SQAT

SOIL QUALITY ANALYSIS TOOL

Deliverable 4.2

Data governance principles

July 2024

Co-funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

Document Information

Delivery Title	D4.2: Data governance principles for use case implementation
Delivery Number	D4.2
Type	R – Document, report
Lead Beneficiary	4 - FARMEYE
Work Package Title	GOVERN: Data management & governance
Work Package Number	WP4
Dissemination level	PU – Public
Due Date	M06

Revision History

Version	Date	Author (Partner)	Remarks
Draft v0.1	22.05.24	Srđan Pavlović, ABE Igor Milosavljević, ABE Jessica Hicks, FARMEYE Brendan Allen, FARMEYE	Structural plan
Draft v0.2	05.07.24	Jessica Hicks, FARMEYE	Draft with core sections complete
Draft v0.3	10.07.24	Jessica Hicks, FARMEYE	Complete rough draft
Draft v0.4	15.07.24	Jessica Hicks, FARMEYE	Final draft
Draft v0.5	21.07.24	Srđan Pavlović, ABE Igor Milosavljević, ABE	Revision
v1.0	25.07.24	Jessica Hicks, FARMEYE	Final version
v1.1	22.11.24	Jessica Hicks, Senus	Edited Final Version
v1.2	17.02.25	Brendan Allen	Made changes based on external reviewer's comments.
V2.0	18.02.25	Srđan Pavlović, ABE	Revision

Disclaimer

The author of this document has taken any available measure to ensure that the information contained in this document is accurate, consistent, lawful, and up to date.

Executive Summary

Farms are at the forefront of the data economy, propelled by digitalisation, robotics, and smart algorithms. However, these advancements exacerbate societal pressures on soil health, demanding cleaner water, healthier soils, increased carbon storage and biodiversity. Current solutions are costly and unsuitable for farmers. With this in mind, the EU-funded SQAT project will develop a smart soil mapping service. Combining multi-level, multi-technology approaches, SQAT offers high-resolution soil property maps and tailored solutions for farmers. Using autonomous robot-mounted sensors and innovative in situ analysis tools, the SQAT system enhances productivity while reducing costs. Co-developing with SMEs, SQAT aims to commercialise its solutions, empowering farmers with variable-rate applications for liming, fertilisation, seeding, tillage, and carbon farming.

This document outlines comprehensive data governance principles essential for effective data management in SQAT. Emphasising the critical role of data governance, the framework ensures the accuracy, reliability and security of data throughout its lifecycle, supporting the project's objective of improving soil data precision and quality. Key principles include data quality, security, privacy and compliance with regulatory requirements and establishing robust security protocols and privacy measures is crucial to protect sensitive information and build stakeholder trust.

Detailed data management practices are outlined for managing data from collection to archiving. These include standardising data collection methods, automating data entry and implementing validation checks to ensure accuracy. Regular audits and version control are necessary to maintain data integrity. Advanced security measures, such as encryption and access controls, are critical for data security and privacy. Compliance with data privacy laws, like GDPR, is emphasised to protect sensitive data and maintain legal and ethical standards.

Clear definitions of roles, such as data stewards, custodians, and end-users, ensure accountability and effective data management. The utilisation of the FARMEYE data governance platforms enhances data management capabilities, offering tools for data stewardship, quality control, security and metadata management, ensuring comprehensive oversight of data governance processes.

The framework addresses common challenges such as data silos, regulatory compliance, and resistance to change by fostering a culture of data stewardship, continuous improvement and leveraging advanced technology to streamline data governance processes. It advocates for regular evaluations, stakeholder feedback and staying updated with technological advancements to enhance data governance practices.

By integrating these principles into every aspect of the project, SQAT aims to build a solid foundation for reliable and secure data management, supporting long-term success and value creation for all stakeholders involved.

Table of Contents

Table of Tables

Table of Figures

Abbreviations

ABAC	Attribute-Based Access Control
AES	Advanced Encryption Standard
DGC	Data Governance Committee
DLP	Data Loss Prevention
DPOs	Data Protection Officers
DQM	Data Quality Management
DUAs	Data Use Agreements
ECC	Elliptic Curve Cryptography
ENISA	European Union Agency For Cybersecurity
EO	Earth Observation
ETL	Extract, Transform, Load
GDPR	General Data Protection Regulation
HSMs	Hardware Security Modules
IDPS	Intrusion Detection and Prevention Systems
KPIs	Key Performance Indicators
MDM	Master Data Management
MFA	Multi-Factor Authentication
NPK	Nitrogen, phosphorus, potassium
PII	Personally Identifiable Information
RBAC	Role-Based Access Control
RSA	Rivest-Shamir-Adleman
SatNav	Satellite Navigation
SFTP	Secure File Transfer Protocols
SOPs	Standard Operating Procedures
SQAT	Soil Quality Analysis Tool
VPNs	Virtual Private Networks

1 Introduction

1.1 Project Overview

SQAT is designed to support the new paradigm in agriculture: the regenerative farm. In a regenerative farming system, soil capital is preserved and even enhanced, which is the cornerstone of environmental sustainability in EU farming. This aligns with key EU policies, including the Farm to Fork Strategy, the Biodiversity Strategy for 2030, the upcoming Soil Strategy for 2030 and the new Common Agricultural Policy for 2023-2027, all aimed at implementing the European Green Deal.

Traditionally, agricultural fields have been managed as homogeneous entities despite the inherent variability in soil properties within these fields. Mapping soil properties has historically been an expensive and labour-intensive process, limiting farmers' ability to leverage this information. However, advancements in digitalisation, robotisation of farm machinery, data management, processing and sensing techniques have made it feasible and actionable to measure soil property variability.

SQAT serves as an intervention for partners to accelerate their service concepts into the market by deploying space-based data sources, including Earth Observation (EO) and Satellite Navigation (SatNav). By integrating these with cutting-edge in-field proximal sensor-based measurement technology, SQAT makes the measurement of soil properties accessible and affordable for many farmers. The resulting soil property maps offer direct actionable insights to improve soil quality, maintain ecosystem functions such as water retention and habitat restoration, enhance food production and mitigate negative agricultural externalities like emissions, eutrophication and loss of biodiversity.

The SQAT project utilises a diverse range of data to achieve its objectives. Key types of data collected and processed include:

- **EO Data:** Sentinel-1 radar imagery, Sentinel-2 optical multispectral data and additional satellite imagery (e.g., PLANET), providing the foundation for field stratification and environmental monitoring.
- **Sensor Data:** On-ground measurements of soil properties such as pH, nitrogen, phosphorus, potassium (NPK) and microelements, collected using advanced sensors and robotic platforms.
- **Geospatial Data:** Farm boundary data and raster zones derived from EO imagery (e.g., Vegetation Index and Bare Soil Index) for precise field mapping and stratification.
- **Derived Data Sets:** Soil analysis maps and crop stratification data generated from sensor and EO inputs to guide variable-rate applications.
- **Personal Data:** Limited stakeholder information, managed in compliance with GDPR and other relevant regulations.

In the SQAT project, autonomous soil sampling equipment is developed for the smart farmer and data-driven advisor. With the tipping point in precision agriculture adoption surpassed, the demand for high-quality, actionable soil data has never been greater. This project capitalises on Europe’s investments in space technology and data innovations to meet this demand.

The overall objective of SQAT is to improve the precision, quality and affordability of soil data, addressing the significant bottleneck in the agricultural data economy. To achieve this, the project aims to develop and commercialise a Copernicus-driven, automated, sensor-based system for high-resolution soil mapping, enabling new and improved Smart Farming Applications.

1.2 Purpose of the Document

The purpose of this document is to establish a comprehensive framework for data governance and outline the key principles essential for managing data effectively within SQAT. As data becomes increasingly vital in driving decisions and enhancing agricultural practices, it is crucial to ensure that data within SQAT is managed with the highest standards of quality, security, and compliance.

This document aims to:

1. **Define the Importance of Data Governance:** highlighting the necessity of a structured approach to managing data to ensure its accuracy, consistency, and reliability. By establishing clear data governance policies, SQAT can maintain the integrity of its data and enhance its usability for various stakeholders.

2. **Outline Data Governance Goals and Principles:** setting forth the primary objectives of data governance in the context of SQAT, including data quality, security, privacy, and lifecycle management. It also delineates the fundamental principles that guide data governance practices.

3. **Provide Detailed Data Management Practices:** offering guidelines and best practices for managing data throughout its lifecycle, from collection and storage to usage and archiving. These practices aim to optimise data quality and ensure that data is used effectively and ethically.

4. **Establish Roles and Responsibilities:** defining the roles and responsibilities of various stakeholders involved in data governance, including data stewards, custodians, and end users. By clearly outlining these roles, the document ensures accountability and effective management of data assets.

5. **Introduce Data Governance Policies and Procedures:** presenting the policies and procedures necessary to implement and maintain robust data governance. This includes policies related to data stewardship, ownership, sharing, and classification.

6. **Highlight Technologies and Tools:** discussing the technologies and tools that support data governance efforts, such as data quality tools, metadata management systems, and data security solutions.

7. **Address Challenges and Solutions:** Identifying common challenges in data governance and providing practical solutions and best practices to overcome these obstacles, ensuring continuous improvement in data management practices.

By adhering to the guidelines and principles outlined in this document, SQAT aims to foster a culture of data excellence, ensuring that its data assets are managed effectively to support high-quality soil analysis and informed decision-making. This document serves as a foundational resource for all stakeholders involved in data management within SQAT, promoting a unified and standardised approach to data governance.

1.3 Document Scope

The scope of this document encompasses all aspects of data governance related to SQAT. It includes the policies, procedures, roles and responsibilities that ensure the effective management, security and utilisation of data within the SQAT ecosystem. This document addresses the following key areas:

1. **Data Collection and Input:** Guidelines and standards for the collection of soil quality data from various sources, including field measurements, laboratory analyses and remote sensing technologies.

2. **Data Storage and Management:** Strategies for secure and efficient storage of soil quality data, including database management, data warehousing and cloud storage solutions. It also covers data retention policies and archiving practices.

3. **Data Quality Assurance:** Procedures to ensure the accuracy, completeness, consistency and timeliness of data. This includes data validation, cleaning and quality control measures.

4. **Data Security and Privacy:** Measures to protect data from unauthorised access, breaches and other security threats. It also includes compliance with relevant data privacy laws and regulations, data encryption, access controls and anonymisation techniques.

5. **Data Usage and Analysis:** Guidelines for the proper use and analysis of soil quality data to support decision-making processes. This includes the application of analytical tools, data visualisation and reporting standards.

6. **Data Sharing and Access:** Policies for sharing data with internal and external stakeholders, including data access permissions, sharing agreements and collaboration frameworks.

7. **Roles and Responsibilities:** Definition of roles and responsibilities for all individuals involved in data governance, including data stewards, custodians, users and the data governance committee.

8. **Monitoring and Reporting:** Mechanisms for ongoing monitoring of data governance practices and the reporting of key performance indicators (KPIs) to ensure continuous improvement and compliance.

9. **Compliance and Risk Management:** Strategies to ensure compliance with legal and regulatory requirements related to data governance, as well as risk management practices to identify and mitigate potential data-related risks.

2 Data Governance Framework

2.1 Principles of Data Governance

Data governance is a comprehensive framework encompassing the practices, policies and procedures designed to manage, protect and optimise the value of data assets throughout their lifecycle. It involves establishing a formalised approach to ensuring data quality, security, privacy and compliance with regulatory requirements through assessing key principals and goals (*Fig. 1 & 2*). Effective data governance is critical for any organisation relying on data-driven decision-making, as it ensures that data is accurate, reliable and accessible.

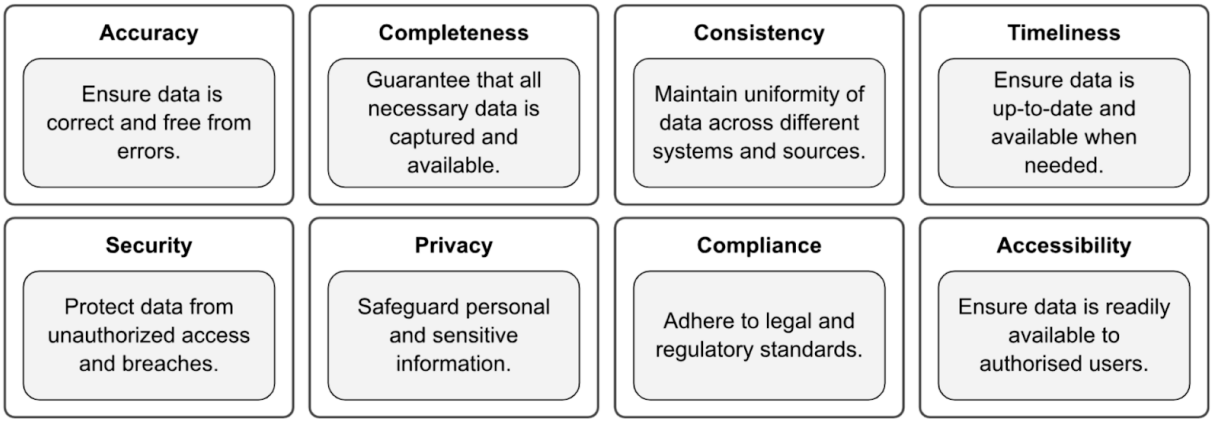


Figure 1 | Matrix of data governance goals.

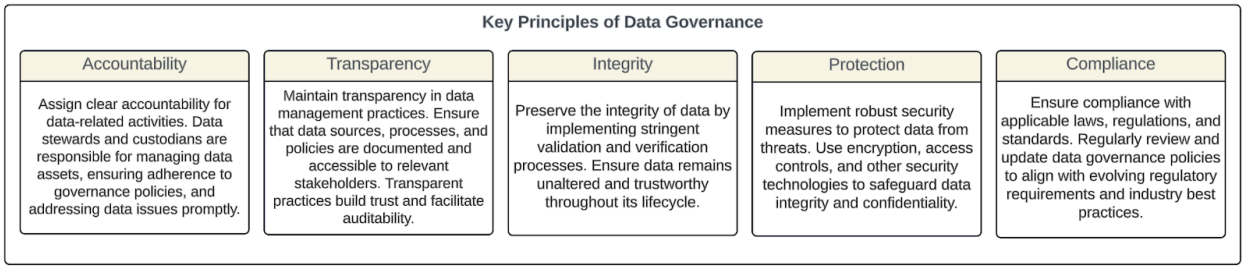


Figure 2 | Matrix of data governance principles.

The importance of data governance lies in its ability to ensure data quality, enhance data security and privacy, improve data management, facilitate regulatory compliance and drive business value. These factors are crucial for making informed decisions and generating actionable insights and reliably accurate analysis depends on maintaining data completeness, consistency and timeliness throughout the project lifecycle, which are essential for achieving successful outcomes.

A key aspect of data governance is the establishment of robust security protocols and privacy measures. These protocols protect sensitive information from unauthorised access, breaches and misuse, ensuring

compliance with legal and regulatory requirements. This not only safeguards the organisation but also builds trust with its stakeholders by demonstrating a commitment to data protection. Additionally, a structured data governance framework streamlines data management processes, encompassing data collection, storage, usage and archiving. By defining clear roles and responsibilities, the framework reduces redundancies, enhances operational efficiency and ensures accountability at every stage of data handling.

In the face of increasing regulatory scrutiny, organisations must adhere to various data protection laws and standards. Data governance ensures that data handling practices align with these regulations, mitigating risks of non-compliance and potential penalties. This includes adhering to regulations such as GDPR and other local data protection laws, which require stringent measures for data privacy and security. Moreover, effective data governance maximises the value derived from data assets. By providing reliable data for analysis, it supports strategic planning, enhances customer experiences and fosters innovation. This reliability allows organisations to leverage data for competitive advantage, driving business growth and operational excellence.

Furthermore, data governance promotes a culture of continuous improvement and innovation. By regularly reviewing and updating data governance policies, organisations can adapt to changing regulatory environments and technological advancements. This proactive approach ensures that data governance remains relevant and effective, enabling organisations to harness the full potential of their data assets.

These aspects apply to SQAT as a consortium across multiple partners, ensuring that the activities within the innovation's data value chain are aligned with regulatory requirements and best practices. By implementing these data governance principles, SQAT can effectively manage data across all partner organisations, fostering a cohesive and compliant approach to data management. This collaborative effort not only supports the project's objectives, but also enhances the overall effectiveness and reliability of the data governance framework.

All of these aspects related to data governance will be further discussed in the following section.

3 Data Management Principles

3.1 Data Quality

3.1.1 Data Accuracy

Accuracy in data management refers to the degree to which data correctly describes the real-world object or event it represents. For SQAT, ensuring the accuracy of soil quality data is critical for reliable analysis

and decision-making. Accurate data allows stakeholders to trust the outputs of the tool, making it essential for operational efficiency and strategic planning.

Accurate data collection begins with the calibration of sensors. Regular calibration of soil quality sensors ensures that data collected from the field accurately reflects the soil conditions, thereby reducing the likelihood of systematic errors. Standardised procedures for data collection also play a crucial role in minimising human error. This involves clear guidelines for sampling methods, the frequency of data collection and data entry protocols. Additionally, providing adequate training for personnel involved in data collection ensures they understand the importance of accuracy and are proficient in using the tools and following protocols.

To maintain accuracy during data entry, it is beneficial to automate data entry processes where possible. Automation tools can directly capture data from sensors or other instruments and input it into the system without human intervention, thereby eliminating manual entry errors. Furthermore, implementing validation checks during data entry helps identify and correct inaccuracies. These checks can include range checks, format checks and consistency checks to ensure that data conforms to expected patterns.

In the data processing stage, error detection algorithms are employed to detect and correct errors. These algorithms can identify outliers or anomalies that may indicate inaccuracies. Cross-verifying data with multiple sources, such as comparing sensor data with laboratory analysis results, also ensures consistency and accuracy.

Regular audits are essential for maintaining data accuracy over time. These audits should be systematic and cover various aspects of the data lifecycle, including collection, entry, processing and storage. Version control of data is also important to track changes and ensure that any corrections are documented and traceable. This helps in understanding the history of data modifications and maintaining its integrity.

Various tools and technologies can assist in ensuring data accuracy. Data quality management software offers features such as data profiling, cleansing and monitoring, which help maintain high data accuracy by providing automated solutions for identifying and correcting errors. Real-time monitoring systems continuously check data for accuracy, providing immediate alerts for any detected discrepancies and allowing for quick corrective actions.

Defining accuracy metrics is a critical step in measuring data accuracy. Common metrics include error rates, validation success rates and data consistency rates. Fostering a culture of continuous improvement, where feedback from data users is regularly collected and used to enhance data accuracy protocols and practices, is also vital. Encouraging collaboration between data collectors, processors and end-users helps in identifying common sources of inaccuracies and developing joint solutions.

3.1.2 Data Completeness

Completeness refers to the extent to which all required data is present and accounted for within a dataset. It is a critical aspect of data quality that ensures the data used in SQAT is comprehensive and covers all necessary parameters for accurate soil quality assessment.

Data completeness involves ensuring that all necessary data fields are filled and no critical information is missing. For SQAT, this includes data on soil composition, moisture levels, pH values, nutrient content and other relevant parameters. Completeness is not just about having data in every field, but having meaningful and relevant data that can contribute to the overall analysis and insights. This requires a clear definition and scope of what constitutes complete data within the context of soil quality analysis.

There are different types of data completeness to consider. Schema completeness ensures all required data fields and attributes as defined by the schema are present in the dataset. Population completeness ensures all records that should be included in the dataset are present, covering the full scope of the study or analysis area. Domain completeness ensures data values fall within the acceptable range or domain of possible values. These types of completeness collectively ensure the dataset is fully representative and usable for its intended purpose.

Ensuring data completeness involves implementing robust data collection methods to capture all required data. This can include standardised data entry forms, automated data collection tools and regular audits. Validation checks at the point of data entry can ensure that mandatory fields are not left blank and that the data entered meets the required criteria. Additionally, data integration processes should ensure that combined datasets are checked for completeness to avoid any gaps or missing values. Regular audits and reviews of datasets are also crucial in identifying and rectifying completeness issues, ensuring ongoing data integrity.

When dealing with incomplete data, imputation techniques can be used to estimate and fill in missing values where appropriate. Common techniques include mean imputation, regression imputation and algorithms like k-nearest neighbours. Data enrichment involves supplementing incomplete data with additional data sources to fill in gaps. For SQAT, this could involve using satellite data, historical records, or data from nearby regions with similar soil characteristics. Training users on the importance of data completeness and providing guidelines on ensuring all necessary data is captured during the data collection process is also essential for maintaining data quality.

The impact of incomplete data can be significant, leading to inaccurate analysis, misleading insights and poor decision-making. For SQAT, this could result in incorrect soil quality assessments, affecting crop management decisions and potentially leading to suboptimal agricultural practices. Ensuring data completeness helps maintain the integrity of the analysis and enhances the reliability of the results produced by SQAT, thereby supporting better decision-making processes.

Metrics for measuring completeness include the completeness ratio, which is the ratio of non-missing data to the total required data points, and the missing data rate, which is the percentage of missing data points within a dataset. Field completeness measures the percentage of fields that have complete data across all records. These metrics help in quantifying and monitoring data completeness, ensuring that the datasets used in SQAT are comprehensive and reliable.

3.1.3 Data Consistency

Consistency in data quality refers to the uniformity and coherence of data across various datasets, systems and applications. Ensuring consistency means that data does not conflict across different sources and maintains the same meaning and value throughout its lifecycle. Consistent data is crucial for reliable analysis, reporting and decision-making. Some key aspects and best practices for maintaining data consistency are as follows:

- **Data standardisation:** Ensures that data is formatted and structured uniformly across all datasets. This includes consistent use of units of measurement, naming conventions and data types. To implement data standardisation, it is essential to establish and enforce data standards and guidelines. Using standardised templates and forms for data collection can help achieve uniformity. Additionally, implementing data validation rules during data entry can ensure that the data adheres to the established standards, thereby maintaining consistency.

- **Master Data Management (MDM):** The process of creating and maintaining a single, authoritative source of truth for key business data. This involves identifying, defining and managing critical data elements to ensure consistency. For effective MDM, it is important to identify key data elements that require consistency, such as soil quality metrics and geographical locations. Developing a centralised master data repository can serve as a reference point for all data-related activities. Regularly updating and synchronising master data across all systems and applications further ensures that data remains consistent.

- **Data integration:** Involves combining data from different sources and ensuring that it is harmonised and consistent. This is particularly important when data is collected from various sensors, devices and external databases. To achieve seamless data integration, ETL (Extract, Transform, Load) processes can be used to integrate data from different sources. Implementing data mapping and transformation rules ensures that data is consistent when brought together. Regular reconciliation and validation of integrated data help detect and resolve any inconsistencies that may arise during the integration process.

- **Metadata management:** Managing data about data, such as definitions, data lineage and usage context. Proper metadata management ensures that data is interpreted and used consistently. Maintaining comprehensive metadata documentation for all data elements is crucial for clarity and uniformity. Using metadata management tools to track data lineage and transformations helps in understanding the data's journey and maintaining its consistency. Ensuring that metadata is regularly updated and accessible to all users further supports consistent data usage.

- **Regular auditing and monitoring:** Help detect and resolve inconsistencies. This includes tracking data changes, identifying anomalies and ensuring compliance with data governance policies. Implementing automated data quality checks and validation processes can help in maintaining consistency. Using data auditing tools to monitor data changes and track discrepancies allows for timely identification of inconsistencies. Regularly reviewing audit logs and reports enables proactive resolution of any issues, thereby maintaining data consistency.

- **Clear data governance policies and procedures:** Essential for maintaining consistency. These policies define roles, responsibilities and processes for managing data consistently. Developing and enforcing data governance policies that emphasise consistency is crucial. Assigning data stewards to oversee and ensure data consistency helps in maintaining accountability. Regularly reviewing and updating policies to adapt to changing data requirements ensures that the governance framework remains relevant and effective.

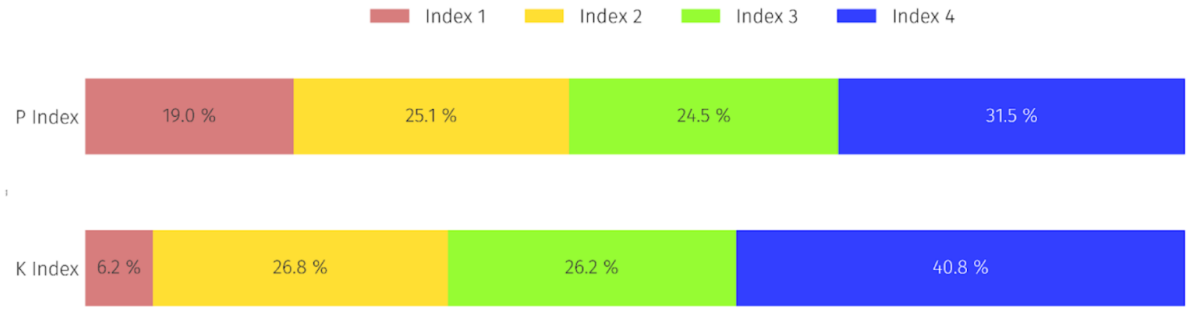
Ensuring that all users understand the importance of data consistency and are trained in best practices is crucial for maintaining consistent data. Conducting regular training sessions on data consistency and quality standards helps in building awareness. Providing resources and documentation to help users maintain consistency supports their efforts and fostering a culture of data quality and consistency within the organisation encourages everyone to prioritise and uphold data consistency in their daily activities.

3.1.4 Data Timeliness

Timeliness ensures data is up-to-date and available when needed. In the context of SQAT, timeliness directly impacts the accuracy of soil assessments, decision-making processes and overall system reliability. Ensuring that the most current and relevant data is used for soil quality analysis helps make informed

decisions, improves responsiveness to soil quality issues and enhances user trust in the tool's reliability and accuracy.

The importance of timeliness in data management cannot be overstated. Access to timely data allows users to make well-informed decisions regarding soil health and agricultural practices. For instance, timely data helps in quickly identifying and responding to soil quality issues, thereby preventing potential negative impacts on crop yield and soil health (*Fig. 3*). Moreover, consistently providing up-to-date data fosters trust among users, as they can rely on the tool for accurate and current information, enhancing the overall user experience and satisfaction.



INDEX	REMARK	P mg/l Grassland	P mg/l Other Crop
P 1	Very Poor	0 - 3.0	0 - 3.0
P 2	Low	3.1 - 5.0	3.1 - 6.0
P 3	Optimum	5.1 - 8.0	6.1 - 10.0
P 4	High	> 8.0	> 10.0

Index	Remark	K mg/l Mineral Soil	K Mg/l Peat Soil
K 1	Very Poor	0 - 50	0 - 100
K 2	Low	51- 100	101 - 175
K 3	Optimum	101 - 150	176 - 250
K 4	High	> 150	> 250

Figure 3 | Example graphic of risk zones for an Irish sample.

To ensure timeliness in data management, several strategies can be implemented. One effective approach is to implement automated data collection processes, which minimise delays caused by manual data entry. Technologies such as sensors and IoT devices can gather soil data in real-time, ensuring that the most current information is always available. Additionally, establishing regular schedules for data updates ensures that the data repository is continuously refreshed with the latest information, whether through daily, weekly, or monthly updates depending on the data type and its usage. Utilising real-time data processing systems to handle and analyse data as it is collected also ensures that users have access to the most current data without significant delays. Furthermore, integrating data from various sources in a timely manner provides a comprehensive view of soil quality by combining data from sensors and satellite imagery.

Regular monitoring and maintenance are essential to ensure data timeliness. Continuous performance monitoring of data collection and processing systems can help identify and resolve any issues that may cause delays. Conducting periodic audits of data quality, with a focus on timeliness, can help pinpoint areas needing improvement (*Fig. 4*). Collecting feedback from users regarding the timeliness of the data they receive can provide valuable insights on how to improve data timeliness further. This feedback loop ensures that the data management process remains user-centric and adaptive to changing needs and conditions.

Figure 4 | Farmeye platform showing risk areas for phosphorus (P) for a test farm.

Maintaining data timeliness can present several challenges. Data latency, or delays in data transmission and processing, can lead to outdated information. To mitigate this, investing in high-speed data processing and transmission technologies is essential. Resource constraints, such as limited computational or human resources, can affect the frequency and speed of data updates. Optimising resource allocation and exploring automation options can help overcome these constraints. Ensuring data synchronisation across various sources can also be complex; implementing robust data integration and synchronisation mechanisms is crucial to maintain consistency and accuracy across data sources.

Adhering to best practices can significantly enhance data timeliness. Setting clear objectives for data timeliness based on the needs of the users and the specific requirements of soil quality analysis provides a focused approach to achieving timely data. Leveraging advanced technologies such as cloud computing, edge computing and real-time analytics can process and deliver data efficiently and promptly. Additionally, regularly reviewing and updating data management processes to incorporate new technologies and methodologies ensures continuous improvement in data timeliness.

3.2 Data Security

3.2.1 Access Controls

Access controls ensure that only authorised users can access specific data within SQAT. These controls help protect sensitive information from unauthorised access, breaches and potential misuse.

Authentication verifies the identity of users attempting to access the system. SQAT employs various authentication methods such as username and password combinations, as well as multi-factor authentication (MFA). Username and password are basic methods requiring users to enter a unique combination to gain access. MFA enhances security by requiring two or more verification methods, such as a password and a code sent to a user's mobile device.

Authorisation determines what an authenticated user is allowed to do within the system. It involves setting permissions and access rights based on user roles. In SQAT, this is managed through Role-Based Access Control (RBAC) and Attribute-Based Access Control (ABAC). RBAC assigns users roles (e.g., admin, data steward, analyst) that grant specific permissions, ensuring users only access data and functionalities pertinent to their role. ABAC bases access decisions on user attributes (e.g., department, job title) and environmental conditions (e.g., time of access, location).

Access Control Lists are used to define which users or system processes have access to specific resources within SQAT. They specify user permissions for data objects, such as read, write, or execute permissions, ensuring that each user’s access is limited to what is necessary for their role.

The least privilege principle ensures users have the minimum level of access—or privileges—necessary to perform their job functions. Implementing least privilege reduces the risk of accidental or malicious data breaches by limiting exposure of sensitive data to only those who need it.

Segregation of Duties is a control measure to prevent conflicts of interest and reduce the risk of fraud. It ensures that no single individual has control over all aspects of any critical process. For example, in SQAT, one user might enter data while another approves it, thereby distributing responsibilities to maintain checks and balances.

Regular access reviews and audits are conducted to ensure that access permissions remain appropriate over time. Periodic reviews involve regularly assessing user access rights to ensure they are still valid and necessary. Audit trails, which are detailed logs of user activities, are maintained and reviewed to detect unauthorised access or suspicious activities. Additionally, it is imperative that each partner organisation is responsible for timely sharing any relevant changes within their team. For instance, if a team member is no longer working on SQAT, the partner must promptly inform the DGC to ensure that the individual's access is revoked. This practice is crucial for maintaining the security and integrity of the system.

Ensuring secure access to SQAT involves implementing encryption and using secure access methods. Data is encrypted both in transit and at rest to protect it from interception and unauthorised access. Virtual Private Networks (VPNs) and other secure channels are used for remote access to the system, ensuring that data remains protected even when accessed offsite.

Regular training and awareness programs are essential to educate users about the importance of access controls and best practices for maintaining data security. These programs help users understand their role

in protecting data and the specific measures they need to follow to ensure compliance with security protocols.

In the context of SQAT, access controls are implemented through a combination of technical solutions and administrative policies. Users are authenticated through a robust MFA system, roles and permissions are clearly defined and managed via RBAC, and regular access reviews are conducted to ensure compliance with security policies. Access controls are fundamental to the security framework of SQAT.

3.2.2 Encryption

Encryption offers a means to secure sensitive and confidential information handled by the SQAT project. This section covers the importance of encryption, types of encryption methods and best practices for implementing encryption.

Importance of Encryption

Encryption transforms readable data into an unreadable format using algorithms and encryption keys. This process ensures that only authorised users with the correct decryption key can access and read the data. Encryption protects data from unauthorised access and breaches, maintaining the confidentiality and integrity of the data throughout its lifecycle. In the European context, encryption is essential for complying with regulations such as the General Data Protection Regulation (GDPR), which mandates stringent data protection measures to safeguard personal data.

Types of Encryption

Symmetric encryption uses a single key for both encryption and decryption. The same key must be securely shared between the parties who need to access the data. Common symmetric encryption algorithms include AES (Advanced Encryption Standard), which is widely accepted in Europe for its robustness and efficiency. Symmetric encryption is efficient for encrypting large amounts of data due to its speed and is suitable for data stored in databases and file systems within SQAT.

Asymmetric encryption, on the other hand, uses a pair of keys – a public key for encryption and a private key for decryption. The public key can be freely shared, while the private key is kept secret. Examples of asymmetric encryption algorithms are RSA (Rivest-Shamir-Adleman) and ECC (Elliptic Curve Cryptography). This method is ideal for secure data transmission and establishing secure channels, such as SSL/TLS for web communications, making it suitable for encrypting data exchanges between SQAT users and servers, particularly those across different European countries with varying data protection laws.

Hybrid encryption combines symmetric and asymmetric encryption to leverage the strengths of both methods. Typically, asymmetric encryption is used to securely exchange a symmetric key, which is then used for encrypting data. This approach provides a balance of security and performance, making it suitable for scenarios in SQAT that require both secure key exchange and efficient data encryption. Hybrid encryption is particularly beneficial for cross-border data transfers within the European Union, ensuring compliance with GDPR and other regional regulations.

Best Practices for Implementing Encryption in SQAT

Using strong encryption algorithms is essential for robust security. Industry-standard and well-vetted algorithms such as AES-256 for symmetric encryption and RSA-2048 or higher for asymmetric encryption

should be used to ensure data protection. European organisations often follow guidelines from the European Union Agency for Cybersecurity (ENISA) for selecting encryption standards.

Key management is another critical aspect of encryption. A secure key management system should be implemented to generate, store, distribute and rotate encryption keys. Keys should be stored in hardware security modules (HSMs) or other secure environments to prevent unauthorised access. European standards for key management often reference frameworks like the ENISA guidelines and ISO/IEC 27001.

Encryption in transit is vital for protecting data from interception during transmission. Secure protocols such as HTTPS, SSL/TLS and VPNs should be used to ensure data integrity and confidentiality during data exchanges between SQAT components and users. Ensuring end-to-end encryption for data in transit is crucial for complying with European regulations that mandate data protection during transfer.

Similarly, encryption at rest is necessary to protect data stored on disk, databases and backup media from unauthorised access in case of physical theft or unauthorised access to storage systems. Sensitive data fields within databases should be encrypted to provide granular protection. Adhering to the GDPR's requirement for data minimisation and protection by design and by default, encryption at rest helps ensure that personal data is not exposed in the event of a breach.

Regular audits and compliance checks are essential to ensure encryption practices comply with relevant European regulations and industry standards, such as GDPR. Security audits and assessments should be conducted regularly, and any identified vulnerabilities should be addressed promptly. European organisations should consider engaging with certified data protection officers (DPOs) to oversee compliance efforts.

Performance considerations should also be taken into account when implementing encryption. Optimising encryption processes can help balance security and performance. Using hardware acceleration and efficient encryption algorithms can minimise the performance impact on SQAT operations, ensuring that the tool remains responsive and efficient for European users.

Finally, user awareness and training are crucial for effective encryption. SQAT users and stakeholders should be educated about the importance of encryption and best practices for handling encrypted data. They should understand how to manage encryption keys and secure data effectively. European organisations can benefit from training programs aligned with ENISA's cybersecurity awareness initiatives to ensure a high level of security awareness among staff.

3.2.3 Data Masking

Data masking is a technique used to protect sensitive information by obfuscating data, making it unreadable and unusable for unauthorised users while preserving its usability for authorised purposes. This method is essential in safeguarding personal data and ensuring compliance with European data privacy regulations such as GDPR.

There are several types of data masking. Static data masking involves the transformation of data in a non-production environment, creating anonymised copies of data for testing, development, or training purposes. The masked data maintains its format and usability but removes sensitive information. Dynamic data masking, on the other hand, applies data obfuscation in real-time as data is accessed by unauthorised users. This approach ensures that sensitive data is masked when viewed or queried by individuals without

proper access rights, providing an additional layer of security for live databases. On-the-fly data masking occurs during data transfer between systems or environments, ensuring that data is masked as it is being moved, preventing sensitive information from being exposed during migration or integration processes.

Various techniques can be employed for data masking. Substitution involves replacing original data with fictitious but realistic-looking values, such as replacing real names with randomly generated ones. Shuffling rearranges data within a dataset to mask sensitive information, like swapping customer addresses within the same column. Masking out hides specific parts of the data, for example displaying only the last three digits of a Land Folio Number (e.g., XXX45F). Tokenisation replaces sensitive data with unique identifiers (tokens) that map back to the original data but are useless if intercepted. While not strictly data masking, encryption can be used in conjunction to enhance security by ensuring data is unreadable without the appropriate decryption key.

When implementing data masking, several considerations must be taken into account. It is essential to conduct a thorough assessment to identify sensitive data elements that require masking, including personally identifiable information (PII), financial data, health records and other confidential information. Clear policies and rules for data masking should be established, specifying which data elements to mask, the masking techniques to use, and the contexts in which masking is applied. Maintaining data integrity is crucial, ensuring that masked data retains its format and referential integrity, remaining useful for testing, development and analysis without compromising data quality or consistency. Compliance with European data privacy regulations such as GDPR is essential, and regular audits of masked data and masking processes should be conducted to ensure ongoing compliance and effectiveness. Continuous monitoring and updating of data masking implementations are necessary to address emerging security threats and regulatory changes, maintaining robust data protection.

The benefits of data masking are significant. Enhanced security is achieved by protecting sensitive information from unauthorised access and data breaches. Regulatory compliance is facilitated by anonymising sensitive data, helping organisations adhere to GDPR and other European data privacy regulations. Masked data provides realistic data for non-production environments, supporting testing, development and training without exposing sensitive information. Furthermore, data masking reduces the risk of data exposure and misuse, ensuring that even if data is accessed, it remains unreadable and unusable.

3.3 Data Privacy

3.3.1 Regulatory Compliance

Ensuring regulatory compliance in data privacy is essential for the effective governance of soil quality data. This involves adhering to comprehensive data protection laws and frameworks that mandate the secure and responsible handling of personal and sensitive information.

Legal requirements mandate that organisations must comply with stringent data protection laws that govern data collection, processing and storage. These regulations emphasise the need for obtaining explicit consent from individuals before collecting their data, ensuring transparency about how their data will be used and providing them with the right to access, correct and delete their data. Organisations must implement procedures that guarantee that these legal requirements are met to protect the rights and

privacy of individuals. For instance, in Ireland, the Data Protection Act 2018 aligns with the GDPR, emphasising the protection of personal data and individuals' rights. Similarly, Belgium and the Netherlands also follow GDPR guidelines, ensuring strict adherence to data privacy and protection standards.

Compliance involves following key data protection principles. Lawfulness, fairness and transparency are critical, meaning that data must be processed lawfully, fairly and in a transparent manner. Organisations should provide clear information about data processing activities to the individuals concerned. Purpose limitation dictates that data should be collected for specified, explicit and legitimate purposes and not further processed in a manner incompatible with those purposes. Data minimisation requires that only data necessary for the intended purpose should be collected and processed. Accuracy is also essential, requiring steps to ensure that data is accurate and, where necessary, kept up to date. Storage limitation means that data should be retained only as long as necessary for the purposes for which it was collected. Integrity and confidentiality require appropriate security measures to protect data from unauthorised access, disclosure, alteration, or destruction. The Netherlands has stringent data minimisation requirements, reflecting the broader GDPR mandates and emphasises the need for organisations to document their data processing activities comprehensively.

Individuals have specific rights regarding their data. The right to access allows individuals to request access to their data to understand how it is being used and to verify its accuracy. The right to rectification enables individuals to request corrections to their data if it is inaccurate or incomplete. The right to erasure, also known as the right to be forgotten, allows individuals to request the deletion of their data under certain conditions. The right to restriction of processing permits individuals to request the restriction of data processing in specific circumstances. The right to data portability allows individuals to request the transfer of their data to another service provider in a structured, commonly used and machine-readable format. The right to object enables individuals to object to the processing of their data for certain purposes, such as direct marketing. In Serbia, the Law on Personal Data Protection aligns with GDPR principles, ensuring that individuals have robust rights regarding their personal data.

Organisations must demonstrate compliance with these principles through accountability and governance measures. Documentation involves keeping detailed records of data processing activities to provide evidence of compliance. Conducting data protection impact assessments for high-risk processing activities helps to identify and mitigate risks to individuals’ privacy. Appointing data protection officers (DPOs) is another crucial step; DPOs oversee compliance efforts and act as points of contact for data subjects and regulatory authorities. These measures ensure that organisations are accountable for their data processing activities and compliant with data protection regulations. In Ukraine, the Law on Personal Data Protection requires similar compliance measures, including the appointment of DPOs and the conduct of impact assessments.

Cross-border data transfers require special attention to ensure that equivalent data protection measures are in place. Transferring data to other regions often involves implementing standard contractual clauses, binding corporate rules, or other mechanisms approved by regulatory authorities. These measures ensure the continued protection of data when it is transferred outside the region, maintaining compliance with data protection laws and safeguarding individuals' privacy. Belgium and the Netherlands have specific guidelines for cross-border data transfers, ensuring that data protection is not compromised when data moves beyond their borders.

3.3.2 Anonymisation

Anonymisation is the process of removing PII from data sets to ensure that individuals cannot be identified directly or indirectly. This is an important aspect of data governance, especially in contexts where sensitive information is involved, such as soil quality analysis that might be linked to specific landowners or regions.

Anonymisation is essential for several reasons. Firstly, it safeguards individual privacy by ensuring that personal data cannot be traced back to an individual. Secondly, various data protection laws and regulations mandate the protection of personal data. Anonymisation helps organisations comply with these regulations. Thirdly, anonymised data can be shared more freely for research, analysis and collaboration without the risk of exposing personal information. Finally, it reduces the risk of data breaches and misuse, as anonymised data is less valuable to malicious actors.

Several techniques can be used to anonymise data, and the choice of method depends on the type of data and the desired level of anonymity. Common methods include data masking, which involves replacing sensitive data with fictitious but realistic data, such as replacing real names with randomly generated names. Aggregation summarises data at a higher level, such as reporting average soil quality metrics for a region instead of individual plots. Suppression involves removing certain data points or attributes that could lead to identification, like omitting specific geographic coordinates. Randomisation alters data slightly to prevent re-identification while preserving overall trends and patterns. Pseudonymisation replaces private identifiers with pseudonyms or codes, allowing for some degree of data linkage without revealing actual identities.

To effectively anonymise data in SQAT, the following steps should be followed. First, identify PII to determine which data elements are considered personally identifiable or sensitive. Next, choose appropriate anonymisation techniques based on the nature of the data and the level of anonymity required. Then, apply the chosen techniques to anonymise the data, which may involve using software tools designed for data anonymisation. It is essential to validate the anonymisation process to ensure it has effectively removed or masked all PII, which can be done through testing and peer reviews. Finally, regularly review the anonymisation process to ensure it remains effective, especially as new data is collected or as data privacy regulations evolve.

Despite its benefits, anonymisation comes with challenges and considerations. Even anonymised data can sometimes be re-identified by combining it with other data sets, so continuous monitoring and updating of anonymisation techniques are necessary to mitigate this risk. There is also a need to balance the need for privacy with the need to maintain data utility for analysis, as over-anonymisation can render data useless for its intended purpose. Additionally, it is crucial to ensure that anonymisation practices comply with legal requirements and ethical standards.

3.3.3 Consent Management

Consent management ensures that data collection, processing and usage comply with legal and ethical standards. It involves obtaining explicit permission from individuals before collecting their data and managing those permissions throughout the data lifecycle. This section outlines the principles, practices and technologies involved in effective consent management.

Importance of Consent Management

Consent management is vital for several reasons. Firstly, it ensures compliance with regulations, adhering to data protection laws and regulations such as GDPR and other regional data privacy laws. Secondly, it builds trust among users by demonstrating a commitment to data privacy and transparency. Lastly, it promotes the ethical use of data by respecting user preferences and autonomy.

Principles of Consent Management

Key principles of consent management include transparency, where users are clearly informed about what data is being collected, why it is being collected, how it will be used and who it will be shared with. Explicit consent should be obtained through a clear affirmative action, ensuring that users fully understand what they are agreeing to. Consent should be granular, allowing users to give consent for specific purposes and types of data processing, rather than a blanket consent. Revocability is crucial, giving users the ability to withdraw their consent at any time, and the process to do so should be simple and accessible. Additionally, all consents obtained should be documented, ensuring that there is a clear record of what consents have been given and for what purposes.

Practices for Effective Consent Management

To provide clear and concise information, easily understandable information about data collection and usage practices should be provided. Avoiding jargon and ensuring that the information is accessible to all users is essential. User-friendly consent mechanisms should be implemented, using intuitive and straightforward mechanisms for obtaining consent, such as pop-up consent forms, checkboxes, or sliders on the SQAT interface. Granular consent options should allow users to consent to different types of data collection and processing activities separately. For example, users can consent to soil sample data collection but opt-out of location data collection. It should be easy for users to withdraw their consent, with clear instructions and accessible options within SQAT for managing their consent preferences. Regular reviews and updates of consent management practices are necessary to comply with evolving regulations and best practices. Users should be informed of any significant changes to consent practices, and re-consent should be obtained if necessary.

Technologies for Consent Management

Consent management platforms can be utilised to automate and manage consent collection, storage and compliance, ensuring that consent practices adhere to legal requirements. Data privacy management software can be implemented to monitor and manage data privacy, including consent preferences and data subject requests. Maintaining detailed audit trails and logs of all consent-related activities is crucial, as it helps demonstrate compliance and provides a record in case of audits or disputes.

Integration with SQAT

A user-friendly consent interface should be designed within SQAT, where users can easily view and manage their consent preferences. Real-time consent verification mechanisms should be implemented to ensure that data processing activities comply with the current consent status of users. Notifications and alerts should be set up to inform users of any changes to data practices or their consent status, ensuring ongoing transparency and trust.

3.4 Data Lifecycle Management

3.4.1 Data Collection

Data collection is the foundational step in the data lifecycle (*Fig. 5*), crucial for generating reliable and accurate insights from SQAT. This process involves gathering soil quality data using autonomous soil sampling technology, ensuring consistency and accuracy through standardised methods. Effective data collection is critical for producing high-quality data that supports informed decision-making in agricultural and environmental contexts. The objectives of data collection are to ensure accuracy, completeness, consistency and timeliness. Accurate data reflects the true state of soil conditions, comprehensive data captures all necessary data points to provide a full view of soil quality, consistent data uses standardised methods to maintain uniformity across datasets, and timely data is collected at appropriate intervals to ensure its relevance for analysis.

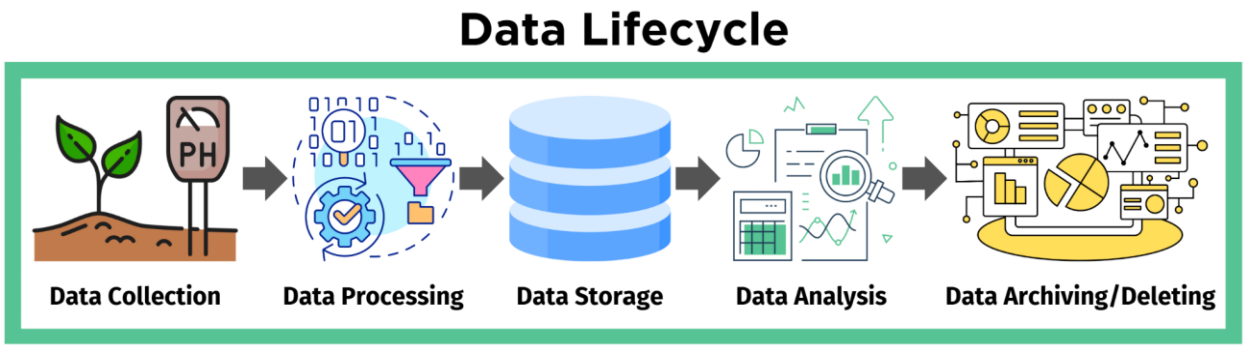


Figure 5 | Data lifecycle graphic.

The autonomous soil sampling technology enhances data collection by automating the process of field sampling, which involves collecting soil samples from various locations and depths. This technology integrates advanced sensor networks to continuously monitor soil conditions, capturing real-time data on parameters like pH, temperature, moisture and nutrient levels. By deploying space-based data sources, including EO and SatNav, and integrating these with in-field proximal sensor-based measurement technology, it ensures high precision and coverage.

The methods used for data collection are designed to meet the specific requirements of soil quality analysis. The autonomous system follows predefined paths and schedules to collect samples, reducing human error and increasing efficiency. Remote sensing techniques involve processing high-resolution imagery from satellites or drones using algorithms to extract soil quality parameters, while the sensor data logging transmits data to a central database at regular intervals, allowing for continuous monitoring and real-time updates.

Standardised protocols are crucial for ensuring the quality and reliability of the data collected by the autonomous system. These protocols include clear guidelines for sample collection, such as the depth, location and frequency of sampling. Using standardised containers and labelling systems ensures traceability and prevents contamination of samples. The technology is programmed to follow these protocols precisely, maintaining data quality and consistency.

To ensure data quality, various quality assurance measures are implemented. These include duplicate sampling, regular calibration of sensors and validation of remote sensing data against ground truth measurements. Regular audits and reviews of the data collection process help identify and address any inconsistencies or errors, maintaining the integrity and accuracy of the data collected.

Ethical considerations are also an important aspect of data collection. It is essential to ensure that data collection methods comply with local regulations and ethical guidelines, especially when collecting data on private land or in environmentally sensitive areas. The autonomous system operates with necessary permissions and informed consent from landowners and stakeholders, maintaining ethical standards and fostering trust with the community.

Finally, thorough documentation of the data collection process is necessary to provide context for data analysis and interpretation. Detailed records are maintained, including methodologies used, date and time of collection, location coordinates and any observations made during sampling. Any anomalies or deviations from standard protocols are documented to provide a comprehensive understanding of the data collected. This documentation supports the transparency and reproducibility of the data collection process.

3.4.2 Data Storage

Data storage involves the systematic organisation, protection and retention of data to ensure its availability, integrity and security throughout its lifecycle. Proper data storage practices are essential for maintaining data quality and supporting the analysis and decision-making processes.

The objectives of data storage in SQAT are multifaceted. Primarily, data must be readily accessible to authorised users whenever needed for analysis and reporting. This ensures that data availability is prioritised to facilitate timely decision-making. Additionally, maintaining data integrity is crucial, meaning that data should remain accurate, consistent and reliable over time without corruption or loss. Enhancing data security is also a key objective, involving the implementation of measures to protect data from unauthorised access, breaches and other security threats. Finally, supporting compliance with regulatory requirements and industry standards related to data storage and retention is vital to avoid legal and financial penalties.

When considering storage architecture, organisations can choose between centralised and distributed storage solutions. Centralised storage solutions, such as data warehouses or cloud-based storage, provide a single repository for all data, simplifying management and access control. In contrast, distributed storage systems, such as data lakes or hybrid cloud solutions, offer scalability and flexibility, allowing data to be stored across multiple locations, thus accommodating diverse data types and volumes. It is also essential to consider the physical location of servers, especially from the perspective of GDPR compliance. Under GDPR, data must be stored within the EU or in countries that provide adequate data protection, ensuring that data sovereignty requirements are met.

Various storage technologies are available to meet different data storage needs. Relational databases are suitable for structured data, using tables to organise data and supporting complex queries and transactions. For unstructured or semi-structured data, NoSQL databases offer scalability and flexibility, handling large volumes of diverse data types efficiently. Cloud storage solutions, such as Amazon S3, Google Cloud Storage, or Azure Blob Storage, provide scalable and cost-effective options with robust

security features. Organisations with specific regulatory or security requirements may opt for on-premises storage solutions, which offer complete control over data storage and access.

To ensure data availability and reliability, redundancy and backup strategies are essential. Implementing redundancy strategies, such as data replication and mirroring, ensures that data is not lost in the event of hardware failures or other disruptions. Regularly scheduled backups are crucial for data recovery in cases of accidental deletion, corruption, or disasters. These backups should be stored in geographically separate locations to protect against local incidents.

Data security measures are paramount in safeguarding stored data. Encrypting data both at rest and in transit protects it from unauthorised access and breaches, with encryption keys managed securely. Implementing strict access controls ensures that only authorised users can access or modify data, utilising RBAC and MFA as recommended practices. Regularly auditing and monitoring data access and usage activities help detect and respond to potential security threats promptly.

Data retention policies play a crucial role in managing data storage efficiently. Defining and enforcing retention periods based on regulatory requirements, organisational policies and data usage needs ensures that data is retained only as long as necessary for its intended purpose. Archiving infrequently accessed data to lower-cost storage solutions optimises storage resources while preserving data for long-term access. Securely deleting data that is no longer needed ensures it cannot be recovered or misused, thereby protecting sensitive information.

Best practices for data storage in SQAT include adopting a multi-tier storage strategy, using different storage tiers based on data access frequency and importance. Frequently accessed data can be stored in high-performance storage, while less critical data can be moved to cost-effective solutions. Implementing data classification based on sensitivity and importance allows for the application of appropriate storage and security measures, ensuring that sensitive data receives stringent protection and access controls. Leveraging automation tools to manage data storage tasks, such as backups, archiving and monitoring, reduces the risk of human error and enhances efficiency. Regularly reviewing storage needs and adjusting storage solutions accordingly ensures that storage capacity and performance align with evolving data volumes and usage patterns. Lastly, staying updated with regulatory changes and industry standards related to data storage ensures compliance, thereby avoiding legal and financial penalties.

3.4.3 Data Usage

Data usage encompasses all activities related to accessing, manipulating, analysing and interpreting data to extract valuable insights and support decision-making processes. Effective data usage is critical for maximising the benefits of SQAT and achieving its objectives. By utilising data appropriately, stakeholders can enhance soil management practices, improve agricultural productivity and make informed decisions based on accurate and timely information.

Objectives

The primary objective of data usage within SQAT is to extract meaningful insights into soil quality and health. This involves utilising data to understand soil conditions, identify trends and predict future outcomes. Supporting decision-making is another key objective, as data-driven insights can guide agricultural practices, environmental management and research activities. Additionally, optimising

resources is crucial, as effective data usage can lead to more efficient and effective soil management practices, ultimately improving crop yields and environmental sustainability.

Best Practices for Data Usage

To ensure effective data usage, it is important to define clear objectives for data analysis. Establishing specific goals helps focus efforts and ensures that the analysis is relevant and aligned with the overall goals of the organisation or project. Ensuring data accuracy is also critical, as accurate and reliable data is the foundation of meaningful analysis. This involves implementing validation checks and data cleaning processes to maintain high data quality.

Utilising advanced analytical tools and techniques, such as machine learning, statistical analysis and geospatial analysis, can enhance data processing capabilities and provide deeper insights. Promoting data accessibility is equally important, as relevant data should be accessible to authorised users when needed. This can be facilitated through user-friendly interfaces and data visualisation tools.

Maintaining data security and privacy during usage is essential to protect sensitive information. This involves implementing robust security measures, such as encryption and access controls, and ensuring compliance with data privacy regulations and user consent agreements. Documenting data usage processes, including methodologies, tools and results, helps ensure consistency and repeatability. Creating and updating standard operating procedures (SOPs) can further enhance this consistency.

Finally, fostering a data-driven culture within the organisation encourages data-driven decision-making and enhances data literacy and analytical skills among users. Providing training and support can help achieve this objective.

Data Usage Workflow

The data usage workflow (*Fig. 6*) begins with data collection, which involves gathering data from various sources, such as soil sensors, satellite imagery and manual sampling. Ensuring that data is collected in a structured and consistent manner is crucial for subsequent analysis.

Following data collection, the preprocessing stage involves cleaning and preprocessing the data to remove errors, fill missing values and standardise formats. Preprocessing techniques enhance data quality and prepare it for analysis.

During the data analysis stage, appropriate analytical methods are applied to extract insights from the data. This may involve using statistical, machine learning and geospatial techniques to analyse soil quality indicators. Interpretation of analytical results is the next step, where meaningful conclusions and insights are derived from the data. Data visualisation tools can be used to present findings in an understandable and actionable format.

Decision support is a critical component of the data usage workflow. Insights from data analysis are utilised to support decision-making processes, providing recommendations and action plans based on analytical findings. The final stage involves collecting feedback from data users to continuously improve data usage processes. This feedback can be used to update and refine analytical methods and tools based on user input and technological advancements.

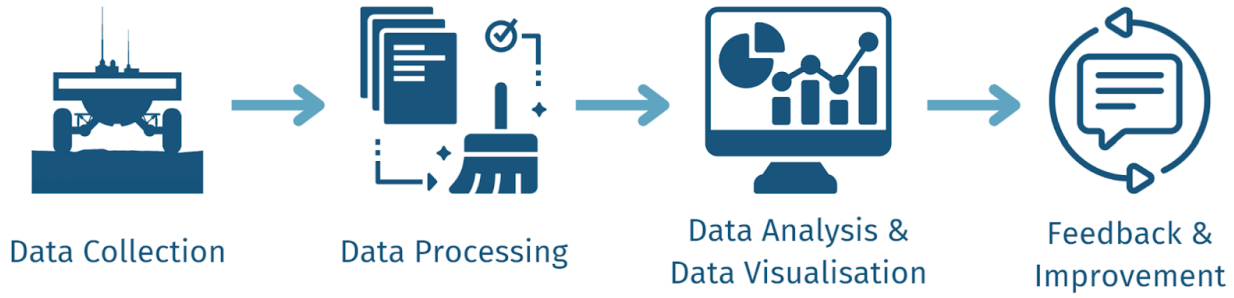


Figure 6 | Data usage workflow graphic.

3.4.4 Data Archiving & Deletion

Effective data archiving and deletion ensure that data is managed responsibly, complying with regulatory requirements and organisational policies. Proper practices in this area help optimise storage, reduce risks associated with data breaches and maintain data quality over time.

Data Archiving

Data archiving involves transferring inactive data from primary storage to a secure and long-term storage system. This process is essential for managing data growth and ensuring that only relevant data is readily accessible while preserving historical data for future reference, compliance, or analysis. Establishing clear retention policies based on legal, regulatory and business requirements is vital. These policies should define the duration for which different types of data should be retained before being archived. It is also important to ensure that archived data remains accessible for authorised users when needed by implementing indexing and search capabilities to facilitate easy retrieval. Utilising cost-effective and secure storage solutions, such as cloud storage, offsite storage facilities, or dedicated archival systems, can help manage archived data efficiently.

Robust security measures, including encryption, access controls and regular security audits, should be in place to protect archived data from unauthorised access, corruption and breaches. Adhering to industry standards and regulatory requirements related to data archiving ensures that the archiving process aligns with data privacy laws and organisational policies. Implementing automated archiving solutions can streamline the process and reduce the risk of human error, while regular audits of archived data help ensure compliance with retention policies and identify any potential issues. Classifying data based on its importance, sensitivity and regulatory requirements can also determine appropriate archiving strategies.

Data Deletion

Data deletion is the process of permanently removing data that is no longer needed from all storage systems. Proper data deletion practices are essential to free up storage space, reduce costs and mitigate risks associated with retaining unnecessary data. Establishing comprehensive data deletion policies that define the criteria and timelines for deleting different types of data is crucial, ensuring these policies comply with legal and regulatory requirements. Secure deletion methods, such as data wiping, shredding and degaussing for physical storage media, should be used to ensure that data cannot be recovered.

Maintaining audit trails of data deletion activities provides accountability and transparency, which is particularly important for compliance and legal purposes. Implementing procedures for notifying relevant stakeholders and obtaining necessary approvals before data deletion helps prevent accidental deletion of important data.

It is also important to ensure that backup copies of data are deleted in accordance with the data deletion policies to prevent unauthorised recovery of deleted data. Utilising automated tools and processes can manage data deletion efficiently and consistently across all storage systems. Training employees on data deletion policies and secure deletion methods ensures proper handling of data throughout its lifecycle. Regularly reviewing and updating data deletion policies to reflect changes in regulations, business needs and technological advancements is also essential.

Challenges and Solutions

Addressing the risk of data recovery by using certified data destruction methods and tools and ensuring compliance with industry standards for secure data deletion is important. Staying informed about changing regulations related to data retention and deletion and adapting policies and practices accordingly ensures ongoing compliance. Identifying and managing dependencies between data sets helps avoid unintentional deletion of data that is still in use or required for other purposes.

4 Data Governance Policies

4.1 Data Ownership

Data ownership defines who is accountable for specific data assets within the organisation. Clear data ownership ensures that data is managed properly throughout its lifecycle, from collection to disposal. In the context of SQAT, establishing data ownership involves identifying the individuals or roles responsible for different types of data and ensuring they have the authority and resources needed to manage the data effectively.

Definition of Data Ownership

Data ownership refers to the legal rights and complete control over a data set. The data owner is responsible for ensuring the accuracy, integrity and security of the data. In the context of SQAT, data ownership involves responsibilities such as defining data standards, managing data quality and ensuring compliance with relevant regulations. This comprehensive control allows data owners to make decisions about how data is collected, used, stored and shared, ultimately ensuring that the data serves its intended purpose effectively and ethically.

In scenarios where SQAT operates as a commercial service provider, the concept of data ownership becomes more complex, particularly concerning the data collected from farmers. Farmers, as the primary generators of the data, have inherent rights over the data they produce. This includes the right to access their data, control how it is used, and ensure its protection. Farmers should have the ability to opt out of specific uses of their data, such as training algorithms, if they choose.

SQAT, on the other hand, needs to clearly define its rights and responsibilities concerning the data it collects and manages. While SQAT has a role in maintaining the data's integrity and security, its rights to use the data for purposes such as improving services or training algorithms must be transparently communicated to the farmers. This includes obtaining explicit consent from farmers regarding how their data will be used and providing options for them to opt out of certain uses.

Furthermore, SQAT must adhere to all relevant data protection regulations, ensuring that the data is used ethically and legally. Any data sharing or usage agreements must be clearly outlined, respecting the rights of farmers while allowing SQAT to leverage the data for service improvement and innovation. This balance of rights and responsibilities ensures that data ownership is managed fairly and transparently, fostering trust and cooperation between SQAT and the farming community.

Responsibilities of Data Owners

Data owners have several key responsibilities that are essential to the effective governance of data. Firstly, they are responsible for data quality management, which involves ensuring that data is accurate, complete and reliable. This responsibility includes establishing data quality standards and continuously monitoring data to maintain these standards. Additionally, data owners are tasked with protecting data security, which involves implementing measures to safeguard data from unauthorised access, breaches and other security threats. This includes adopting appropriate security protocols and ensuring that these measures are regularly updated and enforced.

Another critical responsibility of data owners is ensuring compliance with relevant laws, regulations and policies. Data owners must stay informed about legal requirements and ensure that data handling procedures adhere to these regulations, thus avoiding legal penalties and maintaining the organisation’s reputation. Data accessibility is also a key responsibility, involving the management of data access to ensure that authorised users can retrieve and use data efficiently while preventing unauthorised access. This includes defining access permissions and maintaining an audit trail of data access activities. Furthermore, data owners oversee the entire lifecycle of the data, from creation and storage to archiving and disposal, ensuring that data retention policies are defined and adhered to, and that obsolete data is securely deleted. Lastly, data owners are responsible for addressing and resolving any conflicts or issues related to data ownership, usage, or access, acting as the primary point of contact for any disputes involving their data sets.

Identifying Data Owners

Identifying data owners within SQAT involves assigning ownership based on roles, expertise, and areas of responsibility, in alignment with the Grant Agreement (Article 16). Typically, data ownership is assigned to individuals or teams who generate or collect data, use and analyse data, or manage data systems. For instance, individuals or departments that generate or collect data should have ownership responsibilities. This could include a soil scientist who collects field data, making them the owner of that specific data set due to their direct involvement and expertise in data collection processes.

Similarly, teams or individuals who use data for analysis and decision-making may be designated as data owners, particularly if they play a significant role in data interpretation and reporting. This ensures that those who rely on the data for critical insights are directly involved in its management. Additionally, IT

personnel or departments that manage databases and data systems can also be assigned data ownership roles, especially for technical aspects of data management.

Data Ownership and Collaboration

Data ownership does not imply isolation; effective data governance requires collaboration among various stakeholders, including data owners, data stewards and end users. Data owners must work closely with these stakeholders to ensure data is used effectively and responsibly. Collaboration includes working with data stewards, who are responsible for maintaining data quality and integrity on a day-to-day basis, acting as intermediaries between data owners and data users. Data owners should also provide training and support to ensure that all stakeholders understand their roles and responsibilities regarding data handling and usage. Establishing feedback mechanisms is essential for continuous improvement of data governance practices. Data owners should actively seek input from data users and other stakeholders to identify areas for improvement, thus fostering an environment of shared responsibility and continuous enhancement of data governance processes.

Documentation and Transparency

Maintaining clear documentation of data ownership is essential for transparency and accountability. This includes creating and maintaining a data ownership register that lists all data owners, their responsibilities and the data sets they manage. Such a register ensures that there is a clear and accessible record of who is responsible for each data set, facilitating better management and accountability. Additionally, documenting policies and procedures related to data ownership, including guidelines for transferring ownership and handling disputes, is crucial. These documents serve as a reference point for all stakeholders and ensure consistency in data governance practices. Conducting regular reviews to ensure that data ownership assignments are up-to-date and reflect any changes in roles or responsibilities within the organisation is also important. Regular reviews help in maintaining the relevance and accuracy of the data ownership register, ensuring that it continues to serve its purpose effectively.

4.2 Data Sharing and Access

Effective data sharing policies ensure that the right individuals and systems have timely access to the data they need while maintaining data security and privacy. This section outlines the guidelines and protocols for data sharing and access.

The principles of data sharing within SQAT emphasises transparency, accountability, compliance, security and minimization. Transparency ensures that all stakeholders understand how and why data is shared. Accountability involves clearly defined responsibilities for data sharing to ensure adherence to data governance policies. Compliance with relevant regulations, such as GDPR and other industry-specific standards, is essential. Security measures must be in place to protect data from unauthorised access and breaches during the sharing process. Minimisation means sharing only the necessary amount of data required for a specific purpose.

Data access within SQAT is categorised into three levels: public access, restricted access and confidential access (*Table 1*). Public access data is non-sensitive and intended for public use, such as general soil quality trends and publicly available research data. Restricted access data is sensitive and requires authorization

for access, including specific soil sample data linked to private land or proprietary research data. Confidential access data is highly sensitive and requires strict access controls, such as personal data of individuals involved in soil sampling or confidential research results.

Data sharing protocols within SQAT include a formal request process, Data Use Agreements (DUAs), secure data transfer methods, access controls and monitoring and auditing. The request process involves a standardised request form, approval workflow, and documentation of granted permissions. Before sharing data, a DUA should be signed by the recipient, outlining the terms and conditions for data use, including confidentiality clauses and data protection measures. Secure methods for data transfer must be used, such as encrypted email, secure file transfer protocols (SFTP), or secure cloud storage solutions. Implementing RBAC ensures that only authorised users can access specific data sets, with permissions regularly reviewed and updated. Regular monitoring and auditing of data access logs help detect and respond to unauthorised access attempts or breaches, with anomalies investigated promptly.

	Public	Restricted	Confidential
Soil Quality	●	●	●
Open Datasets	●	●	●
User-specific Soil Data	✕	●	●
Soil Samplers	✕	✕	●
Research Results	✕	✕	●

Table 1 | Data access for all SQAT-related data.

Data sharing scenarios within SQAT include internal sharing, external sharing and public data releases. For internal sharing, data should be shared based on the principle of least privilege, ensuring that employees have access only to the data necessary for their roles. When sharing data with external partners, collaborators, or researchers, it is essential to have DUAs in place and share data securely. Anonymizing or pseudonymised data where possible protects privacy. For data intended for public release, proper anonymization and aggregation should be ensured to prevent the identification of individuals or sensitive locations.

Several challenges are associated with data sharing, including data sensitivity, compliance risks, security threats and the potential misuse of data. To address data sensitivity, carefully assess the sensitivity of data before sharing and implement robust anonymization techniques to protect privacy. To mitigate compliance risks, stay updated on regulatory requirements and ensure all data sharing practices are compliant through regular audits and compliance checks. Address security threats by using advanced security measures, such as encryption and secure access controls, and training staff on security best practices. Clearly define acceptable use policies in DUAs and monitor compliance to prevent the misuse of data, taking corrective actions against any violations.

Best practices for data sharing within SQAT include thorough documentation, regular reviews, stakeholder engagement and continuous improvement. Maintain thorough documentation of all data sharing agreements, access requests and permissions granted. Conduct regular reviews of data sharing policies and practices to ensure they remain effective and compliant with evolving regulations. Engage with stakeholders to understand their data needs and address any concerns related to data sharing and access. Implement a feedback loop to continuously improve data sharing protocols based on lessons learned and emerging best practices.

4.3 Data Classification

Data classification involves categorising data based on its level of sensitivity, value and criticality to the organisation. Effective data classification helps ensure appropriate levels of protection and access controls, facilitating compliance with legal and regulatory requirements while optimising data management processes. The primary purpose of data classification is to identify and categorise data to manage it effectively according to its sensitivity and value. The benefits of data classification include enhanced data security by applying appropriate protection measures, improved compliance with legal and regulatory requirements, better data management, optimised storage solutions, facilitated data sharing and collaboration within appropriate access controls, and increased efficiency in data handling and processing.

Data within SQAT should be classified into distinct levels based on sensitivity and criticality. The proposed classification levels are as follows: Public Data, Internal Data, Confidential Data and Restricted Data. Public Data is information intended for public access, such as general soil quality reports and non-sensitive research findings, posing minimal risk if disclosed. Internal Data is meant for internal use within the organisation, including operational data, internal communications and non-sensitive business processes, where unauthorised disclosure may cause minor inconvenience. Confidential Data includes sensitive information that could harm the organisation or individuals if disclosed, such as detailed soil quality analysis results, proprietary algorithms and non-public research data. Access to this data should be restricted to authorised personnel only. Restricted Data comprises highly sensitive information requiring the highest level of protection, including personal data subject to privacy regulations, critical business secrets and data protected by law, where unauthorised access could result in severe legal and financial repercussions.

The data classification process involves several key steps to ensure accurate and consistent categorisation. It begins with a thorough inventory of all data assets within SQAT to identify and document data types and sources. Establishing clear criteria for each classification level based on sensitivity, value and regulatory requirements is crucial. Data assets are then assigned classification tags using metadata and labelling tools, which help automate data handling processes and enforce access controls. Implementing access control measures based on data classification levels ensures that only authorised users can access sensitive data. Periodically reviewing and updating data classifications to reflect changes in data sensitivity, regulatory requirements and organisational priorities is also essential.

Successful data classification requires collaboration across various roles within the organisation. Data Stewards oversee the classification process and ensure adherence to classification policies. Data Owners are accountable for the classification of data they manage, ensuring data is accurately categorised based

on established criteria. IT and Security Teams implement and maintain technical measures to enforce classification and access control policies. End Users must follow data classification guidelines and handle data according to its designated classification level.

Utilising the right tools and technologies can streamline the data classification process. Metadata management tools help automate the tagging and categorization of data assets based on predefined criteria, while data discovery tools assist in identifying and inventorying data assets across various sources and systems. Access control systems enforce classification-based access controls, ensuring that data is accessed and handled according to its classification level. Data loss prevention (DLP) solutions monitor and protect sensitive data from unauthorised access and breaches.

Implementing data classification can pose several challenges, including managing large volumes of data, which can complicate classification efforts. Automating classification processes can help address this challenge. Ensuring consistent classification across different data types and sources requires clear guidelines and ongoing training. Regularly updating classifications to reflect changes in data sensitivity and regulatory requirements is essential. Best practices for effective data classification include establishing and communicating clear classification policies and procedures to all stakeholders, conducting regular training sessions to ensure all users understand and adhere to classification guidelines, and continuously reviewing and refining classification processes to enhance accuracy and effectiveness.

5 Roles and Responsibilities

5.1 Data Governance Committee

The establishment of a Data Governance Committee (DGC) will be an important step in developing and maintaining a robust data governance framework for SQAT. This committee, to be developed and established in future WP4 progress, will play a pivotal role in overseeing the implementation of data governance policies, ensuring compliance with data management principles and promoting a culture of data stewardship within the project.

- *Composition of the Data Governance Committee*

The DGC will be composed of different members amongst SQAT partners with diverse expertise to ensure comprehensive oversight of data governance activities:

- **Data Stewards:** Responsible for managing and ensuring the quality of data within their respective work packages or areas.
- **Data Custodians:** IT personnel tasked with technical data management, including storage, security and infrastructure.
- **Compliance Officer:** Ensures adherence to legal, regulatory, and industry standards in all data management practices.
- **Business Analysts:** Provide insights into data governance needs and challenges, leveraging data for strategic decision-making.

- **End User Representatives:** Offer practical insights into data usability and accessibility based on regular interaction with project data.

- *Responsibilities of the Data Governance Committee*

The DGC will be responsible for developing and enforcing data governance that align with GDPR and the project's strategic goals. This includes establishing and maintaining a comprehensive framework that defines roles, responsibilities and processes for managing data assets while ensuring compliance with legal and regulatory standards. The committee will oversee data quality by implementing validation and monitoring tools, address risks through proactive compliance measures and manage data-related incidents effectively. Additionally, the DGC will foster a culture of data stewardship by providing best practices training, evaluating and recommending technologies to support governance activities, and engaging with stakeholders to align data management practices with project objectives. Regular updates to policies and the governance framework will ensure adaptability to evolving project needs and external requirements (*Fig. 7)*.

Figure 7 | DGC responsibilities.

- *Governance and Operations*

The DGC will meet monthly to review activities, address issues and plan initiatives. Minutes from each meeting will be recorded and shared with relevant stakeholders to ensure transparency. Open communication with other governance bodies within the project consortium, such as IT governance and risk management committees, will promote alignment and collaboration.

The DGC will have the authority to make decisions related to data governance policies and standards, with major decisions requiring senior management approval. Regular reviews and updates to the project framework will ensure adaptability to changing project needs, regulatory requirements and technological advancements.

5.2 Data Stewards

Data Stewards are important in ensuring the effective management and governance of data within the SQAT project. Their responsibilities include overseeing data quality, compliance with regulatory requirements, and alignment with the project's objectives. These roles are assigned at both the WP level and the project-wide level, ensuring comprehensive coverage across all aspects of the project.

At the project-wide level, Senus serves as the Lead Data Steward, coordinating data management efforts across all WPs to maintain consistency in data governance practices. Senus is responsible for implementing the data governance principles outlined in this document, resolving cross-WP data challenges, and liaising with the DGC to align project activities with regulatory requirements such as GDPR. By providing oversight across the entire project, Senus ensures that data governance remains uniform and effective.

At the WP level, each WP leader is tasked with appointing a dedicated Data Steward to oversee the data management activities specific to their Work Package. These WP-level Data Stewards ensure data quality and compliance within their domains while actively collaborating with the Lead Data Steward to address any interdependencies or challenges. Together, the WP-level and project-wide Data Stewards form a cohesive team that supports the consistent implementation of data governance practices across the SQAT project.

This structure ensures a clear and scalable approach to data stewardship, fostering accountability at both local and central levels while leveraging Senus's expertise as the project's Lead Data Steward to maintain alignment with overall project objectives.

5.3 Data Custodians

Data custodians also have significance in the data governance framework by ensuring the secure and efficient management of data assets. They are responsible for the technical environment where the data resides, including databases, servers and storage systems. Data custodians work closely with IT teams to implement and maintain the infrastructure necessary for data storage, ensuring that it meets organisational standards and regulatory requirements. Their duties encompass a wide range of activities, including data backup, recovery and archiving, which are essential for data integrity and availability.

One of the primary responsibilities of data custodians is to enforce data security measures. This includes setting up and managing access controls to ensure that only authorised personnel have access to sensitive data. Data custodians implement encryption protocols to protect data at rest and in transit, safeguarding it from unauthorised access and breaches. They also regularly monitor the system for security vulnerabilities and apply necessary patches and updates to prevent potential threats. By maintaining robust security practices, data custodians help protect the organisation's data from cyber threats and ensure compliance with data protection regulations.

Data custodians are also tasked with ensuring the data's accuracy and reliability. They oversee data validation processes to detect and correct errors, ensuring that the data used for analysis and decision-making is accurate and consistent. Additionally, they manage metadata, which involves documenting the

data's source, usage and structure. This documentation is crucial for maintaining data quality and providing context to data users, enabling them to understand the data's lineage and trust its validity.

Collaboration with other data governance roles is a key aspect of the data custodian's job. They work in conjunction with data stewards to understand the data requirements and quality standards set by the organisation. Data custodians also liaise with data users to address their technical needs and support their data access requests. This collaboration ensures that the data infrastructure supports the organisation's overall data governance strategy and meets the needs of all stakeholders.

5.4 End Users

End users are individuals who interact with SQAT to access, analyse and utilise soil quality data for various purposes. These users play a crucial role in the data governance framework as they are the primary consumers of the data generated and managed. Their responsibilities include adhering to data governance policies, ensuring data accuracy and maintaining data security in their operations.

End users can range from agricultural professionals, agri-service third parties, researchers, environmental scientists and policy makers to farmers and land managers. Each user group has unique needs and requirements when it comes to data access and usage. Agricultural professionals and researchers might use SQAT data for in-depth analysis and research projects, whereas farmers and land managers might rely on the tool for practical, day-to-day decision-making related to crop management and soil health improvement.

A key responsibility of end users is to follow the established data access controls and security protocols to prevent unauthorised access and misuse of sensitive data. This includes using strong authentication methods, respecting data privacy guidelines, and reporting any anomalies or security breaches promptly. By adhering to these protocols, end users help maintain the integrity and confidentiality of the data, ensuring that it can be trusted for critical analysis and decision-making.

Furthermore, end users are encouraged to contribute to data quality by providing feedback on data accuracy and completeness. They can report discrepancies or suggest improvements, which helps data stewards and custodians to continuously enhance the data governance processes. Active participation from end users in this feedback loop is essential for maintaining high data quality standards and ensuring the data remains relevant and useful.

In addition to their operational responsibilities, end users must stay informed about updates to data governance policies and procedures. Regular training and awareness programs should be provided to keep users updated on best practices, new tools and any changes in data governance protocols. This continuous education helps users stay compliant and effectively leverage the capabilities of SQAT.

6 Data Governance Processes

6.1 Data Quality Management

The SQAT project places a strong emphasis on maintaining high data quality standards to ensure the reliability, accuracy and usability of project datasets. The data quality framework focuses on establishing and maintaining standards for accuracy, completeness, consistency and timeliness while incorporating advanced tools and processes for ongoing assessments and improvements.

To establish these data standards, the project utilises both external tools and an in-house developed data quality and cleaning tool built by Senus. This proprietary tool enables automated data validation, profiling and error correction, ensuring that raw data collected from various sources—such as Sentinel satellites, sensors and in-field inputs—meets predefined quality thresholds. These standards are essential for generating accurate soil property maps, reliable field stratification data and actionable insights for precision agriculture.

Maintaining data quality standards involves robust processes for validation and cleaning. Real-time data validation checks are integrated at the point of collection to minimise errors, particularly for in-situ data from sensors and field devices. The Senus platform ensures consistency through automated workflows that detect anomalies, flagging potential inconsistencies for review. Additional tools such as GeoNetwork can be employed to manage geospatial metadata, ensuring compliance with FAIR principles and interoperability with external data sources like Copernicus datasets.

Ongoing data assessments play a critical role in monitoring and improving quality. KPIs such as error rates, data completeness and timeliness are regularly measured to evaluate the effectiveness of data processes. Advanced analytics tools, coupled with periodic audits, provide actionable insights to refine data management practices further. For example, discrepancies in soil chemical properties or delays in data reporting can be quickly identified and resolved to maintain alignment with project goals.

Moreover, data quality management is a collaborative effort that involves various stakeholders, including data stewards, data custodians and end users. Data stewards are responsible for managing data quality on a day-to-day basis, ensuring that data entry and processing adhere to established standards. Data custodians, typically IT personnel, ensure that the technical infrastructure supports data quality efforts, including data storage, backup and recovery processes. End users play a crucial role by reporting any data quality issues they encounter, thus providing valuable feedback for continuous improvement.

Training and education are also integral components of DQM. Ensuring that all stakeholders are aware of data quality standards, processes and their importance helps in fostering a culture of quality within the organisation. Regular training sessions, workshops and updates on best practices help maintain high data quality standards.

6.2 Data Security Management

Data security management involves implementing measures to protect data from unauthorised access, breaches and other security threats.

One of the primary aspects of data security management is the implementation of access controls. Access controls are designed to restrict data access to authorised users only. This involves defining user roles and permissions, ensuring that only those with the necessary credentials can access sensitive data. Access controls help prevent unauthorised access, thereby reducing the risk of data breaches and ensuring that data is only used for its intended purposes.

Encryption is another crucial element of data security management. Encryption involves converting data into a coded format that can only be deciphered by authorised users with the appropriate decryption key. By encrypting data both in transit and at rest, SQAT ensures that even if data is intercepted or accessed by unauthorised individuals, it remains unreadable and secure. This adds an additional layer of protection to sensitive data, safeguarding it from potential threats.

Data masking is also an important practice in maintaining data security. Data masking involves hiding or obfuscating sensitive information within data sets to prevent unauthorised access to specific details. This technique is particularly useful in scenarios where data needs to be shared for testing or development purposes without exposing sensitive information. By masking data, SQAT can ensure that critical information remains protected while still allowing for necessary data processing and analysis.

In addition to these technical measures, data security management also involves developing and enforcing policies and procedures that govern data security practices. These policies outline the responsibilities of users, data stewards and IT personnel in protecting data. They include guidelines for handling data securely, reporting security incidents and responding to data breaches. By establishing clear policies and procedures, SQAT creates a structured approach to data security, ensuring that all stakeholders understand their roles and responsibilities in safeguarding data.

Regular security assessments and audits are integral to maintaining robust data security. These assessments help identify vulnerabilities and potential risks within the data management system. By conducting regular audits, SQAT can proactively address security gaps, implement necessary improvements and ensure compliance with regulatory requirements. Security assessments also provide valuable insights into the effectiveness of existing security measures and help in making informed decisions about future security strategies.

6.3 Incident Management

Incident management should be designed to handle data-related incidents promptly and effectively. An incident refers to any event that compromises the integrity, confidentiality, or availability of data. This can include data breaches, data loss, unauthorised access and data corruption. A well-defined incident management process ensures that these issues are addressed swiftly to minimise impact and restore normal operations as quickly as possible.

The first step in incident management is the identification and detection of incidents. This involves continuous monitoring of data systems and the use of advanced tools to detect anomalies that may indicate a potential incident. It is important to have predefined criteria for what constitutes an incident, as well as clear channels for reporting such events. Employees should be trained to recognize signs of data incidents and understand the proper reporting protocols.

Once an incident is identified, the next step is containment. The goal of containment is to limit the damage and prevent further impact. This might involve isolating affected systems, suspending compromised accounts, or shutting down certain operations temporarily. Swift containment actions are critical to prevent the incident from escalating and affecting additional data or systems.

After containment, the focus shifts to eradication and recovery. Eradication involves identifying the root cause of the incident and removing any traces of malicious activity. This could include deleting malware, closing security gaps and restoring corrupted data from backups. Recovery involves restoring affected systems and data to normal operations. This step may require coordination with various teams to ensure that all systems are securely back online and functioning correctly.

Communication is a vital part of the incident management process. During an incident, it is important to keep all stakeholders informed about the status and impact of the incident, as well as the steps being taken to resolve it. Clear and timely communication helps manage expectations and maintain trust. After the incident is resolved, a detailed incident report should be created, documenting the incident, the response actions taken and the lessons learned.

Finally, post-incident review and analysis are essential to improve the incident management process. This involves analysing the incident to understand what went wrong and identifying opportunities for improvement. The findings from the review should be used to update incident response plans, enhance detection tools and provide additional training to employees. Continuous improvement helps ensure that the organisation is better prepared for future incidents.

6.4 Risk Management

Risk management focuses on identifying, assessing and mitigating risks associated with data handling and processing. Effective risk management ensures that data-related risks do not compromise the tool's functionality, data integrity, or regulatory compliance.

The first step in risk management is to identify potential risks that could affect data governance processes. These risks can stem from various sources, including data breaches, data corruption, unauthorised access and non-compliance with data privacy regulations. Identifying these risks involves conducting comprehensive risk assessments that consider the entire data lifecycle, from data collection and storage to usage and disposal.

Once risks are identified, the next step is to assess their potential impact and likelihood. This assessment helps prioritise risks based on their severity and the probability of occurrence. For SQAT, this might involve evaluating the impact of data inaccuracies on soil quality analysis results, the likelihood of cyber-attacks targeting the tool's database and the consequences of non-compliance with data protection laws.

After assessing the risks, appropriate mitigation strategies must be developed and implemented. Mitigation strategies for SQAT may include enhancing data encryption protocols to protect sensitive data, implementing strict access controls to limit data access to authorised personnel only, and regularly updating software to patch vulnerabilities. Additionally, regular training sessions for staff on data security best practices and regulatory requirements can significantly reduce the risk of human errors leading to data breaches or non-compliance.

Monitoring and reviewing the effectiveness of risk mitigation strategies is an ongoing process. This involves setting up continuous monitoring systems to detect and respond to data governance issues in real-time. For SQAT, automated monitoring tools can alert administrators to unusual data access patterns or potential security breaches, allowing for prompt action to mitigate any identified risks. Regular audits and reviews of data governance practices also ensure that risk management strategies remain effective and up-to-date with evolving threats and regulatory changes.

7 Technology and Tools for Data Governance

7.1 Data Governance Platform

The effectiveness of data governance within SQAT relies heavily on the use of a robust data governance platform. These platforms provide comprehensive capabilities to manage data assets, ensure data quality and maintain data security. For SQAT, we will be utilising our own Farmeye platform (*Fig. 8*), which has been tailored specifically to meet the unique needs of agricultural data management.

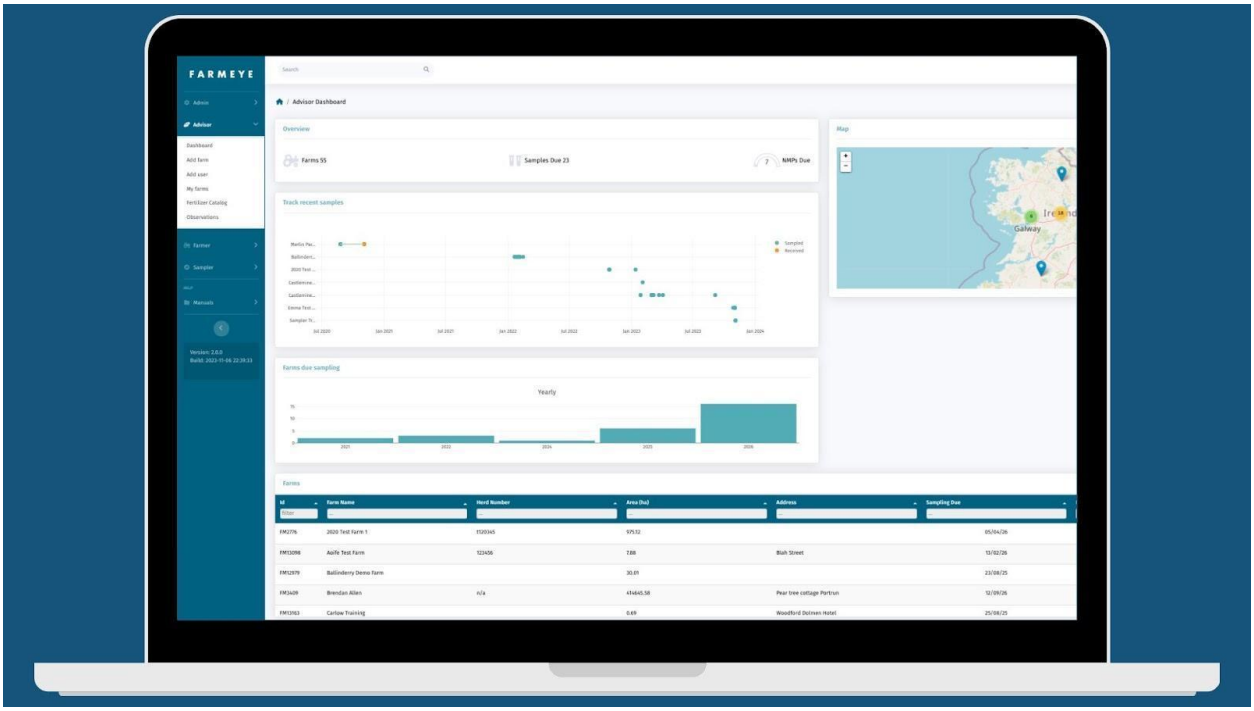


Figure 8 | Example screenshot of an admin view of the Farmeye platform.

Farmeye is designed to offer a centralised solution for data governance, enabling seamless integration with SQAT's functionalities. The platform provides a suite of tools that facilitate data stewardship, ensuring that data is accurate, consistent and reliable. By leveraging Farmeye, we can implement strict data quality controls that automatically validate and cleanse data, reducing the likelihood of errors and inconsistencies that could impact soil quality analysis.

Security is a cornerstone of Farmeye's data governance capabilities. The platform employs advanced security measures, including access controls, encryption and data masking, to protect sensitive

information from unauthorised access and breaches. With Farmeye, we can enforce role-based access controls, ensuring that only authorised personnel have access to specific data sets, thereby enhancing data privacy and compliance with regulatory requirements.

Furthermore, Farmeye supports comprehensive metadata management, allowing us to maintain detailed records of data definitions, lineage and usage. This transparency is critical for effective data governance, as it provides a clear understanding of where data originates, how it has been transformed and how it is being utilised within SQAT. This level of detail helps in auditing and ensures that data governance policies are being followed rigorously.

Farmeye also offers robust reporting and monitoring tools that enable continuous oversight of data governance processes. KPIs can be tracked in real-time, providing insights into the effectiveness of data governance initiatives. These reporting mechanisms are essential for identifying areas that require improvement and for ensuring that data governance practices evolve in response to new challenges and regulatory changes.

7.2 Data Quality Tools

Data quality tools are designed to ensure that data collected, processed and analysed is accurate, complete, consistent and timely, which are crucial attributes for making informed decisions in soil quality assessment.

One of the primary functions of data quality tools is data profiling, which involves examining data sources to understand their structure, content and quality. By analysing the data, these tools can identify patterns, anomalies and potential quality issues. For SQAT, data profiling helps in understanding the soil data characteristics, detecting outliers, and ensuring that the data conforms to expected formats and standards.

Data cleansing is another critical function provided by data quality tools. This process involves detecting and correcting errors or inconsistencies in the data to improve its quality. For instance, in SQAT, data cleansing can correct incorrect soil pH values, fill in missing nutrient information, or standardise units of measurement across different datasets. By automating these tasks, data quality tools significantly reduce the manual effort required and enhance the reliability of the soil analysis results.

Additionally, data quality tools offer data enrichment capabilities, allowing organisations to enhance their existing data by integrating additional relevant information. For SQAT, this could mean augmenting soil sample data with land usage patterns or historical crop yield information. Data enrichment provides a more comprehensive view, enabling more accurate and insightful soil quality assessments.

Data validation is also a critical feature of data quality tools. These tools enforce data quality rules and constraints to ensure that data meets predefined standards before it is used in analysis. In the context of SQAT, data validation can ensure that all soil samples are within the expected geographic coordinates, that soil properties fall within realistic ranges, and that all required fields are populated. This step is vital to prevent erroneous data from skewing the analysis results and recommendations.

Moreover, data quality tools include functionalities for ongoing data monitoring and quality assessment. Continuous monitoring ensures that data quality remains high over time, even as new data is collected

and integrated into SQAT. These tools can automatically detect and alert stakeholders about emerging quality issues, allowing for timely intervention and correction.

Finally, many data quality tools offer comprehensive reporting and dashboard capabilities. These features provide stakeholders with insights into the current state of data quality, trends over time and areas that require attention. For SQAT users, such dashboards will be used to ensure data quality standards are achieved. Specific tools such as the Farmeye platform with its data analytics features will ensure key metrics like the percentage of complete soil records, the frequency of data anomalies and the success rate of data cleansing efforts are transparent at all times. This transparency helps in maintaining a high standard of data governance and demonstrates the value of data quality initiatives to the organisations. Other tools such as QGIS will also play a role in data quality assurance.

7.3 Metadata Management Tools

Metadata management tools help manage and maintain metadata, which is data about data, providing essential context, definitions and relationships that enhance the usability and integrity of data assets. Again platforms such as Farmeye can deliver the required metadata management for the SQAT project. However, each partner will investigate the most efficient method for their own metadata management if Farmeye is not suitable.

Effective metadata management enables SQAT users to understand the origin, structure and meaning of the data they are working with. By leveraging metadata management tools, organisations can ensure that data is well-documented, searchable and easily retrievable. This improves data transparency and fosters a deeper understanding of the data among all stakeholders, from data stewards to end users.

One of the primary functions of metadata management tools is to provide a centralised repository where all metadata can be stored and accessed. This repository serves as a single source of truth for metadata, reducing redundancy and ensuring consistency across the organisation. It allows users to quickly locate and utilise the necessary data elements, thereby improving efficiency and reducing the likelihood of errors.

These tools also facilitate data lineage tracking, which is crucial for maintaining data integrity and compliance. Data lineage provides a detailed view of the data's journey from its origin to its current state, including all transformations and processes it has undergone. This transparency is vital for auditing purposes, regulatory compliance and troubleshooting data issues.

Furthermore, metadata management tools support the creation and maintenance of data dictionaries and business glossaries. Data dictionaries provide detailed descriptions of data elements, including their formats, allowable values and relationships. Business glossaries, on the other hand, define business terms and concepts in a consistent manner, ensuring that all users have a common understanding of the terminology used within SQAT.

By integrating with other data governance tools, metadata management tools enhance overall data quality and governance efforts. They enable automated metadata capture from various data sources, reducing manual effort and ensuring up-to-date metadata. Additionally, these tools often come with

features for metadata versioning and change tracking, which are essential for managing metadata over time and maintaining historical records.

7.4 Data Security Tools

In the realm of data governance, ensuring the security of data is paramount. Data security tools are essential for protecting sensitive information from unauthorised access, breaches and other cyber threats. These tools encompass a variety of technologies and practices designed to safeguard data throughout its lifecycle.

Encryption is one of the cornerstone technologies in data security. It involves converting data into a coded format that is unreadable without the proper decryption key. This ensures that even if data is intercepted or accessed without authorization, it remains unintelligible and useless to the intruder. Encryption is particularly crucial for protecting data at rest (stored data) and data in transit (data being transmitted across networks).

Access controls are another vital component of data security. These controls regulate who can view or use resources in a computing environment. They are typically enforced through authentication and authorization mechanisms. Authentication verifies the identity of users attempting to access the system, often through passwords, biometrics, or multi-factor authentication. Authorization, on the other hand, determines what authenticated users are allowed to do, ensuring that they can only access data necessary for their roles.

Data masking is a technique used to hide sensitive data by obscuring it with modified content. This process allows organisations to use realistic but fictitious data in non-production environments, such as during testing or development, without exposing actual sensitive information. By masking data, organisations can prevent unauthorised users from accessing critical data while still performing necessary operations.

Additionally, intrusion detection and prevention systems (IDPS) play a crucial role in identifying and mitigating security threats. These systems monitor network traffic for suspicious activity and can take actions to prevent potential breaches. IDPS tools help organisations detect and respond to unauthorised access attempts, malware infections and other security incidents in real time.

Moreover, implementing robust data security policies and training programs is essential. Security policies provide a framework for managing and protecting data, outlining the responsibilities and procedures for maintaining security. Regular training and awareness programs ensure that employees understand the importance of data security and are equipped to recognize and respond to potential threats.

8 Monitoring and Reporting

8.1 KPIs

Effective data governance requires continuous monitoring and evaluation to ensure that data management practices align with organisational goals and standards. KPIs serve as critical metrics for

assessing the success and efficiency of data governance initiatives. These KPIs provide insights into various aspects of data quality, security, compliance and overall governance effectiveness, enabling stakeholders to make informed decisions and drive improvements.

One of the primary KPIs in data governance is data accuracy. This metric measures the correctness and reliability of the data collected and used by SQAT. High accuracy levels indicate that the data is free from errors and can be trusted for analysis and decision-making. Regular audits and validation processes are essential to maintain and improve data accuracy, ensuring that users have confidence in the results generated by the tool.

Another crucial KPI is data completeness. This indicator evaluates whether all necessary data is captured and available for use. Incomplete data can lead to incorrect conclusions and suboptimal decisions, particularly in soil quality analysis where comprehensive datasets are vital. Monitoring data completeness involves checking for missing values, ensuring that all required fields are populated and verifying that the data covers the intended scope.

Data consistency is also a key KPI, focusing on the uniformity of data across different sources and systems. Consistent data ensures that there are no discrepancies when data from multiple sources is integrated or compared. This KPI is vital for maintaining the integrity of analysis results and supporting seamless data integration within SQAT. Regular consistency checks and standardisation practices help achieve and sustain high levels of data consistency.

Timeliness is another important KPI, reflecting how up-to-date the data is. Timely data is crucial for accurate soil quality analysis and timely decision-making. This KPI measures the lag between data collection and its availability for use. Shorter lag times indicate that the data governance processes are efficient and responsive to the needs of the users. Implementing real-time or near-real-time data collection methods can significantly improve the timeliness of data.

In addition to these data-specific KPIs, the overall effectiveness of data governance processes can be measured through KPIs such as compliance rates and incident response times. Compliance rates assess adherence to data governance policies, regulatory requirements and best practices. High compliance rates demonstrate that the organisation is effectively managing its data assets and minimising risks associated with data misuse or breaches. Incident response times, on the other hand, measure the efficiency of the processes in place to handle data-related incidents. Quick response times indicate robust incident management protocols and the organisation's capability to mitigate potential data risks swiftly.

8.2 Reporting Mechanisms

Effective data governance requires robust reporting mechanisms to ensure transparency, accountability and continuous improvement. These mechanisms provide insights into the performance of data governance policies, highlight areas for improvement and ensure compliance with regulatory requirements.

To begin with, regular data governance reports should be generated and distributed to relevant stakeholders. These reports should include KPIs that measure the effectiveness of data governance

activities. Examples of KPIs include data accuracy rates, the number of data incidents, compliance with data privacy regulations and the success of data quality improvement initiatives. By tracking these metrics, organisations can identify trends, assess the impact of data governance strategies and make informed decisions to enhance data management practices.

Moreover, automated reporting tools can play a significant role in streamlining the reporting process. These tools can collect data from various sources, perform real-time analysis and generate comprehensive reports with minimal manual intervention. This not only saves time but also reduces the likelihood of human error. Automated reports can be scheduled to run at regular intervals, ensuring that stakeholders receive timely updates on the state of data governance.

Figure 9 | Example dashboard view for SQAT KPIs.

In addition to automated tools, dashboards provide an interactive and visual way to monitor data governance performance (*Fig. 9*). Dashboards can display real-time data on KPIs, allowing stakeholders to quickly identify issues and take corrective actions. They can be customised to show different views for different roles, ensuring that each stakeholder has access to the information they need. For example, data stewards might focus on data quality metrics, while IT personnel might prioritise data security indicators.

Furthermore, incident reporting systems are crucial for managing data-related issues. These systems enable users to report data incidents, such as breaches or quality issues, in a structured manner. Incident reports should be logged, categorised and tracked until resolution. This process helps organisations respond promptly to data issues, mitigate risks and prevent future occurrences. Analysing incident reports over time can also reveal patterns and systemic issues that require attention.

Finally, regular audits and reviews of data governance practices should be conducted to ensure ongoing compliance and effectiveness. Independent audits can provide an objective assessment of data governance processes and identify areas for improvement. Internal reviews, on the other hand, can involve stakeholders from different departments to evaluate the practical implementation of data governance policies and procedures.

8.3 Continuous Improvement

Continuous improvement ensures that data management practices evolve and adapt to changing needs and challenges. Continuous improvement involves regular evaluation and enhancement of data governance processes, policies and tools to maintain high standards of data quality, security and usability.

To foster continuous improvement, it is essential to establish a culture that values feedback and encourages proactive identification of areas for enhancement. This begins with regular audits and assessments of current data governance practices to identify gaps and opportunities for improvement. By systematically evaluating data quality, security measures, compliance with regulations and user satisfaction, SQAT can uncover weaknesses and implement corrective actions promptly.

Incorporating feedback from various stakeholders, including data stewards, custodians and end users, is crucial for continuous improvement. Regularly scheduled meetings, surveys and feedback mechanisms can help gather insights into the effectiveness of current practices and highlight areas that need attention. Engaging stakeholders in this process ensures that the data governance framework remains relevant and responsive to the needs of all users.

Another key element of continuous improvement is staying abreast of technological advancements and industry best practices. As new tools and methodologies emerge, they should be evaluated for their potential to enhance SQAT's data governance framework. This might involve adopting new data quality monitoring tools, enhancing data security protocols, or implementing advanced data analytics techniques to derive more value from the data.

Training and education also play a vital role in continuous improvement. Providing ongoing training for all personnel involved in data governance ensures that they are well-equipped with the latest knowledge and skills. This can help in maintaining high standards of data management and fostering a sense of ownership and accountability among staff.

Finally, setting measurable goals and tracking progress is essential for continuous improvement. Establishing KPIs related to data quality, security and compliance allows SQAT to monitor improvements over time and make data-driven decisions. Regularly reviewing these metrics and adjusting strategies as needed ensures that the data governance framework remains dynamic and effective.

9 Challenges and Solutions in Data Governance

9.1 Common Challenges

Implementing data governance in any organisation comes with its set of challenges, and SQAT is no exception. One of the primary challenges is the existence of data silos. These silos occur when data is isolated within different departments or systems, making it difficult to achieve a unified view of information. This fragmentation can lead to inconsistencies and redundancy, ultimately hampering the decision-making process. Additionally, regulatory compliance presents a significant challenge. With the increasing complexity of data privacy laws and regulations such as GDPR, ensuring that all data management practices are compliant can be daunting. This is compounded by the need to balance compliance with operational efficiency. Another critical challenge is the resistance to change. Implementing new data governance policies often requires altering established workflows and practices, which can be met with resistance from employees accustomed to existing processes. This resistance can slow down the adoption of governance practices and reduce their effectiveness.

Furthermore, ensuring data quality is a persistent challenge. Data quality issues such as inaccuracies, incompleteness, and inconsistencies can undermine the trustworthiness of the data, leading to flawed analysis and decisions. Addressing these issues requires robust data quality management practices, which can be resource-intensive. Data security is another significant concern. Protecting sensitive data from breaches and unauthorised access is paramount, but it requires continuous vigilance and investment in security technologies and practices. Moreover, there is often a lack of clearly defined roles and responsibilities related to data governance. Without clear ownership and accountability, governance initiatives can flounder, leading to gaps in data management and oversight.

Additionally, combining data from multiple organisations and partners introduces another layer of complexity. This is particularly relevant for the consortium involved in research and innovation activities during the project, as well as for the consortium for service provision after the project. Different organisations may have varying data standards, formats and governance practices, making it challenging to integrate data seamlessly. These differences can lead to discrepancies and misalignment, complicating efforts to achieve a cohesive data governance strategy. Effective coordination, harmonisation and standardisation of data practices across all partner organisations are crucial to overcoming this challenge.

Lastly, the challenge of culture change should not be underestimated. Effective data governance requires a cultural shift within the organisation, where data is viewed as a valuable asset, and everyone understands their role in managing and protecting it. This culture change requires strong leadership, ongoing education and engagement across all levels of each organisation. Building a data-centric culture ensures that data governance principles are embraced and practised consistently, thereby enhancing the overall effectiveness of data management efforts.

9.2 Best Practices and Solutions

Implementing effective data governance can be challenging, but adopting best practices and proven solutions can significantly enhance the management and utilisation of data (*Table 2*). One fundamental

best practice is establishing a clear and comprehensive data governance framework. This framework should define roles, responsibilities and processes for data management across the organisation. By creating a structured approach, organisations can ensure accountability and consistency in handling data, which is critical for maintaining data quality and integrity.

Another best practice is fostering a culture of data stewardship within the organisation. Data stewardship involves designating individuals or teams responsible for overseeing data assets, ensuring data quality and managing data-related activities. Data stewards play a crucial role in implementing data governance policies, monitoring compliance and promoting best practices. Encouraging a sense of ownership and responsibility among data stewards helps in maintaining high standards of data governance and drives continuous improvement.

Utilising advanced technology and tools is also essential for effective data governance. Data governance platforms provide comprehensive capabilities for managing data assets, including data quality monitoring, metadata management and data security. These platforms enable organisations to automate and streamline data governance processes, making it easier to enforce policies and track compliance. Additionally, leveraging data quality tools can help identify and rectify data inaccuracies, ensuring that data used for analysis and decision-making is reliable and accurate.

Data security is another critical aspect of data governance and implementing robust security measures is a best practice that cannot be overlooked. This includes establishing access controls to restrict data access to authorised users, employing encryption to protect sensitive data and using data masking techniques to hide confidential information. Regular security audits and assessments can help identify potential vulnerabilities and ensure that data protection measures are up to date.

Moreover, addressing data privacy concerns is essential for compliance with regulations and building trust with stakeholders. Best practices in data privacy include adhering to regulatory requirements, such as GDPR, implementing anonymization techniques to protect personally identifiable information and managing user consents effectively. By prioritising data privacy, organisations can mitigate risks and ensure that data handling practices align with legal and ethical standards.

Consequence	Description	Mitigation Strategy
Data Inaccuracy	Inaccurate data can lead to poor decision-making, affecting soil management practices and crop yields.	Implement regular audits and validation checks to ensure data accuracy.
Data Breaches	Unauthorised access to sensitive data can compromise privacy and lead to financial and reputational damage.	Adopt robust security measures like encryption and access controls.
Non-compliance with Regulations	Failing to comply with regulations like GDPR can result in legal penalties and loss of stakeholder trust.	Stay updated with regulatory requirements and conduct regular compliance checks.
Data Silos	Isolated data systems can hinder data sharing and collaboration, reducing the effectiveness of the data governance framework.	Promote data interoperability and integration across systems.
Resistance to Change	Resistance from stakeholders can slow down the adoption of new data governance practices, impacting overall project	Foster a culture of data stewardship and provide training to stakeholders.

Table 2 | Mapped mitigation strategies for SQAT challenges.

Continuous improvement is a key principle in data governance. Organisations should establish mechanisms for monitoring and reporting on data governance performance, including key performance indicators KPIs that measure the effectiveness of data management practices. Regular reviews and updates to data governance policies and procedures help in adapting to changing needs and emerging challenges. Engaging stakeholders in feedback processes and fostering a culture of continuous learning and improvement can drive innovation and enhance the overall effectiveness of data governance.

9.3 Foundational Data Governance

The principles developed in this document will be reflected in all development activities and the processes of data gathering, storage and utilisation throughout the project. By embedding these principles into the project's lifecycle, we ensure that data governance is not an afterthought, but an integral part of our operations. This approach allows us to maintain high standards of data quality, security and privacy from the outset, thereby fostering trust and reliability in our data management practices.

During the project, we will draw on lessons learned to refine and enhance our data governance framework. Dedicated exploitation activities in WP6 will focus on setting up service provision that adheres to these data governance principles. This will involve establishing a robust and trustworthy data chain that supports the effective exploitation of project results. By continuously evaluating and improving our data governance practices, we aim to adapt to emerging challenges and leverage new opportunities for better data management.

Ultimately, integrating these data governance practices into every aspect of the project will help build a solid foundation for reliable and secure data management. This foundation will support the project's long-term success and value creation by ensuring that data handling practices are robust, compliant and aligned with both organisational goals and regulatory requirements. Through this integrated approach, we can maximise the value derived from our data assets, drive innovation and deliver superior outcomes for all stakeholders involved.

End of document