D2.2 Developed & Integrated DM components, v2

M18/FEBRUARY 2025

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	D2.2 Developed & Integrated DM components Repository architecture, v2
Delivery Time (DOA)	M18
Deliverable Submission Date	28/02/2025
Status	V2
Dissemination Level	PU - Public
Deliverable Lead	Tomáš Řezník (MU)
Author(s)/Organisation(s)	Dajana Snopková (MU), Tomáš Řezník (MU), Paul van Genuchten (ISRIC), Nick Berkvens (EV ILVO)
Contributor(s)	MU, WU, ZALF, EV ILVO, DOMG – VL O, CREA, ISRIC, WE, NP
Peer-Reviewers	Céline Blitz-Frayert (CIRAD) , Giovanni L`Abate (CREA)
Contact	tomas.reznik@sci.muni.cz
Work Package	WP2
Dissemination level	Public
Keywords	data management components, prototype, technical documentation
Abstract	The purpose of this deliverable is to describe the process of co-design and implementation of Data Management components of the SoilWise repository and the current version of Technical documentation comprising a description of the component's functionalities after the first integration and validation cycle.

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's noted that in the grant agreement, the partner is identified by the acronym VL O (Vlaamse Gewest).

List of Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
CI-CD	Continuous Integration and Continuous Delivery / Development
CIRAD	CENTRE DE COOPERATION INTERNATIONALE EN RECHERCHE AGRONOMIQUE POUR LEDEVELOPPEMENT
CREA	CONSIGLIO PER LA RICERCA IN AGRICOLTURA E L'ANALISI DELL'ECONOMIA AGRARIA
CSW	Catalog Service for the Web
DM	Data Management
EJP	European Joint Programme
ESDAC	European Soil Data Centre
ETS	Executable Test Suite
EU	European Union
EUSO	EU Soil Observatory
EV ILVO	EIGEN VERMOGEN VAN HET INSTITUUT VOOR LANDBOUW- EN VISSERIJONDERZOEK
FAIR	Findable, Accessible, Interoperable and Reusable
INSPIRE	Infrastructure for Spatial Information in Europe
ISO	International Organization for Standardization
ISRIC	STICHTING INTERNATIONAL SOIL REFERENCE AND INFORMATION CENTRE

KM	Knowledge Management
M	Month
ML	Machine Learning
MU	Masaryk Univerzity
NP	NEUROPUBLIC AE PLIROFORIKIS & EPIKOINONION
OGC	Open Geospatial Consortium
PSI	public sector information
PU	Public
RDF	Resource Description Framework
SOS	Sensor Observation Service
STAC	Spatio Temporal Asset Catalogue
SWR	SoilWise Repository
T	Task
DOMG – VL O	VLAAMSE GEWEST
WE	WETRANSFORM GMBH
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
WP	Work Package

WPS	Web Processing Service
WR	STICHTING WAGENINGEN RESEARCH
WU	WAGENINGEN UNIVERSITY
ZALF	LEIBNIZ-ZENTRUM FUER AGRARLANDSCHAFTSFORSCHUNG

Table of Contents

List of Tables and Figures

Executive Summary

The SoilWise project, under the Horizon Innovation Actions, aims to develop an open-access soil knowledge and data repository to support users in safeguarding soil across Europe. With 60-70% of European soils currently considered unhealthy, i.e. are losing their capacity to support food production, biodiversity and the regulation of water, nutrient and carbon cycles. The project addresses the urgent need for reliable and harmonized data to support informed decision-making at various levels, aligning with the EU Mission 'A Soil Deal for Europe' and other related strategies. SoilWise, spanning 48 months and involving 15 partners, is designed to create a scalable, modular repository that leverages AI and ML technologies to make soil data Findable, Accessible, Interoperable, and Reusable (FAIR).

This deliverable, D2.2 Developed & Integrated DM Components, presents the development, integration, and validation of Data Management (DM) components within the SoilWise repository. This deliverable is already a second version reporting on the development works after the first integration and validation project phase out of three foreseen. It builds upon the previous version delivered after the first development project phase (M13). Developments of SoilWise Repository follow architectural and technical specifications, aligning with the broader project objectives and stakeholder requirements. The report details the agile methodology used, emphasizing iterative development cycles, stakeholder co-design, and robust validation processes. Technical documentation available at https://prototype-2-0.soilwise-documentation.pages.dev accompanies this deliverable to provide even more detailed information on the functionality, data flows and interfaces of delivered components.

The second prototype of the SoilWise Repository (SWR) integrates several key DM components, including:

- **Harvester:** Collects and harmonizes metadata and knowledge resources.
- **Catalogue:** Facilitates querying, metadata downloads, and visualization.
- **Metadata Validation:** Ensures metadata completeness and structural integrity.
- **Metadata Augmentation:** Enhances metadata records with additional links and keywords.
- **Transformation and Harmonization:** Standardizes metadata formats and spatial references.
- **Repository Storage:** Provides secure and scalable storage solutions.

Developed through an agile, user-centric approach, these components underwent a rigorous validation process involving technical assessments and stakeholder feedback. Integration followed ISO 19100 standards, progressing from “loose integration” to a hybrid model for enhanced scalability and flexibility.

This deliverable marks a significant milestone in the SoilWise project, laying a robust foundation for future development phases. The next deliverable version (M31) will focus on deepening integration and expanding functionalities to further align with EU soil health objectives.

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 collected at local, national and EU levels to allow **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, making them FAIR (Findable, Accessible, Interoperable and Reusable) 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 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, 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.

1.2 Document scope

This deliverable describes the methodology used in the process of co-designing SoilWise repository (SWR) Data Management (DM) components, as well as their implementation, integration, validation, and delivery of their source code. As such, it documents the developments of the SWR prototype within the first project cycle out of the three planned. The first (described in D2.1) and second prototypes have been designed as a proof-of-concept, partly still building on existing (open source) components and integration of independent developments of SoilWise partners. Nevertheless, all the technical development achievements are tightly coupled with the SoilWise mission, scope, and deliverables that have been published so far. Among others, this deliverable (1) comprises achievements that are framed by the architecture defined in D1.3, (2) respects and follows stakeholders' demands/suggestions described in the D1.1, (3) has been shaped by a number of in-depth discussions with the European Commission, DG REA/JRC/AGRI, and (4) dozens of Mission Soil Horizon Europe projects.

Technical documentation was built in the form of an external web page accompanying this deliverable to capture the complexity of the SoilWise achievements. This documentation consists of a description of functionalities, detailed technical specifications, and interfaces between all modular components that will be employed in the SWR. This fits with one of the work package (WP) 2 goals to detail the design and develop the DM components of SWR, more specifically, within the task (T) 2.1 Design of the data technology components (M7-M37) and T2.2 Implementation and deployment of data components (M7-M46).

Initially, a high-level architectural **design** was established, which was detailed in deliverable D1.3 and further refined through collaborative efforts in development clusters and monthly meetings. The **implementation** phase adopted an agile strategy (see Chapter 2), where the development process was broken down into six iterative cycles, each lasting four weeks. Key activities included populating and refining a Product Backlog, planning sprints, developing components, and conducting reviews and retrospectives to ensure tasks met the Definition of Done (DoD) criteria. Further, we followed with a continuous development approach, addressing requirements and incorporating feedback from technical **validation** activities by User Case (UC) groups. Feedback was gathered through an online form, compiled in an Excel document, and translated into technical issues managed via GitHub. Regular biweekly meetings facilitated collaboration between developers and UC partners. Development of Data and Knowledge Management (KM) components initially occurred in independent environments, with **integration** achieved through open APIs, resulting in a "loose integration" per ISO 19100 standards. Future phases will gradually enhance integration depth, progressing from "mixed integration" to a "hybrid model" balancing tightly integrated core components with loosely integrated modular elements, ensuring scalability and flexibility of the SoilWise Repository (SWR).

Delivery of the second SoilWise Repository prototype comprises the following Data Management components: (1) **Harvester** component responsible for harvesting metadata and knowledge resources, harmonizing metadata, and identifying duplicates, (2) **Catalogue** powered by (3) **pycsw** allowing users to query the harvested resources, download metadata, and access map previews, and also links to knowledge sources, (4) **Metadata Validation** component focusing on evaluating metadata completeness, structure validation, and link liveliness assessment, (5) **Metadata augmentation** component suited to enrich source metadata records, e.g. by adding new links, (6) **Transformation and Harmonization** component handling tasks such as metadata standards transformation, CRS transformation, data restructuring, and format conversion, (7) **Repository Storage** component responsible for the storage of raw and augmented metadata, user-enhanced content, and knowledge graphs. Detailed documentation, source code, and access points are provided for each component in Chapter 3.

The first version of this deliverable (D2.1 Developed & Integrated DM components) was submitted in M13 after the first two phases (co-design and development) of the first project cycle (for more information, see Chapter 2.1). This (second) version reports on the development efforts after the additional integration and validation phase. It was created hand-in-hand with the other SoilWise technical Work Packages to seamlessly document all the technical developments and achievements so far. As such, this D2.2 document, together with D3.2 Developed & Integrated KM components (M18) and D4.2 Repository infrastructure, components, and APIs (M18), delivers the second prototype of SWR. The delivery of the second SWR prototype comprises (1) source code of developed technical components (DM + KM), (2) functionality description in the form of external Technical documentation, (3) Release notes (listed in D4.2), (4) three separate accompanying methodological and management overview documents.

The third version of this deliverable will be delivered in M31 (end of the second project iteration cycle), and the fourth version of this deliverable two months before the end of the project in M46 (end of the third project iteration cycle). All the mentioned deliverable versions will describe the technical developments of the SoilWise

Data Management components. This technical level will be accompanied by a complex mid- and long-term strategy, within the D2.5 Report on strategy for FAIRness on soil data in two iterations (M27 and M42).

1.2.1 Summary of deliverable version updates

The second version of this deliverable is, in comparison to the first version, enriched with two new chapters in the Methodology section. Specifically, Chapter 2.1 SoilWise project approach, where the phases of the project cycles are explained in more detail and Chapter 2.4 Validation and integration of Data Management components. This chapter was added to describe the activities that took place within the third phase of the first project cycle, as the first version of this deliverable was submitted before the start of this phase.

A second version of the SoilWise prototype (its source codes) was released together with this deliverable, thus the Chapter 3 Data Management components was updated correspondingly.

Additionally, minor updates were performed in the chapters: 1.2 Document scope, 1.3 Document structure, and 1.4 Relationship to other project deliverables.

1.3 Document structure

This document is comprised of the following chapters:

- **Chapter 1** provides an introduction to the project and the document
- **Chapter 2** explains the methodology used in the process of designing, implementing, integrating and validating SoilWise Repository Data Management components
- **Chapter 3** presents a list of Data Management components that are present in the first delivered prototype. It includes an external link to the Technical documentation that further details the component's functionality and technical specifications.
- **Chapter 4** details the Technical Documentation

Technical documentation is available at https://prototype-2-0.soilwise-documentation.pages.dev/, and a PDF exported version is also available on demand as a non-editable version saved at the date of the deliverable. The documentation includes links to the current versions of the components' code on Github, or version published on Zenodo, when available.

1.4 Relationship to other project deliverables

This deliverable relates to and complements the following deliverables:

- D2.1, D2.3, D2.4 – Developed & Integrated DM components, v1, v3, v4 (M13, M31, M47)
- D3.1, D3.2, D3.3, D3.4 – Developed & Integrated KM 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 DM components (T2.1 and T2.2)
- T1.4 Define SoilWise Architectural Design – will be considered in the design and implementation of DM components (T2.1 and T2.2)
- T1.5 Define SoilWise Multi-Stakeholder governance model – will be considered in the design and implementation of DM components (T2.1 and T2.2)
- **T2.1 Design of the data technology components**
- **T2.2 Implementation and deployment of data components**
- T2.3 AI and ML for data findability and accessibility – will extend the design and implementation of DM components (T2.1 and T2.2)
- T2.4 Strategy for FAIRness on soil data – will be followed in the design and implementation of DM components (T2.1 and T2.2)
- T3.1 Design of the KM components – will consider and align with the design of DM components (T2.1)
- T3.2 Implementation and deployment of knowledge component – will consider and align with the implementation of DM components (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 DM components (T2.1 and T2.2)
- T4.3 Solutions & repository validation and population – will validate designed and implemented DM components (T2.1 and T2.2)
- T5.2 User Cases implementation and demonstration – will demonstrate the functionality of implemented DM components (T2.2)
- T7.2 Technical and Scientific Management – prepares an Open Science and Data Management Plan (D7.2 – D7.4) describing the data management level of handled (meta)data resources.

2 Methodology

2.1 SoilWise project approach

The development of the SoilWise data and knowledge repository follows a user-centred, agile, and iterative approach, fostering collaboration among stakeholders, including data owners, reusers, managers, technology experts, policymakers, academics, and regional stakeholders. This four-phase process, illustrated on Figure 1, unfolds across three development cycles. **Phase 1** involves co-design and stakeholder engagement to identify innovation needs and align the repository with practical governance frameworks, emphasizing open innovation, co-creation, and multi-disciplinary problem-solving. **Phase 2** builds on these insights, developing technological and knowledge components through adaptive sprints, with feedback loops ensuring synchronization and continuous improvement. **Phase 3** integrates these components in a shared repository, where solutions are tested and tailored to user needs, combining existing datasets with project advancements to achieve desired readiness levels. In **Phase 4**, the solutions are demonstrated and evaluated within user cases, ensuring they address diverse stakeholder requirements, enhance data sharing and governance models, and provide added value for target groups. The evaluation process involves feedback from users, consortium partners, and external evaluators to ensure the repository's broad applicability and effectiveness for real-world challenges.

Figure 1 SoilWise project approach to the delivery of the SoilWise repository based on three development cycles, each comprising four phases

The following chapters describe these phases in more detail from the Data Management components' perspective.

2.2 Co-design of Data Management components

The design of the Data Management component was preceded by a high-level design of the architecture, which took place within the framework of T1.4 Define SoilWise Architectural Design. Its first version was the subject of deliverable D1.3 Repository architecture, v1. Within the framework of these activities, the original plan was already exceeded, and the next level of architecture - the individual functions of the components - was partially detailed in the Technical documentation. The architectural design resulted from the co-design project iteration that is described in Chapter 3 of deliverable D1.1. Usage scenarios, requirements.

After the submission of D1.3, work on the more detailed level of architecture design continued separately in the so-called "development clusters", each having its dedicated technical component. However, the original T1.4 working group continued to meet monthly at combined WP2+WP3+WP4 meetings to ensure the harmony of activities and easier integration of all technical components in the future. Meanwhile, the architectural design and the Technical documentation were updated to reflect the recent design suggestions and development activities.

2.3 Development of Data Management components

The development process for the Data Management components was based on an agile project management practice in one-month sprints. This approach, illustrated in Figure 2 and detailed in Chapter 7 of D1.1 Usage scenarios, requirements, follows an iterative workflow:

1. Adding user stories, requirements, acceptance criteria, and functionality descriptions to the Product Backlog[1]. For the first iteration, the Product Backlog was populated with the vision scenarios (see Chapter 3).
2. Reviewing and prioritising tasks in the Product Backlog.
3. Planning the sprint and defining the Sprint Backlog[2].
4. Developing technical component functionality according to the tasks in the current sprint.
5. Verifying acceptance and DoD (Definition of Done) criteria during the sprint review.
6. Discussing improvement practices during the sprint retrospective meeting.
7. Repeating the process for the next sprint.

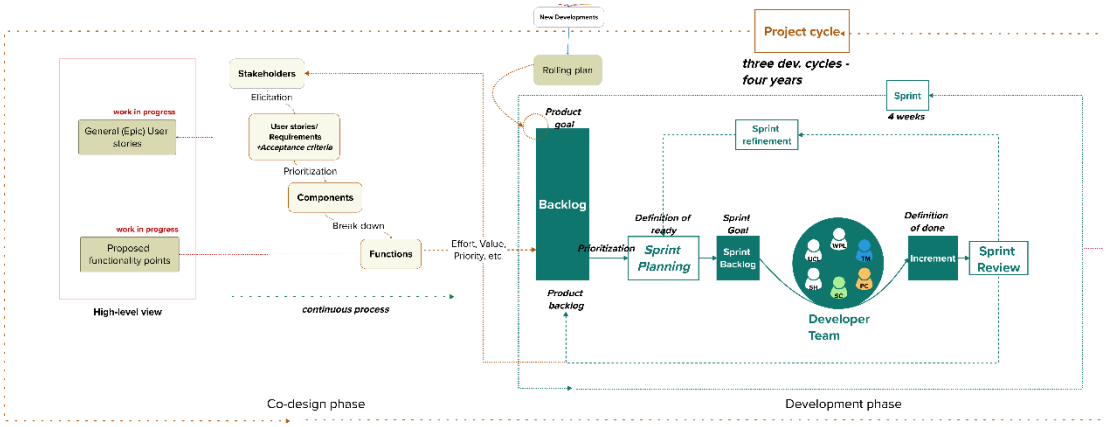

Figure 2 The agile process flow applied for the development of the Data Management components

Coordination of sprints and developments was managed using GitHub and a dedicated GitHub environment at https://github.com/soilwise-he. Separate repositories were created for each technical component, containing both the developed code and tasks (GitHub issues) addressed during the sprints. A GitHub project SoilWise Sprint

[1] a prioritized list of functionality which a product should contain and that drives the developmental work

[2] a list of work items or tasks a development team plans to complete during a project sprint

Backlog (Figure 3) was used to plan and track the developments of each sprint, and another project SoilWise Sprint Refinement was established for group discussion topics during weekly sprint refinement meetings. A project member was assigned responsibility for the development of each project component. Collaborators involved in the component's development were also identified, forming the developmental team for that component. A list of leads, collaborators, and repositories for each component is available at https://github.com/soilwise-he/Soilwise-Project-Backlog/blob/main/components_and_repo.md.

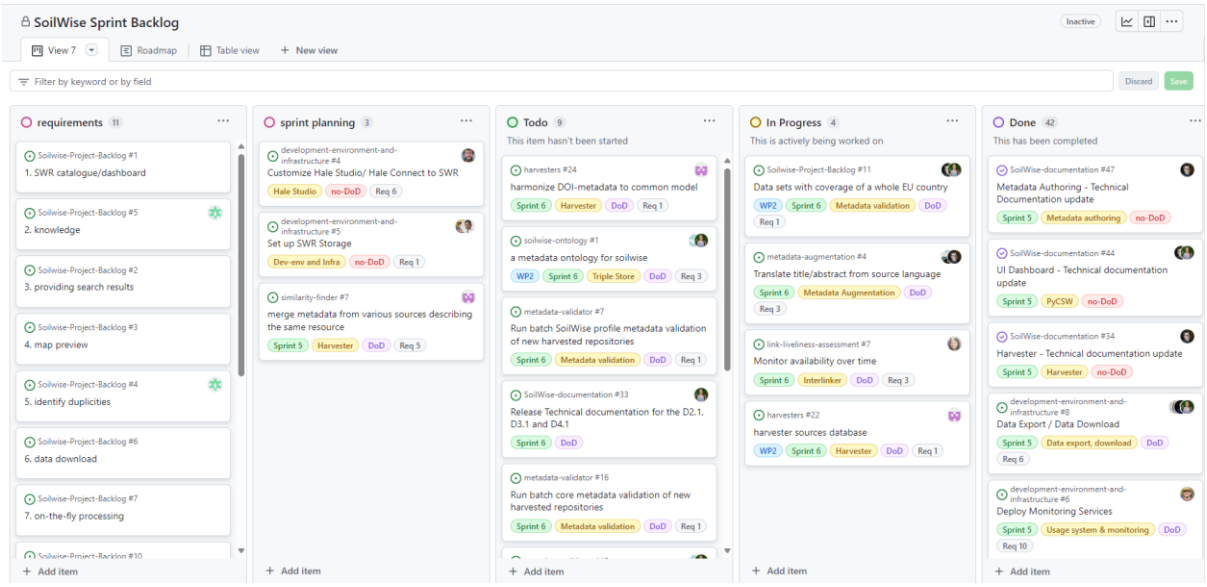

Figure 3 Illustration of the SoilWise Sprint Backlog board containing 11 functionalities (left requirements column) and associated tasks the development team plans to work on or is working on during each sprint (other columns).

Each sprint lasted four or five weeks, with six sprints conducted for the first iteration of development activities (Figure 4). Each sprint began with a Sprint Planning meeting where the component leads selected tasks for the current sprint and added them to the Sprint Backlog. Each task included a description of the problem, the expected developmental work, and one or more DoD criteria to determine task fulfilment. Weekly Sprint Refinement meetings involved component leads discussing developmental issues collectively. At the end of each sprint, a Sprint Retrospective meeting was held to evaluate the status of all developed tasks and confirm that completed tasks met the DoD criteria.

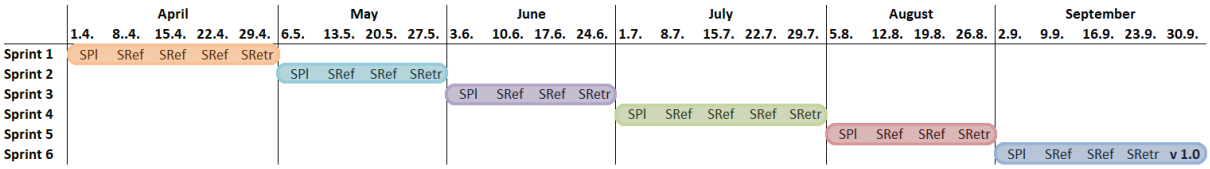

Figure 4 Gantt Diagram of first prototype development sprints, including Sprint Planning meetings (SPI), Sprint Refinement meetings (SRef), and Sprint Retrospective meetings (SRetr)

As project members continued to elicit and refine user stories, requirements, and acceptance criteria, a finalised Product Backlog was not available for the first iteration. The completed tasks in this first iteration pertain to the 10 functionalities defining the first prototype, described in chapter 2.5 of deliverable D1.3 Repository Architecture. An 11th functionality was added during the sprint. These 11 functionalities are documented in the GitHub project SoilWise Sprint Backlog (Figure 3) for guidance. Each functionality is further divided into sub-

functionalities with criteria for evaluating their fulfilment during development. While the 11 functionalities and their derived sub-functionalities drive the development activities in this first iteration, in the next iterations, a fully-fledged Product Backlog comprising the user stories, requirements, acceptance criteria, and functionalities is foreseen.

2.4 Validation and integration of Data Management components

Phase 3 of the project, Integration and Validation (M14 – M18), was characterized by a continuous development approach. On the one hand, developmental activities were driven by lower priority requirements from the prior phase, which focused on developing SoilWise components for knowledge and data management. On the other hand, development activities were equally driven by feedback originating from the technical **validation** activities carried out by the User Case (UC) groups in the task T4.3 Solutions & repository validation and population.

A detailed discussion of these validation activities will be provided in Deliverable D4.5 Repository Data and Knowledge Resources, v1 (M21). These validation activities were conducted in preparation of the individual demonstrators for each of the five UC groups and were coordinated under the scope of WP5. Deliverable D5.3 Deployment and Evaluation Report, v1 (M21), will provide a comprehensive overview of these demonstrations and their outcomes. The demonstrators will be presented to stakeholders and evaluated in Phase 4 of the project cycle.

Feedback from the validation activities was gathered through an online form. Comments were automatically centralized in an Excel document, regularly reviewed, and translated into technical issues. These issues were classified as feature requests or bug reports critical to showcasing the demonstrators, were uploaded to the SoilWise GitHub repository and assigned to developers for further resolution.

To foster collaboration, biweekly meetings and workshops were organized between developers and UC partners. These sessions encouraged discussions on SoilWise shortcomings and identified functionality requirements for the next project iteration. Feedback from these interactions was also gathered via the online form, processed, and addressed in the same structured manner as previously described.

Developments conducted within WP2 (Design and Development of Data Management components) and WP3 (Design and Development of Knowledge Management components) have begun in independent environments of the involved developer teams to manage resources efficiently, i.e., to save time for the iterative development itself rather than focusing on the integration of every component version at all costs. **Integration** of the independently developed components has been performed at the end of the first integration and validation phase (see Figure 1), where applicable, through open APIs, mostly the ones standardized by the World Wide Web Consortium and Open Geospatial Consortium (OGC). As a result, the SWR and its components are characterized by "loose integration" at the end of the first integration and validation phase, as defined by the ISO 19100 series of standards.

The following two development and integration and validation phases will place more emphasis on integration in depth. The second phase will move to the "mixed integration" level as per ISO 19100. The third development phase is expected to become a hybrid from the integration point of view: (1) the selected core components co-defined with SoilWise stakeholder groups will reach the "tight integration" level to provide a seamless proof-of-concept SWR. (2) The remaining components are likely to remain at the "loose/mixed integration" levels. In all cases, the foreseen integration levels support the open, modular, and scalable architecture of the SWR.

3 Data Management components

This chapter lists the Data Management components that are subject to the delivery of the second SoilWise repository prototype, including the links to their source code (GitHub), the latest release (Zenodo) and access point for users (if applicable). Knowledge Management components, which represent another part of the delivered prototype, are listed in Chapter 3 of parallel deliverable D3.2 Developed & Integrated KM components. Since data and knowledge in the SoilWise Repository are closely linked, in some cases, the components cannot be clearly assigned to only one Work Package as the developments have been shared by both WP2 and WP3. Therefore, in the list of Data Management component functions, we indicate that they belong to Knowledge Management components and vice versa.

A more detailed description of the components' functionality is described in the Technical documentation, see more detail in Chapter 4. A summary in the form of release notes is also available in Deliverable D4.2 - Repository infrastructure, components, and APIs, v1. These components are also visible in the high-level architecture schema depicted in Figure 5.

Name:	**Harvester**
Version:	0.2.0
Functions:	Harvest metadata resources Harvest knowledge resources *(part of KM)* Metadata harmonization Metadata RDF Turtle serialization RDF to Triple Store *(part of KM)* Duplication identification *(part of KM)*
Documentation:	https://prototype-2-0.soilwise-documentation.pages.dev/technical_components/ingestion/
Source code:	https://github.com/soilwise-he/harvesters/releases/tag/v0.2.0
Release:	https://doi.org/10.5281/zenodo.14923563
Access point:	not applicable
Name:	**Catalogue User Interface**
Version:	3.0.0.beta1
Functions:	Query catalogue (full text) Query catalogue (facet filters) Data download (AS IS) Map preview Display metadata augmentation results User Engagement
Documentation:	https://prototype-2-0.soilwise-documentation.pages.dev/technical_components/catalogue/

Source code:	https://github.com/soilwise-he/catalogue-ui/releases/tag/v0.2.0
Access point:	https://repository.soilwise-he.eu/
Name:	**pycsw**
Version:	3.0.0.beta1
Functions:	CSW API OGC API-Records STAC API OAI-PMH API OpenSearch
Documentation:	https://prototype-2-0.soilwise-documentation.pages.dev/technical_components/catalogue/
Source code:	https://github.com/geopython/pycsw/releases/tag/3.0.0-beta1
Access point:	https://repository.soilwise-he.eu/cat/csw https://repository.soilwise-he.eu/cat/openapi https://repository.soilwise-he.eu/cat/stac https://repository.soilwise-he.eu/cat/oaipmh https://repository.soilwise-he.eu/cat/opensearch
Name:	**Metadata Validation**
Version:	0.2.0
Functions:	Metadata profile validation Metadata completeness check Link liveliness assessment v1.1.4
Documentation:	https://prototype-2-0.soilwise-documentation.pages.dev/technical_components/metadata_validation/
Source code:	Metadata profile validation: https://github.com/soilwise-he/metadata-validator/releases/tag/v0.2.0/ Metadata completeness check: https://github.com/soilwise-he/metadata-validator/releases/tag/v0.2.0/ Link liveliness assessment v1.1.4: https://github.com/soilwise-he/link-liveliness-assessment/releases/tag/1.1.4
Release:	Metadata validator: https://doi.org/10.5281/zenodo.14924543 Link liveliness assessment: https://doi.org/10.5281/zenodo.14923790
Access point:	Metadata profile validation: https://data.soilwise.wetransform.eu/#/home (authorisation needed, account can be provided on demand) Metadata completeness check: https://dashboards.isric.org/explore/?form_data_key=rLDeTSV5GTFh7loIglovOWGyts1K2iMov5IWtzReKWlj4oMDfmeI_wTtR4Dkw64k&slice_id=337&save_action=overwrite Link liveliness assessment: https://api.soilwise-he.containers.wur.nl/linky/docs

Name:	Metadata Augmentation
Version:	0.2.0
Functions:	Translation module *(part of KM)* Keyword matcher Metadata interlinker
Documentation:	https://prototype-2-0.soilwise-documentation.pages.dev/technical_components/metadata_augmentation/
Source code:	https://github.com/soilwise-he/metadata-augmentation/releases/tag/v0.2.0
Release:	https://doi.org/10.5281/zenodo.14924181
Access point:	not applicable
Name:	**Transformation and Harmonization**
Version:	5.3.0
Functions:	Manual metadata upload Metadata transformation CRS transformation Data restructuring Format transformation Codelist mapping Units of measurements conversion Download interoperable metadata
Documentation:	https://prototype-2-0.soilwise-documentation.pages.dev/technical_components/transformation/
Source code:	https://github.com/halestudio/hale/releases/tag/v5.3.0
Access point:	https://data.soilwise.wetransform.eu/#/home (authorisation needed, account can be provided on demand)
Name:	**Repository Storage**
Version:	Postgres release 12.22 Virtuoso release 07.20.3239
Functions:	Storage of user-enhanced content – GIT *(part of KM)* Storage of raw harvested metadata – PostgreSQL / vector DB Storage of augmented metadata – PostgreSQL / vector DB Storage of augmented, linked metadata, knowledge graph - Triple Store *(part of KM components)*
Documentation:	https://prototype-2-0.soilwise-documentation.pages.dev/technical_components/storage/
Source code:	not applicable

Access point:	GIT: https://github.com/soilwise-he/soilinfohub/discussions PostgreSQL: https://pgadmin.isric.org Triple Store (Virtuoso): https://sparql.soilwise-he.containers.wur.nl/sparql

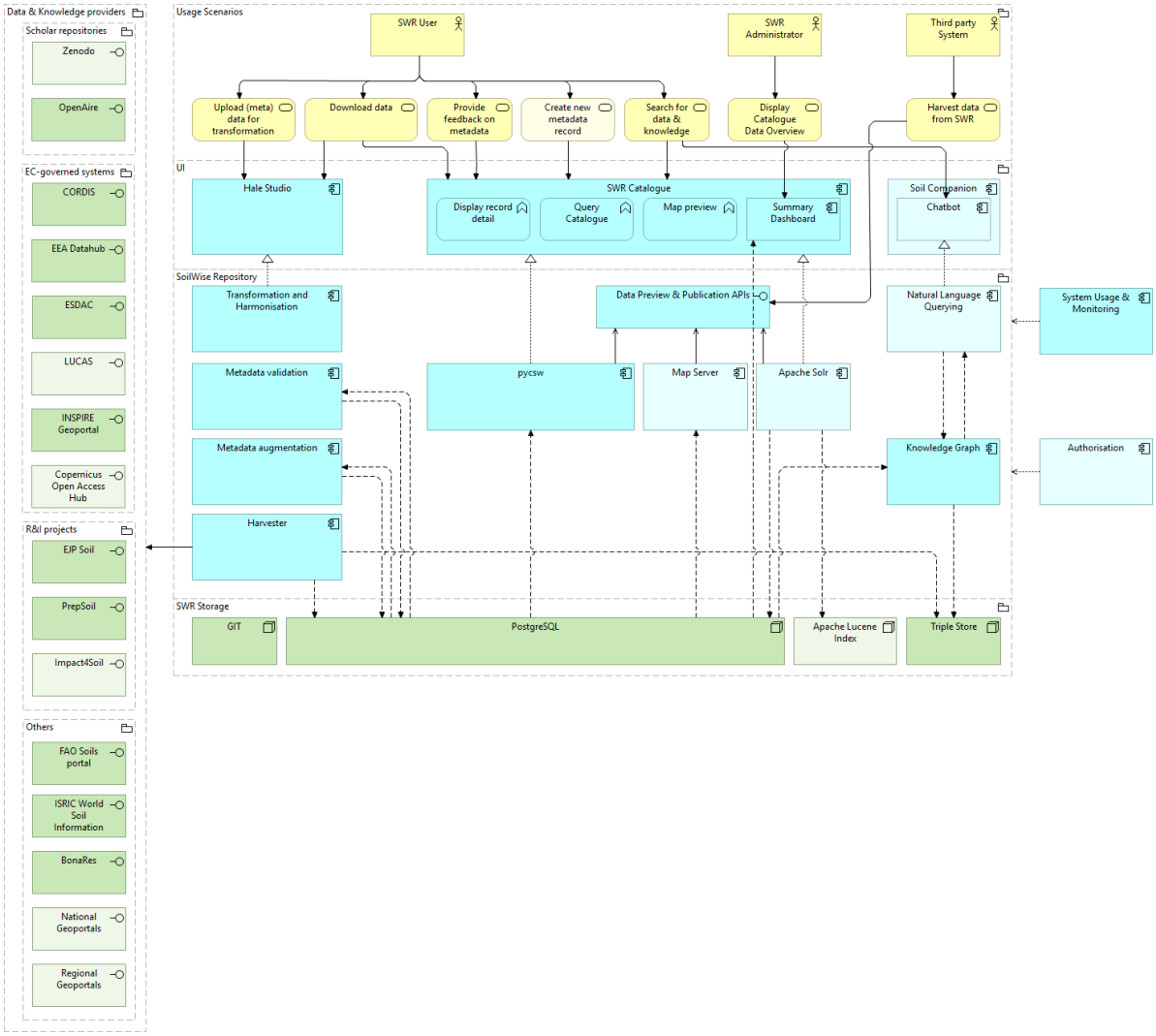

Figure 5 High-level architecture diagram of the SoilWise Repository (light-colored elements are under development and not part of 2nd prototype)

4 Technical documentation

The documentation of SoilWise Repository architecture is maintained in the public GitHub repository: https://github.com/soilwise-he/SoilWise-documentation. It comprises a description of functionality, detailed technical specifications, and interfaces between all modular components and is structured according to the main technical components. Note that the SWR documentation is also a living environment that is continuously updated during the development process. For this deliverable, a stable release is published at: https://prototype-2-0.soilwise-documentation.pages.dev/, and a PDF exported version is also available on demand as the non-editable version made at the date of the deliverable submission.

References

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.