining regular updates, and ensuring adding value services and user-friendly interfaces contribute to an active and dynamic repository.

The expected size of soil data assets reaching the scale of hundreds of terabytes introduces potential challenges related to **storage and processing capabilities**. Traditional infrastructure may struggle to efficiently handle and manage such large volumes of data, leading to potential bottlenecks and performance issues.

An additional critical issue lies in the potential **lack of awareness and training** among researchers regarding Open Science principles and proper data management practices. A major problem is that many researchers might not know much about Open Science principles and how to handle data properly. Though improving, there is a lack of awareness and training in the research community regarding these important aspects. This means that quite a few researchers may not be familiar with the basic principles of Open Science or have the skills needed to manage data effectively. Closing this knowledge gap is crucial to create a research environment where everyone understands and follows Open Science principles and knows how to handle data securely.

Non-compliance with established policies and regulations increases the **vulnerability of a project and SWR to data loss.** This includes **IPR conflict between partners** and failure to adhere to industry standards or legal requirements which may result in severe consequences to the project outcome.

Lastly, an issue could lie in the fact that the **SWR's functionalities are not appropriate for the needs of the target groups.** The aim is that the repository not only meets but exceeds the expectations of its users. Flexibility and adaptability in refining functionalities based on evolving user needs are key elements in maintaining the repository's relevance and ensuring its sustained value within the targeted user community. By prioritizing user-centric design and responsiveness, the partners will be able to change the functionalities or add additional data and knowledge sources to better serve the diverse and evolving needs of their intended audience, fostering a stronger and more impactful connection with SWR.

With a focus on FAIR principles, SoilWise prioritizes technologies and strategies that align with this approach. The commitment extends to integrating the SWR into EUSO, managed by JRC. JRC closely monitors the project's development, providing valuable input. The project acknowledges the fixed budget and strives to optimize costs while adhering to FAIR principles. The **fixed budget may limit the allocation of resources for the SoilWise project,** impacting the extent of features, functionalities, and scalability that SWR can achieve.

8.2 Technical risks

The reliance on standardized metadata models introduces a potential challenge when repositories employ different standards. Diverse metadata models can **impede effective data integration and interoperability,** causing inefficiencies and complicating the harmonization of information across various sources. This is a critical concern, since it could lead to **low quality and interoperability of data**, due to lack of standardized metadata which can, consequently, impede effective sharing and reuse as well. This could also occur due to the lack of collaboration of data holders and their willingness to cooperate. Also, **connecting existing repositories with SWR might face integration challenges** due to varying data formats and structures, therefore lack of standardization poses a lot of different risks to be addressed.

Facing a scenario where the initially **chosen technologies and tools for the SWR fall short of expectations** poses a substantial challenge in executing the project. This situation may arise from unforeseen difficulties in development, necessitating a revaluation of the system's functionalities and user requirements. Such challenges introduce the risk of rethinking technology choices, impacting data accessibility and project timelines. Additionally, it prompts a careful consideration of the contrast between established and new technologies, balancing the reliability of proven solutions against the potential benefits of exploring more innovative options. Furthermore, there is a need to weigh the safety and dependability of these choices, considering the trade-offs between the stability of established tools and the potential flexibility but higher risks associated with newer alternatives. To mitigate this risk, a systematic approach involves assessing the limitations of current technologies, seeking updates or alternative versions when necessary. The SoilWise team will also explore alternative technologies and tools, including open-source solutions, collaborations with other platforms, or the incorporation of emerging technologies, ensuring alignment with SWR's objectives. Thorough testing and pilot implementations play a crucial role in validating the effectiveness and compatibility of alternative technologies before contemplating their full-scale adoption in SWR. This strategy aims to fortify the project against unexpected challenges while maintaining a balance between stability, innovation, and adaptability. This could happen, if the team initially chooses a well-established relational database for the SoilWise Repository due to its proven stability. However, as the project evolves, they identify the need for a more flexible structure to accommodate diverse data types. This prompts a closer examination of newer NoSQL databases, weighing the stability of the established solution against the potential futureproofing and innovation.

Archiving data to Zenodo after project finalization is planned, but potential challenges may arise in ensuring comprehensive metadata and accessibility. **Incomplete or inaccurate metadata could hinder the usefulness of archived data, and retrieval may become challenging if not well-documented.** Social media accounts and website content will be archived to ILVO servers, and domain ownership may be offered to former partners, REA, and JRC after ten years and usage statistics from Google Analytics, SWR services, and Uptimerobot.com are stored temporarily or permanently and archived to Zenodo. Developing a comprehensive archiving process for all this content will be done, ensuring all relevant information is captured. We will document the criteria and process for domain ownership transfer in the project documentation, and regularly review and update the archiving and ownership transfer procedures.

The risk of facing difficulties in **connecting existing repositories with SWR** due to diverse data formats and structures suggests the need for the development and rigorous testing of robust data integration processes. Moreover, establishing proactive communication channels with repository owners is recommended to promptly address any integration issues that may arise during the process. Additionally, the risk associated with incomplete metadata in some data sources hindering effective integration is met with the mitigation strategy of providing guidelines and incentives for data providers to enhance metadata quality, accompanied by the development of tools for automatic extraction and conversion. Furthermore, to tackle the risk of evolving data processing functionalities leading to unexpected challenges, the mitigation involves regular updates to projections based on ongoing developments and subsequent adjustments to storage strategies. These mitigation measures collectively aim to streamline data integration processes, enhance metadata quality, and ensure the adaptability of the SoilWise project to evolving data processing functionalities.

An unexpected **shortfall in data availability** would pose a challenge that demands a strategic and adaptive response to ensure the success of the project. An assessment of existing data sources (see section 2.1) is crucial for the SWR and the project as a whole. Implementing targeted data collection strategies and engaging with stakeholders, firstly by mapping strategies and then engagement practices can help prevent this challenge, ensuring the project has the necessary information to meet its objectives. Prioritizing data quality over quantity is essential, necessitating the implementation of robust quality assurance measures and a data and knowledge governance strategy. Collaboration with data providers and the exploration of data-sharing agreements with relevant parties when licenses are insufficient can enrich the available datasets. Leveraging advanced technologies and tools for data analysis, adopting iterative planning, and fostering transparency in communication about challenges and mitigation strategies are important. Additionally, using data replication techniques to duplicate datasets across geographically dispersed locations could enable data finding. This approach enhances data availability by providing multiple access points, reducing the impact of regional outages or server failures.

The **challenge of manually ingesting data and knowledge** within SWR is amplified by the significantly limited storage capacity. To address this, a mitigation strategy focuses on prioritizing on-the-fly processing, thereby circumventing the need for users to store results within SWR. This approach ensures real-time analysis and utilization of data without overwhelming the constrained storage resources. Additionally, SWR will implement selective data ingestion mechanisms, allowing users to prioritize and ingest only the most critical or relevant data for immediate processing. Archiving policies and automatic data purging will be established to manage storage efficiently, retaining valuable insights while freeing up space. Leveraging compression techniques, cloud-based storage integration, and regular capacity assessments will further optimize storage usage. By providing user notifications and guidance on efficient data handling practices, the SWR aims to navigate the storage limitations effectively, ensuring optimal functionality and sustained performance for data processing and knowledge management.

8.3 Mitigation actions

To mitigate potential challenges and risks associated with the SoilWise project, various strategies have been recommended. Firstly, the divergence from a common DMP structure was identified, and the mitigation involves providing a clear rationale for the chosen structure to ensure stakeholder understanding. Ongoing discussions about the SWR linkage to data spaces were acknowledged as a potential source of uncertainty, and the mitigation suggests regular updates to stakeholders and active communication to address concerns. Standardized metadata models were recognized as a potential challenge, with the mitigation involving the development of conversion tools for compatibility. The expected size of soil data assets, posing challenges in storage and processing capabilities, suggests continuous assessment and infrastructure upgrades, considering cloud-based solutions for scalability. Finally, for sensitive data handling, robust anonymization, encryption practices, regular audits, and GDPR compliance measures were recommended. These strategies collectively aim to enhance the project's efficiency, stakeholder engagement, and compliance with data management and privacy regulations. To sum up all the vital mitigation actions in respect to the above-mentioned issues, implementing measures to address challenges is vital to guarantee the resilience, reliability, and compliance of the project’s organizational data management practices. This means **creating clear guidelines for managing data**

and knowledge and providing the necessary resources and tools to follow those guidelines. To further enhance the mitigation strategies for integration challenges within the SoilWise project, it is recommended to develop further or utilise tools specifically designed to convert metadata from various standards to the **common formats (Dublin Core and further selected ontologies)**. This proactive approach aims to ensure seamless integration by promoting uniformity in metadata representation across diverse sources. These tools would facilitate the conversion process, enabling a more efficient and standardized incorporation of metadata into SWR. By adopting this additional measure, the project can address the potential variability in metadata standards among different sources, ultimately contributing to a more cohesive and interoperable data management system. Automated data validation not only ensures that the metadata aligns with the specified standards but also guarantees the quality and reliability of the underlying data. This approach facilitates a more robust and standardized dataset within SWR. Moreover, the implementation of automated tools encourages the annotation of knowledge sources, promoting transparency and accountability in the data integration process.

Additionally, **automated data validation tools** offer a proactive approach by systematically checking metadata and data for accuracy and completeness and ensure data standardization, minimizing the likelihood of errors and inconsistencies, and encourage annotation of knowledge sources.

For contributing to the project's long-term success and effectiveness, it is essential to encourage **comprehensive training programs and project stakeholder workshops** to educate SWR users/researchers, fostering a culture that emphasizes the benefits of Open Science especially in the context of SWR. By incorporating training programs that specifically address the unique aspects of data handling within the SoilWise context, the project not only mitigates potential risks but also cultivates a community of researchers who are well-versed in the intricacies of SWR data. This proactive approach sets the stage for a sustainable and effective long-term strategy, where researchers play a crucial role in maximizing the value of the repository and ensuring the responsible and informed use of the data it contains. The implementation of such efforts can be correlated with SoilWise's participation in Mission Soil Platform Clusters.

In the context of mitigation, **fostering collaboration with stakeholders** becomes an indispensable element during the reassessment process. Engaging various stakeholders, including technology experts, researchers, and end-users, is essential as it adds a layer of comprehensive insight that aids in identifying potential risks and implementing effective mitigation strategies. These collaborative efforts serve as a proactive measure to address challenges and uncertainties that may arise during the reassessment of technologies within the SoilWise project. The involvement of technology experts ensures a deep understanding of the intricate technical landscape, allowing for the identification of potential pitfalls and the exploration of innovative solutions. Researchers contribute valuable domain-specific knowledge, providing insights into the compatibility of technologies with the scientific objectives of the project. End-users, on the other hand, bring a practical perspective, offering feedback on the usability and real-world applicability of different technologies. This collaborative approach not only widens the scope of expertise but also enhances the practicality of mitigation efforts. By actively seeking input from diverse stakeholders, the SoilWise project can tap into a wealth of collective knowledge, anticipate potential challenges, and develop mitigation strategies that align with the needs and expectations of those directly involved in the project. Additionally, this collaborative reassessment process ensures that technology choices remain grounded in practical considerations, promoting the successful integration and adoption of technologies within the project's ecosystem.

In addressing security concerns, particularly from a technical perspective, the SoilWise project emphasizes the implementation of **robust security measures as a fundamental mitigation strategy.** This involves the incorporation of advanced encryption techniques and stringent access controls to fortify the protection of sensitive data. The utilization of encryption ensures that data remains confidential and secure during storage, transmission, and processing. Concurrently, well-defined access controls are instituted to restrict and regulate the permissions granted to individuals or systems attempting to interact with the data. Furthermore, the project places a strong emphasis on establishing clear and comprehensive protocols for the handling of sensitive data. These protocols are meticulously designed to align with both ethical principles and legal standards, ensuring that all actions taken with regard to sensitive data are conducted in a manner consistent with established regulations. By delineating precise procedures for data access, storage, and processing, the project aims to mitigate risks associated with unauthorized access, data breaches, and potential ethical lapses. This comprehensive technical approach not only safeguards the integrity of the SoilWise project but also underscores its commitment to ethical data management practices and compliance with legal frameworks.

Finally, to **navigate budgetary constraints,** the SoilWise project has devised a comprehensive set of mitigation strategies. First and foremost, the project prioritizes essential features by identifying and focusing on core components that align with its mission and objectives, allowing it to stay within the confines of a fixed budget. Additionally, the team emphasizes efficient resource allocation, ensuring that available funds are maximized by prioritizing initiatives that deliver significant value and align with the primary project goals. Collaborative partnerships are actively fostered to supplement resources and expertise without incurring significant costs. Leveraging open-source technologies helps reduce licensing and development costs, providing cost-effective alternatives while maintaining SWR's capabilities. Overall, these mitigation strategies collectively empower the SoilWise project to navigate budgetary limitations effectively while ensuring the successful development of the SWR.

9 Data Protection / Security

The partners involved in the SoilWise project pledge to uphold the security and confidentiality of all sensitive data and knowledge acquired during the project's duration, ensuring it remains inaccessible to unauthorized individuals. Stringent measures, including appropriate confidentiality practices and adherence to technical security protocols, will be employed in the handling of this information. Compliance with pertinent EU legislation, namely GDPR, Data Act, and AI Act, will be a paramount consideration. It is worth noting that this commitment may necessitate distinct approaches to security and sensitive data and knowledge management to prevent any potential confusion, as not all data and knowledge may require the same level of security measures.

The SoilWise architecture is designed to uphold access control policies, offering versatile and fine-grained methods to allocate permissions. This ensures that access to resources is effectively regulated for any potential scenario. A robust data protection strategy will be employed to securely manage data throughout the entire project lifecycle. Consortium partners will determine specific security controls for each phase, evaluating the level of compliance as the project evolves. Enhanced data protection is ensured through various procedures and technologies, including the utilization of the HTTPS protocol to encrypt all internet transactions.

Sensitive data and knowledge handling and storage strategy are explained in more detail in sections 2.6 Sensitive data and knowledge, and 3.5 Storage of sensitive data and knowledge. Frequent backups of the data stored in the SWR will be implemented to anticipate data recovery. These backups will be stored using a similar storage technology but will be deployed in a distinct ICT environment to that of the SWR.

10 Ethical Aspects

All planned and suggested activities within the framework of the SoilWise project must adhere fully to the regulations outlined in **Article 14 of the Grant Agreement - Ethics and research integrity.** This article explicitly mandates that all actions be conducted in accordance with ethical principles and relevant EU, international, and national laws, as well as the EU Charter of Fundamental Rights and the European Convention for the Protection of Human Rights and Fundamental Freedoms and its Supplementary Protocols. The article clearly emphasizes the importance of adhering to the principle of proportionality, the right to privacy and protection of personal data, the physical and mental integrity of individuals, non-discrimination, and the necessity to ensure environmental protection and a high level of human health protection.

All tasks outlined in the SoilWise project are designed to comply with relevant laws and regulations. Where applicable, project partners are required to secure informed consent from project participants. Project management will ensure that necessary procedures are implemented and followed, particularly focusing on obtaining, collecting, and preserving consent forms from project participants before initiating the data collection process when needed.

The obligations related to maintaining confidentiality, as specified in **Article 13 of the Grant Agreement**, require all partners to comply with EU standards concerning ethics and data management. Activities that pose ethical concerns, such as conducting surveys or interviews involving the acquisition of personal information, must meet additional requirements outlined in **Article 25 of the Grant Agreement.**

While no other forms of ethical issues are anticipated at this project stage, project partners are not exempt from implementing additional measures should unforeseen ethical issues arise during the project's duration.

References

5 Star Open Data website, retrieved from https://5stardata.info/en/

AZcademia website, retrieved from https://www.academia.edu/

Benchmarks website, retrieved from https://soilhealthbenchmarks.eu/

BONARES website, retrieved from https://www.bonares.de/

Bouman, J., & Veerman, C. P. (2022). Developing Management Practices in: "Living Labs" That Result in Healthy Soils for the Future, Contributing to Sustainable Development. Land, 11(12), 2178.

Creative Commons website, retrieved from https://creativecommons.org/

DataCite website, retrieved from https://datacite.org/

DataVerse Project website, retrieved from https://dataverse.org/

DublinCore website, retrieved from https://www.dublincore.org/

EJPSOIL website, rereived from https://ejpsoil.eu/

EOSC Association website, retrieved from https://eosc.eu/

EuroGeoSurveys, GS Soil, retrieved from https://eurogeosurveys.org/projects/gs-soil/

European Commission, CORDIS website, retrieved from https://cordis.europa.eu/projects

European Commission, Destination Earth website, retrieved from https://digital-strategy.ec.europa.eu/en/policies/destination-earth

European Commission, European data repository, retrieved from https://data.europa.eu/en

European Commission, EU Mission: A Soil Deal for Europe, retrieved from https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/soil-deal-europe_en

European Commission, EU Soil Observatory (EUSO), retrieved from https://joint-research-centre.ec.europa.eu/eu-soil-observatory-euso_en

European Commission, EU Vocabularies, retrieved from https://op.europa.eu/en/web/eu-vocabularies/euroscivoc

European Commission, GPR, Data Protection in the EU, retrieved from https://op.europa.eu/en/web/eu-vocabularies/euroscivoc

European Commission, Infrastructure for Spatial Information in Europe (INSPIRE), retrieved from https://knowledge-base.inspire.ec.europa.eu/index_en

European Commission, INSPIRE GeoPortal, retrieved from https://inspire-geoportal.ec.europa.eu/srv/eng/catalog.search#/home

European Commission, Integrated Administration and Control System (IACS), retrieved from https://agriculture.ec.europa.eu/common-agricultural-policy/financing-cap/assurance-and-audit/managing-payments_en

European Commission, Joint Research Center, European Soil Data Center, ESDAC, retrieved from https://esdac.jrc.ec.europa.eu/

European Commission, Open Science, retrieved from https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en

European Commission, Open Research Europe, retrieved from https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en

European Environment Agency, retrieved from https://www.eea.europa.eu/en/analysis

European Open Science Cloud website, retrieved from https://eosc-portal.eu/

European Soil Partnership website, retrieved from https://www.europeansoilpartnership.org/

FAO, AGROVAC Multilingual Thesaurus, retrieved from https://agrovoc.fao.org/browse/agrovoc/en/

FAO, Global Soil Partnership, retrieved from https://www.fao.org/global-soil-partnership/en/

Green Deal DataSpace website, retrieved from, https://green-deal-dataspace.eu/

International Data Spaces, Dataspace Protocol 2024-1, retrieved from https://docs.internationaldataspaces.org/ids-knowledgebase/v/dataspace-protocol/overview/readme

International Science Council website, retrieved from https://council.science/member/iuss-international-union-of-soil-sciences/

International Science Council, CODATA, retrieved from https://codata.org/

ISLANDR website, retrieved from https://islandr.eu/

ISO, ISO 19115-2:2019, Geographic information Metadata, retrieved from https://www.iso.org/standard/67039.html

ISO, ISO 28258:2013 Soil quality, retrieved from https://www.fao.org/global-soil-partnership/areas-of-work/soil-information-and-data/en/

ISRIC website, retrieved from https://www.isric.org/

ISRIC, World Data Centre for Soils (WDC-Soils) https://www.isric.org/

JISC, Directory of Open Access Repositories, OpenDOAR, retrieved from https://v2.sherpa.ac.uk/opendoar/

Langner, T., Kamoun, S., & Belhaj, K. (2018). CRISPR crops: plant genome editing toward disease resistance. Annual review of phytopathology, 56, 479-512.

MARVIC website, retrieved from https://www.project-marvic.eu/

Montaldo, S. (2022). The Green Deal and the Case for a Soil Health Framework Directive. European papers: a journal on law and integration, 7(2), 527-532.

Panagos, P., Montanarella, L., Barbero, M., Schneegans, A., Aguglia, L., & Jones, A. (2022). Soil priorities in the European Union. Geoderma Regional, 29, e00510.

Research Gate website, retrieved from https://www.researchgate.net/

Open AiRE website, retrieved from https://www.openaire.eu/

Open Data Commons website, retrieved from https://opendatacommons.org/

Open Geospatial Consortium, OGC API – Records, retrieved from https://ogcapi.ogc.org/records/

ORCaSa website, retrieved from https://irc-orcasa.eu/

SIEUSOIL website, retrieved from https://www.sieusoil.eu/

Soil O-live website retrieved from https://soilolive.eu/

SoilWise Github website, retrieved from https://github.com/soilwise-he

SoilWise website, retrieved from https://soilwise-he.eu/

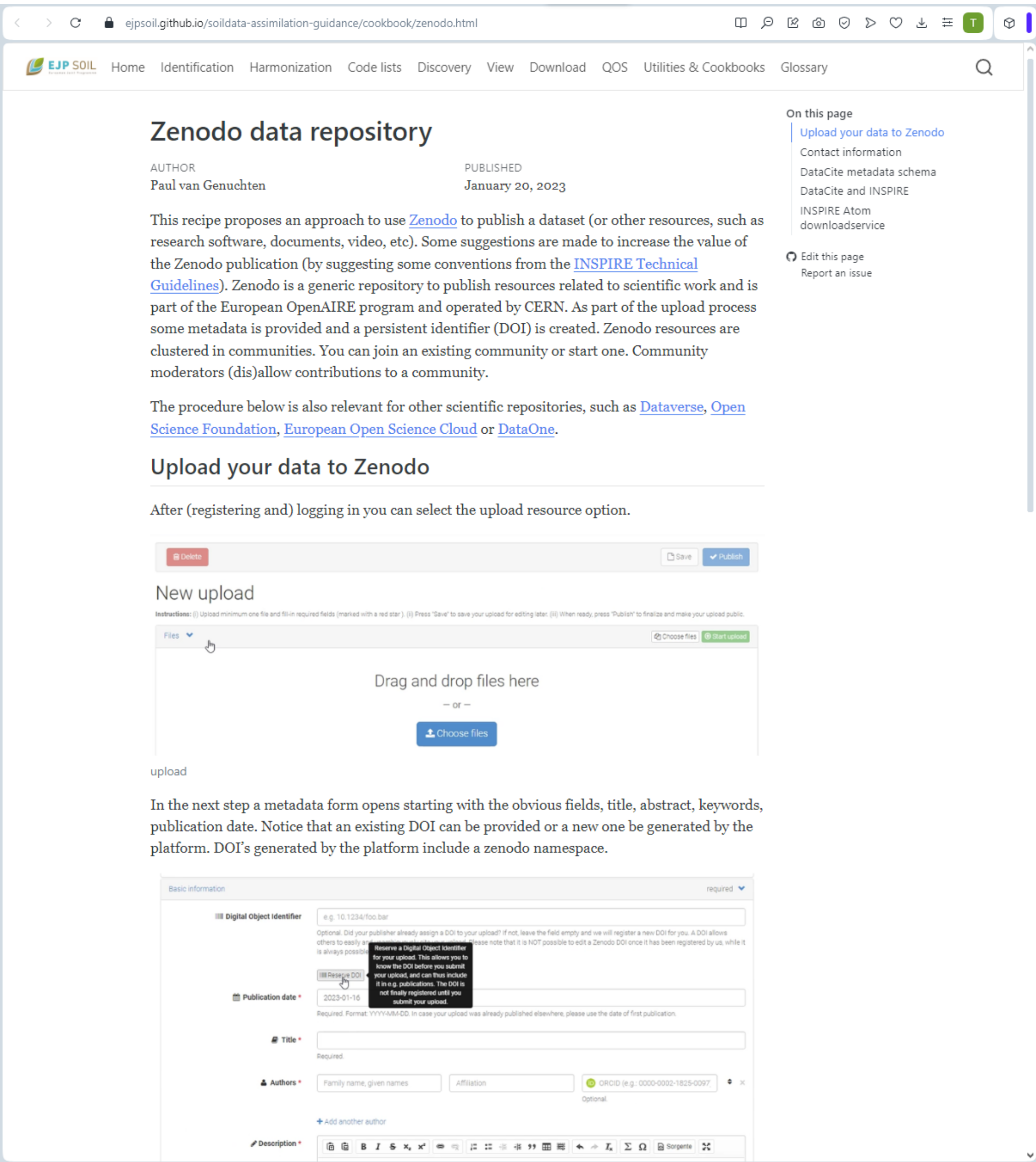
STAC, SpatioTemporal Asset Catalogs, retrieved from https://stacspec.org/en

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. (2016). The FAIR Guiding Principles for Scientific Data Management and Sewardship. Scientific Data 3:e1002295. http://dx.doi.org/10.1038/sdata.2016.18

Wenodo website, retrieved from https://zenodo.org/

11 Annex I - Upload data to Zenodo manual

Example of a guidance on how to load data to Zenodo, created within HE project EJP SOIL, can be found here.



ejpsoil.github.io/soildata-assimilation-guidance/cookbook/zenodo.html

EJP SOIL Home Identification Harmonization Code lists Discovery View Download QOS Utilities & Cookbooks Glossary

Zenodo data repository

AUTHOR
Paul van Genuchten

PUBLISHED
January 20, 2023

This recipe proposes an approach to use Zenodo to publish a dataset (or other resources, such as research software, documents, video, etc). Some suggestions are made to increase the value of the Zenodo publication (by suggesting some conventions from the INSPIRE Technical Guidelines). Zenodo is a generic repository to publish resources related to scientific work and is part of the European OpenAIRE program and operated by CERN. As part of the upload process some metadata is provided and a persistent identifier (DOI) is created. Zenodo resources are clustered in communities. You can join an existing community or start one. Community moderators (dis)allow contributions to a community.

The procedure below is also relevant for other scientific repositories, such as Dataverse, Open Science Foundation, European Open Science Cloud or DataOne.

Upload your data to Zenodo

After (registering and) logging in you can select the upload resource option.

Delete Save Publish

New upload

Files Choose files Start upload

Drag and drop files here

– or –

Choose files

upload

In the next step a metadata form opens starting with the obvious fields, title, abstract, keywords, publication date. Notice that an existing DOI can be provided or a new one be generated by the platform. DOI's generated by the platform include a zenodo namespace.

Basic information required

Digital Object Identifier e.g. 10.1234/foo.bar

Reserve DOI

Reserve a Digital Object Identifier for your upload. This allows you to know the DOI before you submit your upload, and can thus include it in e.g. publications. The DOI is not finally registered until you submit your upload.

Publication date * 2023-01-16

Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.

Title *

Required.

Authors * Family name, given names Affiliation ORCID (e.g.: 0000-0002-1825-0097)

Optional.

+ Add another author

Description *

On this page
Upload your data to Zenodo
Contact information
DataCite metadata schema
DataCite and INSPIRE
INSPIRE Atom downloadservice

Edit this page
Report an issue