D1.3 Repository architecture

M08/APRIL 2024

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	D1.3 Repository architecture
Delivery Time (DOA)	M08
Deliverable Submission Date	30/04/2024
Status	V1
Dissemination Level	PU - Public
Deliverable Lead	Tomáš Řezník (MU)
Author(s)/Organisation(s)	Dajana Snopková (MU), Tomáš Řezník (MU), Paul van Genuchten (ISRIC), Rob Lockers (WR), Luis De Sousa (ISRIC), Nick Berkvens (EV ILVO)
Contributor(s)	MU, WU, ZALF, EV ILVO, VL O, CREA, ISRIC, WE, NP
Peer-Reviewers	Nikos Kalatzis (NP), Somakanthan Somaligam (WE)
Contact	tomas.reznik@sci.muni.cz
Work Package	WP1
Dissemination level	Public
Keywords	architecture, ArchiMate, technical components, functionality
Abstract	The purpose of this deliverable is to describe the process of co-definition of repository architecture and the current version of documentation for the architecture design comprising ArchiMate diagram of repository application layer and description of component's functionalities as foreseen.

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

List of Abbreviations

CREA	CONSIGLIO PER LA RICERCA IN AGRICOLTURA E L'ANALISI DELL'ECONOMIA AGRARIA
EUSO	EU Soil Observatory
EV ILVO	EIGEN VERMOGEN VAN HET INSTITUUT VOOR LANDBOUW- EN VISSERIJONDERZOEK
ISRIC	STICHTING INTERNATIONAL SOIL REFERENCE AND INFORMATION CENTRE
MU	Masarykova univerzita
NP	NEUROPUBLIC AE PLIROFORIKIS & EPIKOINONION
PU	Public
SWR	SoilWise Repository
T	Task
VL O	VLAAMSE GEWEST
WE	WETRANSFORM GMBH
WP	Work package
WR	STICHTING WAGENINGEN RESEARCH
WU	WAGENINGEN UNIVERSITY
ZALF	LEIBNIZ-ZENTRUM FUER AGRARLANDSCHAFTSFORSCHUNG

Table of Contents

List of Tables and Figures

1 Introduction

1.1 Project summary

Now more than ever, soil health is an issue that needs to be addressed urgently, as recent assessments state that 60-70% of European soils can be considered unhealthy (Bouman, 2022). The EU Mission 'A Soil Deal for Europe', the EU Soil Strategy and the proposal for 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 defining SoilWise repositiory (SWR) architecture and presents documented architecture diagram in ArchiMate notation. Technical documentation was built in the form of an external web page accompanying this deliverable. This documentation consists of 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) 1 goals to deliver a high-level prototype architecture of SoilWise, more specifically within the task (T) 1.4 Define SoilWise Architectural Design (M3-M46). The original plan foreseen in the Grant Agreement for first iteration of this deliverable was outperformed, the activities behind the architectural design went already beyond the high-level scope, partially overlapping with the activities related to T2.1 Design of the data technology components, and T3.1 Design of the KM components. All this for the purpose of better linking both tasks and not isolating development activities. Moreover, the presented architecture is aligned with the requirements of various stakeholder groups and those

originating from the EU Soil Observatory (EUSO) development. Requirements of various stakeholder groups within and beyond the SoilWise project, including other Mission Soil projects, were considered.

The content presented within this deliverable and accompanying Technical documentation can also be understood as a bridge between the SoilWise stakeholder groups and the technical core of the project for a joint understanding of the SWR development. The evolution process went as follows. As the first, the **"Kick-off architecture"** was established following requirements from the SoilWise Grant Agreement and of the Joint Research Centre of the European Commission. The "Kick-off architecture" was later elaborated on by the SoilWise technical partners. The **"User stories derived architecture"** was elaborated in parallel out of 64 user story epics from various stakeholder groups (for more information on user story epics, see D1.1 Usage scenarios, requirements, v1.0). The **"Bridging architecture"** was created as a further step to provide a synthesis of the "Kick-off architecture" and "User stories derived architecture". Moreover, the "Bridging architecture" was further validated during online meetings with individual Use Cases working groups. The core of this deliverable represents the so-called **"SWR Repository Architecture 1.0 - Brugges"**, which came out of the in-person Technical meeting in Brugges and focuses on implementing the SoilWise core product. At the same time, it is based on a consensus between stakeholder groups and technical viewpoints, providing an overarching vision for the upcoming SWR development.

This deliverable is the first release out of two foreseen and serves as a baseline to initiate the development activities needed to set up the repository governance model (WP1), the infrastructure (WP4) and develop Data (WP2) and Knowledge (WP3) Management Components. Following the agile approach, the architecture design will be continuously updated reflecting (1) new or updated requirements from the use cases stakeholder groups, (2) new details discovered during development, and (3) novel scientific and technology advances, all according to the product backlog and SoilWise Repository Rolling plan. The second, and final, version of this deliverable will be delivered a half year before the end of the project (M42).

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 and software used in the process of defining SoilWise Repository architecture and presents a list of envisioned modular functionalities of the SoilWise Repository core product
- **Chapter 3** presents the architecture diagram in ArchiMate notation, overview of all technological components together with external link to a short description of their foreseen functionality

Technical documentation is available at: https://soilwise-documentation.pages.dev/, and a PDF exported version is also available on demand, as non-editable version saved at the date of the deliverable. Interactive version of architecture diagram is available at: https://soilwise-architecture.pages.dev/

1.4 Relationship to other project deliverables

This deliverable relates to and complements the following deliverables:

- D1.4 – Repository architecture, v2 (M42)
- D1.1, D1.2 – Usage Scenarios, Requirements, v1, v2 (M6, M36)
- D1.5, D1.6 – Repository GM, v1, v2 (M21, M42)
- D2.1, D2.2, D2.3, D2.4 – Developed & Integrated DM components, v1, v2, v3, v4 (M13, M18, 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)
- D4.5, D4.6, D4.7 – Repository Data and Knowledge Resources, 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 architecture design (T1.4)
- **T1.4 Define SoilWise Architectural Design**
- T1.5 Define SoilWise Multi-Stakeholder governance model – will define the governance model of the repository based on the architecture design
- T2.1 Design of the data technology components – will elaborate the architecture design (T1.4) in more detail
- T2.2 Implementation and deployment of data components – will follow the architecture design (T1.4)
- T3.1 Design of the KM components – will elaborate the architecture design (T1.4) in more detail
- T3.2 Implementation and deployment of Knowledge component – will follow the architecture design (T1.4)
- T4.1 Repository digital infrastructure for deployment and delivery – will follow the architecture design (T1.4)
- T4.2 Interfaces for access, sharing, population and integration with EUSO – will follow and extend the architecture design (T1.4)

2 Methodology

2.1 Co-designing architecture repository

Following the quadruple helix framework (e.g. Carayannis & Campbell, 2010), SoilWise is driven by an evolutionary, agile, well-delineated and lean approach, which incorporates an open-innovation, co-creation, multi-stakeholder, and problem-solving approach. Whole architecture design process and its relation to requirements elicitation process, business analysis and development activities is depicted in Figure 1.

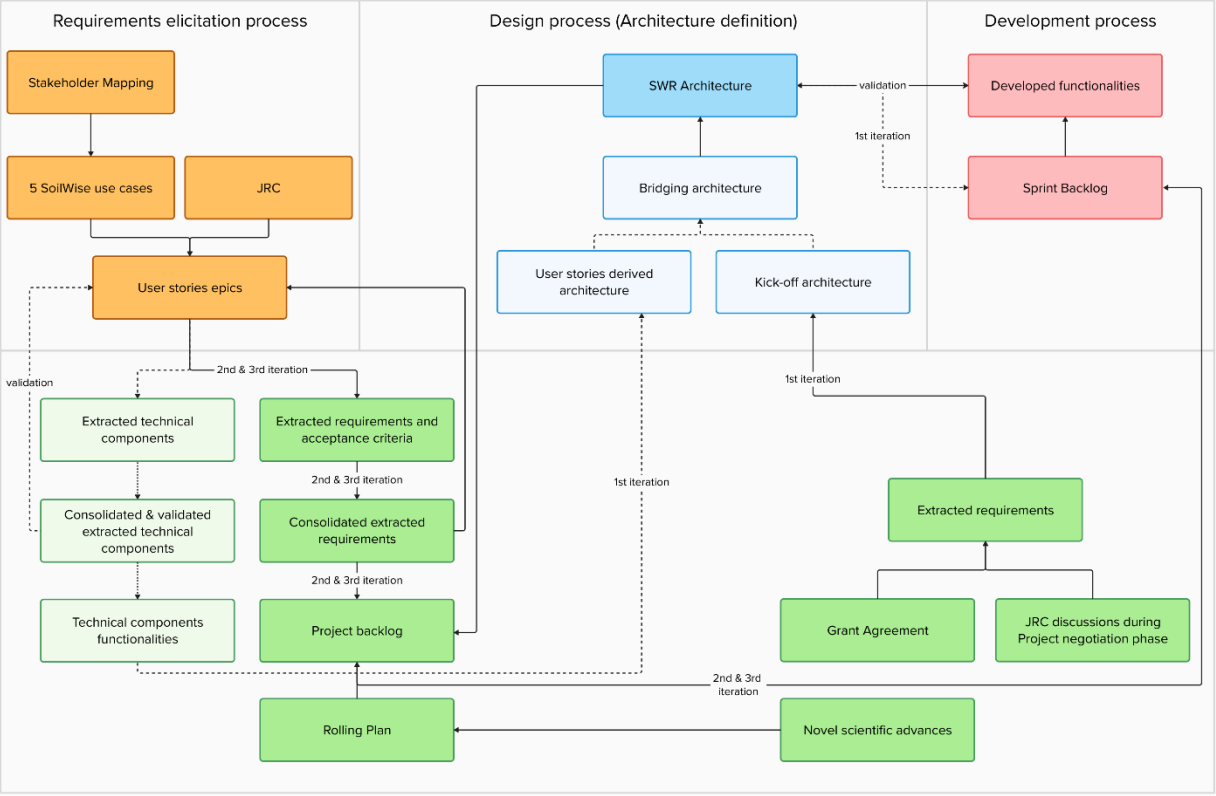


Figure 1 Overview of the co-designing architecture definition process

The process of requirement elicitation was partially described in the D1.1 Usage scenarios, requirements v1. The presented diagram on Figure 1 is, however, centralized around the Architecture design process. In the **first project iteration cycle** the Architecture design process was driven by two main forces. The first one emerged directly from the development team led by T1.4 (Define SoilWise Architectural Design) working group who elaborated a first draft of architecture (Kick-off architecture) based on requirements extracted from Grant Agreement and JRC discussions during the project negotiation phase. The second force emerged from the WP1 working group via stakeholder mapping and elaboration of user stories and epics describing the user-desired SoilWise Repository functionality. These user story epics were further analysed and technical components were extracted, consolidated, and finally incorporated in a structured technical description – the "User stories derived architecture". The two existing architecture drafts (Kick-off architecture, User stories derived architecture) were mutually compared and synthesized to the one common design – the "Bridging architecture". The final, and

most detailed architecture design arose from the "Bridging architecture" during the technical meeting in Brugges and it is based on a consensus between stakeholder groups and technical viewpoints, providing an overarching vision for the upcoming SWR development. Moreover, the SWR architecture together with the Technical documentation will serve as a base for populating the Project and Sprint Backlog used for development planning.

In the **second and third project iteration cycles**, we anticipate slightly different design processes. The Project and Sprint Backlog will be filled primarily with extracted and consolidated requirements emerging from the stakeholder-defined user story epics and user stories. Any updates related to the architecture design will be directly reflected in the final SoilWise D1.4 Repository architecture design.

2.2 Kick-off architecture

The very first steps of phase 1 co-designing a user driven SoilWise Repository, were taken already at the in-person **Kick-of meeting in September 2023** in Brussels, where the Architecture working group was set up. During the first day of the workshop, following open discussions with consortium partners, a first draft of architecture design was proposed. This "Kick-off architecture design" version (see Figure 2) was based on the experience and knowledge of the workshop participants and the description of the SWR as promised in the Grant agreement. Sticky notes represent individual technical components in different level of detail: yellow sticky note was used in the first round for the component title, orange for description of functions and relationships. On the following day, the proposed "Kick-off architecture" design was validated against the requirements extracted from the Grant Agreement (white magnetic notes), and initial discussions with JRC (blue magnetic notes). More information on the requirements extraction process is available in the Deliverable D1.1 Usage scenarios, requirements v1, Chapter 6.2.

Figure 2 Kick-off architecture design.

The initially defined “Kick-off architecture” was later digitalised using a virtual collaborative board called Miro, where the discussions of the Architecture working group continued during weekly online meetings. The final version of “Kick-off architecture” design was also elaborated to the ArchiMate notation (see Chapter 3.2) to document the co-designing process, see Figure 3.

In regard to the ArchiMate notation, the “Kick-off architecture” design did not distinguish between components and functionalities and also did not comprise a detailed functional description of the components. This was finalised later in the SoilWise Repository Architecture, see Figure 6. The main aim of the “Kick-off architecture” was to facilitate discussions in the T1.4 (Define SoilWise Architectural Design) working group until the requirements elicitation process (see Figure 1) proceeded.

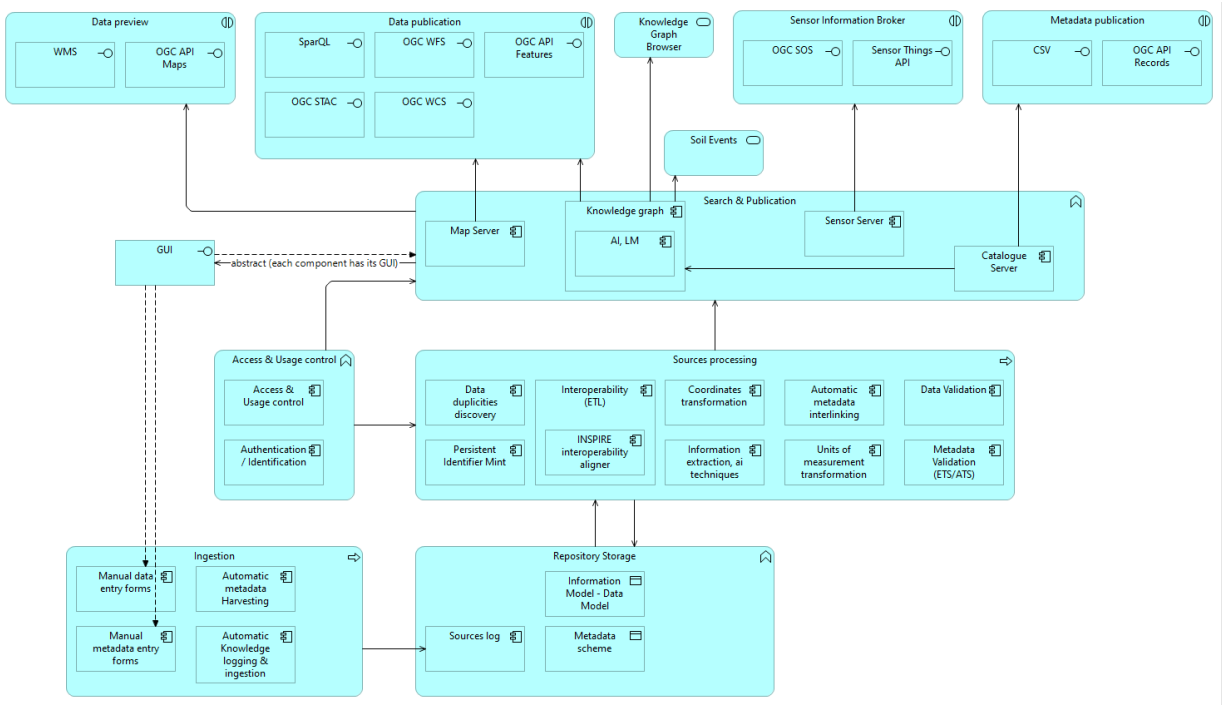


Figure 3 Final version of Kick-off architecture in ArchiMate notation

2.3 User stories derived architecture

In parallel, selected members of the Architecture working group attended to the series of **SoilWise use case stakeholder meetings**. These meetings resulted in a ranked list of varyingly detailed user stories epics, which described the desired functionality from the stakeholder's perspective. Adequately detailed user story epics were further analysed, and a list of technical components was extracted for each of them (see Table 6 in Deliverable D1.1 – Usage Scenarios, Requirements, v1 for more information). Technical components extracted from the user story epics were used to create a first version of the “User stories derived architecture” design of SWR as perceived by the stakeholders, see Figure 4.

Similar to “Kick-off architecture” design, in regard to the ArchiMate notation, the “User stories derived architecture” design did not distinguish between components and functionalities and also did not comprise a detailed functional description of the components. This was finalised later in the SoilWise Repository

Architecture, see Figure 6. The main aim of the “User stories derived architecture” was to summarize the User story epics into structured technical description.

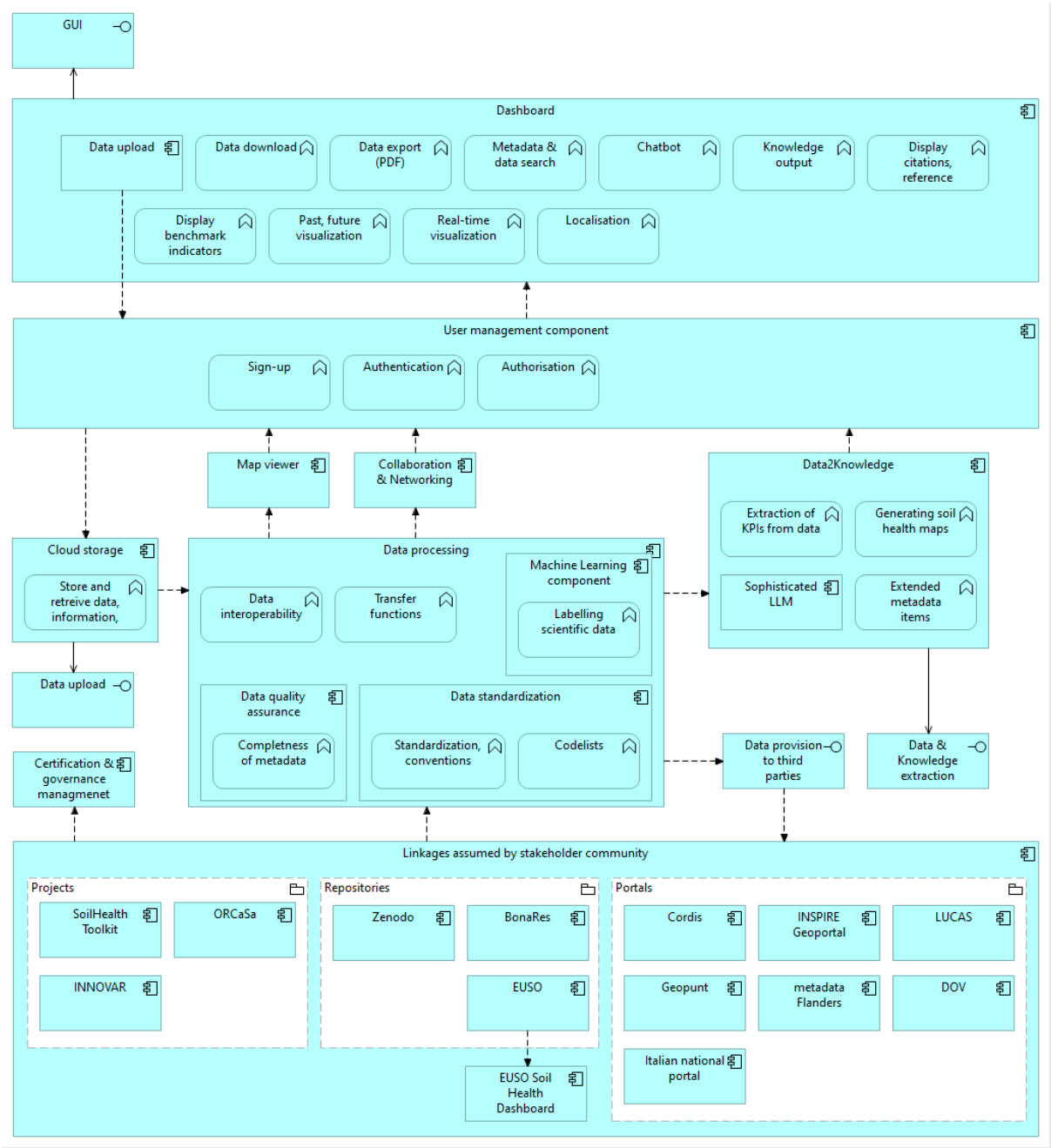


Figure 4 User stories derived architecture in ArchiMate notation

2.4 Bridging architecture

In the next phase, both versions of the architecture (Kick-off and User stories derived) in ArchiMate notation were compared with each other, and their synthesis resulted in the "**Bridging architecture"**, see Figure 5. This architecture was again validated with the Use cases working group.

As in previous cases, in regard to the ArchiMate notation, the "Bridging architecture" design did not distinguish between components and functionalities and also did not comprise a detailed functional description of the components. This was finalised later in the SoilWise Repository Architecture, see Figure 6. The main aim of the "Bridging architecture" was to interconnect both existing architecture designs (Kick-off and User stories derived). This version was further cross validated with the Use case leaders and authors of the prioritised User story epics to avoid any misinterpretation.

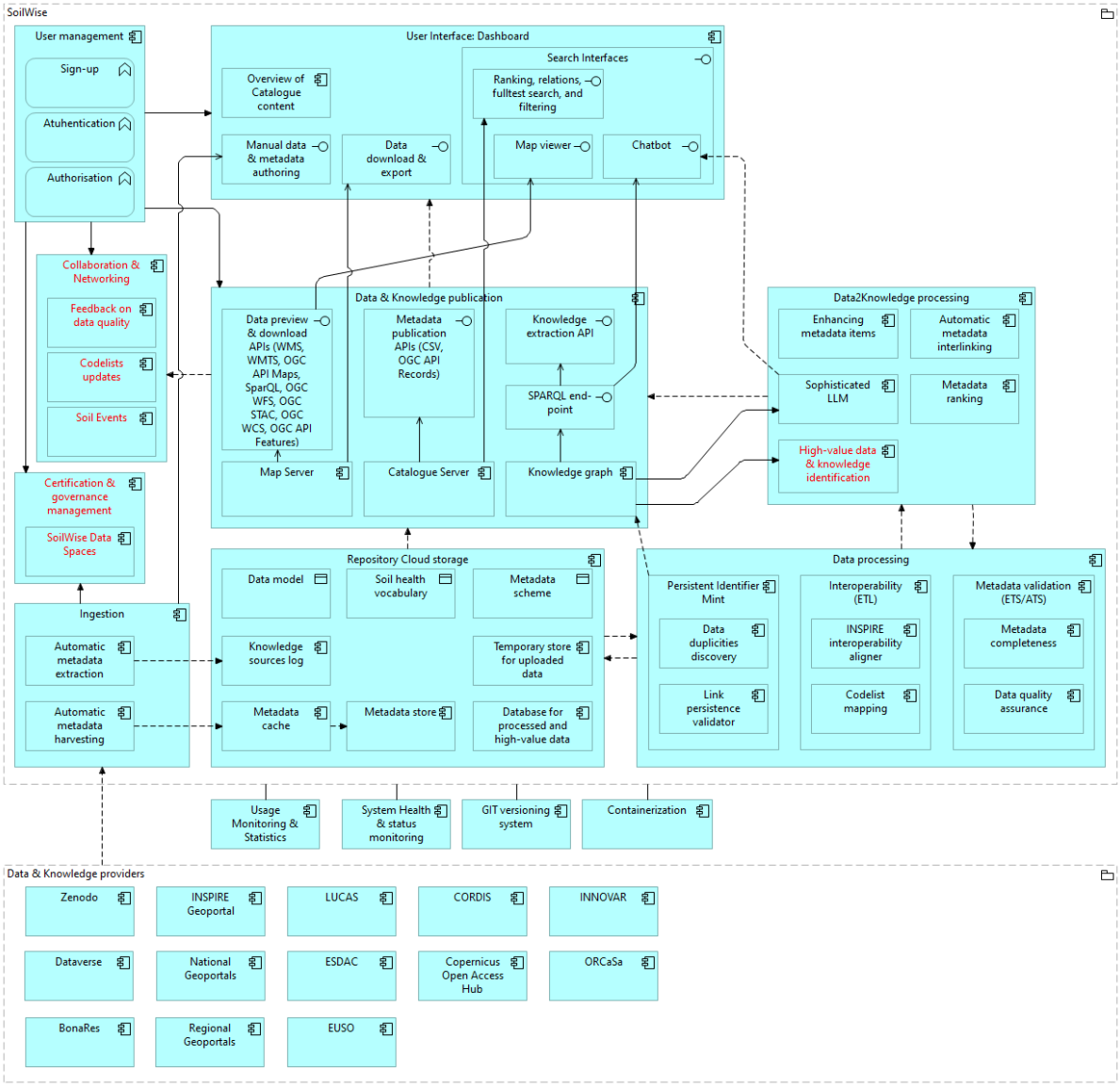


Figure 5 Bridging architecture in ArchiMate notation. Components, that were not considered a part of the SoilWise core product are highlighted in red.

2.5 Vision scenarios

In order to steer the development of the SWR in the first iteration cycle, four **generalized usage scenarios** were formulated based on the user story epics and use cases:

- a user searches for data & knowledge in the SoilWise Repository to evaluate their applicability for a given purpose;
- a user downloads or displays soil data & knowledge upon her/his request in “as is” or in interoperable data format for further analyses;
- a user uploads some data at the SoilWise Repository to be processed / transformed within the SoilWise Repository to improve their findability and interoperability;
- a third-party system connects to the SoilWise Repository to reuse data findable through the SoilWise Repository (or stored at the SoilWise Repository) to generate added value for its business;

These were further converted to a **list of modular functionalities**, which are foreseen as a vision for the SoilWise core product:

1. **The SWR Catalogue will be powered by a GIT repository, relational database and triple store containing standardised metadata of external data/knowledge resources. Standardised means a unified structure despite various input metadata structures from the underlying repositories, such as CORDIS, INSPIRE Geoportal, Zenodo, BonaRes, OpenAire, etc.**
2. **The Triple store will be extended with links to knowledge assets from external resources, for the 1st iteration extracted from CORDIS. To a user, the knowledge will be made available, among others, based on interlinked metadata, i.e., metadata linked to relevant projects and project deliverables.**
3. **When providing the search results, the SWR Catalogue will display a dataset or knowledge source to a user and show in which external repositories are, e.g., such a dataset available. The SWR also provides metadata, data & knowledge validation, among others, checks automatically persistence of available data/knowledge (e.g. if a dataset uploaded into Zenodo still exists or if a Web service endpoint is still alive and providing data).**
4. **The SWR Catalogue will display a map preview of a resource (dataset/knowledge/service/…) from the source graphic/WMS/…, if applicable.**
5. **Thanks to a Interlinker component, that will be powered by a Triple store, the SWR will identify duplicates based on metadata.**
6. **SoilWise Repository will provide data download in two modes, “as is” or in an “interoperable way”. “As is” means the SWR is a broker connecting a user to a relevant data source. The “interoperable way” means the SWR is a mediator that converts external data into, e.g. INSPIRE compliant through an ETL-like tool, e.g. HALE Studio.**
7. **SoilWise Repository will enable user to manually upload data for their on-the-fly processing within the SWR. Capacity of the SWR will be significantly limited; however, a demonstration of manual data upload and their processing, e.g. transformation of coordinate systems or measurements units.**
8. **In the first iteration, the SWR will primarily demonstrate functionality available to a non-registered user. Nevertheless, a lightweight proof-of-concept on functionality available to an authorised user will be present.**
9. **The SoilWise Repository functionality will be offered via open, standardised APIs (e.g. CSW, OGC REST API – Records, WMS, etc.) and a demonstration on how SWR can be used by other Soil Health**

Mission projects will be performed (e.g. BENCHMARKS and BonaRes already offered their participation).

10. **The SoilWise Repository will be equipped with a Usage monitoring & statistics module, automatically counting e.g. which soil datasets are the most searched for, which datasets are the most downloaded, how many visitors the SWR has etc.**

2.6 SoilWise Repository architecture

During the Technical in-person meeting in Brugges on the 25th and 26th of March 2024, the T1.4 Architecture working group again discussed and cross-checked the "Bridging architecture" design against the general usage scenarios, the list of modular functionalities and the current developments related to the role of catalog services in data exchange, using Data Spaces Support Center – Bluprint v1 as this convergence all community efforts, including the CSAs in Agricultre (AgriDataSpace project) and Green Deal (Great project). Moreover, to initiate the development activities of WP2, WP3, and WP4, the team needed to clearly distinguish components from their functionalities, which were overlapping in the "Bridging architecture". The final **"SoilWise Repository Architecture 1.0 - Brugges"** resulting from these discussions depicted in ArchiMate notation is illustrated in Figure 6 and can also be accessed at: https://soilwise-architecture.pages.dev/. A detailed functional description of components can be found at: https://soilwise-documentation.pages.dev/.

This version of architecture is being used as a joint blueprint for detailed solution designs, on top of which specific applications and services will be built. Note that the findings of the subsequent iterations of the stakeholder elicitation and validation process (depicted in Figure 1), as well as novel scientific advances resulting in architecture design updates, will be reflected directly in the "SoilWise Repository Architecture", all according to the product backlog and SoilWise Repository Rolling plan.

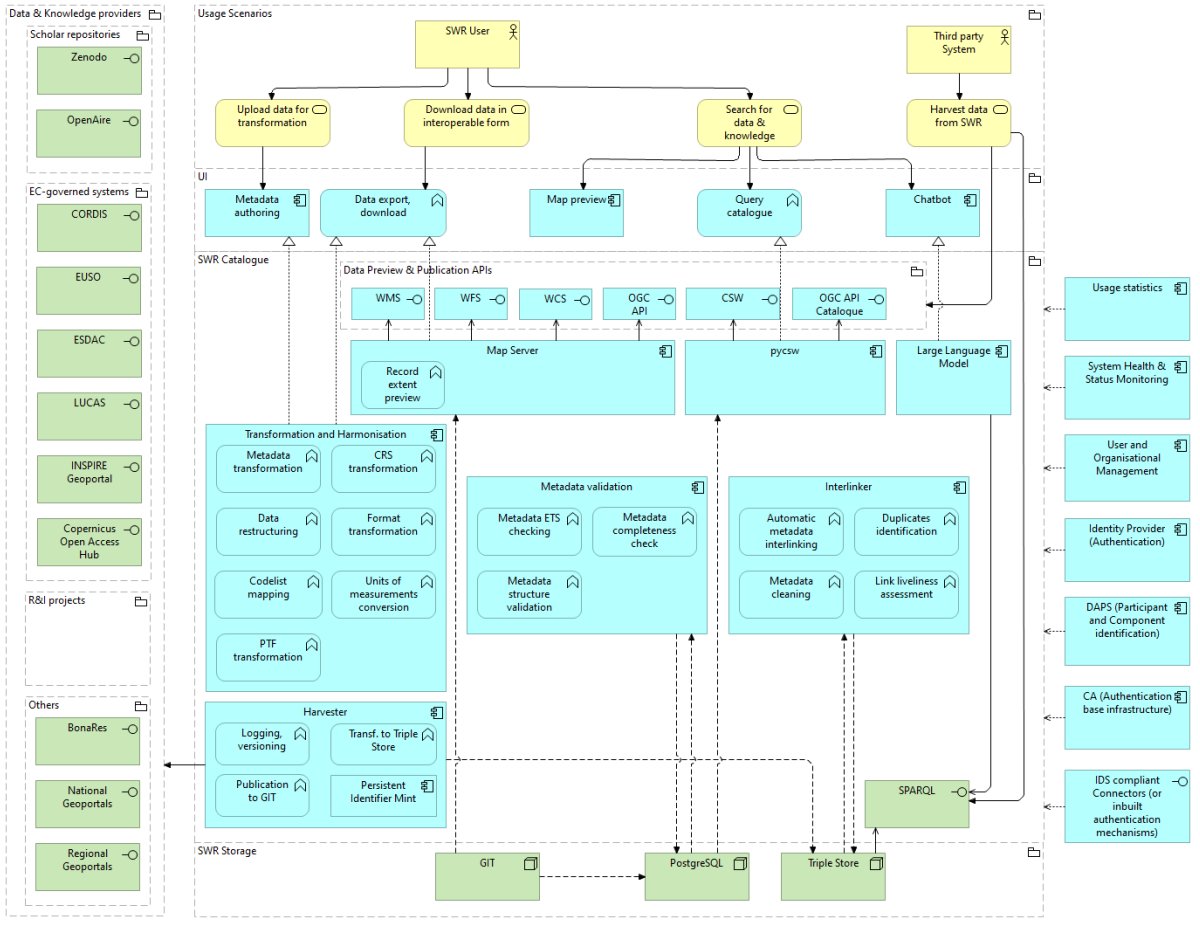


Figure 6 SWR Architecture 1.0 – Brugges in ArchiMate notation

In addition, it would be desirable to have an instantiation of the SWR Architecture 1.0 based on the “generalized usage scenarios”. This can be realised with message sequence diagrams specifying indicative interactions among actors and SWR components for each scenario scene.

3 SoilWise Repository Architecture – documentation

3.1 Introduction

The architecture design was developed using Archi software and is available at the public GitHub repository: https://github.com/soilwise-he/soilwise-architecture. An interactive read-only version fixed for this deliverable and corresponding to the Figure 6 is published at: https://soilwise-architecture.pages.dev/.

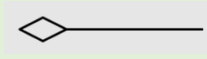
The documentation of SoilWise Repository architecture is maintained in the public GitHub repository: https://github.com/soilwise-he/SoilWise-documentation. It comprises description of functionality, detailed technical specifications and interfaces between all modular components and is structured according 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://soilwise-documentation.pages.dev/, and a PDF exported version is also available on demand, as non editable version made at the date of the deliverable submission.

The following Tables (Table 1, Table 2, Table 3, Table 4) summarize the ArchiMate specification used in the architecture diagrams (Figure 3, Figure 4, Figure 5, Figure 6).

3.2 ArchiMate notation

This chapter provides a guideline to reading architecture in the ArchiMate notation. ArchiMate is an open and independent enterprise architecture modelling language to support the description, analysis and visualization of architecture within and across business domains in an unambiguous way. ArchiMate is a technical standard from The Open Group and is based on concepts from the now superseded IEEE 1471 standard. It is supported by various tool vendors and consulting firms. ArchiMate is also a registered trademark of The Open Group.

Table 1 Relationships

Relationship	Definition	Notation
Structural Relationships		
Composition	Represents that an element consists of one or more other concepts.	
Aggregation	Represents that an element combines one or more other concepts.	
Assignment	Represents the allocation of responsibility, performance of behaviour, storage, or execution.	
Realization	Represents that an element plays a critical role in the creation, achievement, sustenance, or operation of a more abstract element.	
Dependency Relationships		

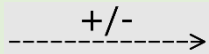
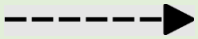
Serving	Represents that an element provides its functionality to another element.	
Access	Represents the ability of behaviour and active structure elements to observe or act upon passive structure elements.	
Influence	Represents that an element affects the implementation or achievement of some motivation element.	
Association	Represents an unspecified relationship, or one that is not represented by another ArchiMate relationship.	
Dynamic Relationships		
Triggering	Represents a temporal or causal relationship between elements.	
Flow	Represents transfer from one element to another.	
Other Relationships		
Specialization	Represents that an element is a particular kind of another element.	

Table 2 Business Layer Concepts

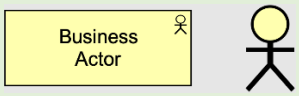
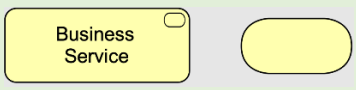
Element	Definition	Notation
Business Actor	Represents a business entity that is capable of performing behaviour.	
Business Function	Represents a collection of business behaviour based on a chosen set of criteria such as required business resources and/or competencies and is managed or performed as a whole.	
Business Service	Represents explicitly defined behaviour that a business role, business actor, or business collaboration exposes to its environment.	

Table 3 Application Layer Concepts

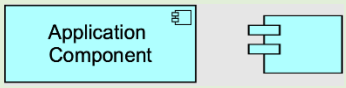
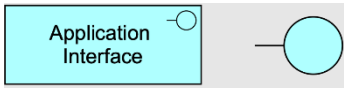
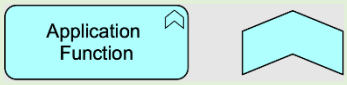
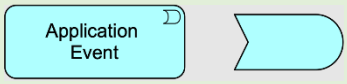
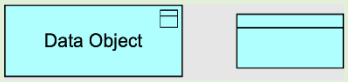
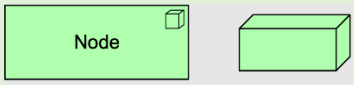
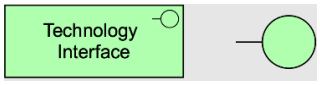
Element	Definition	Notation
Application component	Represents an encapsulation of application functionality aligned to implementation structure, which is modular and replaceable.	
Application interface	Represents a point of access where application services are made available to a user, another application component, or a node.	
Application function	Represents automated behaviour that can be performed by an application component.	
Application process	Represents a sequence of application behaviours that achieves a specific outcome.	
Application event	Represents an application state change.	
Application service	Represents an explicitly defined exposed application behaviour.	
Data object	Represents data structured for automated processing.	

Table 4 Technology Layer Concepts

Element	Definition	Notation
Node	Represents a computational or physical resource that hosts, manipulates, or interacts with other computational or physical resources.	
Technology Interface	Represents a point of access where technology services offered by a technology internal active structure can be accessed.	

References

Carayannis, E. G. & Campbell, D. F. (2010). Triple Helix, Quadruple Helix and Quintuple Helix and How Do Knowledge, Innovation and the Environment Relate To Each Other? : A Proposed Framework for a Trans-disciplinary Analysis of Sustainable Development and Social Ecology. International Journal of Social Ecology and Sustainable Development (IJSESD), 1(1), 41-69. http://doi.org/10.4018/jsesd.2010010105