D3.5 Report on strategy for effective soil knowledge management

M27/Dec2025

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	D3.5 Report on strategy for effective soil knowledge management
Delivery Time (DOA)	M27
Deliverable Submission Date	30/11/2025
Status	v1
Dissemination Level	PU - Public
Deliverable Lead	Rob Lokers (WR)
Author(s)/Organisation(s)	Valeh Khaledi (ZALF), Rob Lokers (WR), Cenk Dönmez (ZALF), Dajana Snopková (MU), Tomáš Pavelka (MU), Paul van Genuchten (ISRIC), Nick Berkvens (EV ILVO), Radu Giurgiu (ILVO).
Contributor(s)	WR, MU, WU, ZALF, EV ILVO, DOMG – VL O, CREA, ISRIC, WE, NP
Peer-Reviewers	Céline Blitz Frayret (CIRAD); Marc Rosiers (ELO)
Contact	Cenk.Doenmez@zalf.de & Valeh.khaledi@zalf.de
Work Package	WP3
Dissemination level	Public
Keywords	FAIR, Knowledge management, Metadata, Semantic
Abstract	The purpose of this deliverable is to present a strategy for effective soil knowledge management that ensures soil knowledge resources are FAIR (Findable, Accessible, Interoperable, Reusable) through the SoilWise Repository, and to provide guidance by outlining key practices such as metadata harvesting, DOIs, and controlled vocabularies to enhance findability, clarifying the roles of data providers, repository governance, and maintainers, and detailing how to prepare and publish soil knowledge, covering content structure, metadata standards, licensing, repository selection, and versioning.

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.

In this document, the acronym 'DOMG – VL O' is used to refer to the Department of the Environment and Spatial Development, Flanders, Belgium, as per the partner's request for clarification. It is noted that in the grant agreement, the partner is identified by the acronym VL O (Vlaamse Gewest)

List of Abbreviations

SWC	SoilWise metadata Catalogue
KM	Knowledge management
FAIR	Findable, Accessible, Interoperable, Reusable
INSPIRE	Infrastructure for Spatial Information in the European Community
DOI	Digital Object Identifier
GIS	Geographic Information System
LTE	long-term field experiments
ZALF	Leibniz Centre for Agricultural Landscape Research
OGC	Open Geospatial Consortium
GDI-DE	Geodateninfrastruktur Deutschland
EOSC	European Open Science Cloud
OGC CSW	Open Geospatial Consortium Catalogue Service for the Web.
CC-BY	Creative Commons Attribution
RDA	Research Data Alliance
GA no	Grant Agreement number
CC0	Creative Commons Zero
GEMET	General Multilingual Environmental Thesaurus
KG	Knowledge Graph
ML	Machine Learning
EC	European Commission

Table of Content

List of Figures and tables

Executive Summary

Purpose

This deliverable (D3.5) provides a strategic guide for managing soil knowledge within the SoilWise project, a 48-month Horizon Europe "A Soil Deal for Europe" initiative. SoilWise aims to build an open-access European soil data and knowledge repository because currently 60–70% of EU soils are assessed as unhealthy. In response, D3.5 defines methods to capture, structure, and share diverse soil information (ranging from structured datasets to grey literature) so that it is FAIR (Findable, Accessible, Interoperable, Reusable) and aligned with ISO 30401 knowledge-management standards. The deliverable outlines how to embed these practices into SoilWise's workflows, ensuring leadership commitment and stakeholder engagement in creating a trusted, long-term knowledge infrastructure.

Intended audience

The guidance is aimed at all stakeholders involved in soil data and knowledge creation, curation, and use. This includes knowledge managers and data providers within SoilWise and related projects, soil researchers, and land managers (e.g. agriculture, forestry, urban planning) who generate or use soil information. It is also relevant for policymakers and decision-makers responsible for land-use planning, environmental management, or soil policy at national and EU levels. In short, anyone interested in understanding or applying effective soil knowledge management practices, from technical teams developing the repository to experts interpreting soil data for policy or land management, should find this report valuable.

Description of the main activities

The report provides practical guidance on managing, curating, and publishing knowledge through trusted repositories that are interoperable with SoilWise. It builds upon the FAIR Data Strategy (D2.5) and outlines how project outputs, including structured datasets and grey literature, can be made discoverable and reusable through harvesting, metadata enrichment, and alignment with ISO 30401. The methodology emphasizes organizational integration, sustainable practices, and support for semantic tools, licensing, and persistent identifiers.

To support implementation, the report details the following key knowledge management components:

- **Knowledge preparation and publication** – Guidance for knowledge providers
- **Metadata and semantic alignment** – Ensuring discoverability, interoperability, and consistent linkage across repositories.
- **Repository linkage and validation** – Integrating Horizon Europe and Mission Soil projects through standardized metadata.
- **Knowledge reuse and accessibility** – Promoting discovery, citation, and reusability through SoilWise's catalogue and semantic tools.
- **Continuous improvement and alignment** – Supporting the evolution of knowledge workflows in line with ISO 30401 and FAIR principles.
- **Large Language Models (LLMs)**– An experimental chatbot allows to bridge between natural language to conceptualised knowledge-based results

Each component is supported by links to technical documentation, source code, and development repositories, ensuring transparency and practical implementation.

Key Results:

- **Result 1: Integrated guidance on publishing diverse knowledge types**
  The deliverable provides a consistent and standards-aligned methodology for publishing validated knowledge (not just datasets), including grey literature, software, and models, increasing visibility and usability.
- **Result 2: Strengthened discoverability and reusability of soil knowledge**
  Through structured metadata, persistent identifiers, and semantic enrichment, the strategy enables cross-platform linking and enhanced user access, improving the FAIRness of soil knowledge.
- **Result 3: Operationalisation of knowledge workflows within SoilWise**
  The defined processes and components are already being embedded into the SoilWise technical stack, allowing pilot use and continuous improvement of knowledge flows across the platform.

Research and Practice Implications:

The strategy enables researchers to systematically manage and publish a broader range of outputs, from policy briefs to models, with persistent identifiers and standard metadata. This helps ensure that research efforts remain visible, citable, and reusable long after project completion. In practice, the SoilWise knowledge workflows help land managers and applied stakeholders locate relevant tools, case studies, and contextual documents more efficiently.

Policy Implications:

The deliverable supports the knowledge needs of soil-related policy initiatives, including the Soil Monitoring and Resilience Directive. It helps operationalise the collection and cataloguing of best practices, guidance documents, and evidence needed to inform implementation of EU and national soil policies.

Conclusion:

By combining FAIR data principles, ISO-aligned knowledge practices, and semantic tools, D3.5 provides a solid foundation for an interoperable knowledge ecosystem. It lays the groundwork for SoilWise to scale and serve diverse users across research, policy, and practice, contributing to the overarching Mission Soil goal of improving soil health by 2030.

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 a Soil Monitoring and Resilience Directive (5 July 2023), aims to have 75% of EU soils healthy or significantly improved by 2030 and all soils healthy in 2050. Reaching such an ambition requires, among others, access to reliable, harmonised existing and new data and knowledge, including diverse sources such as grey literature (e.g. technical reports, policy papers, and sectoral documentation), collected at local, national and EU levels. Integrating this broader spectrum of information is key to enabling **informed decision-making at all scales to support the proposed 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, while offering guidance and tools for knowledge providers to publish their outputs in a FAIR (Findable, Accessible, Interoperable, and Reusable) way. 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 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 Catalogue (SWC) 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, 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 upholding the highest standards of data privacy and security for our users.

1.2 Document scope

This deliverable provides guidance for knowledge providers on preparing, publishing, and managing soil related knowledge in alignment with FAIR principle and ISO 30401, complementing the technical developments described in earlier deliverables. This deliverable outlines the first iteration of the SoilWise strategy for effective Soil Knowledge Management, building on the technical developments and methodological groundwork laid out in D3.1 and D3.2. It addresses strategic and practical questions faced by knowledge providers aiming to make their knowledge resources accessible, reusable, and impactful within the soil health and wider environmental research communities, particularly in alignment with the goals of the EU Mission "A Soil Deal for Europe."

This D3.5 document complements the technical perspective presented in earlier deliverables with a higher-level, guidance-oriented strategy. It focuses on helping knowledge providers understand **what knowledge is**, **how it can be prepared and published effectively**, and **how SWC supports the FAIR (Findable, Accessible, Interoperable, Reusable) principles** in this process. It also addresses how to search, discover, and reuse

knowledge through the SWC. It is loosely based on concepts from the ISO 30401 standard on knowledge management systems.

As such, this document contributes to Work Package 3 goals by supporting the design and implementation of effective KM practices, as part of Task 3.1 (Design of KM Components) and Task 3.2 (Implementation and Deployment of KM Components). It also serves as a visionary and strategic layer to the more technical deliverables, in line with the development roadmap of the SWC outlined in D1.3 and implemented iteratively through D3.1 and D3.2.

More specifically, this deliverable:

- Defines core concepts and practices around knowledge provisioning and how these relate to the SoilWise mission.
- Explains how the SWC infrastructure supports FAIR knowledge practices through its modular architecture, including harvesting, metadata enrichment, and repository storage (as introduced in D3.1 and D3.2).
- Provides guidelines for preparing knowledge for publication, including structure, formats, metadata standards, licensing, repository selection, and keyword tagging using controlled vocabularies with a particular focus on requirements set by the Mission Soil initiative.
- Demonstrates publishing workflows, including examples using Zenodo, and explains interoperability mechanisms such as CORDIS-to-repository linkages and versioning practices.
- Introduces the user-facing SWC interface, highlighting how metadata and tagging influence the findability and usability of shared resources.

In this first iteration (M27), D3.5 provides a foundation for best practices, targeting knowledge providers from across the soil health research and innovation ecosystem. A second version (to be delivered in M42) will refine and expand the strategy based on lessons learned, community feedback, and further integration of SoilWise services.

This document is designed to bridge the gap between technical components and stakeholder engagement, ensuring that knowledge contribution is not only possible but also impactful and aligned with long-term strategic goals of sustainable soil management in Europe.

1.3 Document structure

This deliverable is structured as follows:

- **Chapter 1 - Introduction (What is knowledge)**
  - Background, purpose, and scope of D3.5 in the context of SoilWise.
- **Chapter 2 - Knowledge management principles**
  - Core KM concepts, SoilWise strategy, opportunities, and implementation
- **Chapter 3 - Access: How to prepare knowledge for publication**
  - Preparing and publishing knowledge with FAIR principles, metadata, licensing, and vocabularies.
- **Chapter 4 - Management: How to publish and manage knowledge**
  - Repository workflows, linkage for Mission Soil projects, and updating knowledge assets.
- **Chapter 5 - Data Reuse, Finding and Accessing Knowledge in SoilWise**

- Knowledge reuse, practical examples, future directions, and accessing content via SoilWise

1.4 Relationship to other project deliverables

This deliverable relates to and complements the following deliverables:

- D3.1, D3.3, D3.4, D3.5 – Developed & Integrated KM components, v1, v3, v4 (M18, M31, M47),
- D2.1, D2.2, D2.3, D2.4 – Developed & Integrated DM components, v1, v2, v3, v4 (M13, M18, M31, M47),
- D4.1, D4.2, D4.3, D4.4 – Repository infrastructure, components and APIs, v1, v2, v3, v4 (M13, M18, M31, M47)
- D1.3 – Repository architecture, v1, v2 (M08, M42)
- D1.1, D1.2 – Usage Scenarios, Requirements, v1, v2 (M6, M36)
- D1.5, D1.6 – Repository GM, v1, v2 (M21, M42)
- D4.5, D4.6, D4.7 – Repository Data and Knowledge Resources, v1, v2, v3 (M21, M34, M46)
- D5.3, D5.4, D5.6 – Deployment and Evaluation Report, v1, v2, v3 (M21, M34, M46)
- D7.2, D7.3, D7.4 – Open Science and Data Management plan, v1, v2, v3 (M6, M27, M48)

1.5 Relationship to project tasks

This deliverable relates to the following project tasks:

- T1.3 Requirements, Validation framework and Rolling plan will feed and update the design and implementation of KM components (3.1 and T3.2)
- T1.4 Define SoilWise Architectural Design – will be considered in the design and implementation of KM components (T3.1 and T3.2)
- T1.5 Define SoilWise Multi-Stakeholder governance model – will be considered in the design and implementation of KM components (T3.1 and T3.2)
- T2.1 Design of the data technology components – will consider and align with the design of KM components (T3.1)
- T2.2 Implementation and deployment of data components – will consider and align with the implementation of KM components (T3.2)
- **T3.1 Design of the KM components** – is addressed in this deliverable
- **T3.2 Implementation and deployment of knowledge components** – is addressed in this deliverable
- T3.3 AI and ML for open and accessible knowledge – will extend the design and implementation of KM components (T2.1 and T2.2)
- T2.4 Strategy for efficient KM – will be followed in the design and implementation of KM components (T2.1 and T2.2)
- T4.1 Repository digital infrastructure for deployment and delivery – will integrate implementation of DM components (T2.2) and KM components (T3.2)
- T4.2 Interfaces for access, sharing, population and integration with EUSO – will be considered in the design and implementation of KM components (T3.1 and T3.2)
- T4.3 Solutions & repository validation and population – will validate designed and implemented KM components (T3.1 and T3.2)
- T5.2 User Cases implementation and demonstration – will demonstrate the functionality of implemented KM components (T3.2)

2. Knowledge Management Principles

2.1 Knowledge in SoilWise

SoilWise develops the SWC, that focuses on bringing together the currently fragmented data and knowledge in the soil health domain. In the context of the SoilWise project, knowledge represents the processed, contextualized, and actionable information, derived among others from data, that contributes to understanding and decision-making in soil health and management. While knowledge and data are often discussed as separate topics, there are clear relationships and many overlaps between knowledge management and data management. For the general public and stakeholders, knowledge may refer broadly to practical insights, experience-based information, or interpretive materials that support action. For specialists in knowledge management, as considered in this deliverable, knowledge typically refers to structured or semi-structured content (e.g. reports, videos, infographics, models) that conveys meaning beyond raw data. Compared to data, such knowledge assets are generally less structured and governed by fewer standardized formats. Nevertheless, like with datasets, it is common practice to describe the content of knowledge assets through metadata and to deposit these in repositories.

Knowledge in the context of SoilWise includes:

- **Research Outputs, other than data:** Such as publications, reports, policy briefs, videos, infographics, source codes (software or models) that have been validated and are ready for publication.
- **Semantics and Metadata layers:** Structured descriptions and contextual information that define relationships, provenance, and meaning of knowledge assets. These enable interoperability and intelligent discovery across systems.
- **AI models:** e.g. Large Language Models that support generating natural language answers to natural language questions. Tools such as chatbots that allow users to pose natural language questions and receive contextualized, knowledge-based answers.
- **Referenced Knowledge Hubs and Project Repositories:** External repositories and portals (e.g. from Horizon Europe projects or national initiatives) can be considered knowledge sources when they expose structured, harvestable metadata or link to relevant outputs. While SoilWise typically indexes the individual resources within these hubs, the hubs themselves are also referenced as entry points to valuable knowledge ecosystems.

2.2 Soil Knowledge Management Principles

The effective functioning of the SWC, to serve the soil health community in addressing the challenges defined by the Mission Soil requires a strong foundation for developing, publishing, and archiving soil-related knowledge. While data typically refers to structured, often numerical or categorical information (e.g. measurements, observations), knowledge encompasses broader formats such as reports, videos, policy briefs, and interpretative insights that are often unstructured and context dependent. This makes managing knowledge more complex: it varies widely in format, is less standardized, and is often scattered across diverse platforms with inconsistent metadata. Despite these challenges, the same core principles can and should be applied. Therefore, the SoilWise knowledge management strategy builds on two central frameworks: the FAIR principles (Findability, Accessibility, Interoperability, and Reusability), and ISO 30401 (ISO 30401:2018), the international standard for knowledge management, which emphasizes embedding knowledge within organizational processes and

ensuring stakeholder alignment. These provide a structured yet adaptable approach to enabling discovery, reuse, and long-term preservation of diverse knowledge types in SoilWise.

With respect to FAIR, the FAIR principles apply to knowledge assets as much as to data, with an exception to source codes. Knowledge should be organized and documented to facilitate its discovery and reuse by both humans and machines, to enable use in new research contexts, enabling validation, replication, and extension of findings. To make knowledge Reusable, it is essential that it is Findable and Accessible and as far as applicable, Interoperable. In that respect, knowledge in the form of semantics (e.g. expressed in vocabularies) is a way to improve the FAIRness of both data and (unstructured) knowledge and even allows the interlinkage of semantically related data and knowledge assets.

The ISO 30401 standard for knowledge management is a recognized standard for setting up knowledge management systems for organisations. Its concepts are guiding principles to develop the knowledge management strategy. Our KM strategy aims to address the management of the full lifecycle of knowledge assets including their creation, validation, publication and maintenance to ensure that the repository remains up-to-date and useful over time. From the perspective of ISO 30401, the following aspects need to be assessed and integrated in a KM strategy that can be deployed:

- Context of the organization, its stakeholders and knowledge needs
- Leadership commitment and KM policy adoption
- Setting objectives and identifying opportunities and risks
- Implementation of KM processes and their support and integration into the organisation
- Monitoring, evaluation and improvement

SoilWise promotes alignment with FAIR and ISO 30401 knowledge management principles where feasible, while recognizing that the quality and compliance of external data and knowledge remain the responsibility of the original providers. SoilWise aims to transform isolated points of data and knowledge into a cohesive body of knowledge that supports sustainable soil management practices and policies.

2.3 Scope of the ideal knowledge management strategy

As ISO 30401 focusses on organisations, a critical point is to decide what we see as “the organisation” in SoilWise. What is the scope of organisations and stakeholders that need to be aligned with regard to the management of knowledge in order for SWC to become the sustainable knowledge hub on soil heath. This is crucial, as it defines the scope of the processes that need to be in place. Moreover, and more importantly, because ISO 30401 presumes leadership commitment, it allows defining where responsibilities lie for the enforcement of standards and the fostering of their adoption beyond the SoilWise project. Any knowledge management system is ineffective without support and engagement at the executive level. Therefore, the scope of the knowledge management strategy and guidelines developed by SoilWise primarily targets the Mission Soil community and its ecosystem of providers, users, and managing organisations. While broader adoption by research laboratories beyond Mission Soil is desirable, it is acknowledged that the full application of ISO 30401 may not be feasible in all contexts due to resource constraints. In such cases, the strategy promotes selective alignment with ISO principles, focusing on practical, scalable elements, rather than full compliance.

From a broader ecosystem perspective, the SWC represents one component within a much larger knowledge management landscape. The effective processes under SoilWise’s control depends on alignment and coordination with the wider ecosystem, including the Mission Soil Initiative; the Mission Soil Cluster on

knowledge and data management, EUSO, EOSC and other national and European soil-related stakeholders. SoilWise proposes a strategy and guidelines for this larger context, but it's full implementation and deployment go beyond the project's tasks and requires ownership at another level. Besides, beyond a broadly adopted framework of structural and procedural measures, fostering a culture of collaboration and knowledge sharing is essential, encouraging community engagement, and recognition mechanisms through active participation by stakeholders. E.g., to maximize its value, the SWC will align with relevant external national and international knowledge initiatives (e.g. EJP SOIL, the INSPIRE directive), thereby supporting broader semantic and institutional interoperability.

This translates into two levels of knowledge management

1. **The ecosystem level**
   This level describes how the "soil health community" should operate and collaborate to ensure that knowledge produced on the EU and national level is handled in such a way that they are FAIRified and made available for effective reuse and exploitation. A strategy on this level should ideally be broadly adopted by the community. While not imposed by SoilWise, it reflects a shared vision for coordinated knowledge stewardship. However, it is acknowledged that many research organisations currently lack dedicated governance or resources for FAIR-aligned knowledge management outside of project-funded contexts. Therefore, sustained adoption would require both executive-level commitment and long-term funding mechanisms.

2. **The SWC level**
   This part of the strategy describes the aspects of efficient knowledge management that are in direct control of the organisation maintaining SWC, currently the SoilWise project and in the future JRC/EUSO. It is obvious that the effectiveness of this strategy level is highly dependent on how for instance the implementation of standards and ways of working with knowledge by the community are aligned and adopted.

2.4 Organisational context

2.4.1 Scope and objectives

For the definition of the scope of a sound knowledge management strategy on soil health, we define the organisational context as the "Soil Health Community"**.** We see this as the ecosystem of stakeholders that works in the soil health domain and directly benefits from the services offered by the SWC and particularly benefits from knowledge that is produced and flows in such a way that it can be efficiently shared, discovered and reused. In this ecosystem view, we focus on the scope that was defined in the SoilWise and the group of stakeholders that are relevant in that respect:

- The EC, and particularly the Mission Soil platform, JRC, REA and DG Agri
- The Mission Soil projects
- The (other) stakeholder groups defined as part of SoilWise: land managers, researchers, policy makers, businesses, etc.
- The SoilWise consortium

Knowledge needs of the soil health ecosystem are defined by the strategic objectives of the Mission Soil and the derived needs of EU Member States, research organisations, businesses and other involved parties to implement

those objectives. For that cause, a range of research and other initiatives are deployed that are supposed to bring this forward. An important condition is that relevant existing soil health knowledge (and data) becomes easily available for further reuse in a FAIR manner, implementing knowledge flows that are "FAIR by design". To ensure that newly produced knowledge on soil health is FAIR by design, a sound strategy and clear guidelines are needed. These should explain and guide stakeholders that generate knowledge in the best way to develop, document and publish knowledge, so it is formatted and documented according to standards and easily discoverable for reuse.

The main objectives of the SoilWise knowledge strategy are:

- To clarify the context of knowledge management, and its importance to forward the Mission Soil and eventually improve soil health in Europe.
- To define a sustainable workflow for the production, documentation and publication of FAIR soil health knowledge, focussing on reuse.
- To advise on the implementation of good practices to FAIRify knowledge that support that workflow, e.g. by advising standards for formatting, metadata, archiving etc. knowledge that support that workflow, e.g. by advising standards for formatting, metadata, archiving etc.
- To describe the role and functioning of SWC in this workflow and the role and responsibilities of its maintainers.

2.5 Opportunities and Challenges

The benefits of a well-functioning knowledge strategy for the soil health community are obvious. Supported by a network of stakeholders and strong executive leadership, it can create a sustainable stream of FAIR knowledge from high quality resources and make it easily discoverable, accommodated to the specific demands and requirements of its consumers. This will facilitate reuse, catalyse innovation and result in better, more sustainable soil management practices and eventually increase soil health.

It is becoming common practice to deposit data in repositories and describe it with standardised metadata. Contrary to data, the standardisation around knowledge, particularly where it concerns unstructured knowledge, is less developed. Knowledge assets are still residing more scattered over the Internet. They are hosted on heterogeneous resources like repositories of scientific publishers and funders of research programmes, individual project websites and knowledge bases, video platforms like YouTube etc. In addition, valuable grey literature and aggregated data, often produced by private companies, industry associations, or advisory networks, represent a significant but under-integrated source of actionable knowledge for stakeholders in the food and soil value chain. Around the topic of soil health, and particularly the Mission Soil, there is a tendency to focus more on recognized repositories like Zenodo, but also national research repositories. On the other side, many individual research projects and other initiatives, including also quite some of the Mission Soil projects, are still creating their own isolated knowledge bases. This suggests that there is currently no strong strategy to avoid such fragmentation towards the future. A negative consequence is that part of the generated knowledge might remain under unclear and uncertain future governance and management, with limited guarantees for long-term persistency and accessibility. Moreover, due to the often limited or lacking support of standards for metadata and metadata harvesting, such assets cannot be easily brought together and made discoverable by domain centred initiatives like SWC.

Semantics, and particularly structured vocabularies can be highly supportive in organising knowledge and making it easier to discover assets that fit the needs of the user. There are ontologies (e.g. GLOSIS) that focus

on structuring and harmonising soil data. At the same time there is no coordinated development of a controlled soil health vocabulary. In practice this means that tagging and linking resources happens on a case-by-case basis, using individually selected terminology, or at best using terms from existing vocabularies like Agrovoc or Gemet that offer only a very limited support for the soil health domain.

2.6 Implementation of KM processes and their support and integration into the organisation

The strategy developed by SoilWise towards the implementation of knowledge management (KM) processes within the soil health community is designed to ensure that knowledge assets are not only collected but also systematically supported and embedded across the organisation. Building on FAIR principles and ISO 30401 guidance, the approach proposed by SoilWise integrates resources, competencies, awareness, communication, and documentation into a coherent framework that serves both data providers and knowledge users. While the implementation of the knowledge management strategy is taken up by SoilWise during the projects' lifetime, the transition of the SWC to the European Commission will require the development of a post-SoilWise strategy for the adoption of ownership and the embedding of governance as well as securing resources for sustainable operation.

A first priority is the **allocation of resources**, ensuring that sufficient technological infrastructure, workforce capacity, and financial support are in place. This guarantees that processes such as metadata harmonisation, semantic enrichment, and knowledge graph management remain reliable and scalable over time.

Equally important is **competence development**. Contributors and users of the SWC are supported through training, guidelines, and capacity-building activities. These measures build the necessary skills for publishing, curating, and reusing soil health knowledge. Competence is addressed at multiple levels: from repository maintainers responsible for system integrity, to researchers ensuring the quality of submitted content, and stakeholders who rely on SoilWise outputs for decision-making.

Awareness and engagement are fostered through clear communication of KM policies and benefits. All stakeholders are encouraged to understand their role in contributing to, and benefitting from, the repository. This includes awareness of licensing obligations, the impact of FAIR-aligned practices, and the added value of semantic technologies for interoperability.

Communication flows are structured both internally and externally. Internally, collaborative tools and knowledge-sharing practices (e.g. communities of practice, documentation guidelines) ensure effective coordination among project partners. Externally, SoilWise communicates through user-facing interfaces, documentation, and training resources, providing transparent access to repository functions and outputs.

Finally, **documented information and process control** underpin integration into the organisation. Version-controlled documentation of methodologies, schemas, and workflows ensures consistency and reliability. This documentation is openly accessible and maintained to support both replication and improvement.

Through this integrated approach, the SoilWise approach embeds KM processes as part of its organisational fabric, ensuring that soil-related knowledge is systematically supported, continuously improved, and effectively delivered to the European soil health community.

2.6.1. Keyword and vocabularies

To further enhance discoverability, SoilWise integrates a **Keyword Matcher** component, which automatically aligns submitted keywords with a predefined vocabulary set. This improves metadata coherence, reduces redundancy, and enables more accurate search filtering. Additionally, metadata records are enriched through **Metadata Augmentation**, which adds related terms and semantic links, allowing users to uncover hidden connections across fragmented datasets.

Implementation Guidance for Knowledge Providers

- Each metadata record should include **3 to 7 well-targeted keywords**.
- Terms should represent both thematic content (e.g., “soil degradation”) and methodological relevance (e.g., “GIS modelling”).
- Free-text or institution-specific terms should be avoided to maintain consistency and interoperability.
- Repository curators can assist in validating keyword choices during the metadata submission process.

By encouraging the use of harmonized vocabularies and supporting semantic enrichment, SoilWise significantly improves the precision and relevance of knowledge discovery. This means that users, from researchers to policymakers, can more easily find, compare, and combine knowledge from different disciplines and sources, even if they use different terms. It ultimately helps turn scattered data into coherent insights for soil health research and decision-making.

2.7 Monitoring, evaluation and improvement

The knowledge management system proposed by SoilWise is subject to systematic monitoring and evaluation to ensure its continued relevance, effectiveness, and alignment with Mission Soil objectives. Monitoring focuses on both compliance with FAIR and ISO 30401 principles, and on the demonstrable value delivered to stakeholders. It has to be mentioned that while SoilWise provides tools, workflows, and monitoring mechanisms to support quality and consistency, the success of knowledge management ultimately relies on the active engagement of knowledge providers. Accurate metadata, adherence to standards, and timely updates remain the responsibility of the data owners. Technical interventions can only complement, not replace, the foundational diligence required from contributors and the coordination enabled through governance structures.

Monitoring and evaluation should be performed at least at two levels. Monitoring of the performance of the SWC allows evaluation of processes and their gradual improvement. More importantly, monitoring the use and perceived usability of the repository will provide input for evaluation and learning on the functioning of knowledge management over the soil health community, contributing to the broader objectives of the Mission Soil. It will provide a wealth of information on how the soil health community manages its knowledge, reveal existing gaps and allow the further steering and improvement of knowledge FAIRification.

Monitoring & Evaluation and improvement of the SoilWise Catalogue workflows

Performance monitoring includes tracking key indicators such as the number and diversity of knowledge resources integrated, frequency of reuse, quality of metadata, uptake of semantic services, and user satisfaction. Methods combine automated analytics (e.g., usage statistics, link integrity checks, KG queries) with stakeholder

feedback gathered through consultations and surveys. Results are analysed at regular intervals to identify trends and assess whether KM processes meet defined objectives.

Management reviews ensure that the system remains responsive to changing scientific, technical, and policy contexts. Reviews assess progress against strategic objectives, address open risks, and evaluate opportunities for innovation, such as AI-enabled discovery or integration with external infrastructures. This ensures the repository continues to meet the evolving needs of researchers, policymakers, and practitioners.

When nonconformities arise, such as broken metadata links, semantic mismatches, or user-reported errors, SoilWise applies corrective actions aimed at both immediate resolution and systemic prevention. Corrective actions are documented and reviewed for effectiveness to ensure recurrence is avoided.

Finally, SoilWise embraces a principle of continuous improvement. This includes refining harvesting pipelines, enhancing semantic alignment, expanding interoperability with external systems, and incorporating new user requirements. Through this iterative cycle, the repository evolves from a discovery platform into a knowledge ecosystem that grows in accuracy, usability, and impact over time.

M&E and improvement of the soil health community knowledge management processes

Analytics from continuous monitoring of the SWC workflows and their performance are foreseen to be one of the major inputs for the community as a whole to learn and improve on knowledge management. Comparing it with external statistics from various sources, it can be used as a valuable resource to draw conclusions and formulate directions on various knowledge management aspects that are out of scope of the SWC workflows.

While SoilWise will deliver basic analytics as a foundation, further development will be needed based on operational experiences. Examples of such analytics are:

- The percentage of knowledge resources that are technically identifiable as relevant (i.e. containing a Grant Agreement (GA) number and soil-related keywords in the title or abstract) and are accessible for harvesting
- The quality of provided metadata and specific metadata elements
- The compliance with specific metadata and knowledge standards (such as Dublin Core, ISO 19139, and adherence to FAIR principles)
- The formulation of end user search queries
- The statistics on knowledge items accessed by end users

Similarly, once SWC is fully operational as part of EUSO, internal audits and management reviews will result in conclusions and recommendations that relate to the functioning of knowledge management processes as a whole. It is foreseen that the scope of many of them will be at least partly beyond the scope and responsibility of the Soil Wise project and its operational processes and workflows.

3. Access: How to prepare knowledge for publication

3.1 Preparing knowledge for discoverability through SoilWise

This section defines the starting point in the process of accessing soil-related knowledge through the SoilWise platform. SoilWise serves as a centralized discovery interface built under the European Soil Mission framework, designed to harmonize access to high-quality, knowledge assets and interoperable inputs relevant to soil and related research domains. It actively harvests metadata from repositories and portals (e.g. CORDIS, OpenAIRE, INSPIRE, EJP SOIL, BonaRes) to build a local metadata catalogue for efficient querying.

This deliverable (D3.5) is structured as a practical strategy to demonstrate how this access process works. It builds directly on the technical groundwork established in Deliverable D4.5, which focused on repository population and metadata interoperability. D4.5 detailed the harvesting of nearly 20,000 metadata records and their standardization under Dublin Core and ISO 19139:2007. These records were made openly available as a CSV publication on Zenodo (DOI: https://doi.org/10.5281/zenodo.14851857). D3.5 focuses on the practical outcome: how to improve the FAIRness of knowledge and can be discovered and retrieved as usable knowledge through the SoilWise platform. This section clarifies some of the main principles that the SWC implements to process, harmonise and integrate knowledge from heterogeneous sources into one FAIRified soil health knowledge base, and what this implies for knowledge providers.

Ensuring Findable and Accessible Knowledge: SoilWise's mission is to make soil-related knowledge easily discoverable and reusable across Europe. For knowledge providers, this means every resource they contribute should be **Findable, Accessible, Interoperable, and Reusable (FAIR)**. In practice, providers must **assign persistent identifiers** (e.g. DOIs or other stable URIs) to each knowledge resource so that it can be reliably located and cited over the long term: Relying on temporary links (such as project websites that might disappear) is discouraged instead, content should be deposited in trusted repositories (like Zenodo, DataVerse, etc.) that issue PIDs and guarantee long-term availability. By using persistent identifiers, knowledge providers ensure their outputs remain accessible through the SoilWise catalogue and beyond, strengthening trust and enabling others to confidently reference these resources. This commitment to stable identification directly supports SoilWise's goal of an integrated knowledge base, linking to distributed knowledge resources that are maintained at the source, that is widely searchable and **accessible to all stakeholders.**

SoilWise enables users to locate soil health–related knowledge by applying precise filters based on geography, topic, experiment type, and more. For example, a search using the term "erosion" with the filter set to "Europe" returns 480 curated entries. Each result is linked to the source repository, such as the Zenodo, BonaRes Repository, enabling immediate access to different format of knowledge from documents like journal article (Figure 3.1a) to video (Figure 3.1b), metadata description, license information, journal articles and download options.

The user interface supports advanced search functionalities, including spatial filtering through map-based selection and thematic filtering via facets. These features are designed to accommodate various stakeholder needs—from policymakers looking for regional data to researchers performing spatially explicit modelling.

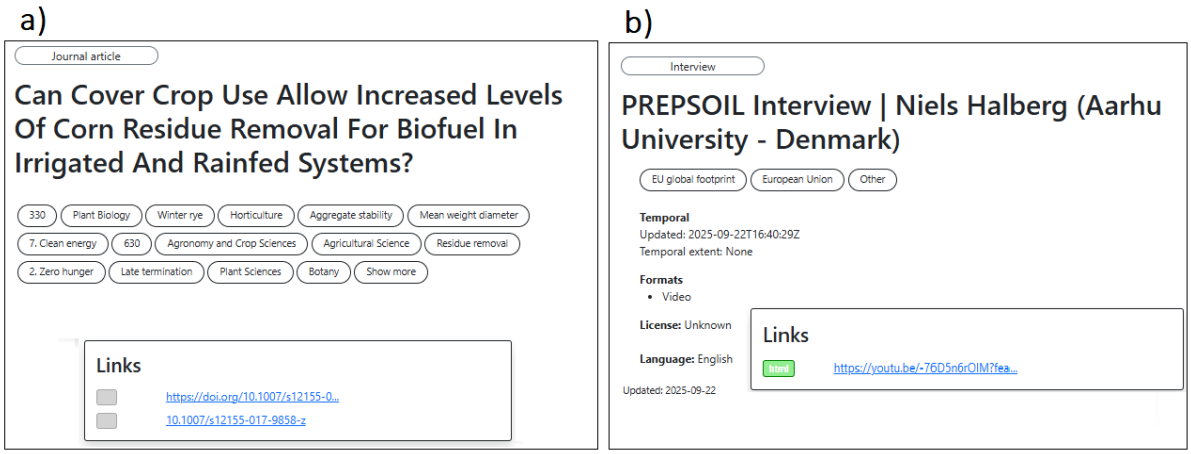

Figure 3.1 SoilWise interface showing filtered results for journal article (a). and video (b) as different sorts of knowledge under the keyword search for soil erosion Europe.

High-Quality Metadata and Semantic Tagging: Alongside persistent IDs, knowledge providers are responsible for **supplying rich, high-quality metadata** for each resource. This includes clear titles and descriptions, correct units of measure, authorship and provenance details, and usage or licensing information. Complete and accurate metadata makes it much easier for others to understand and reuse the knowledge product. It is also indispensable for the SWC workflows to process and index the resources in such a way that they can be easily discovered through end-user interfaces. SoilWise can augment and complement metadata with smart tools, but ultimately the onus is on the provider to ensure the record is thorough. A key part of this metadata is **semantic tagging** using controlled vocabularies. Where possible, providers should apply standard terms from established vocabularies relevant to soils and the environment such as AGROVOC (agriculture), GEMET (environment), or INSPIRE-compliant terminology. Where this is not possible, the limitation should be clearly documented to ensure transparency and traceability. Using such **controlled vocabulary** terms in the metadata helps connect and interlink resources across different systems and disciplines. It improves discoverability because search engines and catalogues can recognize these standardized terms and group related content. SoilWise in fact encourages building on existing thesauri and classification systems (rather than inventing new terms) to describe content. By adhering to common metadata schemas (e.g. Dublin Core or DataCite) and tagging with well-known terms, knowledge providers make their contributions **interoperable,** they can be seamlessly discovered alongside other datasets, documents, and maps in the SWC and even in external data portals. In short, **descriptive, standardized metadata** is the bridge that turns an isolated report or dataset into a findable piece of the larger soil knowledge puzzle.

Providing Context and Usability for Target Audiences: Knowledge in SoilWise is not just raw data or technical reports, it also encompasses guidance, best practices, and explanatory materials that various stakeholders can apply. Knowledge providers should therefore aim to **present information in a clear, contextualized way** that serves the needs of different target user groups. Rather than treating a knowledge resource as a data dump, consider its **guidance role**: for example, a researcher uploading a study could include a summary or factsheet highlighting practical implications for land managers, or a policymaker's briefing might distil key evidence from scientific data for decision-makers. The SoilWise strategy underscores that the repository is a common ground to support informed decision-making for **land managers, policymakers, researchers, and industry alike.** This means a knowledge resource should be packaged with its audience in mind, using language and formats appropriate to that audience, so that it truly transfers insight and know-how, not just information. By crafting resources that are understandable and relevant to non-experts (when appropriate), knowledge providers help

broaden the impact and "reach" of soil knowledge. To maximize practical value, such knowledge should also be linked to specific soil indicators, regional assessments, or decision-support tools where it can directly inform sustainable land management practices. In practical terms, this could involve providing an executive summary, including visual aids or examples, or linking the resource to related data and tools (so users can easily find background or follow-up materials). Such user-oriented framing ensures that the knowledge content in SoilWise acts as guidance, enabling others to learn from it and apply it within SoilWise, which is ultimately the repository's purpose. Importantly, this approach complements the technical data: while other project deliverables handle data specifics, the knowledge resources should interpret and contextualize those data, turning them into actionable wisdom for the community.

To further support discoverability and cross-linking of related knowledge, SoilWise incorporates a **Knowledge Graph** (KG), published and accessible via Zenodo DOI: 10.5281/zenodo.14936020 (Figure 3.2). The KG semantically interlinks knowledge resources, datasets, and soil-related concepts with relevant taxonomies, vocabularies, and indicators. Although it is not directly integrated into the SoilWise Catalogue, it complements the platform's search functionality by enabling semantic enrichment and supporting the **Keyword Matcher** component.

The KG is built using RDF graph structures and is accessible through a public SPARQL endpoint. It enhances contextual understanding by linking related entities, such as soil health indicators, experimental data, and thematic concepts, thereby helping users interpret relationships between different knowledge resources even when these concepts are described differently across disciplines.

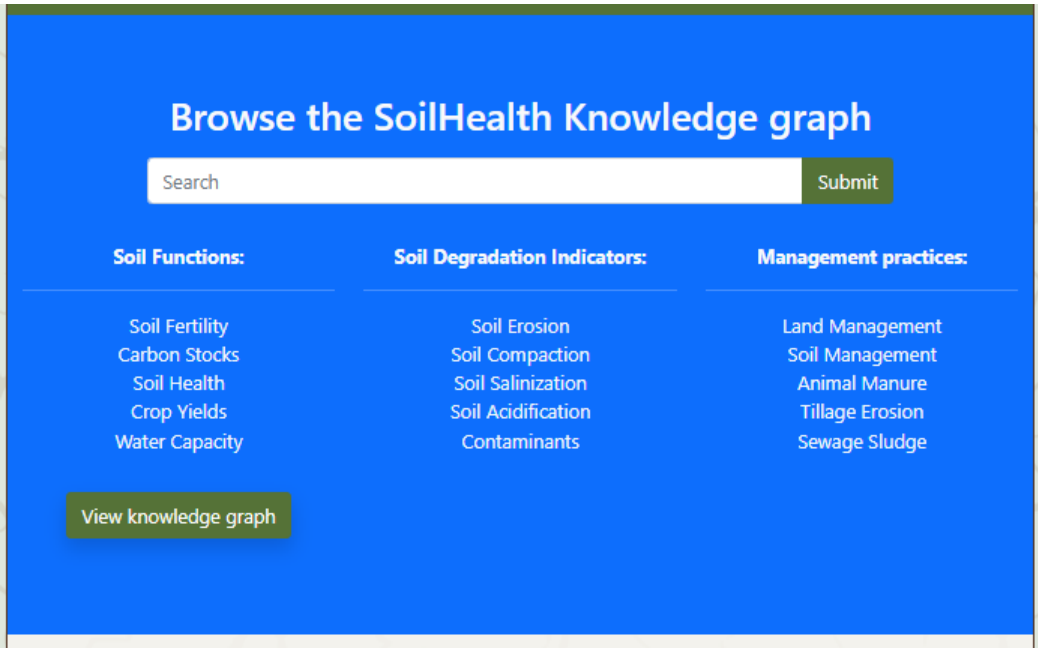

Figure 3.2. SoilWise repository the knowledge graph section

Moreover, search performance within SoilWise is actively improved through backend enhancements such as the keyword-matcher module, which helps cluster and clean metadata entries to reduce inconsistencies and

improve filter precision. These enhancements make SoilWise not only a point of access but also a continuously improving platform for metadata alignment and semantic enrichment.

In summary, SoilWise provides a robust and flexible starting point for accessing soil-related knowledge. By combining standardized metadata aggregation, semantic structuring through the KG, and a user-friendly interface with advanced search functionality, it ensures that users can discover, filter, and access the most relevant resources efficiently. This capability is essential for enabling subsequent steps in knowledge management and reuse, as discussed in the following sections.

Alignment with Governance and Long-Term Maintenance: Finally, knowledge providers must align their contributions with the overall governance and maintenance strategy defined by the EC in consultation with the soil health community, that is supported by the workflows of SoilWise. The repository's longevity and coherence depend on everyone following agreed standards and processes. For example, **governance guidelines** may specify quality checks or metadata profiles that providers should adhere to, ensuring consistency across thousands of entries. Knowledge providers should stay engaged with these guidelines, using the recommended metadata templates or workflows, so that their resources integrate smoothly into the system. Moreover, since SoilWise itself is a time-bound project (and not a perpetual host for content), providers are encouraged to plan for **long-term preservation** of their knowledge assets, where feasible, using trusted national or institutional archiving services. In practice, this circles back to using established repositories and archiving methods. SoilWise will harvest and catalogue the metadata, but the content's durability relies on the provider's chosen repository commitments. By depositing in **trusted digital repositories** and using persistent links, knowledge providers ensure that even after SoilWise's project period, the knowledge remains accessible and doesn't vanish with temporary websites or project endpoints. This is a critical responsibility, and one that requires not just good practice at the point of publication, but active governance and oversight.

Looking ahead, the long-term sustainability of this process must be embedded in governance structures beyond the SoilWise project. This includes:

- **Monitoring Horizon-funded projects** for proper deposition and metadata compliance, ideally through automated workflows (e.g. OpenAIRE harvesting by GA number, where applicable, does not apply to all grey literature, which requires separate identification and ingestion strategies).
- **Integrating grey literature** by identifying key aggregators and establishing ingestion protocols (e.g. monitoring outputs from industry bodies, NGOs, or public–private consortia).
- **Defining ownership** of maintenance roles, whether within the European Commission, the Soil Mission Secretariat, or a future coordinating body.

In essence, the role of a knowledge provider is not one-off: it involves **ongoing responsibility** to curate and update the information as needed (e.g. versioning a report if new findings emerge or updating metadata if a resource moves). Coordination with data providers is also key, oftentimes a knowledge resource (like a case study or a model) is linked to underlying datasets, so working together to maintain those links and context benefits everyone. SoilWise contributes by structuring the metadata flow and enabling machine-actionable discovery. However, institutional actors will need to carry forward this governance framework to ensure that the knowledge ecosystem remains reliable, complete, and up to date. By collaborating within the SoilWise community and following the governance best practices, knowledge providers help maintain a coherent, durable knowledge base. This paves the way for the following sections of the strategy, which address the complementary roles of data providers, platform governance, and long-term maintenance, to ensure that soil knowledge is managed as a sustainable, connected resource for the future.

3.2. Guidance for knowledge providers (SoilWise)

This section provides practical guidance for researchers and knowledge providers stakeholders on how to prepare and publish Knowledge resources, not just datasets, so that they are findable, accessible, interoperable, and reusable within SoilWise and related infrastructures. It follows the process from structuring and formatting knowledge, to assigning metadata, licences, and keywords, ensuring that every output can be seamlessly harvested and discovered through SoilWise and OpenAIRE.

3.2.1. Structure: Preparing Knowledge for Publication

Knowledge in SoilWise covers a wide range of research outputs, including reports, policy briefs, publications, models, methods, visualisations, videos, and other validated materials. Each item should follow a consistent and transparent structure to ensure clarity, traceability, and reuse.

A well-structured document should begin with front matter that includes the title, author(s) and affiliation(s), ORCID identifiers, contact details, abstract, keywords, version/date, DOI (if assigned), and any related DOIs linking to associated datasets or models. In line with Mission Soil requirements, publications from Mission Soil-funded projects must also display the Grant Agreement (GA) number and project acronym on the cover page and within the metadata. A brief executive summary written in plain language should be included for non-technical readers.

The main body of the document should clearly outline objectives. methods, results, limitations, and practical reuse guidance, ending with acknowledgements and references. Where applicable, annexes may include supplementary tables, code fragments, coordinate reference information, or glossaries. Version control should be applied systematically recording updates in a changelog (e.g., v1.0 → v1.1 for minor edits). For composite resources, such as a model accompanied by guidance materials or datasets, a README file should describe each component and how they interrelate.

3.2.2 File formats: What you receive and why it matters

Knowledge intended to be discovered through SoilWise should be published externally in open, non-proprietary, and archival formats to support long-term preservation and machine processing. When deposited in trusted repositories documents should be submitted in PDF/A for archiving, accompanied by editable source files (DOCX, ODT, or Markdown). Analytical notebooks may be shared as Jupyter Notebooks (.ipynb) with HTML or PDF exports, and visual materials such as maps or diagrams should use PNG or SVG formats with clear legends and coordinate references. Publishing knowledge in this FAIR-compliant manner ensures that SoilWise can optimally index, interpret, and connect it across the broader ecosystem.

For models or software, include the full source code, a LICENSE file, a README explaining usage, and an environment specification (e.g., environment.yml). All files should follow a predictable and descriptive naming convention (e.g., projectA_report_soilhealth_v1_20251001.pdf). As a good practice, source code should be maintained in a sovereign software repository (e.g., institutional or European GitLab) and archived in Software Heritage to ensure long-term preservation. While Software Heritage is not currently harvested by OpenAIRE, exploring direct harvesting pathways or metadata bridging solutions could enhance integration in the future.

Beyond single outputs, more complex knowledge packages—such as multi-year experimental syntheses, scenario-based model results, or linked variable networks—are provided in structured archives (e.g., SQL

databases or bundled repositories). These collections preserve relationships across variables and time, reflecting the ontology-based organization of knowledge.

Crucially, each package is enriched with explanatory materials such as methodological notes, workflow diagrams, or conceptual models. Together, these components ensure that what is shared goes beyond information—it becomes actionable knowledge, ready to support reasoning, interoperability, and evidence-based decisions.

Repositories connected to SoilWise typically provide customized access to their knowledge resources. Within these repositories, users can filter, subset, or reformat content based on spatial or thematic relevance, ensuring that knowledge assets align with specific research or policy needs. This on-demand flexibility, combined with rich metadata and standardized formats, enhances transparency, supports reproducibility, and facilitates multidisciplinary integration across the SoilWise ecosystem.

3.2.3 Metadata: Describing Knowledge for Discovery

Accessing knowledge begins not only with accessibility and findability through platforms like SoilWise but with the foundation of well-structured, high-quality metadata. Metadata as a type of knowledge acts as the bridge between stored data and its meaningful reuse, it provides the descriptive and technical context that allows users to interpret, cite, and apply knowledge confidently.

Within the SoilWise framework, metadata harvesting and integration are built around internationally recognized standards, primarily Dublin Core and ISO 19139:2007. These standards ensure compatibility across diverse repositories and data infrastructures. The metadata schema used by domain-specific repositories like BonaRe and IPCHEM captures both general and domain-specific information, for example, experimental design and soil parameters for long-term field experiments (LTEs), pollutant concentrations and measurement protocols for environmental chemistry, or taxonomic identifiers for biodiversity observations. SoilWise does not replicate or reinterpret these domain-specific elements; rather, it harvests and preserves them as references within its harmonized metadata structure. Where possible, key descriptors (e.g., location, temporal coverage, dataset type) are mapped to the common SoilWise metadata template to enable cross-domain discovery, while links to the source repositories are maintained for full access to detailed, discipline-specific metadata and analysis. This approach ensures interoperability across communities while respecting the integrity and granularity of each domain's standards.

To enhance semantic alignment and searchability, metadata records in SoilWise are enriched using controlled vocabularies and taxonomies, including AGROVOC, GEMET, and custom soil health indicators from the European Environment Agency. These vocabularies enable consistent keyword assignment and facilitate more accurate discovery, especially when datasets come from different countries or disciplines. Always include the GA number, project acronym, and related DOIs for linked outputs in the metadata. Use the repository's versioning function when updates are made so all versions remain traceable and citable.

The Knowledge Graph (KG) is being developed to enhance semantic connectivity across SoilWise resources, enabling users to explore thematic and contextual relationships between datasets and other knowledge objects. Once fully implemented, this structure helps unify fragmented definitions of soil health across research, policy, and land management domains and enables advanced querying through a public SPARQL endpoint. Practical demonstrations illustrating the added value of semantic querying through the Knowledge Graph compared to traditional search will be provided in subsequent outreach or training materials.

Metadata quality is actively monitored and improved during harvesting through components like the keyword matcher and link liveliness Assessment component (LLA), and Metadata Completeness Validator. The Keyword Matcher harmonizes heterogeneous terms, while the LLA checks link accessibility and flags inactive references (green = active, red = inactive, grey = unchecked) in the SoilWise Catalogue. In addition, the Completeness Validator ensures that all mandatory fields are populated, and INSPIRE records undergo schema validation before integration. Together, these checks maintain the accuracy, consistency, and interoperability of metadata harvested in SoilWise.

SoilWise supports metadata harvesting from a wide range of sources, including BonaRes, INSPIRE Geoportal, FAO, EEA, ISRIC, and CORDIS. Harvested records are enriched using OpenAIRE when DOIs are available, enabling structured metadata from scientific publications to be linked back to datasets and deliverables. This bidirectional linkage between data and literature is key to supporting reproducibility and research traceability. Overall, this structured and standards-based approach ensures that metadata in SoilWise is machine-readable, semantically rich, and interoperable across repositories. These qualities are foundational to building a resilient and user-centric data infrastructure for soil health, ensuring that datasets can be accessed, cited, and reused effectively by scientists, policymakers, and data providers alike.

3.2. 4 Select a license: Enabling Legal Reuse

Licensing clarifies how and under which conditions others can reuse knowledge, e.g. ensuring that authors receive proper credit. Within SoilWise, the Creative Commons Attribution (CC BY 4.0) licence is the preferred default for publications, reports, and other openly accessible outputs, supporting the principles of open science. However, in justified cases—such as when outputs contain sensitive or personal data subject to GDPR or confidentiality restrictions—SoilWise also supports the use of restricted licences or controlled access. In all cases, the corresponding metadata should remain publicly available under CC0 to ensure the resource is discoverable and indexed, even if the full content is access restricted.

SoilWise and its harvested repositories,such as BonaRes, Zenodo, and others, follow well-established, machine-readable licensing standards. This balance between openness and author attribution makes CC BY the most FAIR-compliant option.

Metadata records are often published under the Creative Commons Zero (CC0) license, which places them in the public domain. This facilitates full integration with harvesting platforms (e.g., OpenAire, Data Europe, INSPIRE...) and supports automation-friendly discovery processes.

An example for restricted access is the publication (Soil Microbial Biomass Carbon and Nitrogen As Affected by Cropping Systems) with closed-access license and is referenced in SoilWise. Although the full text is under **Closed Access**, its metadata remain publicly available and searchable through SoilWise. This illustrates how SoilWise maintains the visibility, citation, and contextual linkage of research outputs even when the underlying publication is not openly accessible. This approach may also support the discoverability of valuable grey literature and decision-support tools held by private entities, where exposing metadata is feasible even if full content access is restricted.

Evidence from SoilWise Records

A license distribution analysis of ~20,000 metadata records harvested in SoilWise reveals strong adoption of open licensing (Table 2.1):

Table 3.1 License type and records in SoilWise

License Type	Records
Open Access	5,016
Closed Access	4,058
CC BY	961
Not Available / Unspecified / No Conditions	988
Restricted	254
CC BY-NC-SA 4.0	138
CC-BY-4.0 (alternative formatting)	113
No limitations to public access	109

While over half of the entries are open access licensed, a substantial portion still lack machine-readable licensing metadata or remain restricted. This illustrates the importance of applying standardized licensing during knowledge and data publication.

When publishing through repositories like BonaRes:

- **Default to CC BY** for datasets unless constraints require otherwise.
- **Use CC0** for metadata to maximize machine-readability.
- **Ensure license tags are embedded** both in metadata and in repository interfaces.
- **Request support from data stewards** if complex legal or institutional requirements exist.

3.2.5. Selecting a Repository: Where to Publish

To ensure visibility in SoilWise, knowledge should be deposited in trusted repositories that expose standard metadata and persistent identifiers. For academic and research outputs, this typically includes repositories linked with OpenAIRE. For governmental or operational datasets, equivalent trusted endpoints such as data.europa.eu, Copernicus, or other institutional repositories can serve as valid sources, provided they ensure interoperability and long-term accessibility.

Zenodo is the recommended general-purpose repository for open publications, software, figures, and teaching materials. It supports Grant Agreement numbers and automatically provides DOIs and version control. Zenodo is linked with OpenAire and resources will therefore also be harvested by the SWC.

BonaRes is the preferred domain repository for agricultural and soil-related outputs, offering expert curation, discipline-specific metadata, and integration with SoilWise. Other repositories interoperable with OpenAIRE (e.g., Dataverse, EJP SOIL, ISRIC) may also be used.

3.2.6. Keywords and Vocabularies: Enhancing Discoverability

The assignment of high-quality, standardized keywords is a critical component of metadata creation, directly impacting the discoverability, interoperability, and reuse of research outputs. Within SoilWise and its harvested repositories such as BonaRes, the use of controlled vocabularies is foreseen as a key feature to ensure alignment with FAIR principles, particularly those relating to *Findability* and *Interoperability*. SoilWise already supports the use of controlled vocabulary keywords through the keyword matcher component, which processes and aligns terms provided via repositories such as Zenodo or Dataverse.

Use of Standardized Vocabularies

Knowledge providers are required to select keywords from established, domain-relevant vocabularies, including:

1. **ISO** 11074:2018 for soil-specific concepts (ISO 11704:2018).
2. **AGROVOC** (FAO) for agricultural terminology;
3. **GEMET** (EEA) for environmental terms;
4. These sources ensure consistency in metadata indexing and facilitate cross-repository discovery via semantic search tools and harvesting systems such as EjpSoil, Impact4Soil, Prepsoil and the BonaRes infrastructure.

3.3. Mission Soil–Specific Requirements

Knowledge providers contributing under the **Mission Soil framework** must follow specific additional requirements. Each published record must include the **Grant Agreement number** and project acronym in the metadata, ensuring automatic linkage with **CORDIS** and **OpenAIRE**. At this stage, the automated harvesting setup is confirmed for Mission Soil GA numbers; support for other Horizon Europe or national projects may depend on future repository configurations or harvesting expansions.

Public-facing documents must include a plain-language Executive Summary and acknowledgements referencing the funding source. The Executive Summary should summarize the key knowledge content of the document and will be used by SoilWise for indexing and making the document better discoverable.

Where applicable, CORDIS URLs should be listed under related identifiers, and any official Mission Soil templates (once available) should be applied to ensure consistency across projects and repositories.

3.3.1. Quick Pre-Publication Checklist

Before publishing, verify that your knowledge record meets all FAIR and Mission Soil requirements:

- Document structure complete (front matter, methods, results, annexes, executive summary).
- Open and archival formats used (PDF/A, CSV, HTML, code + licence).
- Dublin Core metadata complete;
- GA number and DOIs entered.
- Appropriate licence selected (CC BY for content, CC0 for metadata).
- Repository chosen (e.g. Zenodo, BonaRes, Dataverse) and record created with version noted.

- Adding soil keywords selected from AGROVOC, GEMET, or INSPIRE vocabularies in title or abstract description, and as metadata keywords.
- Contact information and changelog provided for traceability.

Following these steps ensures that your knowledge becomes a **fully discoverable, reusable asset** within the European soil knowledge ecosystem. Properly structured and licensed outputs are harvested by **OpenAIRE** and indexed in **SoilWise**, automatically linking them to related datasets, publications, and Mission Soil projects. This process not only ensures compliance with FAIR principles but also maximizes visibility, citation, and long-term value of your research.

By aligning with these practices, knowledge providers contribute to a more coherent, transparent, and sustainable European soil knowledge commons.

4. Management: How to publish and manage knowledge

This chapter provides practical guidance on how to publish, link, and manage knowledge within the SoilWise ecosystem. It builds upon the FAIR Data Strategy (D2.5) and describes how project outputs can be made discoverable, accessible, and reusable through trusted repositories that are interoperable with SoilWise. It outlines the publication workflow (Section 4.2), specific guidance for Mission Soil and Horizon Europe (HE) projects (Section 4.3), principles for updating and versioning (Section 4.4), and forward-looking aspects of knowledge evolution and integration (Section 4.5).

4.1 Publishing Knowledge in SoilWise: Repositories and Workflows

(Connected to Deliverable D2.5 – Data FAIR Strategy)

SoilWise does not host or accept direct submissions. Instead, it indexes and enriches metadata from **trusted repositories** (e.g., BonaRes, Zenodo, Dataverse, ISRIC) through harvesting and semantic augmentation. This includes peer-reviewed publications as well as other knowledge types such as grey literature (e.g., reports, policy briefs, technical notes), provided they are deposited in OpenAIRE-compatible repositories with persistent identifiers and appropriate metadata.

This workflow aligns project-level requirements (such as inclusion of Grant Agreement numbers), repository best practices (e.g., DOIs, licensing, controlled vocabularies), and SoilWise components (harvesting, metadata augmentation, keyword matching) to ensure that all research outputs, regardless of origin, become Findable, Accessible, Interoperable, and Reusable (FAIR).

Generic Workflow for All Knowledge Providers

1. Collecting results from Mission Soil projects

The workflow begins with the publication of results from officially recognised Mission Soil projects. Each project must include its Grant Agreement (GA) number, which corresponds to the main identifier linking datasets, publications, and tools across systems such as CORDIS and OpenAIRE. When the GA number is correctly entered in repository metadata, project outputs become automatically discoverable across European knowledge infrastructures and in the SoilWise catalogue.

2. Metadata harvesting and cataloguing

Once outputs are deposited in connected repositories, SoilWise harvests the metadata and integrates it into its central catalogue. The discoverability of a resource depends on the completeness and quality of its metadata. To ensure interoperability, metadata must comply with recognized standards such as Dublin Core or ISO 19139, and every resource should have a persistent identifier (DOI or similar) linking directly to its content. SoilWise augments harvested metadata through translation, normalization, and keyword harmonization, improving alignment across languages and disciplines.

- Each record must have a **persistent identifier (DOI)**.

- Metadata must meet **Dublin Core or ISO** standards.
- Enrichment processes (translation, keyword harmonization, semantic linking) improve interoperability and visibility.

3. From metadata to knowledge discovery

Harvested metadata is indexed and semantically enriched so that users can search by geography, topic, or project. Records with high-quality, complete metadata are prioritized in the search interface. The combination of persistent identifiers, semantic tagging, and ontology references allows SoilWise to transform static metadata into actionable knowledge, connecting related datasets, reports, and models.

4. Ensuring reusability

Publishing through trusted repositories such as Zenodo ensures wide accessibility and compliance with FAIR and Open Science principles, while domain-specific platforms like BonaRes provide expert curation and discipline-specific metadata. SoilWise then links and enriches these records through its Knowledge Graph and keyword-matching components, ensuring that deposited knowledge becomes reusable and discoverable beyond its original repository.

4.2 Repository Linkage and Data Management for HE and Mission Soil Projects

Ensuring effective publication and reuse of research outputs under Horizon Europe (HE) and Mission Soil projects requires a consistent approach to repository linkage and metadata management. SoilWise facilitates this process by integrating project results from trusted repositories into a central discovery environment.

1. Find a connected repository

- Mission Soil projects should deposit their results in repositories that are interoperable with SoilWise and connected via **OpenAIRE**.
- Examples of sources from which SoilWise harvests metadata include: **BonaRes, Zenodo, ISRIC, EJP Soil, Impact4Soil, PREPSOIL, CIRAD Dataverse, RDG Dataverse, Springer, ScienceDirect, FAO, EEA**, and others. Note that while some platforms like Zenodo or BonaRes allow direct deposition by users, others such as EJP Soil, Impact4Soil, or PREPSOIL serve primarily as aggregators or project-specific portals and may not accept external submissions.
- Depositing in these repositories ensures that outputs, datasets, publications, models, or other resources, can be harvested and linked into SoilWise.

2. Provide complete and standardized metadata

- Metadata is the key enabler of discoverability. SoilWise harvests and indexes metadata through OpenAIRE and other connectors.
- Critical fields include:
    - **Grant Agreement Number (GA No.)** – mandatory to identify and filter Mission Soil results.
    - **Persistent Identifiers (DOIs)** linking to datasets, documents, or tools.
    - **Licensing information** (e.g., CC-BY, CC0) for clarity on reuse conditions.
    - **Executive summaries** for documents, ensuring accessibility to non-specialists.

- **Keywords**, preferably selected from recognized vocabularies
- Metadata must conform to widely recognized standards (e.g., **Dublin Core**) to ensure interoperability across repositories.

3. Metadata harvesting and quality assurance

- Deposited outputs are harvested via SoilWise tools and indexed in the **metadata catalogue**.
- Outputs that have been assigned a **Mission Soil GA No.** will be recognizable, filterable and when applicable prioritized in search results.
- The **quality and completeness of metadata** (e.g., detailed descriptions, controlled keywords, correct licensing) determine how easily results can be found and reused.
- Poorly described outputs risk low visibility, whereas well-curated knowledge objects gain higher discoverability in SoilWise and across platforms such as EOSC and FAIRagro.

4. Linking repositories to project identities

- Linking project results to **CORDIS** through their GA number enables automated association of outputs with their parent project.
- This linkage strengthens accountability, supports reporting obligations, and ensures visibility within both the Mission Soil community and the wider HE ecosystem.

5. Continuous improvement and user feedback

- SoilWise provides mechanisms for users to **give feedback** when records are missing or incomplete.
- This feedback loop allows continuous refinement of metadata harvesting and repository integration, ensuring that SoilWise evolves as a reliable and user-oriented knowledge hub.

4.3 Publishing and Updating Knowledge Assets

Repositories like **BonaRes** and **Zenodo** provide robust infrastructure for publishing and versioning knowledge products. Each project is assigned a **persistent DOI** upon initial publication, which remains valid across all future versions.

For Mission Soil projects and similar Horizon Europe efforts, this traceable versioning supports transparency, reproducibility, and compliance with open science mandates. BonaRes and other repositories data stewards' teams assist researchers in managing version control, ensuring that all legal, technical, and metadata quality criteria are met during resubmission.

5. Reference

1. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. https://doi.org/10.1038/sdata.2016.18
2. Donmez, C., Schmidt, M., Cilek, A., Grosse, M., Paul, C., Hierold, W., & Helming, K. (2023). Climate change impacts on long-term field experiments in Germany. *Agricultural Systems*, *205*, 103578.
3. Donmez, C., Sahingoz, M., Paul, C., Cilek, A., Hoffmann, C., Berberoglu, S., ... & Helming, K. (2024). Climate change causes spatial shifts in the productivity of agricultural long-term field experiments. *European Journal of Agronomy*, *155*, 127121.
4. *ISO 30401:2018 – Knowledge management systems – Requirements.* International Organization for Standardization, Geneva.
5. *ISO 11704: 2018 – Water quality — Gross alpha and gross beta activity — Test method using liquid scintillation counting.* International Organization for Standardization, Geneva.
6. Xuemeng, T., Rolf, S., Serkan, I.M., Robert, Minarik., Yufeng, H., Tom, H. (2024). *Spatiotemporal prediction of soil organic carbon density (SOCD) for pan-Europe (2000–2022) in 3D+T* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13774197

6. Annex (Knowledge Reuse)

6.1. Knowledge reuse in practice

Knowledge reuse is the ultimate objective of the SoilWise Knowledge Management strategy, ensuring that research outputs are not only accessible but also applied in new contexts. The examples below demonstrate how repositories harvested by SoilWise already enable reuse across scientific and policy domains.

6.1.1 BonaRes Repository: Long-Term field Experiment (LTE) Knowledge Integration

The BonaRes Repository, integrated within SoilWise, showcases how domain-specific infrastructures transform curated data into reusable knowledge. Its Long-Term field Experiment (LTE) collection provides harmonized metadata, experimental documentation, and persistent identifiers (DOIs) that support comparative analyses, climate modelling, and educational reuse. Through SoilWise harvesting, these LTE resources become discoverable alongside other Mission Soil outputs, enabling cross-project and multi-scale knowledge synthesis.

Beyond technical data, BonaRes exemplifies how **metadata-rich, interoperable experiments** enable knowledge transfer from research to policy, education, and practical soil management. The **LTE Overview Map**, developed under the BonaRes initiative, further enhances reuse by providing harmonized access to site information, experiment types, and geographic metadata, all of which can be **discovered and connected through SoilWise's harvesting framework** (Figure 6.1)

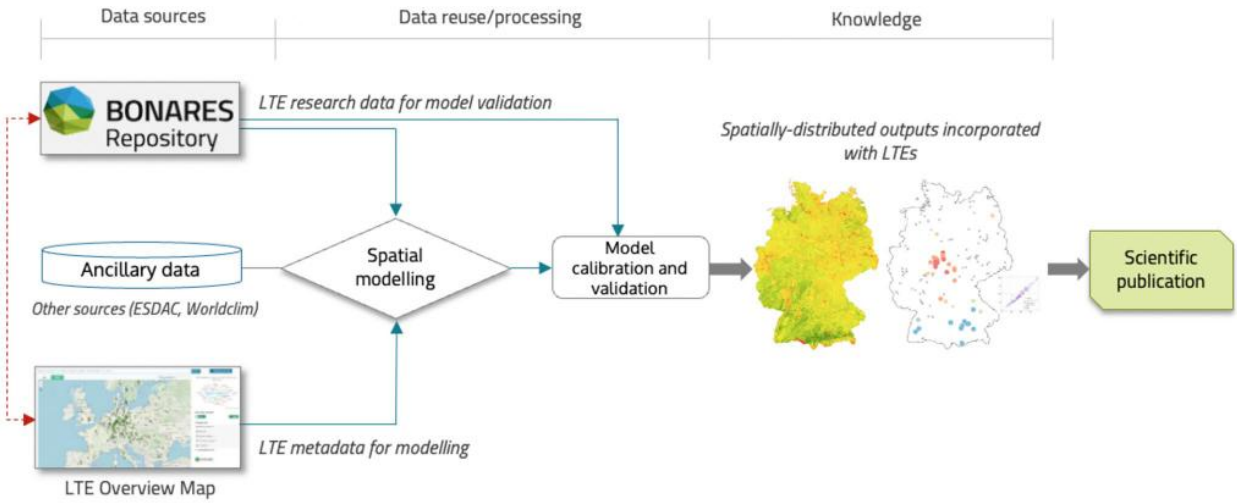

Figure 6.1 Workflow of LTE data reuse from the BonaRes Repository and LTE Overview Map for spatial modelling and publication (adapted from Dönmez et al. 2024; CC BY 4.0).

6.1.2 Zenodo: Pan-European Soil Health Knowledge Sharing

Zenodo demonstrates how open repositories contribute to large-scale reuse. The AI4SoilHealth project's European Soil Type maps, shared under an open license with structured metadata, are discoverable through SoilWise and support research, policymaking, and AI applications. This example highlights how standardised

metadata, persistent identifiers, and FAIR-aligned publication ensure broad reusability beyond the original project scope.

6.2 Future Directions for Knowledge Reuse

As **SoilWise** evolves, knowledge reuse will increasingly benefit from:

- **Semantic technologies and knowledge graphs** linking publications, datasets, and external ontologies (e.g., EuroSciVoc, INSPIRE Soil).
- **AI-assisted discovery** to recommend related resources or workflows based on user searches.
- **Integrated interfaces** that present related datasets, publications, and experimental metadata together, enabling users to navigate knowledge by topic, region, or project rather than searching separately.
- Improved support for software and model reuse through standard citation practices, dependency archiving, and integration with trusted code repositories (e.g., institutional GitLab, Software Heritage).

6.3. Finding and Accessing Knowledge in SoilWise

Building on the access framework introduced earlier, SoilWise operates as a **public discovery hub** that connects users to authoritative repositories (e.g., BonaRes, Zenodo, Dataverse nodes, EJP SOIL, ISRIC). SoilWise **does not store originals**; it separates **discovery** (in SoilWise) from **access and download** (at the source repository), preserving ownership while improving findability and trust.

What SoilWise adds on top of basic search is **structured, machine-actionable context**. Records are indexed with persistent identifiers (e.g., DOIs) and **standards-aligned metadata** (INSPIRE/DataCite/Schema.org profiles), enabling reliable citation, cross-platform harvesting, and interoperability with external ecosystems (EOSC, FAIRagro, Google Dataset Search).

For users, this means:

- **Targeted discovery** that quickly narrows large collections to relevant, research-ready resources (with both map and list views for orientation and inspection).
- **Context in one place**—where available, entries surface related publications, projects, and derived outputs, reducing fragmentation between literature and underlying resources.
- **Direct hand-off** to the source repository for licensing details, usage terms, and customizable downloads.

For contributors, SoilWise strengthens **visibility and traceability**: PID-based records are more easily cited and discovered across platforms, with usage signals (where available) feeding back into institutional and project reporting.

Looking ahead, development focuses on **semantics-driven discovery** and assistive features:

- **Knowledge Graph–based linking** to relate datasets, publications, projects, and external ontologies (e.g., EuroSciVoc), improving thematic navigation and context.
- **AI-supported recommendations** to surface adjacent resources and reduce search friction for non-expert users.

In short: SoilWise bridges the gap between scattered sources and **actionable knowledge**, combining standardised metadata, persistent identifiers, and semantic enrichment to support research, policy, and practice across the European soil community.

6.4. Outlook and Integration with AI Systems

The SoilWise initiative is actively working to extend and refine its role as a trusted hub for soil-related data and knowledge. Several forward-looking actions are already in progress to strengthen the FAIRness (Findability, Accessibility, Interoperability, and Reusability) of research outputs and ensure alignment with European and international standards:

- **Joint Metadata Template with JRC**
  SoilWise, in collaboration with the Joint Research Centre (JRC), is developing a standardized metadata template. This template will improve interoperability across repositories and facilitate consistent documentation of soil-related knowledge, ensuring resources are more discoverable and reusable.
- **Guidelines for FAIR Publication**
  Comprehensive guidelines are under development to support researchers, practitioners, and policymakers in preparing and publishing data and knowledge in line with FAIR principles. These guidelines will provide step-by-step instructions tailored to different resource types.
- **Recommended Repositories**
  SoilWise will issue a curated list of recommended repositories—both general-purpose and domain-specific—to assist knowledge providers in selecting appropriate deposit locations that maximize visibility and reuse potential.
- **Instructions for Metadata Provision**
  An initial framework for metadata provision across diverse resource types (datasets, publications, models, tools, policy briefs, etc.) defining minimum and recommended fields in line with international standards such as Dublin Core. Detailed implementation guidance and templates is further elaborated in Deliverable D2.5 (Data FAIR Strategy) to ensure full alignment across SoilWise components.
- **Recommendations on Vocabularies**
  To support semantic interoperability, SoilWise will publish recommendations on the use of controlled vocabularies and ontologies (e.g., AGROVOC, GEMET, INSPIRE) that enable machine-actionable knowledge discovery across domains and platforms.
- **Formats and Content Elements**
  Outlines general principles for preferred file formats and content elements (e.g., geospatial layers, time-series data, documents, visualizations) to promote consistency and reusability. Specific technical specifications and validation procedures is developed in coordination with D2.5, where they will be consolidated into the FAIR publication guidelines.

Together, these efforts will ensure that SoilWise evolves into a sustainable, user-driven knowledge ecosystem. By combining technical rigor with practical usability, the platform will strengthen its role as a cornerstone of the Mission Soil agenda and provide long-term value to the wider soil health community in Europe and beyond.

6.5 Conclusion

Knowledge reuse transforms repositories like SoilWise, BonaRes, and Zenodo from passive archives into active engines of innovation. By ensuring that outputs are well-described, openly licensed, and semantically linked,

these platforms allow knowledge to be continuously rediscovered and applied in new contexts—whether for research, policy, education, or land management. This ecosystem approach ensures that every published resource contributes to the collective effort of achieving healthier European soils and advancing sustainable practices globally.