
# Climate-Resilient Development Pathways in Metropolitan Regions of Europe

## Deliverable 4.1 – Data Model and Data Management - Interim

February 2025



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

Grant Agreement Number 101137851

| | |
|---|---|
| Deliverable No: | 4.1 |
| Deliverable nature: | Report |
| Work Package and Task: | 4.1 |
| Dissemination level: | PU - Public |
| Authors: | Andrés Carrancá (UFR)<br>Support: Moritz Wussow, (UFR), Dirk Neumann (UFR) |
| Reviewers: | Sorin Cheval, Dana Micu, Laurentiu Ciuhu (MeteoRo)<br>Iphigenia Keramitsoglou (NOA)<br>Emma Ferranti (University of Birmingham)<br>Mirabela Marin, Cezar Ungurean (INCDS)<br>Christos Spyrou (AoA), Marianna Adinolfi (CMCC)<br>Iuliana Parvu (CNC)<br>Steffi Urhausen (MOHC)<br>Hamza Kechiche, Gaelle Descloitres (SDSN) |
| Contractual submission date: | 28 February 2025 |
| Actual submission date: | 5 February 2025 |



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

## Technical References

| | |
|---|---|
| Project title | Climate-Resilient Development Pathways in Metropolitan Regions of Europe |
| Project acronym | CARMINE |
| Grant agreement number | 101137851 |
| Instrument | HORIZON EUROPE |
| Call | HORIZON-CL5-2023-01 |
| Starting date of the project | 01 February 2024 |
| Project duration | 48 Months |

## Document history

| Version | Date | Description |
|---|---|---|
| 1.1 | 17-12-2024 | First draft sent for review |
| 1.2 | 23-01-2025 | Second draft sent for review |
| 1.3 | 28-01-2025 | Final review |
| 1.4 | 05-02-2025 | Final draft with layout review |

## Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. Every effort has been made to ensure that all statements and information contained herein are accurate, however the CARMINE Project Partners accept no liability for any error or omission in the same.

UK participants in this project are co-funded by UK Research and Innovation (UKRI).

The work of Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

# Table of Contents





Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

## Table of Figures





Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

# Abbreviations

**API**: Application Programming Interface
**CKAN**: Comprehensive Knowledge Archive Network
**DCAT**: Data Catalog Vocabulary
**DB**: Database
**DT**: Digital Twin
**DOI**: Digital Object Identifier
**ETL**: Extract, Transform, Load
**FAIR**: Findable, Accessible, Interoperable, Reusable
**FDM**: Federated Data Management
**FDMS**: Federated Data Management Solution
**FTP**: File Transfer Protocol
**GDPR**: General Data Protection Regulation
**HIPAA:** Health Insurance Portability and Accountability Act
**IDSS**: Impact-Based Decision Support System
**JDBC/ODBC**: Java Database Connectivity/Open Database Connectivity
**JSON**: JavaScript Object Notation
**JSON-LD**: JavaScript Object Notation for Linked Data
**KPI**: Key Performance Indicator
**LFDM**: Lightweight Federated Data Management
**LOD**: Linked Open Data
**MiM**: Minimum Interoperability Mechanisms
**OIDC**: OpenID Connect
**RDF**: Resource Description Framework
**REST**: Representational State Transfer
**REST API**: REST-based Application Programming Interface
**SPARQL**: SPARQL Protocol and RDF Query Language
**SQL**: Structured Query Language
**UI/UX**: User Interface/User Experience
**UFR**: University of Freiburg



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

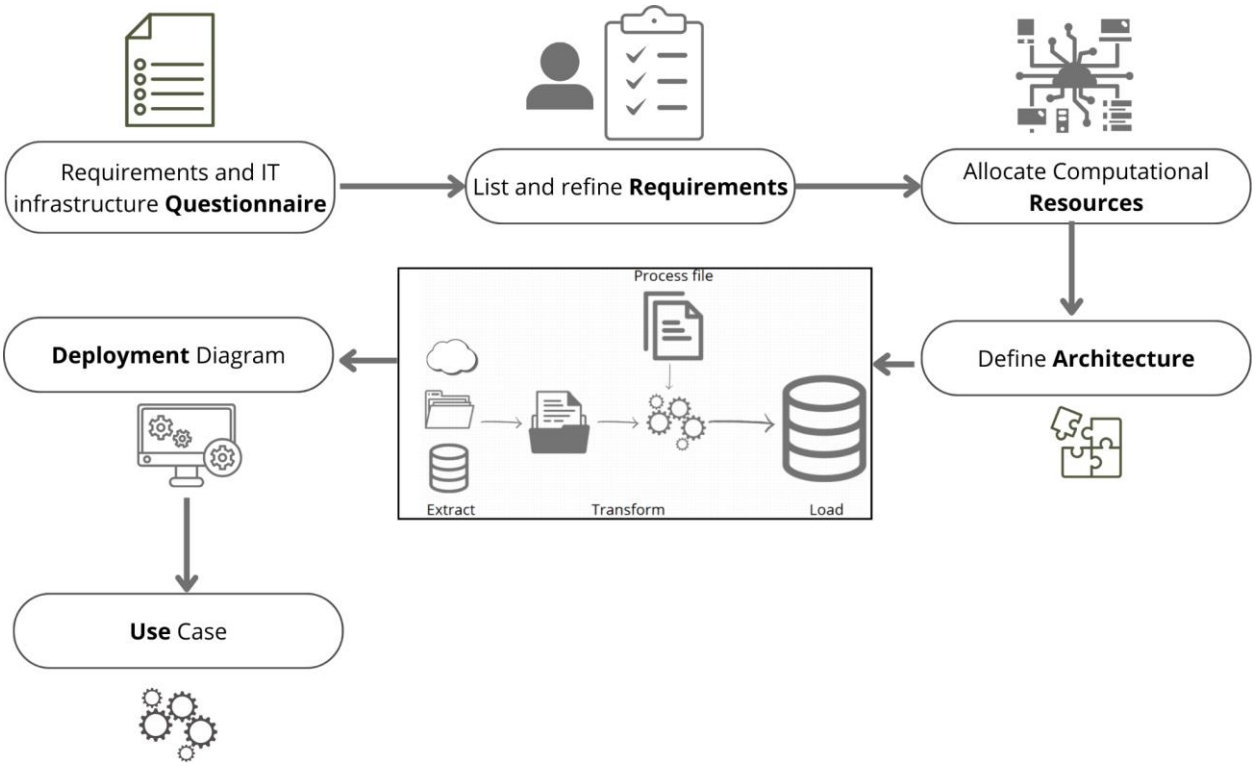
Co-funded by
UK Research and Innovation

**UML**: Unified Modeling Language
**URI**: Uniform Resource Identifier
**URL**: Uniform Resource Locator
**XML**: Extensible Markup Language



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

# Executive Summary




For the **CARMINE Project's** objective of co-creating and co-developing decision-support services and guidelines for enhanced resilience and adaptive capacity, including early warning and disaster risk management systems, it is important to establish a robust and scalable **Federated Data Management Solution (FDMS)** to support collaborative data sharing and analysis across diverse stakeholders. By integrating lightweight federated data management concepts, the project facilitates efficient data harmonization and accessibility while ensuring compliance with regulatory requirements, such as **GDPR**. This report outlines the project's objectives, methodology, and proposed architecture for the FDMS.

The proposed FDMS architecture is the result of investigative work that addresses the needs of the project partners, incorporating features such as metadata repositories, data harmonization processes, and integrations with the **CARMINE Scheme** to ensure interoperability across systems. The architecture emphasizes **batch data access**, as real-time processing is beyond the project's current scope due to infrastructure limitations. However, Real-Time or Near-Real-Time data ingestion will be ultimately dependent on each partner's data infrastructure as they are the ones hosting the data by leveraging **ETL workflows**, the system will process and standardize data periodically to support decision-making and analytical needs.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

This report also addresses compliance, scalability, and interoperability challenges by adopting **open standards**, **minimum interoperability mechanisms (MiM)**, and **modular workflows**. The CARMINE FDMS is positioned as a cost-effective, adaptable solution to enhance data sharing while meeting the technical, security, and regulatory needs of project partners.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

# 1. Approach

In the context of federated data management (FDM), the design of a lightweight Federated Data Management Solution (FDMS) requires careful alignment with the diverse needs and infrastructure of its users and stakeholders. This project aims to develop an FDMS capable of integrating data from various sources hosted by project partners while meeting fundamental requirements of modularity, scalability, and efficiency. To ensure the system aligns with the expectations and constraints of these partners, a systematic requirement-gathering process was undertaken.

To identify the specific needs and technical constraints of the partners, we designed and distributed a detailed questionnaire. The responses provided valuable insights into their existing systems, expected functionalities, and technical limitations. Following the survey, meetings with selected partners were conducted to clarify their answers, ensuring a shared understanding of the project goals and refining the gathered data. This collaborative approach allowed us to iteratively polish the FDMS requirements and define a roadmap for its development.

Using the refined requirements, an initial architecture was proposed. This architecture, while subject to change as the project evolves, serves as a foundational blueprint to guide the development of the system. Furthermore, a deployment diagram was derived from this architecture to visualize how the various components and data flows would interact in a real-world implementation.

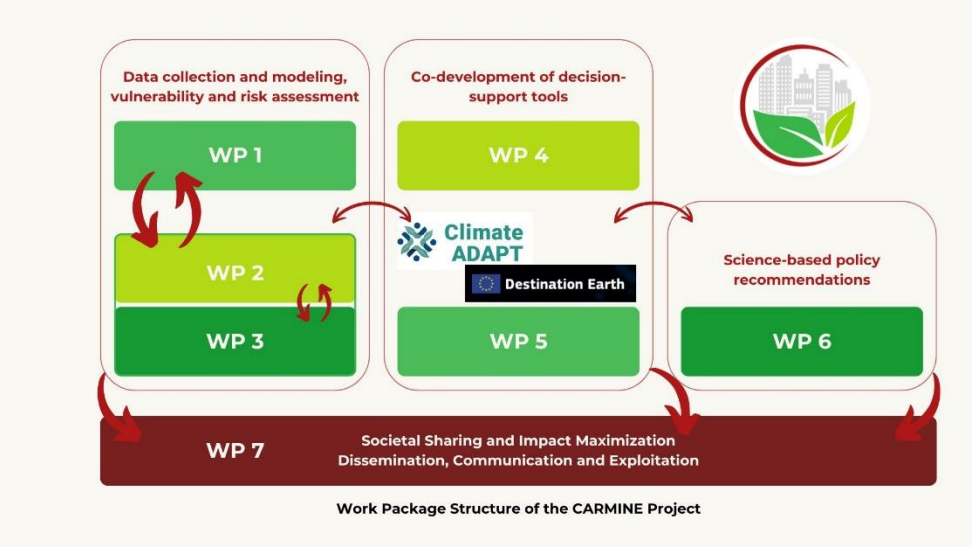



*Figure 1: WPs in CARMINE project*

It can be seen from the image where WP 4 finds itself amidst the overall CARMINE project. It has a close connection with WP5, where the Digital Twins (DT) are developed and with



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

DestinE, an initiative by ESA, ECMWF and EUMETSAT to build a digital model of the earth (https://destination-earth.eu/).

## 1.1 Requirement Analysis

This section outlines the systematic approach for co-developing a Federated Data Management Solution (FDMS) tailored to the requirements of the CARMINE project. The FDMS is designed to serve as the backbone of data harmonization, storage, and analytics, ensuring GDPR compliance and interoperability across various data sources and systems. The system must effectively support the project's overarching goals by integrating data from diverse stakeholders, including DT developers and stakeholders involved in Living Labs, while fostering collaboration across the federation.

A key focus of the requirement analysis is effective communication with project partners to understand their data needs, infrastructure capabilities, and expectations from the FDMS. The system must strike a balance between what the University of Federal Research can produce and host and what the partners require for their workflows. Partners rely on the FDMS to provide seamless data access, federated query capabilities, and robust security measures to facilitate the collaborative aspect of the project. These interactions help ensure that the FDMS aligns with partner needs while meeting the technical and operational constraints of the project.

The analysis begins by evaluating relevant standards, such as GDPR, to ensure compliance with data protection laws and regulations. This is especially critical for federated systems where data sovereignty and cross-border data sharing are key concerns. Additionally, a thorough review of existing frameworks and architectures is conducted to identify reusable components and best practices. This analysis is presented to the partners for feedback, ensuring that the system design aligns with current standards and their specific requirements.

For the FDMS to succeed, it must address key requirements that define its functionality and scope. Data harmonization is crucial, ensuring the system can standardize and integrate information from diverse, heterogeneous sources using common schemas and ontologies. Interoperability is another vital element, enabling seamless communication between distributed nodes through the use of open standards and protocols. Security and compliance mechanisms must be robust, with features such as encryption, role-based access control, and audit trails protecting sensitive data and ensuring adherence to regulatory requirements like GDPR.

Scalability and performance are essential, as the system must support high-volume data flows and complex queries across distributed systems without significant latency. Metadata management also plays a critical role, with centralized metadata catalogs enabling data discoverability and lineage tracking while promoting FAIR principles (Findable, Accessible, Interoperable, Reusable). The FDMS must be lightweight and modular in its design, minimizing infrastructure overhead and allowing for flexibility in adapting to future needs and upgrades.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

By addressing these core requirements, the FDMS will not only meet partner expectations but also enable advanced functionalities such as federated querying, real-time data access, and robust data governance. This requirement analysis provides the foundation for an iterative co-development process that ensures the system is both functional and adaptable to the evolving needs of the project.

## 1.2 Questionnaire Design and Deployment

The complete survey can be found in Annex 1.

After distributing the questionnaire to the responsible partners, the responses were collected, analyzed, and used to refine the pre-established requirements. Additionally, meetings were held with the partners to delve deeper into their specific needs and expectations. These discussions allowed for a more detailed understanding of their requirements, enabling further refinement and establishing the precise specifications that the FDMS must fulfill.

The partners that participated in the questionnaire are involved in WP4 and WP5, they are responsible for the development of the DIgital Twins which will use and produce data that will be part of the FDMS and also for the IDSS, the list of the partners answering the questionnaire is as follows:

- Primăria Municipiului Braşov - Serviciul TI
- Met Office UK
- Meteomatics AG
- Academy of Athens
- National Observatory of Athens / BEYOND
- Ruhr University Bochum
- Mitiga
- Technical University of DenmarkUniversity of Bucharest

## 1.3 Survey Results

The survey recorded the participation of up to 20 persons from said institutions. Here are the most important answers shown in the survey.

### *1.3.1 Defining the Output of the FDMS*

The purpose of this section is to clarify what the output of the FDMS should look like. Understanding the intended outcomes of the FDMS is critical for tailoring the system to meet user needs effectively. The questions in this subsection aim to capture the primary objectives of the FDMS, the expected data access requirements, and the frequency of data processing.

**What are your primary objectives for using the FDMS?**



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

This question seeks to identify the primary goals and motivations behind implementing the FDMS. For example, whether the objective is to enable better decision-making, enhance data sharing among partners, or improve data governance. The answers provide a high-level view of what the consortium partners expect from the system.

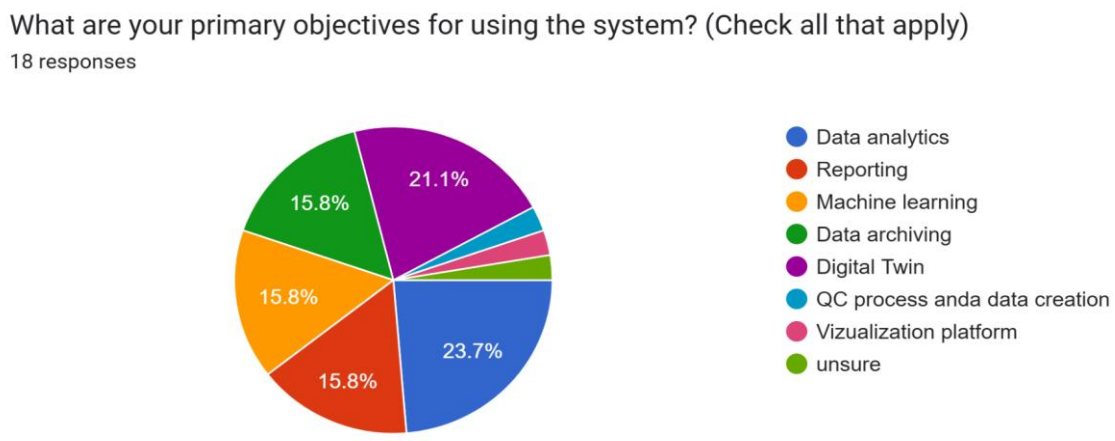



*Figure 2: Statistical diagram for primary objectives of FDMS*

**What are your data access requirements?**

This question helps determine how users plan to access and interact with the data stored in the FDMS. As noted, Real-Time access is excluded from the FDMS due to its high infrastructure demands. Instead, the focus is on batch data access, where data is processed in scheduled intervals.

- Batch Access: This method processes data in chunks over specific periods, which is more resource-efficient and fits the scope of the FDMS.
- Real-Time Access: Although not a part of the FDMS, follow-up discussions may be conducted to evaluate whether certain requests truly require real-time capabilities or if batch processing suffices.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

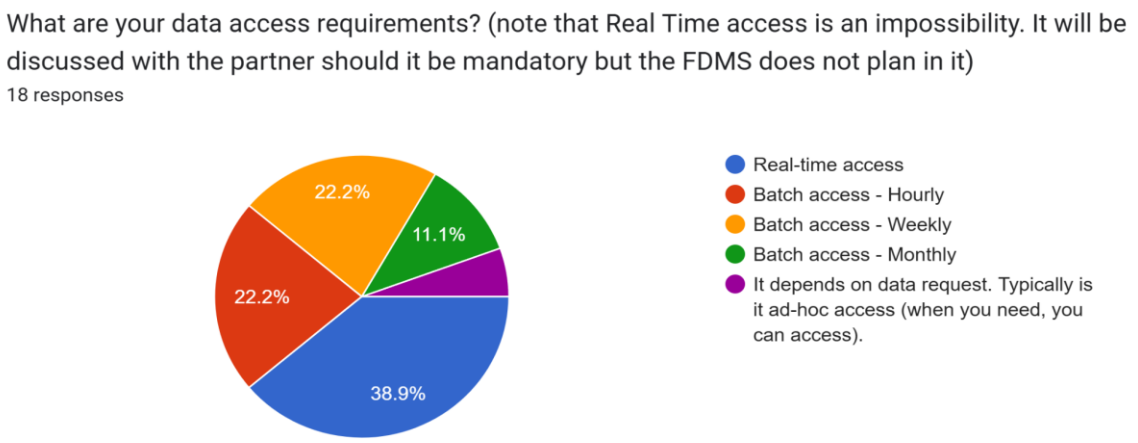



*Figure 3: Data access requirements*

Insights from This Analysis:

The responses to these questions help determine:

- How often data should be processed (e.g., daily, weekly, or monthly).
- Whether the data output is intended for reporting, integration, or analysis.
- The specific access patterns and intervals needed by stakeholders.

By combining this information with other system requirements, the FDMS can be fine-tuned to deliver outputs that align with organizational objectives while optimizing resources.

The survey has shown that most of the partners consider Real-time access as something that they would be interested in having implemented. Therefore, further communication with them to align the scope and capabilities of the project and their needs was needed. Real-time access is not needed for the purpose of the FDMS, but hourly batch access will suffice for training models and test outputs. There are multiple objectives for using the data processed by the FDMS, thus, flexibility is a key concept of the FDMS. The FDMS should provide access to data in a detailed and consistent way so it can be used for all its purposes. It is also important to have a consistent naming of variables throughout the project as data will be shared among multiple partners and multiple fields of study.

### *1.3.2 Data Sources being used*

When assessing an organization's data storage infrastructure, it is important to understand the types of systems that are currently used to manage and store data effectively. This question aims to identify the specific storage solutions employed, such as **Data Warehouses**, **Data Lakes**, **Databases**, or more ad-hoc methods like **project-based storage**. These systems serve different purposes depending on the nature of the data, its volume, and the



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

organization's operational and analytical needs. By understanding which of these systems are in place, we can evaluate how data is stored, accessed, and utilized to support business objectives.

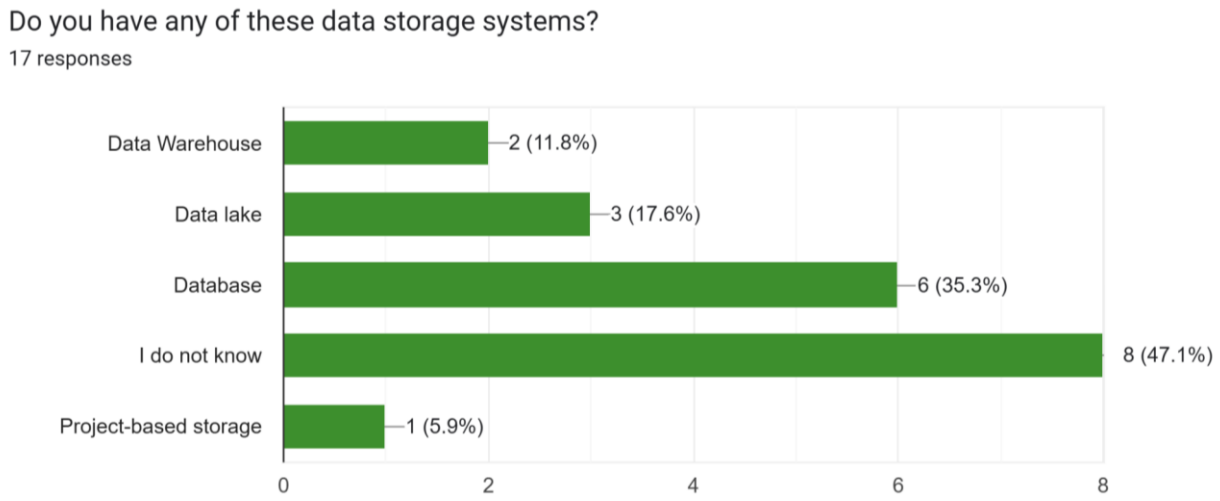



*Figure 4: Available storage system*

This question helps understand how the FDMS connects with the different data sources.

### *1.3.3 Assessing Storage Needs for CARMINE Data*

This subsection focuses on understanding the storage requirements and capabilities of the partners to host CARMINE data. By evaluating the expected data volume, growth rate, and available storage capacity, we can make informed decisions about where and how to host the data within the FDMS.

**What is the expected volume of CARMINE data you will store in the system?** This question helps categorize the anticipated size of data contributions using predefined ranges:

- Small (<100GB)
- Medium (100GB - 1TB)
- Large (1TB - 10TB)
- Very Large (>10TB)

Knowing the volume is essential for selecting appropriate storage solutions and ensuring the infrastructure can handle the load.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

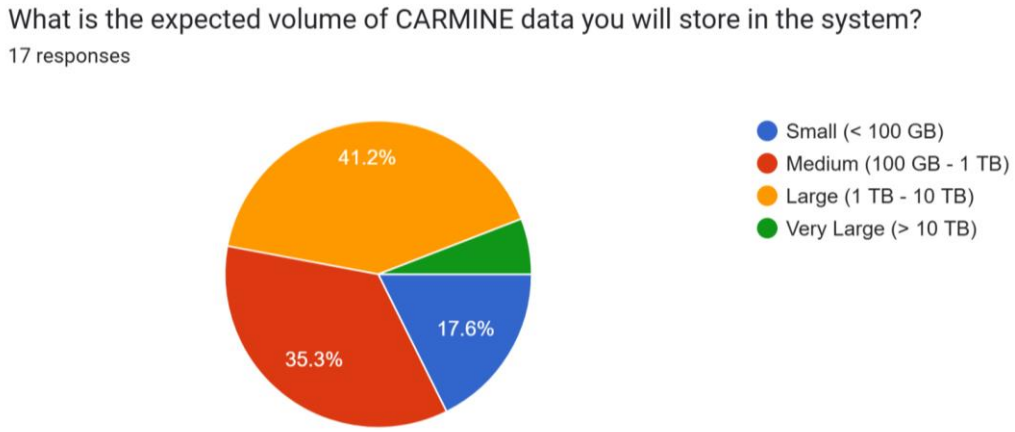



*Figure 5: Expected data volume stored in FDMS*

**What is the expected growth rate of your data volume?**

This question gauges how quickly the stored data is expected to grow, using categories such as:

- Slow (0-10%)
- Moderate (10-30%)
- Fast (30-50%)
- Very Fast (>50%)

Understanding the growth rate ensures the storage system can scale to accommodate future needs without requiring frequent reconfiguration.

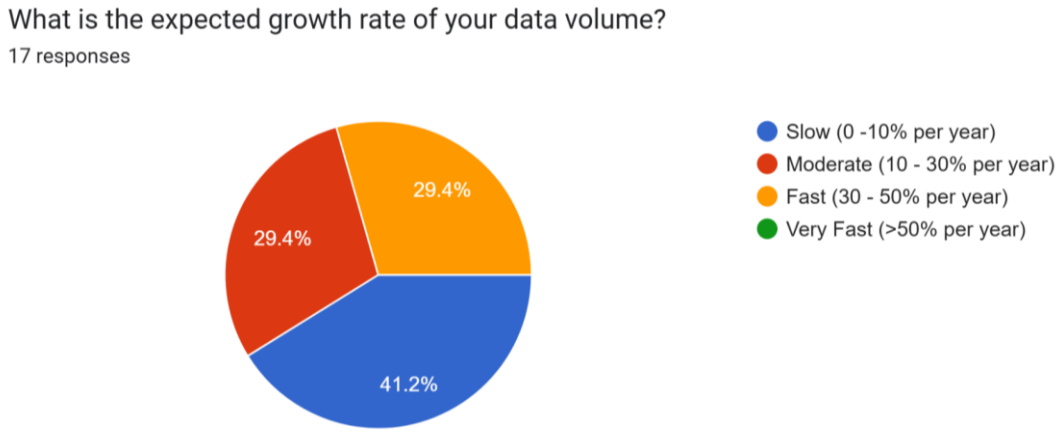



*Figure 6: Growth rate of data volume*

**Do you have sufficient storage capacity for CARMINE data?**



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

This question assesses whether participants currently have enough storage to accommodate their share of the data. Responses indicate whether they:

- Have enough space for only their part of the project.
- Have additional capacity to support other data contributions.
- Lack of sufficient storage for their own data needs.

This helps identify gaps in storage availability across partners.

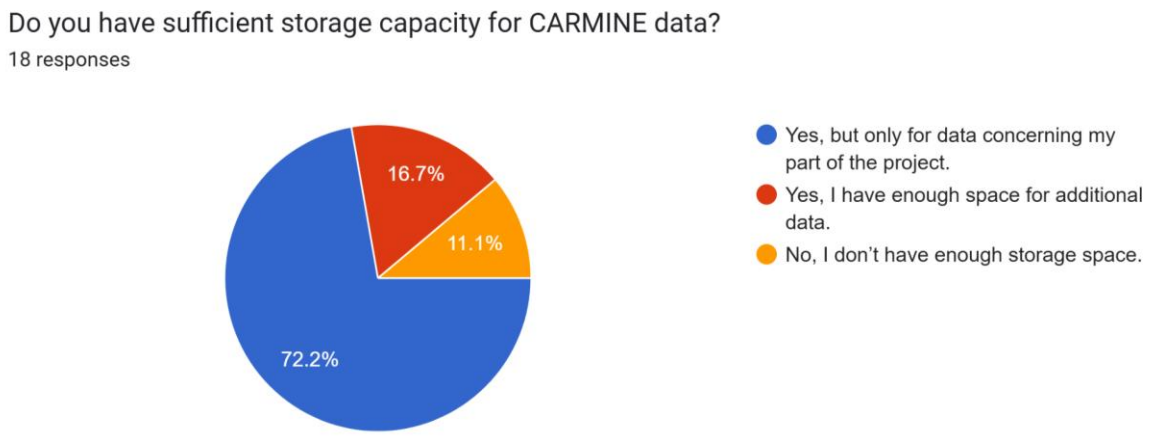



*Figure 7: Storage capacity*

**If you have additional storage capacity, how much free space do you have?** For those with excess storage, this question quantifies the available free space using predefined ranges:

- <100GB
- 100GB - 1TB
- 1TB - 10TB
- >10TB
- No additional space

These responses guide decisions on where to allocate CARMINE data, leveraging available capacity while avoiding overburdening individual stakeholders.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

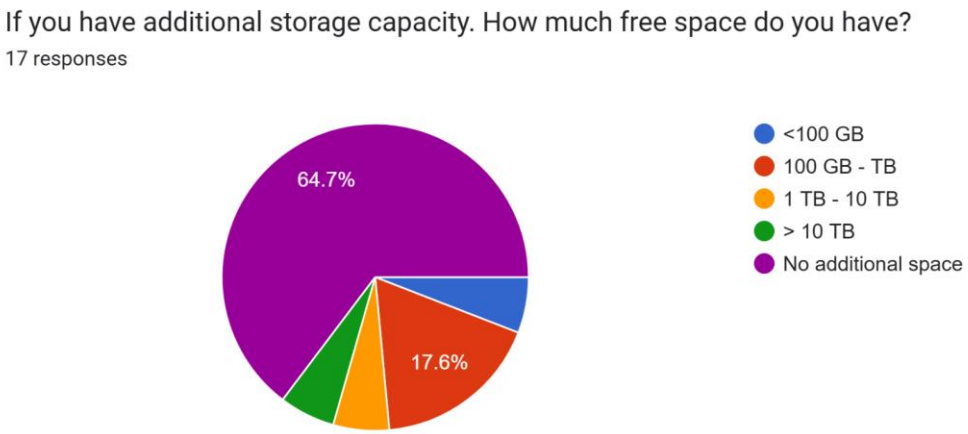



*Figure 8: Additional storage capacity*

By analyzing these responses, we can estimate the total storage requirements for the CARMINE data across the FDMS, ensuring that the system can accommodate the collective data needs of all stakeholders. Additionally, the responses help identify partners who have the capacity to host data or require additional infrastructure support, enabling a more efficient allocation of storage resources. Understanding the expected data growth rates also allows for proactive planning of future storage expansions, ensuring the system remains scalable and avoids potential bottlenecks. This analysis is crucial for designing a balanced and robust storage strategy that meets both current requirements and anticipated future demands of the project.

### *1.3.4 Preferred Methods for Accessing Data*

This question explores the preferred methods for accessing data within the FDMS. By understanding whether stakeholders favor **SQL queries**, **APIs**, **data visualization tools**, or **data file access**, the system can be tailored to provide the most efficient and user-friendly interfaces for interacting with the data. These preferences help determine how to structure the system's data delivery mechanisms, ensuring they align with user workflows and technical requirements. Additionally, this information provides insight into the technical expertise of users, the types of tools they are accustomed to using, and any additional functionality that may be required to support their data access needs. This ensures that the FDMS is designed to optimize accessibility and usability for all stakeholders.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

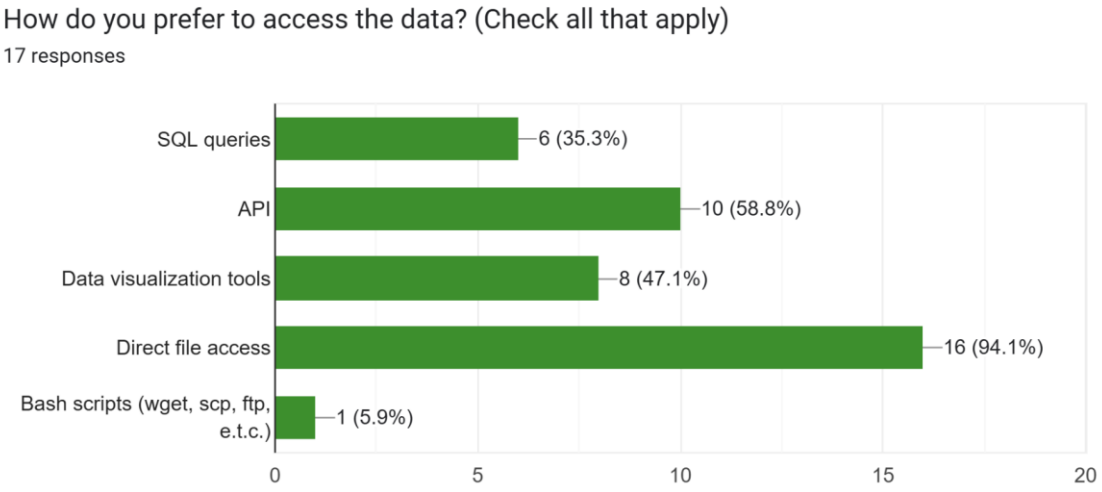



*Figure 9: Data accessing*

### *1.3.5 Compliance Requirements for Data*

This question focuses on identifying any compliance requirements associated with the data managed within the FDMS. Understanding whether stakeholders need to adhere to regulations such as **GDPR**, **HIPAA**, or other standards is critical for ensuring the system is designed to meet legal and ethical obligations. These requirements influence how data is stored, processed, and accessed, as well as the implementation of security measures like encryption and access controls. Additionally, compliance considerations help define the roles and responsibilities of stakeholders in maintaining regulatory standards. Incorporating these requirements into the FDMS ensures the system aligns with legal frameworks and fosters trust among users and partners.

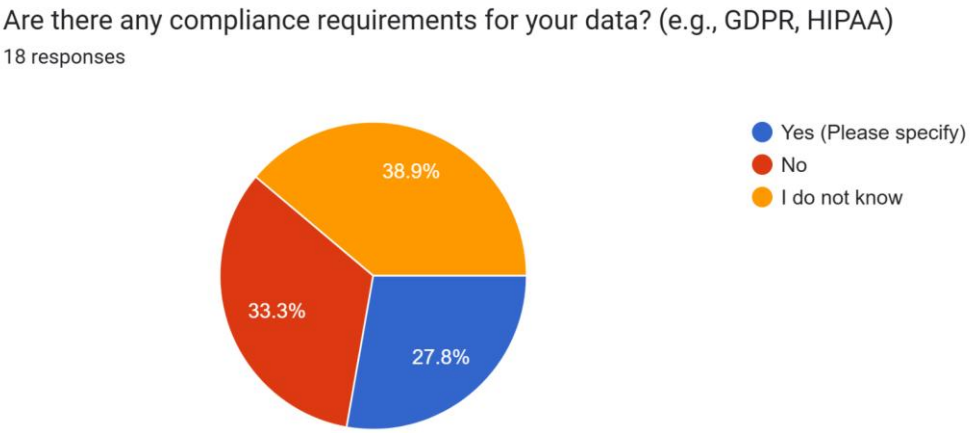



*Figure 10: Compliance requirements - 16 Answers*



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

What are your data security requirements? (Data encryption, access control, Data masking, Audit logging)

### *1.3.6 Data Security Requirements*

This question examines the specific security requirements for the data managed within the FDMS. Understanding whether stakeholders require measures such as **data encryption**, **access control**, **data masking**, or **audit logging** is essential for designing a secure system that protects sensitive information. These requirements help define the necessary safeguards to prevent unauthorized access, ensure data privacy, and maintain a robust audit trail for accountability. By addressing these security needs, the FDMS can be configured to comply with industry best practices and regulatory standards while fostering trust among users and protecting the integrity of the data.

The answers show that the main requirement is access control. So that should be implemented into the FDMS.

### *1.3.7 Refining Requirements*

The refinement of requirements focuses on consolidating survey findings, aligning them with GDPR and interoperability standards, and finalizing the system's specifications. This involves addressing data harmonization needs by identifying common schemas and ontologies to standardize diverse data sources. Metadata is critical, providing details about the structure, relationships, and constraints of the data, such as column names, data types, and table relationships like foreign keys. By leveraging metadata, consistent pre-processing methods for both new and legacy datasets are established, enabling seamless integration into the federated system. Metadata further supports analytics integration, links the FDMS to systems like the IDSS and the Climate-ADAPT Platform, ensures data lineage tracking, and enables interoperable data models. This refinement process results in a comprehensive requirements document and compliance guidelines to ensure the system meets all stakeholder expectations while being secure and interoperable.

<u>Requirements to be Met:</u>

- **Data Harmonization and Standardization**
  - Identify and implement common schemas and ontologies.
  - Use metadata to describe data structures, relationships, and constraints.

- **Metadata Utilization**
  - Ensure consistency in pre-processing new and legacy datasets.
  - Support data discovery through metadata catalogs.
  - Enable lineage tracking to ensure data quality and origin verification.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

- **Integration with Analytics Systems**
  - Connect the FDMS with external systems like IDSS and Climate-ADAPT.
  - Ensure outputs are compliant with geographical and interoperability standards.

- **Compliance and Security**
  - Define encryption protocols, authentication mechanisms, and access controls.
  - Track database changes and ensure GDPR compliance.
  - Align with Minimum Interoperability Mechanism (MiM) guidelines.

- **Data Governance and Accessibility**
  - Provide clear documentation of data organization and usage.
  - Maintain accessibility for authorized users while safeguarding sensitive information.

- **Outcome Deliverables**
  - Comprehensive functional and non-functional requirement specifications.
  - GDPR and MiM-aligned compliance guidelines.
  - Draft architecture proposal based on refined requirements.

Once the requirements that the FDMS must fulfil are agreed, a first draft architecture is proposed.

## 1.4 Use case scenario

A partner requires data from another partner's infrastructure to generate a report. This data is accessed using a Federated Query that connects to the other partner's data system. To properly build the report, the first partner also needs to understand the structure of the received data. For this, they rely on the metadata created and made available for the project.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

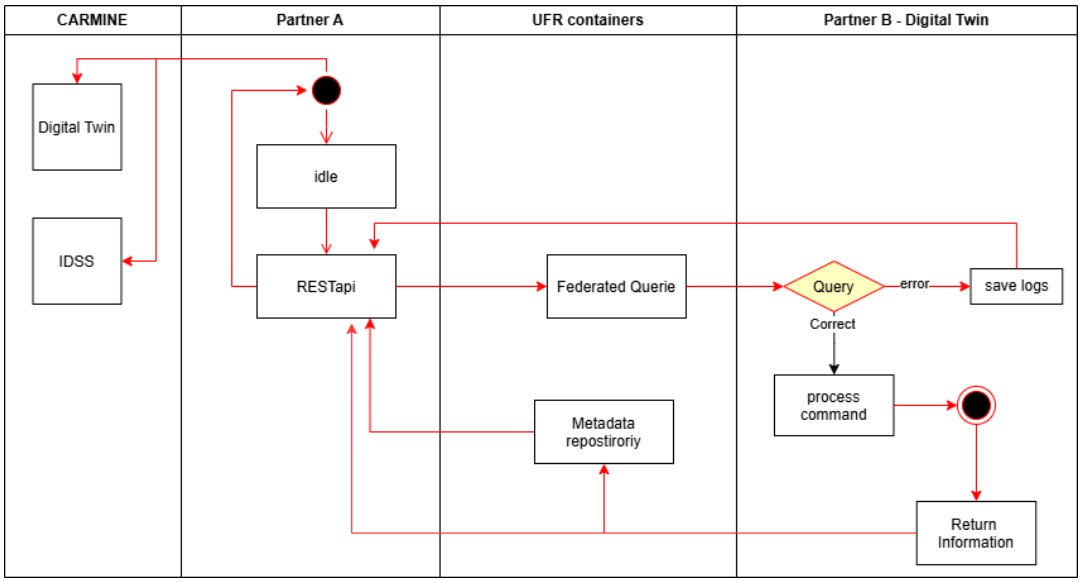



*Figure 11: Use case scenario*

- Partner A uses Federated Queries to access data through the project.
- The project connects to Partner B's data infrastructure to retrieve the required data.
- Partner B provides the data to the project's data handling tools.
- Partner A receives the structured data along with metadata.
- Partner A uses the data and metadata to update its own DT or for the IDSS.

## 1.5 Examples of different data architectures

Data architecture is an abstract discipline where it is important to have the goals and concepts align. There are many different common architectures from which it can be built on.

For the CARMINE project, a basic ETL architecture will be implemented initially. As the project evolves, it will be essential to explore alternative architectures, such as Kappa and Lambda, to accommodate potential future requirements. The decision to adopt a basic ETL approach is driven by the need for simplicity, ensuring clear and structured batch-based extraction and transformation processes. The loading phase, in this context, is primarily focused on making the processed data readily accessible.

### *1.5.1 Basic data warehouse with ETL*

Usually, a data warehouse pulls data from application systems using ETL. That means that the data is pulled by the extract phase handing over to the transformation phase, which cleans the data and ends up in the load phase, thus loading the data into a central repository.

The data is loaded into multiple data marts aimed at specific analytical lines of work.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

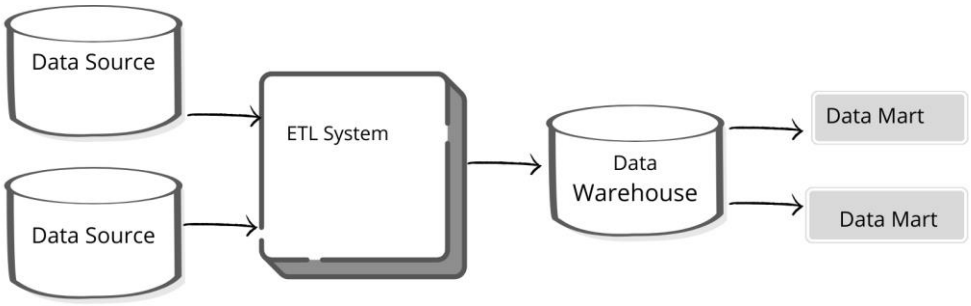



*Figure 12: Use case scenario*

The advantage of such systems is that all the data can be found inside of the centralized data warehouse. It is easier to implement than other methods. It is also ensured that the data handling is done by a data engineer team.

Its main drawbacks also come from its advantages. As it is a centralized data system, it does not yield the most flexibility regarding computational infrastructure. It requires searching big chunks of data for every query, although software has been developed to scan massive amounts of data in parallel allowing high performance.

Another drawback is that since the data is handled by a main centralized team, they might not be as specialized in the data as the team that will use it, hence its transformation and handling tends not to be as quick.

### *1.5.2 ELT - modification for Data Lake*

This is a modification of ETL where Extract takes data from the data sources and directly accommodates it into a data warehouse (or a data lake). This is useful for storing unstructured data. Further Transformations are done pre-Analysis and reporting.

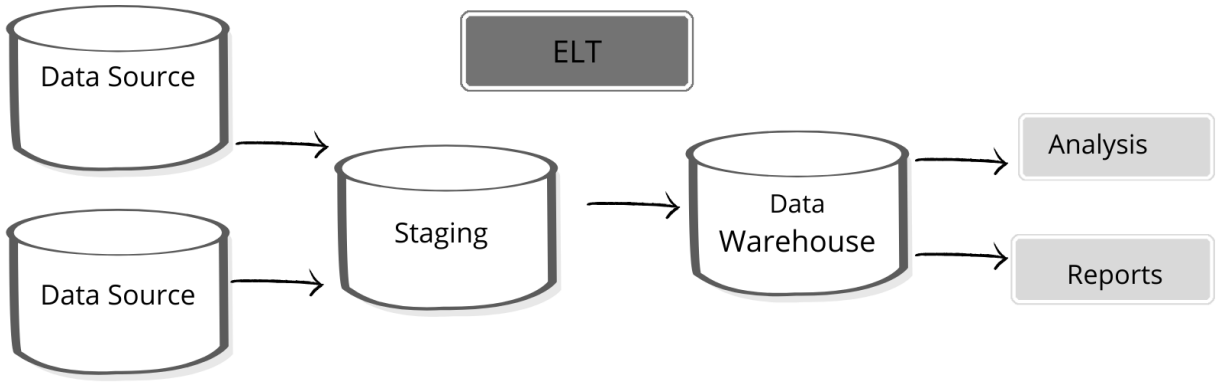



*Figure 13: Data lake*



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

### 1.5.3 Lambda architecture

Lambda architecture integrates batch and streaming data into a unified framework. It operates with three independent layers: batch, stream, and serving. The source system is ideally immutable. The stream processing layer, often referred to as the "speed" layer, is designed for low-latency data handling and typically utilizes a NoSQL database. The serving layer aggregates and queries results from both the batch and stream layers, providing a comprehensive view of the data. However, the main drawback of this architecture is the complexity of managing multiple systems with distinct codebases.

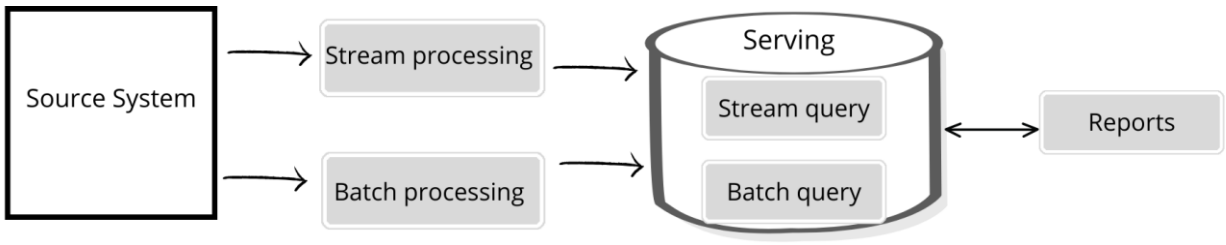



*Figure 14: Lambda architecture*

### 1.5.4 Kappa architecture

This architecture is a second step from lambda architecture. It uses a stream-processing platform for all data handling (ingestion, storage and serving). Thus, it reconciles some extra work done for the lambda architecture.

Real-time and batch processing can be applied to the same data by reading the live event stream directly and replaying large chunks of data for batch processing.

Its main drawback is that in practice it turns out to be more complicated to implement than expected. Also managing everything as a stream is more expensive than batch processing.

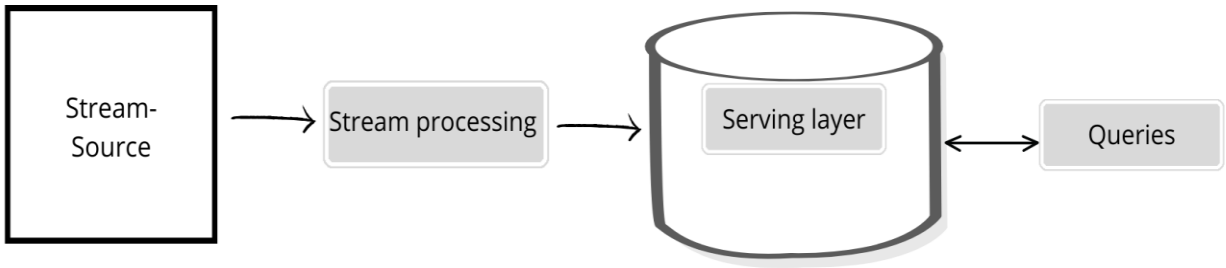



*Figure 15: Lambda architecture*



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

## 1.6 Common Lightweight ETL Architectures

Once the ETL architecture has been selected, it is important to understand its processes and components when it comes to different data sources and data processing. This section explores the most common lightweight ETL architectures, each tailored to specific data processing needs. It includes monolithic ETL setups for simplicity, batch architectures for periodic processing, streaming solutions for real-time requirements, and file-based workflows for straightforward data exchanges. Understanding these architectures is essential for designing ETL pipelines that align with project goals while balancing efficiency and complexity.

### *1.6.1 Monolithic ETL Architecture*

Monolithic ETL architecture relies on a single, centralized process to handle all three stages of the ETL workflow: extraction, transformation, and loading. This architecture is straightforward and typically used for small-scale, uncomplicated data processing tasks. In a monolithic setup, data is extracted from one or more sources, processed using transformation logic—often defined through scripts or tools—and finally loaded into a target system, such as a database or data warehouse. The simplicity of this architecture makes it easy to implement and manage, though it struggles to meet the demands of scalability and flexibility required for larger or more complex datasets.

**Tools and Technologies:**

- Open-Source Examples: Talend Open Studio, Apache Nifi.
- Custom Scripts: Python (using Pandas, NumPy) or SQL-based processing.

**Advantages**:

Monolithic ETL systems are simple to set up and maintain, making them an excellent choice for small-scale ETL tasks with limited complexity. They are particularly useful for projects with minimal data sources and straightforward transformations.

**Limitations:**

The architecture is not designed to scale effectively with growing datasets or diverse sources. Additionally, it offers limited fault tolerance and lacks the flexibility to handle dynamic data processing needs.

### *1.6.2 Batch ETL Architecture*

Batch ETL architecture processes data in grouped intervals or "batches" rather than continuously. This method is ideal for scenarios where real-time updates are unnecessary, as it focuses on efficiency and periodic processing. In a batch ETL workflow, data is first collected and grouped into batches, often defined by time intervals or specific criteria. Transformation is applied collectively to the entire batch before loading the processed data into the target system. While the architecture is less resource-intensive than real-time processing, it introduces latency between data extraction and availability.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

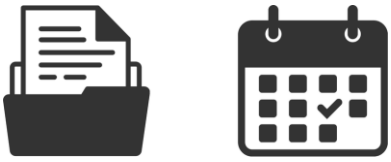


*Figure 16: Representation of batch processing*

**Tools and Technologies:**

- Open-Source Examples: Apache Airflow (workflow scheduling), dbt (transformations).
- Data Integration Tools: Pentaho, Kettle.

**Advantages:**

Batch ETL is efficient for datasets that do not require immediate updates, reducing the computational load compared to continuous processing. It is well-suited for environments where periodic data availability is acceptable, such as nightly report generation or historical data analysis.

**Limitations:**

This architecture is not suitable for real-time or near real-time use cases where data needs to be processed and acted upon immediately. Additionally, there is a delay between data extraction and when the processed data becomes available.

### *1.6.3 Streaming ETL Architecture*

Streaming ETL architecture processes data continuously, enabling **real-time** or near real-time data integration. This architecture is designed for scenarios where time-sensitive insights are critical, such as fraud detection, monitoring systems, or dynamic dashboards. Data is extracted as it is generated—often from sensors, logs, or live APIs—and undergoes transformations in real time before being loaded into the target system. **Streaming** ETL architectures provide low-latency updates and ensure data is always up-to-date in the target environment.

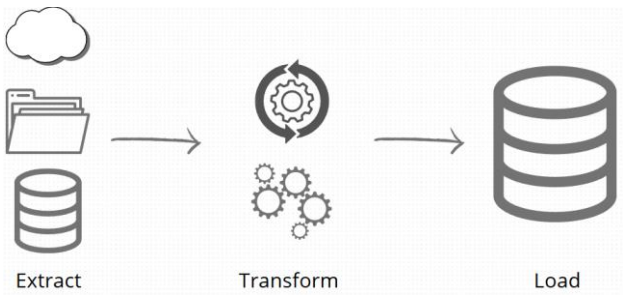



*Figure 17: Streaming ETL Architecture representation*



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

**Tools and Technologies:**

- Open-Source Examples: Apache Kafka, Apache Flink, StreamSets.
- Cloud Platforms: AWS Kinesis, Google Dataflow.

**Advantages:**

Streaming ETL is ideal for applications that require immediate data availability and minimal latency, such as monitoring systems or predictive analytics. It allows businesses to react to events as they occur, providing a significant advantage in time-sensitive operations.

**Limitations:**

Despite its advantages, streaming ETL architectures are more complex to implement and maintain compared to batch processing. They require higher resource consumption and may introduce challenges in fault tolerance and error recovery.

### *1.6.4 File-Based ETL Architecture*

File-based ETL architecture uses flat files, such as CSV, JSON, or XML, to drive the ETL workflow. Data is extracted from the source systems and saved to files, which serve as intermediaries during the ETL process. Transformations are performed directly on these files using scripts or ETL tools, and the processed files are subsequently loaded into the target system. This architecture is particularly useful when integrating data between systems with incompatible APIs or when infrastructure constraints limit the use of more advanced ETL techniques.

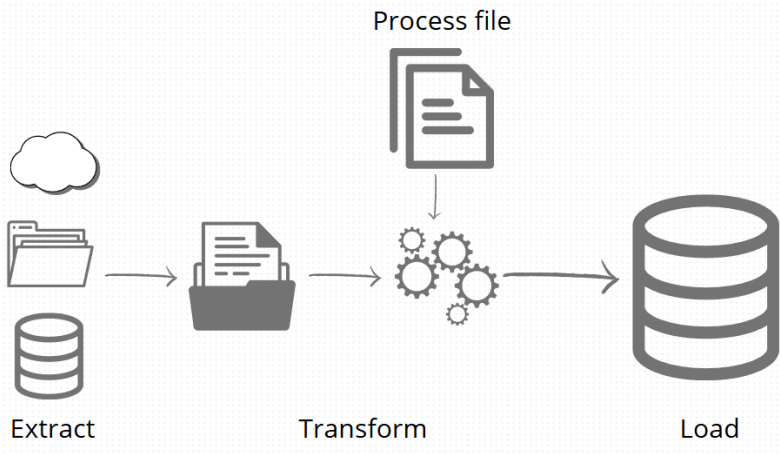



*Figure 18: File based ETL architecture*

**Tools and Technologies:**

- Custom Python scripts (using libraries like Pandas).
- Unix utilities (e.g., awk, sed) for basic transformations.
- Cloud-based file processing (e.g., AWS S3 + Lambda).



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

**Advantages:**

File-based ETL is easy to implement, requiring minimal infrastructure and technical complexity. It is a practical solution for transferring data between systems that cannot directly interface with one another, making it valuable for lightweight and ad-hoc data integration tasks.

**Limitations:**

This architecture can become unwieldy when handling large datasets or requiring frequent processing, as file management overhead increases. Additionally, it lacks the advanced capabilities of more sophisticated ETL architectures, such as dynamic scaling or seamless integration with modern APIs.

## 1.7 Key Features of the Required Architecture

**Scalability**, **interoperability**, **data governance**, and **flexibility** are the defining characteristics of this architecture. Scalability is achieved by accommodating distributed and diverse data sources without the need for centralized control, ensuring the system can grow as new sources are integrated. Interoperability is built into the design through the use of schemas, metadata, and standards like MiM, enabling seamless communication and compatibility among stakeholders and platforms. Data governance is prioritized with robust GDPR-compliant mechanisms for data handling, access control, and security, safeguarding sensitive information. Finally, flexibility allows the architecture to adapt to a wide range of data types and analytics needs by harmonizing both structured and unstructured data through tailored workflows. Together, these features ensure that the architecture is not only robust but also future-proof, capable of meeting evolving requirements and challenges.

## 1.8 Initial architecture based on the refined requirements

This draft architecture provides a foundation for integrating diverse data sources into a harmonized system that supports analytics, decision-making, and knowledge sharing across the consortium.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

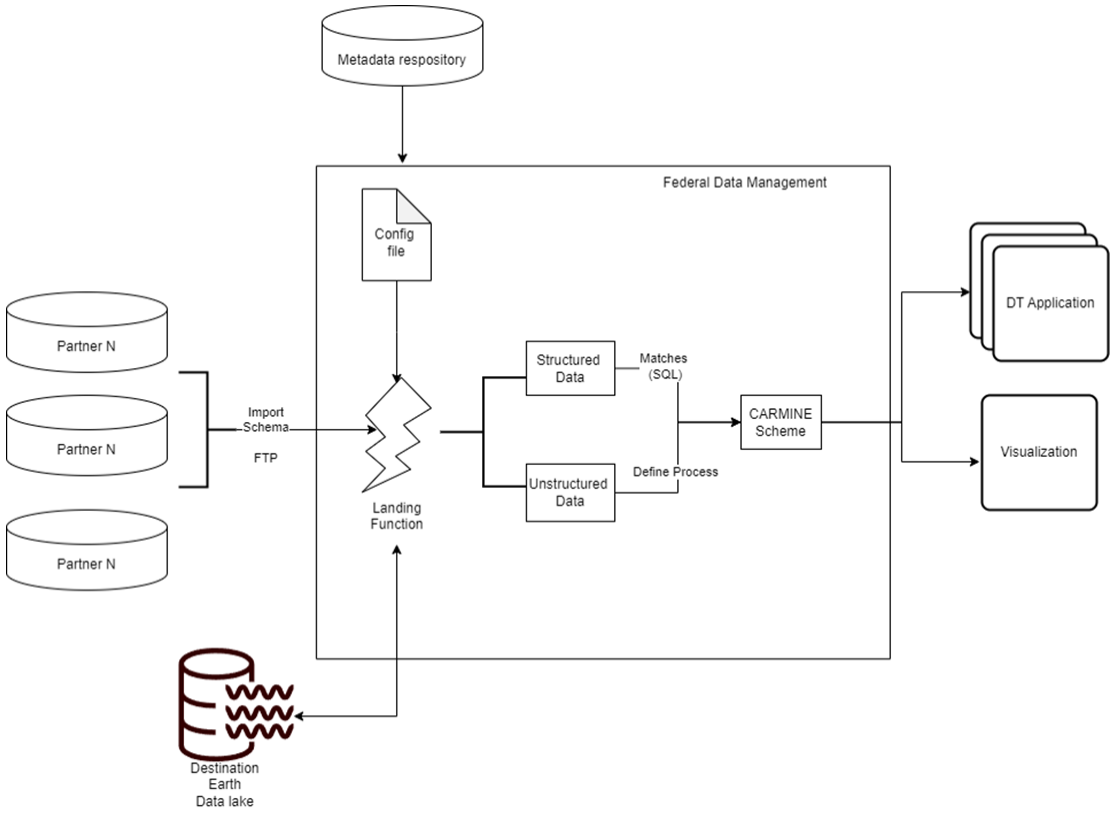



*Figure 19: Initial architecture based on ETL for CARMINE*

The diagram outlines a **lightweight federated data management architecture** designed to integrate, harmonize, and make data available for analytics within the CARMINE project. This architecture connects multiple distributed data sources, applies harmonization processes, and ensures compliance with GDPR and interoperability standards. The key components and their roles are as follows:

This design is built upon a straightforward **ETL** architecture, with data processing managed in **batches**. The transformation phase will be limited to the essential steps required to ensure data harmonization and standardization.

A **metadata** repository will be implemented to maintain consistency and facilitate data discovery. Authentication will be secured through simple key-based access mechanisms.

**Data governance** throughout the process will adhere to **GDPR** standards. Additionally, any federated queries that encounter errors will generate data logs, which will also be stored for monitoring and troubleshooting purposes.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

### 1.8.1 Data Sources

The architecture integrates data from a diverse array of sources, including partner organizations—referred to as Partner N—and external repositories like the Destination Earth Data Lake. These sources provide a mix of structured formats, such as relational databases, and unstructured formats, like multimedia files or text documents. Data is ingested into the system using predefined import schemas, which define the rules and structure for data entry. For larger data transfers, FTP protocols are employed to ensure secure and efficient delivery. This multi-source integration is designed to accommodate the diversity and scale of data needs while maintaining consistency across all inputs.

### 1.8.2 Landing Function

The landing function acts as the **gateway** for incoming data and is critical to the integrity and usability of the system. As data enters the federated system, it undergoes a standardization process guided by configuration files and metadata repositories. These elements ensure that data is correctly interpreted, aligned with the import schema, and prepared for downstream use. The landing function minimizes errors, reduces the manual effort involved in data handling, and supports automation of the data ingestion process, making it a vital component of the architecture.

### 1.8.3 Metadata Repository

At the core of the architecture is a centralized metadata repository that serves as a critical reference point for the system. This repository **stores** detailed **information** about the structure, origin, lineage, and governance of data, enabling robust validation processes for all incoming data. It also enriches datasets by providing additional context and ensuring that harmonization efforts are guided by well-defined rules. By offering a unified source of truth for metadata, the repository simplifies the integration of heterogeneous data sources and supports compliance with governance standards, including GDPR.

### 1.8.4 Data Harmonization

Following ingestion, data is classified into two distinct streams: structured data, which typically includes relational databases, and unstructured data, such as text, images, or videos. Structured data is processed using SQL-based techniques to ensure it adheres to relational models, while unstructured data undergoes tailored workflows to make it compatible with the federated schema. The harmonization process culminates in mapping the data to the CARMINE Scheme, a unified model designed to promote interoperability. This ensures that data from various sources can be used seamlessly, regardless of its original format or structure.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

### 1.8.5 Destination Earth Data Lake

The Destination Earth Data Lake is a centralized repository that plays a dual role in the architecture. It acts as a long-term storage location for processed and harmonized data while also serving as a reference point for analytics and data-sharing initiatives. By linking the data lake to the federated system, bi-directional data exchange is enabled, allowing for real-time updates and synchronization. This setup ensures that the data lake remains a dynamic and integral part of the architecture, supporting both immediate needs and long-term data strategies.

### 1.8.6 CARMINE Scheme

The **CARMINE Scheme** is the linchpin of the architecture, ensuring that harmonized data is stored and managed under a **consistent framework**. Built upon metadata gathered from project partners, the scheme provides a unified structure for accessing and analyzing data. It adheres to GDPR regulations and MiM mechanisms, ensuring that data is both compliant and interoperable. The **CARMINE Scheme** not only facilitates data alignment across sources but also empowers users by offering a clear and consistent model for interaction with the data.

A scheme to share and load data from different data sources involves defining a structured approach to efficiently collect, integrate, and distribute data across multiple systems. This process typically involves several key components and methodologies to ensure data consistency, availability, and reliability.

The CARMINE scheme is defined by the way the data is hosted by each partner, what protocols are there to access the data and to share it. In the concept of this project, the data is to be hosted by each partner and access is granted by them.

The data is to be hosted in databases, data that is locally stored and needs to be accessed by other partners should be made available via REST applications or uploaded onto databases. The scheme to share the data is handled by Federated Queries and made available via REST applications that connect the server and the client. All this connection should be secured and easy to interact with.

Naming conventions are also part of the CARMINE scheme. This means that all data that is related to the same variables should be named likewise, ensuring easier data processing and harmonization.

### 1.8.7 Applications and Visualization

Harmonized data within the federated system is designed to support a variety of applications. MiM applications utilize high-resolution datasets and advanced process models to enable simulations, decision-making, and predictive analytics. Visualization tools enhance accessibility by providing intuitive and interactive platforms for stakeholders to explore data insights. These tools allow users to interpret complex datasets, identify trends, and make informed decisions, ensuring that the architecture delivers actionable value across use cases.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

## 1.9 Deployment Diagram

A deployment diagram is a UML diagram that visualizes the physical deployment of software components on hardware nodes, illustrating the system's hardware devices, the software artifacts deployed on them, and their interactions. It includes elements like **nodes** (e.g., servers or devices), **artifacts** (e.g., executables or libraries), and communication paths that represent **relationships** and data flows.

Deployment diagrams are useful for understanding system architecture, planning deployment strategies, allocating hardware resources, optimizing scalability and performance, and troubleshooting issues by identifying dependencies and interconnections. They bridge the gap between software design and physical implementation, ensuring a clear and efficient deployment process.

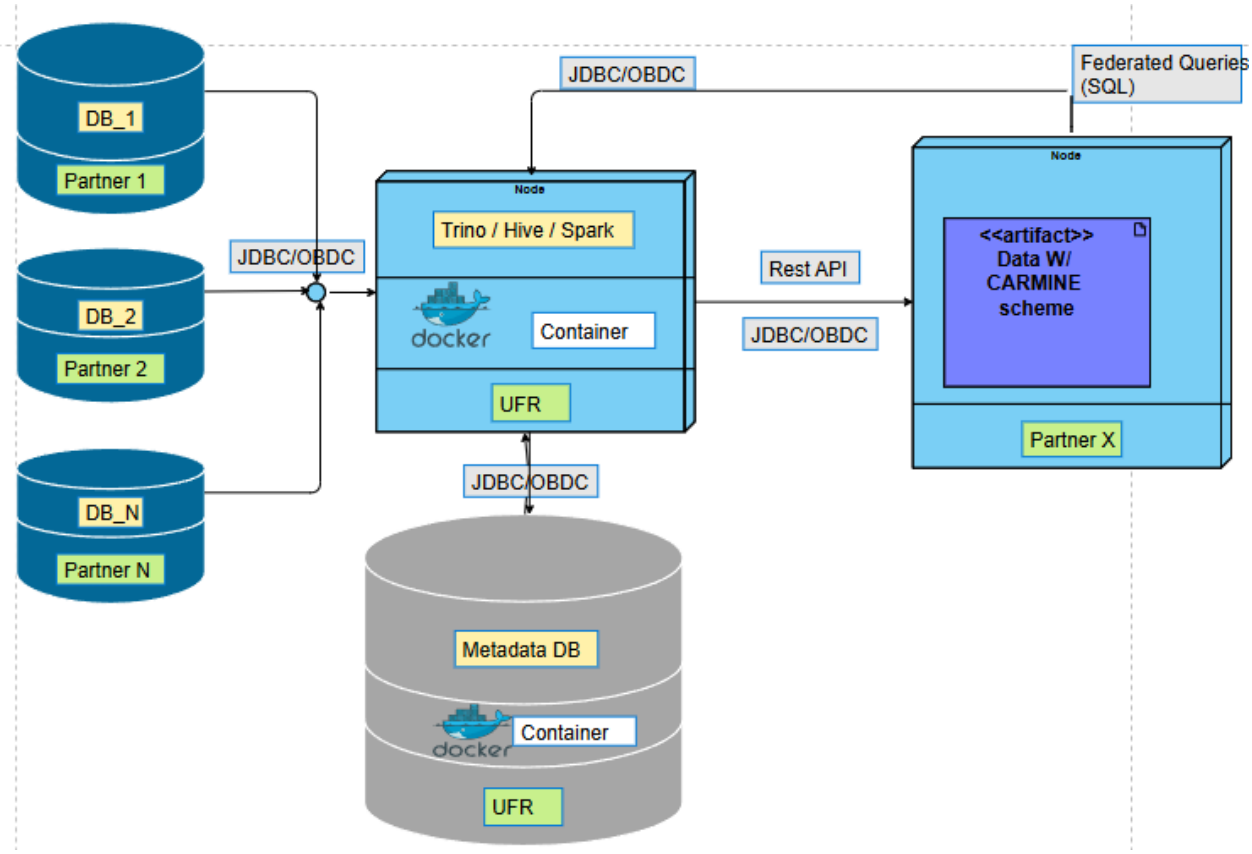



*Figure 20: Deployment diagram*

**Nodes and Components:**

- DB_1, DB_2, DB_N: Represent the databases of different partners. Each database connects to the central node via JDBC/ODBC.
- Central Node: Hosts multiple components:



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

Co-funded by UK Research and Innovation

- Trino/Hive/Spark: Tools or frameworks for federated query processing and big data operations.
- Docker Container: Virtual environment hosting the applications.
- UFR: University of Freiburg hosting.
- Metadata DB: This is a Hive Metadata database, which stores information about the structure and location of the partner databases. It is essential for Trino/Hive to resolve queries correctly. It is also hosted by UFR.
- Partner X: Represents a partner node interacting with the central node using a REST API. It also utilizes federated queries with data structured according to the "CARMINE scheme."

**Communication**:

- Partner databases communicate with the central node and the metadata database using JDBC/ODBC protocols, enabling seamless data exchange. Federated queries allow Partner X to retrieve and combine data from multiple sources efficiently. Additionally, a REST API facilitates specific interactions between Partner X and the central node, supporting flexible and targeted data access.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

# 2. Concepts of Lightweight Federated Data Management

## 2.1 Overview of Lightweight Federated Data Management

LFDM represents a streamlined approach to handling distributed data across diverse organizations and systems. By leveraging the principles of federation, LFDM enables seamless collaboration without requiring centralized control, fostering interoperability and data sovereignty. This concept aligns closely with modern requirements for scalable, secure, and adaptable systems that support data sharing while respecting individual autonomy and compliance frameworks such as GDPR.

Core Principles of Federated Data Management explores the fundamental ideas behind distributed data systems, highlighting the significance of modularity, interoperability, and decentralized control. Building on this foundation, Open-Source Philosophies in Lightweight Systems discusses how leveraging open-source tools and frameworks contributes to the adaptability and efficiency of LFDM solutions. Finally, Benefits, Limitations, and Relevance to Diverse Stakeholders evaluates the practical implications of LFDM, weighing its advantages and challenges while underscoring its importance to a wide range of users, from data providers to analysts and decision-makers. Together, these sections provide a comprehensive understanding of the theoretical underpinnings and practical applications of LFDM.

Lightweight Federated Data Management (LFDM) refers to a distributed approach for managing data across multiple systems, organizations, or platforms while maintaining local data sovereignty and minimizing the complexity of integration. Unlike centralized data management solutions, federated solutions allow stakeholders to share, query, and utilize data without needing to physically centralize it.

In the context of **lightweight systems**, the focus is on achieving this functionality with minimal technical and computational overhead. The approach prioritizes simplicity, cost-efficiency, and accessibility, making it particularly suitable for organizations with limited resources or for applications requiring rapid deployment.

## 2.2 Key characteristics of FDMS include:

FDM is built on the premise of **decentralization**, allowing data to remain within the control of its owners while enabling collaborative access and analysis. The following principles define its core:



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

1. **Data Sovereignty**:

Data sovereignty ensures that data remains stored and managed by the originating entity, thus respecting privacy, security, and local regulations. This approach avoids the need for physical aggregation of data, enabling organizations to retain control over their data while adhering to policies such as the GDPR. By maintaining data in its original location, federated systems avoid potential risks associated with cross-border data transfers, providing greater assurance that data governance policies are followed effectively.

2. **Interoperability**:

Interoperability refers to the ability of disparate systems to work together seamlessly, regardless of differences in data formats, platforms, or protocols. This is achieved through adherence to open standards and consistent metadata practices, which provide a shared understanding of data and its structure. By standardizing communication protocols and data representations, federated systems ensure that data can be integrated and used across diverse environments, making it easier for organizations to collaborate and share information.

3. **Decentralized Collaboration**:

Federated systems enable distributed stakeholders to collaborate on data without requiring central ownership, fostering decentralized collaboration. This approach is particularly valuable in supporting diverse, cross-border partnerships, where each partner can retain control over their own data while contributing to a shared objective. By allowing for collaborative work across different domains and jurisdictions, federated systems promote inclusivity and cooperation among a wide range of stakeholders, including those from various sectors and regions.

4. **Scalability and Adaptability**:

Federated systems are designed to grow alongside increasing data volumes and user bases, avoiding the need for excessive infrastructure changes. This scalability ensures that the system can handle both small datasets and large-scale operations without compromising performance or flexibility. Additionally, federated systems are adaptable to a variety of use cases, offering the ability to support different data types, workflows, and analytics requirements, which makes them suitable for diverse industries and applications.

5. **Security and Privacy**:

Security and privacy are paramount in federated data management systems. These systems ensure secure data sharing through robust protocols, such as tokenized access, encrypted communication, and federated authentication methods like OpenID Connect. By implementing these security measures, federated systems protect sensitive data from unauthorized access and ensure that privacy is maintained. The distributed nature of federated systems means that each participant can implement their own security policies while still enabling secure interactions across the network.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

6. **Findability and Accessibility**:

Metadata-driven discoverability plays a crucial role in enabling users to locate and access datasets across federated repositories. With a well-organized metadata layer, users can search and retrieve data from multiple sources in an efficient and effective manner. In addition, access controls and user interfaces are designed to ensure that only authorized users can retrieve and interact with the data, providing a secure and user-friendly experience. This combination of findability and accessibility ensures that the right data is available to the right users, facilitating data-driven decision-making across the federated network.

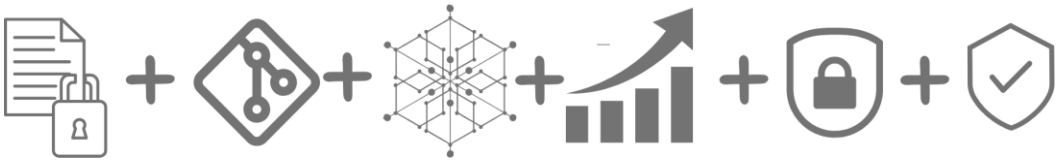


*Figure 21: Data sovereignty, Interoperability, Decentralization, Scalability, Security, Accessibility*

## 2.3 Key characteristics of LFDM include:

- **Decentralization**: Data remains under the control of its respective owners, promoting privacy and compliance with data protection regulations.
- **Interoperability**: Use of open standards (e.g., FAIR principles, JSON-LD) enables seamless integration and data sharing across diverse systems.
- **Lightweight Architecture**: Minimalistic design choices reduce infrastructure and maintenance requirements, focusing on essential functionality without unnecessary complexity.
- **Scalability**: Solutions are designed to accommodate growing datasets and users without major system overhauls.
- **Flexibility**: Support for heterogeneous data formats and protocols, ensuring compatibility with a wide range of use cases.

*Figure 22: Decentralization, Interoperability, Lightweight architecture, Scalability and Flexibility*



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

The **open-source philosophy** plays a critical role in LFDM, fostering transparency, collaboration, and innovation. By leveraging open-source tools, standards, and communities, LFDM solutions can achieve broad adoption and adaptability while promoting equitable access to technology.

This report explores the foundational concepts, relevant standards, and core processes underpinning lightweight federated data management. It also addresses challenges, showcases applications, and provides actionable recommendations for stakeholders to adopt LFDM effectively within an open-source framework.

## 2.4 Open-Source Philosophies in Lightweight Systems

The lightweight federated approach aligns naturally with open-source philosophies, emphasizing collaboration, transparency, and innovation:

**Transparency and Trust**

Transparency and trust are fundamental in federated systems, achieved through visibility into code, architecture, and protocols. By relying on open-source principles, users can independently verify compliance with standards, security protocols, and best practices. This openness fosters confidence among stakeholders, ensuring that systems adhere to agreed-upon norms while reducing concerns around hidden vulnerabilities or proprietary restrictions.

**Cost-Effectiveness**

The use of open-source tools significantly reduces licensing and operational costs, making federated systems more accessible to organizations with limited resources. This cost-effectiveness democratizes the adoption of lightweight federated systems, enabling small and medium enterprises (SMEs), as well as organizations in developing regions, to participate in and benefit from distributed data management frameworks.

**Customizability and Modularity**

Open-source tools used in federated systems are typically designed with high customizability and modularity, allowing stakeholders to adapt solutions to their unique requirements. This flexibility ensures that organizations can address specific technical or business challenges by modifying or extending existing tools without being constrained by rigid, one-size-fits-all architectures.

**Knowledge Sharing**

Open-source ecosystems foster a culture of collaboration and knowledge dissemination. Communities and organizations can build upon shared innovations, enhancing system capabilities and avoiding redundant efforts. This collaborative environment accelerates technological advancements and encourages the development of best practices, benefiting all users.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

**Ethical Technology Use**

The open-source model aligns with ethical practices by promoting equitable access to technology and reducing dependency on proprietary solutions. Federated systems built on open standards enhance fairness and inclusivity by removing barriers such as vendor lock-in, and making advanced data management capabilities available to a broader range of users and stakeholders.

## 2.5 Key Insights into Concepts, Standards, and Processes

This report distils essential insights into the foundational concepts, critical standards, and fundamental processes that underpin **Lightweight Federated Data Management (LFDM)**, offering a roadmap for its effective implementation in open-source ecosystems.

### *2.5.1 Concepts*

1. **Federation vs. Centralization**:
   LFDM operates on a decentralized model, where data is not aggregated into a central repository but remains distributed across various nodes. This enables greater control for data custodians and enhances privacy compliance.

2. **Lightweight Principles**:
   Emphasizing simplicity and efficiency, LFDM systems are designed to minimize computational and infrastructural demands, making them accessible to resource-constrained organizations.

3. **Flexibility and Adaptability**:
   LFDM is inherently modular, allowing for integration with diverse systems, formats, and workflows, tailored to specific use cases.

### *2.5.2 Standards*

**1. FAIR Principles:**

LFDM leverages the principles of being Findable, Accessible, Interoperable, and Reusable to ensure data usability across systems and organizations.

**2. Interoperability Protocols:**

- REST APIs and JSON-LD enable seamless communication and data exchange between systems.
- SPARQL and RDF support querying and structuring semantic data in federated environments.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

3. **Data Governance and Licensing:**
   Adherence to open data standards and clear licensing protocols ensures compliance, transparency, and legal clarity for data sharing.

4. **Metadata Standards:**
   Formats like DCAT and OAI-PMH facilitate data discoverability and cataloging in federated systems.

### *2.5.3 Processes*

1. **Federated Querying and Access**:
   LFDM systems enable users to query data across multiple repositories without the need to migrate or duplicate it.

2. **Integration Workflows**:

   Processes such as **ETL (Extract, Transform, Load)** and data harmonization are streamlined to maintain lightweight operation and ensure consistency across diverse datasets.

3. **Security and Privacy**:

   Robust protocols are integrated to safeguard sensitive data and ensure secure federated access.

4. **Open-Source Collaboration**:
   By relying on open-source tools, LFDM systems encourage innovation and collaboration while reducing proprietary dependencies.

5. **Scalability with Minimal Overhead**:

   Lightweight designs support incremental scaling as data or user demands grow, ensuring long-term viability without major infrastructure investments.

By focusing on these key concepts, standards, and processes, LFDM offers a blueprint for achieving decentralized, efficient, and open-source-driven data management solutions. These insights set the foundation for deeper exploration in subsequent sections of the report.

## 2.6 Benefits, Limitations, and Relevance to Diverse Stakeholders

### *2.6.1 Benefits*

**Stakeholder Empowerment**

Federated systems empower data owners by allowing them to retain full control over their data. This autonomy ensures compliance with sovereignty and privacy regulations, particularly important for industries governed by strict legal frameworks.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

**Interoperability Across Ecosystems**

Open standards and lightweight architectures enable seamless data sharing and interaction across heterogeneous systems and domains. This interoperability reduces friction in collaborative environments and supports cross-industry initiatives.

**Cost Efficiency**

Federated systems minimize infrastructure and operational expenses through lightweight designs and reliance on open-source tools. This affordability lowers the barrier to entry for organizations seeking advanced data management capabilities.

**Rapid Adoption**

The simplicity and low resource requirements of lightweight federated systems facilitate quicker adoption. SMEs, startups, and organizations in resource-constrained environments can easily integrate federated solutions into their workflows.

**Enhanced Collaboration**

By bridging data silos while maintaining individual ownership, federated systems encourage collaboration among diverse stakeholders. This fosters innovation and cooperation across industries, academia, and policy domains.

### *2.6.2 Limitations*

**Complex Coordination**

Federated systems require establishing and managing agreements among stakeholders. These coordination efforts can be complex, involving negotiations around standards, roles, and responsibilities.

**Security Risks**

The distributed nature of federated environments exposes them to potential vulnerabilities. Ensuring robust security measures, such as encryption and access control, is essential to mitigate risks.

**Interoperability Challenges**

While federated systems promote interoperability, achieving seamless integration across highly heterogeneous systems remains a significant technical challenge. Differences in data formats, protocols, and governance models can complicate implementation.

**Performance Trade-offs**

Real-time data querying and processing in federated systems may involve latency compared to centralized systems. This performance trade-off can impact applications requiring low-latency data access.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

### *2.6.3 Relevance to Stakeholders:*

**Policy Makers**

Federated systems enable compliance with regulations while promoting ethical data-sharing practices. They support initiatives that balance innovation with privacy and sovereignty concerns.

**Researchers and Academics**

Federated frameworks facilitate collaboration between institutions, allowing researchers to share data without compromising intellectual property or confidentiality. This promotes innovation and accelerates discovery.

**Developers and IT Teams**

Lightweight federated systems offer a framework for designing secure, scalable, and standards-compliant solutions. Developers benefit from the flexibility of open-source tools to build tailored applications.

**Organizations and SMEs**

Federated systems enable efficient data management without requiring significant infrastructure investments. SMEs can enhance their competitiveness by adopting lightweight, cost-effective solutions that scale with their growth.

By adhering to federated principles and leveraging open-source philosophies, lightweight systems provide a flexible, sustainable, and equitable approach to modern data management challenges.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

# 3. Standards in Lightweight Federated Data Management

Standards are the foundation of LFDM, enabling decentralized systems to achieve interoperability, scalability, and accessibility. In federated systems, data remains distributed across multiple nodes, often owned and managed by diverse stakeholders. Open standards ensure that these nodes can communicate, share, and utilize data seamlessly, regardless of the underlying technologies or domains involved.

By adhering to well-defined protocols and frameworks, LFDM systems align with global best practices for data discoverability, reusability, and ethical handling. Standards like FAIR principles, OpenAPI, and DCAT play a crucial role in structuring data for accessibility and usability, while technologies like SPARQL and OIDC ensure secure and efficient querying and authentication. Moreover, interoperability is further enhanced through tools like JSON-LD and Linked Data, which provide the semantic backbone necessary for machine-readable data exchange.

Equally significant is compliance with open licensing models and data-sharing frameworks, which guarantees transparency, equitable access, and legal clarity. These practices empower federated systems to balance openness and trust while addressing challenges in scalability and governance. This section explores key standards, interoperability mechanisms, and compliance strategies that underpin LFDM, illustrating their critical role in fostering collaborative and ethically sound data ecosystems.

## 3.1 Key Open Standards and Their Significance

Open standards are essential in LFDM as they provide the framework for achieving interoperability, scalability, and accessibility across diverse systems. These standards serve as the guiding principles for the development and implementation of federated architectures, ensuring seamless integration, broad compatibility, and the flexibility to adapt to evolving data management needs.

### *3.1.1 FAIR Principles (Findable, Accessible, Interoperable, Reusable)*

The FAIR principles emphasize the need for data to be discoverable, accessible, and reusable by both humans and machines. These principles underpin the design of federated systems by promoting standardized practices for metadata tagging, cataloging, and data identification. Through the use of persistent identifiers like DOIs and rich metadata schemas, datasets become more structured and aligned with FAIR principles. This alignment enhances the discoverability of resources across distributed environments, fostering collaboration and reuse while supporting compliance with global data-sharing standards.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

### 3.1.2 OpenAPI (RESTful API Specification)

OpenAPI defines a widely adopted standard for designing and documenting RESTful APIs, ensuring interoperability between systems built on diverse architectures. By providing a common language for API specifications, OpenAPI facilitates seamless integration and communication among distributed nodes in federated systems. In LFDM, REST APIs enable secure and standardized interactions for data access, query execution, and process automation, significantly reducing implementation complexity and improving system compatibility across domains.

### 3.1.3 DCAT (Data Catalog Vocabulary)

DCAT offers a standardized vocabulary for describing and sharing data catalogs, playing a critical role in federated environments. It allows for the aggregation and harmonization of metadata from distributed data repositories, providing a unified view of available resources. By adopting DCAT, federated systems ensure that datasets remain easily discoverable and accessible while aligning with global data cataloging practices. This standard is instrumental in supporting large-scale federations that require extensive metadata management.

### 3.1.4 SPARQL (SPARQL Protocol and RDF Query Language)

SPARQL serves as a powerful tool for querying RDF data, enabling federated systems to leverage semantic relationships and Linked Data principles. Its ability to query distributed datasets without requiring centralization is particularly advantageous for preserving data locality and minimizing redundancy. SPARQL supports advanced semantic queries, which are essential for integrating complex datasets across domains while maintaining consistency and adhering to FAIR principles.

### 3.1.5 OIDC (OpenID Connect)

OIDC, an authentication protocol built on OAuth 2.0, provides a secure framework for managing user identities in federated systems. Its token-based authentication mechanisms ensure that users can seamlessly access multiple distributed resources without compromising security. In LFDM, OIDC supports decentralized authentication workflows, enabling robust access control, simplified user management, and enhanced privacy compliance.

## 3.2 Discussion on Interoperability

Interoperability is a cornerstone of LFDM, allowing diverse systems, stakeholders, and technologies to collaborate without compromising their unique infrastructures. Achieving interoperability requires adopting technologies and standards that ensure seamless data exchange and integration, even across highly heterogeneous environments.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

### JSON-LD (JSON for Linked Data)

JSON-LD combines the simplicity of JSON with the richness of Linked Data, enabling datasets to carry semantic context alongside their structure. This format allows data to be both machine-readable and human-friendly, ensuring smooth integration into applications that rely on semantic web standards. In LFDM, JSON-LD is used to exchange annotated data between federated nodes while preserving the underlying semantic meaning. Its lightweight nature and broad compatibility make it an ideal choice for building interoperable systems.

### Linked Data Principles

Linked Data promotes the use of URIs and RDF to create a web of interconnected datasets, enabling seamless integration across domains. Federated systems leverage Linked Data principles to connect distributed knowledge bases, fostering data reuse and collaboration. By maintaining semantic consistency, Linked Data ensures that information shared between nodes remains meaningful and interoperable, even when systems differ in structure or purpose.

### RESTful Protocols

RESTful APIs provide a standardized method for enabling communication between systems using HTTP. Their stateless architecture ensures compatibility and ease of implementation, making them a popular choice for lightweight federated systems. In LFDM, RESTful protocols underpin real-time data transfer, query execution, and process automation, supporting a wide range of use cases while maintaining simplicity and reliability.

## 3.3 Compliance with Open Licensing and Data-Sharing Frameworks

Compliance with open licensing and data-sharing frameworks is critical for building trust, ensuring ethical collaboration, and aligning with global regulations in LFDM. These frameworks provide the foundation for transparent, equitable, and legally sound data-sharing practices.

### Open Licensing Models

Open licensing models, such as Creative Commons (CC), Open Data Commons (ODC), and Open Government License (OGL), empower stakeholders to share data while maintaining clear usage guidelines. These models encourage the reuse and redistribution of datasets by addressing attribution and usage rights, reducing ambiguity, and avoiding legal conflicts. Federated systems that adopt open licenses ensure that all participants understand and adhere to established terms, fostering transparency and collaboration.

### Data-Sharing Frameworks

Frameworks like the European Data Governance Act and Open Data Directive play a pivotal role in promoting transparent and ethical data sharing. They establish guidelines for equitable access, privacy compliance, and regional governance, enabling federated systems to navigate



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

complex legal landscapes. Sector-specific frameworks, such as INSPIRE for geospatial data, further enhance domain-specific data-sharing practices, ensuring consistency and compliance within specialized fields.

**Metadata Licensing and Provenance Standards**

Accurate metadata licensing and provenance tracking are essential for ensuring that datasets remain traceable and legally compliant. Standards like Dublin Core and PROV-O (Provenance Ontology) allow federated systems to describe the origin, ownership, and usage conditions of datasets. This level of detail enhances transparency, accountability, and trust among stakeholders, while also aligning with FAIR principles and open data-sharing practices.

By embracing these open standards, licensing models, and compliance frameworks, LFDM fosters an ecosystem that prioritizes interoperability, ethical data handling, and seamless collaboration across domains.

Adopting and adhering to these standards allows LFDM to achieve seamless collaboration, data integration, and ethical compliance, making it a cornerstone of modern decentralized data management.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

# 4. Processes in Lightweight Federated Data Management

LFDM is built on a foundation of open standards, reproducible workflows, and robust privacy protocols that enable distributed data systems to operate efficiently, securely, and at scale. The inherent modularity and decentralized nature of federated systems offer unparalleled flexibility, allowing organizations to manage and query data without consolidating it in a single location. This approach ensures data sovereignty while maintaining interoperability across diverse datasets and platforms.

To achieve these goals, LFDM systems adopt processes that emphasize modular design, open-source integration, metadata-driven discovery, and scalable workflows. At the same time, ensuring data privacy and security is critical within open and distributed environments. From implementing standardized pipelines for reproducibility to using advanced authentication protocols for data protection, each process plays a vital role in creating efficient and trustworthy federated systems.

This section explores the key standards, workflows, and privacy measures that underpin LFDM. It begins with an examination of open architectures and their significance, highlighting the benefits of modularity, decentralized resource management, federated query engines, and metadata-driven discoverability. Next, it delves into workflows designed to ensure reproducibility and scalability, including the use of containerization, version control, and ETL pipelines. Finally, the section concludes with a detailed overview of privacy and security protocols essential for safeguarding distributed data while complying with legal and ethical standards. Together, these processes define the operational backbone of lightweight federated data management systems.

## 4.1 Key Open Standards and Their Significance

**Open architectures** in federated data management emphasize modularity, flexibility, and adaptability, creating lightweight and scalable systems. The following elements define their significance:

### *4.1.1 Modular and Layered Design*

Federated systems rely on modular and layered architectures to ensure flexibility and adaptability. These systems are composed of distinct components such as data sources, mediators, and clients, each functioning independently while seamlessly integrating with the overall framework. This modularity allows for simplified system maintenance and upgrades, as individual components can be updated or replaced without disrupting the operation of the entire system. Moreover, the layered design promotes scalability and makes it easier to introduce new functionalities or data sources as requirements evolve, supporting lightweight implementation and future-proof adaptability.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

### 4.1.2 Decentralized Resource Management

A cornerstone of federated systems is their decentralized approach to resource management, where data remains under the control of its originating entity. Communication between distributed nodes occurs through open protocols such as REST or WebSocket, ensuring seamless interaction without centralizing ownership. This decentralization enhances data sovereignty, allowing organizations to comply with privacy regulations and maintain control over their sensitive information. Additionally, decentralization reduces bottlenecks associated with centralized architectures, as operations are distributed across nodes, improving system reliability and responsiveness.

### 4.1.3 Federated Query Engines

Federated query engines, such as Apache Drill, Presto, and SPARQL endpoints, enable the querying of multiple distributed datasets without requiring data to be physically moved or centralized. These engines allow users to perform complex queries across heterogeneous data sources, presenting the results as a unified dataset. This approach preserves data locality, which minimizes the risk of data duplication and redundancy. By keeping data within its native repositories and only retrieving relevant results, federated query engines provide an efficient solution for large-scale data analysis while maintaining the principles of lightweight federated systems.

### 4.1.4 Integration with Open-Source Tools

Open-source tools play a pivotal role in the implementation and management of federated systems, fostering community-driven development and innovation. By leveraging tools like Apache Kafka, dbt, or Apache Airflow, organizations can build systems that are cost-effective and highly customizable. Open-source integration simplifies system maintenance, as the community often contributes to regular updates, patches, and extensions. Furthermore, the inherent flexibility of open-source software allows developers to add or modify components to meet specific requirements, making these tools ideal for lightweight and adaptable federated architectures.

### 4.1.5 Metadata-Driven Discoverability

In federated systems, metadata-driven discoverability is essential for locating, accessing, and integrating distributed datasets. Centralized metadata catalogs aggregate information about repositories using standardized formats such as the Data Catalog Vocabulary (DCAT). This approach ensures that datasets remain findable, accessible, interoperable, and reusable (FAIR), even in distributed environments. Metadata facilitates efficient data navigation by providing detailed descriptions of datasets, their origins, and their structures. This ensures interoperability between diverse systems, promotes collaboration across organizational boundaries, and simplifies compliance with global data-sharing standards.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

## 4.2 Workflows Supporting Reproducibility and Scalability

To achieve reliable and scalable federated systems, structured workflows and tools are employed:

### *4.2.1 Standardized Data Pipelines*

Structured data pipelines are fundamental for achieving consistency and reliability in federated systems. These pipelines, particularly through **ETL** processes, ensure uniform handling of data from diverse and heterogeneous sources. By applying a standardized approach, ETL workflows guarantee that data is accurately transformed into compatible formats and efficiently loaded into the target systems. This standardization not only improves reproducibility by enabling the rollback to prior data states when errors occur but also provides a transparent framework for monitoring and auditing workflows. The traceability inherent in these pipelines fosters trust and clarity, essential qualities in federated environments where multiple stakeholders interact with shared resources.

### *4.2.2 Version Control for Data and Metadata*

Version control systems for data and metadata play a critical role in maintaining the integrity and accountability of federated workflows. These systems track changes to datasets, metadata, and configurations over time, creating an auditable history that preserves previous states. By maintaining comprehensive records of alterations, version control ensures reproducibility in data processing, as users can revert to earlier versions when needed. Moreover, this transparency enhances accountability, as every modification is documented and attributed. In federated environments, where datasets evolve across distributed nodes, version control ensures that all participants have access to consistent, up-to-date, and verified information.

### *4.2.3 Containerization and Orchestration*

Containerization and orchestration are key technologies supporting scalable and reliable federated systems. Tools like **Docker** allow workflows to be encapsulated in containerized environments, which package all necessary dependencies and configurations to ensure consistent execution, regardless of the underlying hardware or software. Meanwhile, orchestration tools such as Kubernetes manage these containers, enabling dynamic scaling to meet fluctuating computational demands. This combination ensures horizontal scalability, making it possible to handle growing data volumes and increasing complexity without compromising performance. Containerization also enhances reproducibility, as the standardized environment guarantees that workflows can be executed consistently across different platforms, from development to production in federated setups.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

## 4.3 Privacy and Security Protocols Within Open Environments

The distributed nature of federated systems necessitates robust privacy and security protocols to ensure data safety and compliance:

### *4.3.1 Secure Communication Channels*

Secure communication channels are essential for ensuring the safe transmission of data between federated nodes. Implementing protocols such as **HTTPS** and **SSL/TLS** ensures that data exchanged across the network is encrypted, protecting it from interception or unauthorized access. By securing data in transit, these channels safeguard sensitive information, maintain the confidentiality of communications, and prevent cyberattacks such as man-in-the-middle attacks. In a federated environment, where data often flows across distributed systems, robust encryption mechanisms are critical to establishing trust and protecting the integrity of the system.

### *4.3.2 Federated Authentication and Authorization*

Federated systems rely on sophisticated authentication and authorization mechanisms to control access to distributed resources. Token-based authentication protocols like OAuth 2.0 and OpenID Connect allow for secure, decentralized identity management, ensuring that only verified users gain access to system components. Additionally, Role-Based Access Control (RBAC) enforces fine-grained permission management by restricting user actions based on their assigned roles. These practices prevent unauthorized data access, minimize security risks, and support compliance with stringent data privacy regulations. By combining these approaches, federated systems maintain a strong security posture while enabling seamless interaction across multiple nodes.

### *4.3.3 Data Anonymization and Masking*

To protect sensitive data, federated environments implement anonymization and masking techniques such as k-anonymity, differential privacy, and attribute masking. These techniques ensure that personally identifiable information (PII) and other confidential attributes are obfuscated or generalized before sharing. Data anonymization protects user privacy while still allowing meaningful insights to be derived from datasets. By applying these methods, federated systems ensure that data remains secure even during analysis or cross-node interactions, aligning with legal and ethical standards like GDPR and HIPAA while fostering trust among users.

### *4.3.4 Audit Trails and Logging*

Comprehensive audit trails and logging systems are vital for maintaining accountability and detecting anomalies within federated environments. These processes involve tracking user activities and system operations in real time, using tools like the ELK Stack (Elasticsearch, Logstash, Kibana) or Grafana for visualization and monitoring. Audit trails help organizations



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

identify unauthorized activities, investigate incidents, and ensure compliance with regulatory requirements. By maintaining detailed logs, federated systems foster transparency, enhance system reliability, and provide a foundation for security audits and operational reviews.

### *4.3.5 Data Sovereignty Protocols*

Data sovereignty protocols are designed to ensure compliance with regional regulations and maintain ethical data-handling practices. Federated systems achieve this by adhering to frameworks like GDPR and HIPAA, implementing geofencing techniques to restrict data movement across borders, and ensuring that data storage aligns with the legal requirements of specific jurisdictions. These protocols empower organizations to maintain control over their data, reduce legal risks, and respect the sovereignty of the regions where the data is generated or stored, which is especially critical in globally distributed systems.

### *4.3.6 Intrusion Detection and Prevention Systems (IDPS)*

Intrusion Detection and Prevention Systems (IDPS) provide an active layer of defense for federated environments by continuously analyzing network traffic for anomalies or malicious activity. Tools like Snort and Suricata detect potential threats in real time, enabling rapid response to security incidents. By monitoring and safeguarding the network, these systems help prevent data breaches, unauthorized access, and other cyber threats. In federated setups, IDPS enhances the overall security framework, ensuring the integrity and reliability of distributed data systems.

This refined structure provides a clear and systematic explanation of the components, workflows, and security measures critical for lightweight federated data management systems.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

# 5. Data Concepts: Definitions and Applications

The concepts and definitions introduced in this chapter lay the groundwork for understanding the mechanisms underlying LFDM. A federated approach to data management relies on the seamless integration, interoperability, and accessibility of data distributed across diverse systems. To achieve these goals, it is essential to grasp the fundamental components of data systems, including data sources, databases, and methodologies for querying and processing data. These concepts not only provide a technical vocabulary but also clarify the logical structures and workflows required to manage, harmonize, and utilize data effectively within federated environments.

Data sources, databases, and integration tools such as import schemas and landing functions ensure that raw data from multiple origins can be harmonized into a coherent framework. Metadata and data classification (structured versus unstructured) establish the organizational principles that make federated queries and data governance achievable. Additionally, core technologies like SQL, federated queries, and protocols such as FTP demonstrate how federated systems overcome the technical and logistical challenges of distributed data access and integration.

Understanding these foundational concepts is critical for addressing the challenges of building and maintaining an LFDM system. LFDM requires lightweight, adaptable solutions that can scale across distributed environments while respecting data sovereignty, complying with standards such as GDPR, and supporting interoperability across heterogeneous systems. As this chapter progresses into its subsections, it elaborates on each concept, linking it directly to its role within LFDM. This ensures a comprehensive understanding of how these definitions and principles enable the practical implementation of federated data systems in complex, multi-stakeholder projects.

## 5.1 Data Sources

Data sources are the origins of information that feed into a system, encompassing a broad spectrum of structured, semi-structured, and unstructured data. Structured data sources include relational databases such as MySQL and PostgreSQL, which store data in well-defined rows and columns, ensuring consistency and ease of query. Semi-structured data sources include JSON files and XML-based APIs, which, while not rigidly structured, contain embedded metadata that facilitates data parsing and integration. Unstructured data sources, such as multimedia files, social media content, and raw text, lack a predefined format and require advanced techniques for processing and extraction of meaningful insights.

**Examples:**

- **Structured:** Relational databases like MySQL, PostgreSQL.
- **Semi-structured:** JSON files, XML APIs.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

- **Unstructured:** Multimedia files, social media feeds.

**Relevance to LFDM:**

In a federated data management system, data sources are inherently distributed across different partners or organizations. Ensuring seamless interoperability between these heterogeneous data sources is crucial. LFDM must account for the integration of diverse formats, offering a unified access layer while respecting data sovereignty. By accommodating the varying natures of structured, semi-structured, and unstructured data, LFDM enables robust analytics and decision-making processes, even in highly decentralized environments.

## 5.2 Database

A database is an organized collection of data stored electronically, managed using a Database Management System (DBMS). Relational databases, such as PostgreSQL and SQLite, utilize structured tables and rely on SQL to perform operations like querying, updating, and aggregating data. These databases are ideal for transactional systems and use cases requiring consistency and integrity. NoSQL databases, including MongoDB and Cassandra, handle unstructured or semi-structured data, offering flexibility and scalability, particularly for big data and real-time applications.

*Figure 23: Database*

**Types:**

- Relational Databases: Use structured tables and SQL (e.g., PostgreSQL, SQLite).
- NoSQL Databases: Handle unstructured or semi-structured data (e.g., MongoDB, Cassandra).

**Relevance to LFDM:**

Databases form the backbone of federated systems by acting as core repositories where distributed data is stored and accessed. In LFDM, they ensure fast, reliable, and secure data storage while maintaining compatibility with various data formats. Relational databases are often leveraged for structured data harmonization, while NoSQL systems cater to more complex and varied datasets. The role of databases in LFDM extends to enabling efficient querying, metadata management, and interoperability, thus serving as critical components of the system's infrastructure.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

## 5.3 Data Lake

A data lake is a centralized repository that stores large amounts of raw data in its original format, encompassing structured, semi-structured, and unstructured data. Unlike databases, which enforce schema constraints during data storage, data lakes adopt a schema-on-read approach, applying structure only when the data is queried. This flexibility allows data lakes to handle massive datasets and support advanced analytics, such as machine learning and artificial intelligence workflows.

*Figure 24: Data lake*

**Relevance to LFDM:**

Data lakes complement LFDM by acting as a unified storage solution for data harmonized from multiple federated sources. They facilitate large-scale analytics by providing a single repository where federated data can be accessed for visualization, modeling, and decision-making. For federated systems, the bi-directional integration of data lakes ensures that harmonized data can flow seamlessly between partners while supporting scalable and high-performance analytics tasks.

## 5.4 Data Warehouse

A data warehouse is a system designed specifically for the storage and analysis of structured data. Unlike data lakes, data warehouses operate with a schema-on-write approach, meaning that data must conform to a predefined schema before storage. This makes data warehouses ideal for high-performance analytics, reporting, and business intelligence applications, as the structured nature of the data enables efficient querying.

*Figure 25: Data Warehouse*



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

**Relevance to LFDM:**

Data warehouses serve as a high-performance layer within LFDM architectures, particularly for use cases requiring structured, aggregated, and historical data analysis. They integrate with federated systems by providing fast, consistent, and reliable access to harmonized data. In scenarios where federated systems must support decision-making through analytical dashboards or complex simulations, data warehouses prove indispensable.

## 5.5 Import Schema

An import schema defines the structure, format, and rules for ingesting data into a system. By standardizing the ingestion process, it ensures compatibility and consistency across diverse datasets. Import schemas are integral to ETL pipelines, enabling raw data from various sources to be transformed into structured formats that align with the target system's requirements.

**Relevance to LFDM:**

In federated systems, import schemas play a vital role in harmonizing data from distributed and heterogeneous sources. They ensure that incoming data aligns with predefined standards, making it easier to integrate and query within the federated ecosystem. By adhering to import schemas, LFDM can handle diverse data formats while ensuring consistency, interoperability, and compliance with regulatory standards.

## 5.6 FTP (File Transfer Protocol)

FTP is a standard network protocol used for transferring files between a client and a server. It is widely employed for moving large datasets, either as part of batch processing workflows or real-time pipelines. The simplicity and reliability of FTP make it an enduring choice for data transfer, particularly when other APIs or protocols are unavailable.


*Figure 26: File Transfer Protocol*

**Relevance to LFDM:**

In the context of LFDM, FTP provides a lightweight and straightforward mechanism for data exchange across distributed sources. It enables partners to upload or retrieve datasets from



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

federated repositories without requiring complex integration. When integrated into the LFDM architecture, FTP ensures efficient and secure file transfers, supporting collaboration across geographically and organizationally distributed partners. Its lightweight nature aligns with the principles of LFDM, making it suitable for integrating data from both legacy and modern systems.

## 5.7 Landing Functions

Landing functions are essential processes involved in staging raw data before it undergoes further transformations in an ETL pipeline. These functions typically include data validation and cleansing to ensure that only high-quality, usable data progresses through the system. Temporary storage in a designated "landing zone" allows the system to handle raw data securely and in an organized manner, serving as a buffer for subsequent processing steps.

**Relevance to LFDM:**

In a federated data management system, landing functions play a pivotal role in standardizing and preparing data from distributed sources. Since federated systems integrate data from multiple partners, the quality and format of incoming data can vary significantly. Landing functions address this variability by enforcing standardized ingestion processes, enabling seamless integration across nodes. This ensures that the federated system can operate efficiently and reliably, even when handling diverse datasets originating from different organizations.

## 5.8 Metadata

Metadata is data that describes other data, offering critical context about its structure, origin, and usage. It is categorized into three main types: descriptive metadata, which explains data content through tags or summaries; structural metadata, which defines how data is organized, such as schemas or hierarchies; and administrative metadata, which includes details like access rights and data provenance. Metadata serves as a blueprint, providing the necessary information to interpret and manage data effectively.

*Figure 27: Representation of Metadata*

**Types:**

- **Descriptive Metadata:** Explains data content (e.g., tags, summaries).
- **Structural Metadata:** Defines the organization of data (e.g., schema, hierarchy).
- **Administrative Metadata:** Includes access rights and provenance.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

**Relevance to LFDM:**

Metadata is foundational to the success of federated data management systems. It ensures data discoverability by enabling users to locate datasets across distributed repositories. Moreover, it supports interoperability by providing consistent definitions and context for data exchange between partners. In compliance with standards like FAIR, metadata is critical for federated queries and harmonization, allowing LFDM to achieve seamless integration of data from diverse sources.

## 5.9 Structured Data

Structured data is information organized in a predefined format, typically stored in rows and columns within relational databases. Its standardized organization makes structured data highly accessible and easy to manipulate using SQL-based tools and systems. Examples include transactional records, sensor readings, or inventory tables.

*Figure 28: Representation of Structured data*

**Relevance to LFDM:**

In federated systems, structured data is the easiest to query and integrate due to its inherent standardization. This enables partners to harmonize their datasets effectively, facilitating interoperability within the LFDM. Since structured data is often central to decision-making processes, LFDM can use it to support applications such as business intelligence, predictive modeling, and analytics with minimal additional processing.

## 5.10 Unstructured Data

Unstructured data lacks a predefined organization or format, encompassing diverse content types like videos, images, emails, social media posts, and free-form text. Unlike structured data, it cannot be directly stored in rows or columns, requiring advanced processing techniques, such as natural language processing or machine learning, to extract insights.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

*Figure 29: Representation of unstructured data*

**Relevance to LFDM:**

Unstructured data is increasingly valuable in federated systems due to its prevalence and potential insights. However, its integration poses significant challenges, requiring specialized workflows to align it with federated schemas. LFDM addresses these challenges by leveraging tools and technologies designed for unstructured data processing, enabling it to coexist with structured datasets. This capability broadens the applicability of LFDM to include multimedia analysis, social media monitoring, and AI-based applications.

## 5.11 SQL (Structured Query Language)

SQL is a standardized programming language for interacting with relational databases. It facilitates operations like querying, updating, and managing structured data. SQL supports both simple tasks, such as retrieving specific fields, and complex ones, like joining datasets or creating hierarchical structures.


*Figure 30: SQL*

**Relevance to LFDM:**

SQL is widely employed in federated data management systems due to its simplicity, flexibility, and universal acceptance. It enables LFDM to perform efficient queries across distributed databases, ensuring timely and accurate data access. The use of SQL in federated systems also supports interoperability, as it provides a common language for interacting with structured data across different partner nodes.

## 5.12 Queries

Queries are requests issued to a database or system to retrieve or manipulate data. They range from simple queries, which fetch specific records or fields, to complex queries, which involve operations such as aggregations, joins, or nested subqueries. Queries form the backbone of data access and analysis.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

*Figure 31: Representation of Queries*

**Types:**

- **Simple Queries:** Retrieve specific fields or records.
- **Complex Queries:** Join multiple datasets or perform aggregations.

**Relevance to LFDM:**

Efficient querying is essential for federated data systems, where data resides across multiple nodes and databases. In LFDM, queries ensure that users can interact with distributed datasets seamlessly, retrieving the necessary information for analysis without requiring centralized storage. The ability to handle both simple and complex queries makes LFDM versatile and applicable to diverse stakeholder needs.

## 5.13 Federated Queries

Federated queries are specialized SQL-like requests that access and combine data from multiple distributed sources within a federated system. Rather than centralizing data, federated queries distribute execution across participating nodes, where the results are gathered and merged into a single dataset for the user.

*Figure 32: Representation of Federated Queries*

**Relevance to LFDM:**

Federated queries are the cornerstone of Lightweight Federated Data Management, enabling seamless analysis of data across diverse, distributed systems. By allowing users to query multiple sources without needing to centralize the data, federated queries uphold data sovereignty while promoting collaboration. They ensure efficient, real-time access to integrated datasets, supporting analytics, reporting, and decision-making in a federated environment. This functionality embodies the core principles of LFDM, such as scalability, interoperability, and decentralized collaboration.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

This foundational understanding of data concepts provides the building blocks for implementing robust, efficient, and scalable federated data management systems.



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

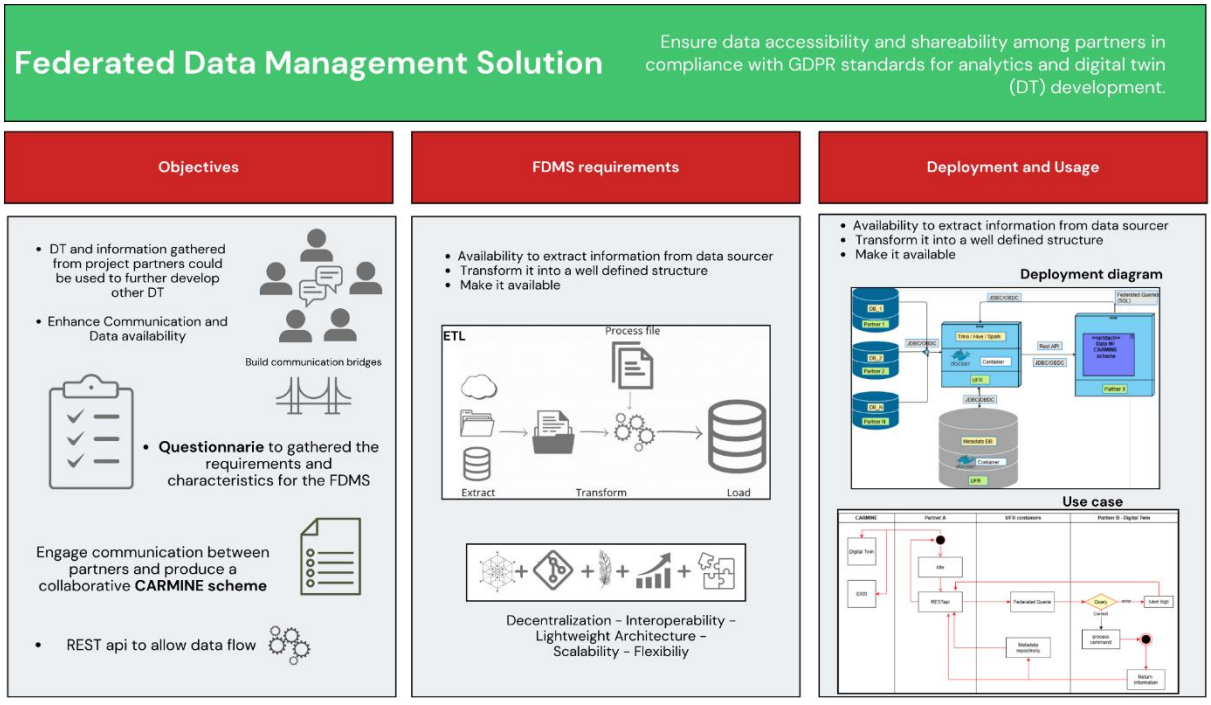
Co-funded by
UK Research and Innovation

# 6. Conclusion

In conclusion, after defining and describing the key components of a Federated Data Management System (FDMS), we leveraged the refined requirements gathered through surveys and interviews to propose an initial architecture (defined in section 4.8) for the Lightweight Federated Data Management Solution (LFDM). This architecture is designed with a lightweight approach to address the limited storage and processing capacities available among partners, avoiding the need for a centralized data warehouse. Instead, it ensures data remains hosted by each partner while integrating a metadata repository and the CARMINE scheme to enable seamless data access and retrieval across distributed sources. A simple use case scenario has been presented in section 4.4, that exemplifies how the FDMS would work once this project has been developed, this means that the communication between partners and their data sources will be swift and secure under the newly defined CARMINE scheme.

The architecture employs a basic batch ETL pipeline, as real-time processing exceeds the computational scope of this project. Batch processing within a time frame of minutes was deemed sufficient by most partners during the requirement-gathering phase. To facilitate data integration, federated queries are employed to access data directly from distributed sources, using open-source tools to ensure transparency, flexibility, and cost-effectiveness.

Compliance with GDPR is a cornerstone of the solution, ensuring adherence to data protection regulations. Security is further reinforced by implementing robust access controls to manage and restrict data access among partners. This architecture lays a strong foundation for an open, scalable, and secure federated system tailored to the needs and constraints of the participating stakeholders.




Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

# 7. ANNEX

**CARMINE Data Survey**

This form will be used for the creation of a new IT system for the CARMINE project. We are gathering information from all project partner institutions and we want to hear from you!

The idea of the Survey is to gather as much information as possible regarding each partner's Data Structure.

We want to understand:

What kind of data each partner is working with?

Where is this data stored?

How large is this data (and its expected growth)?

How do you handle the data and how would you like to handle it?

Do you have metadata for it ?

Are there any specific requirements for your Data  Etc?

Please enter all information to the best of your ability. The more information you provide, the easier it will be for all of us in the project.

Should you have any question, please contact andres.carranca@is.uni-freiburg.de

* Indicates required question



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

1. Email *

## 1. User Information

2. Name
3. Organization/Department
4. Position
5. Contact Information (Email/Phone)

## 2. General Data Needs

This section addresses data structure at input, expected delivery format, and its intended use, covering the practical aspects of the data.

6. What types of data do you currently use in your work? (Check all that apply) *
   - Tick all that apply.
   - Structured data (e.g., databases)
   - Semi-structured data (e.g., JSON, XML)

     Unstructured data (e.g., text files, images) Time

     Series
   - Geotagged data
   - Other:
7. What are your primary objectives for using the system? (Check all that apply) *

   *Mark only one oval.*
   - Data analytics
   - Reporting
   - Machine learning
   - Data archiving
   - Digital Twin
   - Other:
8. Please specify if **other** is selected or add any comment if necessary *



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

9. How frequently do you access data for your work? *

*Mark only one oval.*

- Daily
- Weekly
- Monthly
- Occasionally

10. What are your main sources of data? (Check all that apply)

*Tick all that apply.*

- Internal databases
- External APIs
- Third-party datasets
- Data from sensors or IoT devices
- Other:

## 3. Data Details

This section focuses on current and future data volume, aiming to determine the total available storage for the project.

11. Are there any public datasets that you use? If so, please provide the link to the dataset here.

12. Do you have any of these data storage systems?

*Tick all that apply.*

- Data Warehouse
- Data lake
- Database
- I do not know
- Other:

13. What is the expected volume of CARMINE data you will store in the system?

*Mark only one oval.*



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

- Small (< 100 GB)
- Medium (100 GB - 1 TB)
- Large (1 TB - 10 TB)
- Very Large (> 10 TB)

14. What is the expected growth rate of your data volume?

*Mark only one oval.*

- Slow (0 -10% per year)
- Moderate (10 - 30% per year)
- Fast (30 - 50% per year)
- Very Fast (>50% per year)

15. Do you have sufficient storage capacity for CARMINE data? *

*Mark only one oval.*

- Yes, but only for data concerning my part of the project.
- Yes, I have enough space for additional data.
- No, I don't have enough storage space.

16. If you have additional storage capacity. How much free space do you have?

*Mark only one oval.*

- <100 GB
- 100 GB - TB
- 1 TB - 10 TB
- > 10 TB
- No additional space

17. Do you have Metadata for your data?

(Database metadata refers to data that describes the structure of a database, including information about the tables, columns, indexes, constraints, and other elements that make up the database.)



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

*Mark only one oval.*

- Yes
- No

18. **(if yes)** Please describe the kind of metadata.

*Note*: A database schema defines the structure and organization of data, including table names, column data types, constraints, relationships between tables (e.g., foreign keys), and defined indexes and keys. It also specifies the physical file locations of the database, permissions and access controls for users, and a history of modifications, such as added or altered tables and columns.

## 4. Data Usage and Access

19. Who needs access to the data? (Check all that apply)

*Tick all that apply.*

- Individual users
- Teams
- Departments
- External partners

20. What are your data access requirements? (Note that Real Time access is an impossibility. It will be discussed with the partner should it be mandatory but the FDMS does not plan in it)

*Mark only one oval.*

- Real-time access
- Batch access -
- Hourly Batch access -
- Weekly Batch access -
- Monthly
- Other:

21. How do you prefer to access the data? (Check all that apply)

*Tick all that apply.*



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

- SQL queries
- API
- Data visualization tools
- Direct file access
- Other:

## 5. Data Security and Compliance

22. What are your data security requirements? (Check all that apply)

*Tick all that apply.*

- Data encryption
- Access control
- Data masking
- Audit logging

23. Are there any compliance requirements for your data? (e.g., GDPR, HIPAA)

*Mark only one oval.*

- Yes (Please specify)
- No
- I do not know

24. **(If yes)** Please specify compliance requirements for your data

## 6. Data Quality and Governance

25. Do you have any data governance policies that need to be followed?

*Mark only one oval.*

- Yes (please specify)
- No
- I do not know

26. **(If yes)** Please specify any data governance policies to be followed



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation

## 7. Additional Requirements and Comments

27. Are there any additional features or capabilities you require from the CARMINE IT system?

28. Do you have any other comments or suggestions regarding your data needs?

Thank you for your time!



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. UK participants in this project are co-funded by UK Research and Innovation (UKRI).
The work of the Meteomatics partner (Switzerland) has received funding from the Swiss State Secretariat for Education (SERI).

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

Co-funded by
UK Research and Innovation