CRMs DataSpace

RFCS PROJECT

Common European Data Space on Critical Raw Materials for the Green Deal: Recovery Potential from Closed Extractive Coal Waste Facilities

Grant Agreement 101216677

Deliverable 3.1: High-level data model conceptual design

Co-funded by the European Union

Research Fund for Coal & Steel

Authors

María del Carmen Feria, SISTEMAS AVANZADOS DE TECNOLOGÍA, S.A.

Shahved Katoch, SISTEMAS AVANZADOS DE TECNOLOGÍA, S.A.

Carlos Olivares, SISTEMAS AVANZADOS DE TECNOLOGÍA, S.A.

David Rojas, SISTEMAS AVANZADOS DE TECNOLOGÍA, S.A.

Deliverable 3.1	
Due date of Deliverable	30.09.2025
Start - End Date of Project	01.07.2025 – 30.06.2028
Duration	36 months
Deliverable Lead Partner	SATEC
Dissemination level	Public
Work Package	WP3
Digital File Name	D3.1 High-level data model conceptual design
Keywords	Critical raw materials, waste recovery, conceptual data model, European data spaces, semantic interoperability

Disclaimer

The information and photographs in this Deliverable remain the property of the CRMsDataSpace Project or its Partners. You must not distribute, copy or print this information.

Table of contents

EXECUTIVE SUMMARY 8

1 INTRODUCTION 9

2 DEFINITION OF THE REFERENCE MODEL 11

2.1 PRELIMINARY ANALYSIS OF STANDARDS AND REGULATIONS 11
2.1.1 EU REGULATIONS AND INITIATIVES 11
2.1.2 THEMATIC NORMS AND STANDARDS 12
2.1.3 STANDARDS FOR SPATIAL DATA AND METADATA MANAGEMENT 14
2.2 REFERENCE MODEL 16
2.2.1 THE ISO 19101 REFERENCE MODEL AND INSPIRE 17
2.2.2 RULEBOOK FOR SEMANTIC INTEROPERABILITY 18

3 MAIN COMPONENTS OF THE SOLUTION 20

3.1 DATA LIFECYCLE OVERVIEW 20
3.2 BASIC DATA MODEL CONCEPTS 21
3.3 INTERNATIONAL BENCHMARKING 22

4 DEFINITION OF METADATA PROFILE 26

5 HIGH-LEVEL APPLICATION SCHEMA 28

5.1 DESIGN METHODOLOGY 30
5.2 CONCEPTUAL DATA MODEL FOR INDUSTRIAL WASTE 30

6 GOVERNANCE AND INTEROPERABILITY 36

6.1 CONTROLLED VOCABULARIES 37
6.1.1 APPLICATIONS OF CONTROLLED VOCABULARIES 37
6.2 DATA ONTOLOGY 38
6.3 INTEROPERABILITY IN DATA SPACES 38
6.3.1 MECHANISMS FOR DATA CONSISTENCY 39
6.3.2 MECHANISMS FOR DATA INTEGRITY 40

7 KEY SUCCESS FACTORS & DEPENDENCIES WITH OTHER TASKS 42

7.1 AWARENESS OF THE MODEL AS A WORK IN PROGRESS 42
7.2 UNDERSTANDING THE IMPLICATIONS FOR STANDARD ADOPTION 42

7.3 DEPENDENCIES WITH OTHER PROJECT TASKS 43

8 EXPECTED BENEFITS & USE CASES 46

8.1 EXPECTED BENEFITS 46
8.1.1 ECONOMIC BENEFITS 46
8.1.2 ENVIRONMENTAL AND SOCIAL BENEFITS 46
8.1.3 GOVERNANCE AND POLICY BENEFITS 47
8.2 USE CASES 47
8.2.1 PROSPECTING AND FEASIBILITY ASSESSMENT FOR PROJECT PROMOTERS 47
8.2.2 REGIONAL DEVELOPMENT AND JUST TRANSITION PLANNING 48
8.2.3 ACADEMIC RESEARCH AND ENVIRONMENTAL IMPACT MODELLING 50
8.2.4 AI-POWERED PREDICTIVE MODELLING FOR RESOURCE ESTIMATION 51

9 CONCLUSIONS & NEXT STEPS 54

10 GLOSSARY 56

11 REFERENCES 60

List of Figures

Figure 2-1. High-level model components 17
Figure 3-1. High-level data lifecycle diagram 20
Figure 5-1. General Feature Model (INSPIRE/ISO 19109) 28
Figure 5-2. Adapted model of Observations and Measurements (based on INSPIRE D2.9 and ISO 19156, applied to the industrial waste domain) 29
Figure 5-3. Main thematic domains of the data model for Industrial Waste 31
Figure 5-4. Preliminary conceptual model of Industrial Waste (based on ISO 19109, INSPIRE, GeoSciML and O&M) 33
Figure 7-1. Dependencies of task T3.1 with other tasks 43
Figure 8-1. Sequencing diagram for prospecting and feasibility assessment use case.. 48
Figure 8-2. Sequencing diagram for regional development use case..... 49
Figure 8-3. Sequencing diagram for environmental impact modelling use case..... 50
Figure 8-4. Sequencing diagram for AI-Powered predictive modelling use case..... 51

List of Tables

Table 2-1. Application of ISO 19109 to the rulebook of semantic interoperability 19
Table 4-1. Main metadata and their application to Data Catalogue 26

Executive Summary

Deliverable 3.1 presents the **high-level conceptual design of the data model** for the CRMsDataSpace project. Positioned within Work Package 3, this deliverable establishes the foundations for the data catalogue dedicated to critical raw materials (CRMs), aligning with EU regulations, standards and interoperability frameworks. Since this is an early stage, the design should be considered a preliminary version subject to expansion and refinement as the project progresses.

The data model is developed after a thorough review of **widely used standards and European regulations** (INSPIRE Directive, ISO 19100 family, European Data Strategy and the Critical Raw Materials Act, among others). It defines fundamental data concepts while introducing a metadata profile aligned with ISO 19115 and DCAT-AP. The design is informed not only by normative frameworks but also by real-world sample datasets from waste deposits, thus ensuring its applicability to operational contexts.

The conceptual design outlined in this document covers the following key components:

- **A reference model** for semantic interoperability and standardized terminology.
- **A data lifecycle** from gathering and transformation to cataloguing, sharing, and consumption.
- **Controlled vocabularies and ontologies** to ensure semantic consistency.
- Preliminary **international benchmarking** against current initiatives on CRMs recovery.
- **Dependencies with related project tasks** to ensure continuity from conceptual design to implementation, quality assurance, governance, and platform development.

The project will deliver **economic** benefits (stimulating investment, improving market efficiency, and reducing import dependency), **environmental and social** benefits (enabling circular economy practices, supporting just transition in coal regions, and advancing land remediation), **and governance benefits** (streamlining compliance, supporting evidence-based policymaking, and fostering cross-border collaboration). Some **use cases** are included to illustrate this, including prospecting and feasibility studies for project promoters; regional development and just transition planning; academic research and environmental modelling; and AI-powered predictive modelling for CRMs estimation.

Deliverable 3.1 marks a **critical step in laying the conceptual and semantic foundation of the project**. While preliminary, it sets the stage for interoperable, standards-based data catalogues that will unlock scientific, industrial, and policy value across Europe. **Continuous refinement will ensure alignment** with project requirements, evolving standards, and stakeholder needs.

1 Introduction

Among the several challenges identified in the project proposal, **Challenge 1** is "To develop the first **data catalogue** that leverages metadata and data management tools focused on critical raw materials".

The goal of **Work Package Nº3** is the development of this data catalogue, its specific objectives being as follows:

1. Develop a high-level conceptual design after analysing European norms and standards applicable to the data and data spaces that are the object of this project.
2. Define the fundamental components of the data model.
3. Characterise the different phenomena with their attributes, relations, restrictions, controlled vocabulary and multimedia.
4. Develop a rulebook for the data and metadata sets to increase consistency and ensure comparability.
5. Publish detailed catalogues addressing data consumers and suppliers, providing interoperability support for the data and services, and a catalogue of requirements and mechanisms to guarantee data quality, integrity, and consistency.

Deliverable 3.1 describes the fundamental components of the data model that will serve as the initial basis for the creation of the databases and data space that are the object of this project.

This deliverable is an outcome of **Task 3.1 – High-level conceptual design**, within which SATEC is developing a high-level conceptual design after analysing norms and standards, mainly European, applicable to the data and data spaces that are the object of this project (ISO 19100 family, INSPIRE Directive and European Data Strategy), defining the fundamental components of the data model: phenomenon, property, attribute, inheritance, relation, restriction, observation, measurement, digital object, and defining the metadata profile.

In addition to the established norms, standards and regulatory frameworks, **the use of real-world sample datasets obtained from waste deposit sites will be of critical importance for informing and refining the data model**, thus ensuring a comprehensive and practical design. Empirical data not only support the validation of theoretical structures and assumptions, but also facilitate the identification of edge cases, inconsistencies, and context-specific variables that are unlikely to emerge from abstract or normative references alone.

This will enable the development of a more robust, context-aware, and operationally relevant data model, better aligned with real-world conditions, especially the

complexities and heterogeneity inherent in actual waste management practices. All of this will ultimately enhance the accuracy, relevance and usability of the data model.

SATEC will lead this task. **GIG-PIB, UNIOVI, INSEMEX and other parties and collaborators will cooperate** with SATEC based on their expertise in coal mining.

Finally, it must be noted that the data model conceptual design submitted at this stage is preliminary and may be subject to changes as the project progresses. As development advances and requirements evolve, adjustments may be necessary to ensure alignment with project needs, data availability and technical constraints. This model should therefore be **considered a first version rather than a final specification**.

2 Definition of the reference model

A reference model is an abstract, conceptual framework that outlines the key entities, relationships, and rules relevant to a particular domain, serving as a common foundation for the design of specific data models. It promotes consistency, interoperability, and alignment with domain standards by providing reusable structures and terminology, while remaining flexible enough to be adapted or extended based on the specific requirements of a given implementation.

In the context of CRMs, the reference model is a domain-specific conceptual framework that establishes a standardized vocabulary, structure and set of relationships for representing data related to the lifecycle management of CRMs. It serves as a guide to the development of interoperable, compliant, and scalable data models, ensuring alignment with regulatory frameworks, environmental reporting requirements and industry best practices. The reference model facilitates data integration across stakeholders (mining operators, environmental agencies, policymakers, etc.) by promoting semantic consistency throughout the CRMs value chain.

2.1 Preliminary analysis of standards and regulations

Before defining the reference model, it is necessary to conduct a preliminary analysis of the standards and EU regulations related to data spaces, CRMs, data and metadata management and geographical information. This will contribute to ensure data quality and multi-level interoperability and will obviously provide the essential base for the development of subsequent activities further in the project.

2.1.1 EU regulations and initiatives

The main EU regulations and initiatives providing the legal framework for management of thematic and spatial information, data space, CRMs, waste and digital services are as follows:

1. **European Geological Data Infrastructure (EGDI)**, a pan-European platform designed to provide access to geological and geospatial data collected and maintained by EuroGeoSurveys, aiming to support research, policy-making and industry by making harmonized, interoperable geological data available on a European scale.
2. **Directive 2007/2/EC – INSPIRE**, establishing an infrastructure for enabling cross-border data sharing and provide interoperable services in support of policy makers and implementers, as well as business, science, and citizens.
3. **European Data Strategy, which envisions the creation of Common European Data Spaces** in strategic domains, meant to enable secure, interoperable data sharing among stakeholders while respecting EU values.

4. **Regulation 2024/1252 - Critical Raw Materials Act (CRMA)**, establishing a framework to ensure a secure, resilient and sustainable supply of critical raw materials for the EU for a secure and sustainable supply of critical raw materials.
5. **Directive 2006/21/EC on the Management of Waste from Extractive Industries**, establishing the legal requirements for how waste generated by extractive operations (prospecting, extraction, treatment, storage, working of quarries) should be managed to protect human health and the environment.
6. **Directive 2019/1024 on Open Data and the re-use of public sector information**, aiming to strengthen the EU data economy by making more public sector data available for reuse, under conditions that facilitate innovation, transparency and competition.
7. **Regulation 910/2014 eIDAS on electronic identification and trust services for electronic transactions**, establishing a legal framework for electronic identification (eID) and trust services across the EU internal market, and aiming to ensure that electronic transactions have a legal status equivalent to paper-based ones, with sufficient security and cross-border interoperability.

2.1.2 Thematic norms and standards

The main thematic norms and standards of global use in geological and mining data modelling, with areas focused on mineral resources and waste management, are as follows:

1. **INSPIRE Directive Data Specifications Technical Guidelines**. The INSPIRE Directive includes several Implementing Rules on interoperability of spatial data sets and services and Technical Guidelines (Data Specifications) that specify common data models, code lists, map layers and additional metadata on the interoperability to be used when exchanging spatial datasets.
   The INSPIRE Directive includes 34 spatial data themes, each one having its specific Technical Guideline, whose implementation maximises the cross-border and cross-thematic interoperability of data and services, even with other sectors.
   In this project, the main Data Specifications Technical Guidelines of interest are:
   a. **D2.8.II.4 – Geology.**
   b. **D2.8.III.8 – Production and Industrial Facilities.**
   c. **D2.8.III.11 – Area Management/Restriction/Regulation Zones and Reporting Units.**
   d. **D2.8.III.21 – Mineral Resources.**

   Other Data Specifications Technical Guidelines will be consulted when needed. Guidelines for the use of Observations & Measurements and Sensor Web Enablement-related standards will be consulted as well.

2. **OGC GeoSciML (Geoscience Markup Language)** is an international, open standard developed by the Open Geospatial Consortium (OGC) to enable the exchange and interoperability of geoscientific information, including geological and mining data. It provides a common vocabulary and data model for representing complex geological features such as rock units, mineral occurrences, faults, and boreholes.
   GeoSciML facilitates sharing, integration, and visualization of geological datasets across different organizations and software platforms, supporting applications in mineral exploration, environmental assessment, resource management, and research. Its structured, machine-readable format ensures consistent, detailed, and semantically rich descriptions of subsurface geology essential for informed decision-making in mining and geoscience.

3. **EarthResourceML (ERML)** is a specialized data exchange standard developed to represent and share information about earth resources, including mineral deposits, mineral occurrences, commodities, mining activities, and exploration results.
   Built on the OGC's GeoSciML framework, ERML enables the interoperable transfer of resource-related geoscientific data between geological surveys, governments, and industry stakeholders. It supports standardized, detailed descriptions of resource types, deposit models, economic attributes, spatial locations, and associated exploration or extraction activities.
   ERML is especially valuable for cross-border mineral resource assessments, facilitating harmonized reporting and integration in international databases such as those used by the EU's Raw Materials Information System (RMIS) and the European Geological Data Infrastructure (EGDI).

4. The **Exploration, Mining, Metals & Minerals Vertical (EMMM)** Forum is a global, vendor-neutral environment where members create Enterprise Architecture standards and reference models for the exploration, mining, metals and minerals industries.
   This initiative focuses on developing and promoting interoperable, standards-based data models and ontologies for the entire lifecycle of mineral resource development.
   The EMMM Forum brings together industry stakeholders, technology providers, standards bodies, and research organizations to collaboratively define and align data interoperability frameworks, with the goal of enabling seamless data exchange, integration, and automation across systems used in the mining and minerals value chain.

5. The **United Nations Framework Classification for Resources (UNFC)** is a globally applicable classification system developed by the United Nations Economic

Commission for Europe (UNECE). It provides a standardized framework for classifying and reporting natural resources based on three key criteria: Economic and social viability (E-axis), Field project feasibility (F-axis) and Geological knowledge (G-axis).
This 3-dimensional system (E-F-G) allows for a structured, transparent evaluation of a resource's potential to be developed sustainably and responsibly.

2.1.3 Standards for spatial data and metadata management

The main standards for spatial data and metadata management, globally used in catalogues and interoperable web portals and services, are as follows:

1. The **ISO 19100 family** is a comprehensive suite of international standards that define methods, models, and services for the representation, management, exchange, and use of geographic (spatial) information across various domains, including geology, mining, environment, and infrastructure.
   These standards ensure semantic and syntactic interoperability of geospatial data and systems; provide standardized vocabularies, metadata schemas, coordinate reference systems and data models; enable integration of geospatial data from multiple sources and domains (environmental, geological, cadastral, etc.); and facilitate discovery, access, sharing and re-use of geospatial data.
   Regarding geological and mining data, the ISO 19100 series provide the foundation for geoscience-specific models like **GeoSciML** and **EarthResourceML**, ensuring compatibility with broader ecosystems; support data quality, traceability, and documentation, which are crucial in exploration, environmental impact assessments and regulatory reporting; and enable the integration of mining data into spatial data infrastructures (), such as **INSPIRE** or **EGDI**, through common schemas and metadata.
   The ISO 19000 standards of interest are ***ISO 19115 Geographic information – Metadata, ISO 19156 Geographic information — Observations, measurements and samples*** and ***ISO 19157 Geographic information – Data quality***. *ISO 19101 Geographic information — Reference model* and *ISO 19109 Geographic information — General feature model and rules for application schema* will also be analysed for more specific purposes.

2. The **Open Geospatial Consortium (OGC)** is an international, not-for-profit standards organization that develops and promotes open, interoperable standards for geospatial content and services, including geographic information systems (GIS), remote sensing, and location-based technologies. It brings together governments, industry, academia, and s to advance the use of open standards that enable the sharing and integration of geospatial data across platforms, domains, and borders.

Its most important contributions are the development of widely adopted standards such as support for interoperability and real-time data exchange between spatial data infrastructures (SDIs) and promotion of modular and extensible data models, used in specialized domains like geology, hydrology, urban planning, and environmental monitoring.
Regarding geological and mining data, OGC standards underpin **GeoSciML** and **EarthResourceML**, enabling interoperable geological and mineral resource data exchange; empower cross-border data integration via SDIs like **EGDI** or **INSPIRE;** support mining workflows including resource mapping, environmental impact monitoring, and regulatory reporting; and facilitate machine-readable, API-based access to geospatial and sensor data in exploration and mining contexts.
Consequently, the OGC provides the technical foundation for sharing and leveraging geospatial data across systems, making it a critical enabler for data-driven innovation in the geological and mining sectors.

3. The **Dublin Core Metadata Initiative (DCMI)** is an international organization that develops and maintains the Dublin Core Metadata Element Set, a standardized set of metadata terms used to describe digital and physical resources in a simple, consistent, and interoperable way.
   Originally developed for web resources, Dublin Core has become a foundational metadata standard used across many domains including libraries, archives, research data repositories, and increasingly, geospatial and scientific data infrastructures. It features several core metadata elements and it is designed to support interoperability, discoverability and reuse of information assets.
   Dublin Core is commonly used to describe datasets, maps, reports, samples, and other digital assets in geoscience data catalogues. It also forms the foundation for metadata profiles in systems such as **INSPIRE** metadata specifications, **EGDI**, and **GeoSciML** and **EarthResourceML**-based catalogues.

4. **W3C DCAT (Data Catalog Vocabulary)** is a metadata standard developed by the World Wide Web Consortium (W3C) to facilitate interoperable publishing, sharing, and discovery of datasets across the web, particularly in the context of data catalogues and open data portals. DCAT enables datasets from different sources (including governments, research institutions, and industry) to be described using a common, machine-readable vocabulary, making them easier to discover, access, and link.
   It defines standard classes and properties for catalogues, datasets and data services, enabling semantic interoperability and integration with Linked Open Data. In addition, it is aligned with other metadata standards such as **DCMI**. DCAT is widely used in national and international geoscience data portals, such as **INSPIRE** Geoportal, **EGDI** and **EU Open Data** Portal.

Concerning geological and mining data, it supports standardized, discoverable metadata for datasets related to geological maps, mineral occurrences, boreholes, mining waste, resource assessments and environmental monitoring. It enables integration with broader data space initiatives through semantic web technologies and facilitates automated harvesting and federated search across multiple catalogues.
Therefore, DCAT Ensures datasets are findable and accessible across institutional and national boundaries; supports data governance, provenance, licensing, and versioning; and promotes data reuse and interoperability. This makes DCAT a cornerstone standard for making geological and mining datasets visible and usable on the web, particularly in multi-stakeholder, cross-border data ecosystems.

2.2 Reference model

The reference model is defined in *ISO 19101 Geographic information — Reference model* as “a framework for understanding significant relationships among the entities of some environment, and for the development of consistent standards or specifications supporting that environment”. This model provides the structured conceptual foundation for standardisation in geographic information, describing how data, metadata, services, and reference systems are interrelated.

In line with the project proposal, Figure 2 of ISO 19101, where the High-Level Model Components, is adopted as the starting point for the conceptual modelling of this project and has been adapted to its thematic scope. This diagram illustrates the relationships between the different semantic levels defined by ISO standards, from metamodels and feature catalogues to application schemas and datasets.

It also shows how semantics flow from abstract concepts (meta levels) to data instances, ensuring consistency between conceptual definitions and their implementation in interoperable systems.

For the purposes of this project, the original ISO 19101 framework has been adapted to the project's subject area, while maintaining its structural logic. The resulting model (see figure below) shows how real-world phenomena are abstracted into entities or ‘features’ with properties and attributes, which are then organised into application schemas and materialised as datasets accessible through the data catalogue and associated services.

This high-level reference model, based on ISO 19101, supports the understanding and development of geographic information systems and services. It provides a coherent conceptual structure that identifies the areas of standardisation required and

establishes the relationships between the different components of the model: metamodels, feature catalogues, application schemas and datasets.

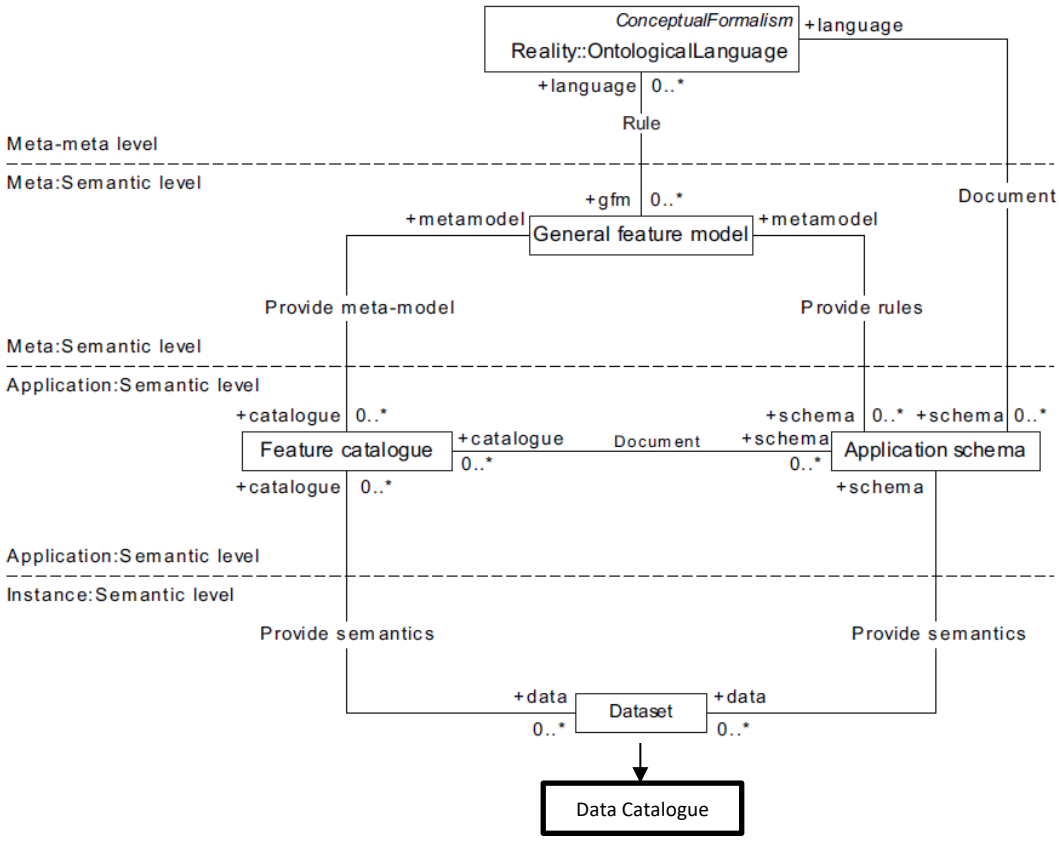


Figure 2-1. High-level model components

This framework also promotes technical and semantic interoperability between different domains and systems, ensuring that the data and services developed within the project can be effectively integrated into broader spatial infrastructures, in accordance with the principles and standards defined by INSPIRE and the European Data Strategy.

2.2.1 The ISO 19101 reference model and INSPIRE

The description of the reference model in ISO 19101 includes the fundamental concepts and principles for modelling geographic information, as well as a structured framework for standardization, which models must be created and how they relate to each other. In addition, ISO 19101 defines the types of standards required, including data models, metadata, spatial and temporal reference systems, quality principles and service

architectures. All these components are coordinated within what ISO defines as the General Reference Model (part of the ISO 19101 Reference Model).

The General Reference Model distinguishes between semantic levels, providing a logical structure for the flow of semantics from abstract concepts to data instances. It also integrates several complementary perspectives that define its conceptual structure:

- **Enterprise:** purpose, scope and policies (why the system exists).
- **Information:** structure and semantics of the geographic data.
- **Computational:** functional decomposition and service interactions.
- **Engineering:** infrastructure and communications.
- **Technology:** specific technologies used.

The Reference Model also sets out the relationships between different ISO 19100 standards, such as *ISO 19109 Geographic information — General feature model and rules for application schema* (for modelling thematic data) and *ISO 19115 Geographic information – Metadata* (for discovery services). The guidelines of ISO 19109 are particularly considered in the following sections:

- **7.3:** principles for the definition of feature types and their conceptual structure.
- **7.4:** rules for inheritance and specialization, enabling the modelling of hierarchies between entities.
- **7.5:** definition of properties, distinguishing between attributes and associations, and introducing constraints to ensure consistency.

The combination of the Reference Model from ISO 19101 (see previous figure) and the General Feature Model from ISO 19109 (sections 7.3, 7.4, and 7.5) constitutes the reference model adopted by this project, ensuring conceptual coherence, interoperability, and compatibility with both INSPIRE and the European Data Strategy.

The INSPIRE Directive adopts and implements this reference model based on ISO standards to build a pan-European spatial data infrastructure. Its architecture, data specifications and services are structured in accordance with ISO 19101, ensuring international compatibility and the effective exchange of spatial information across Europe.

2.2.2 Rulebook for semantic interoperability

A rulebook will be developed to:

- Provide a **common guide** within the project ecosystem.
- Minimize ambiguities and ensure the **semantic consistency** between different systems.

- Serve as a reference for **data harmonization** and the development of interoperable tools.

Several sections of ISO 19109 will be used as a guide during the making of the rulebook, as explained in the following table:

Table 2-1. Application of ISO 19109 to the rulebook of semantic interoperability

Rulebook section	ISO 19109 section	Application of ISO 19109
Default reference systems and units of measure	7.4 The General Feature Model	Definition of accepted spatial and temporal reference systems (CRS, datum, projections); standard units of measurement (SI, derived units).
Data content	7.3 Features and the application schema	Basic UML notions: explanation of UML as a modelling language. Application schemas: description of application models, feature catalogue and code lists. Minimum dataset contents: specification of minimum contents that a dataset must have as a deliverable.
Data Quality	7.5 Attributes of feature types	Quality parameters (positional accuracy, completeness, logical consistency, timeliness, etc.), and methodologies for quality control and assurance methodologies.

This Rulebook, aligned with ISO 19109, will ensure that the conceptual model developed in this project can be effectively implemented in interoperable data catalogues and services, while maintaining semantic integrity, technical consistency, and compliance with European standards such as INSPIRE.

3 Main components of the solution

3.1 Data lifecycle overview

The following figure features a high-level representation of the data lifecycle in the context of the project.

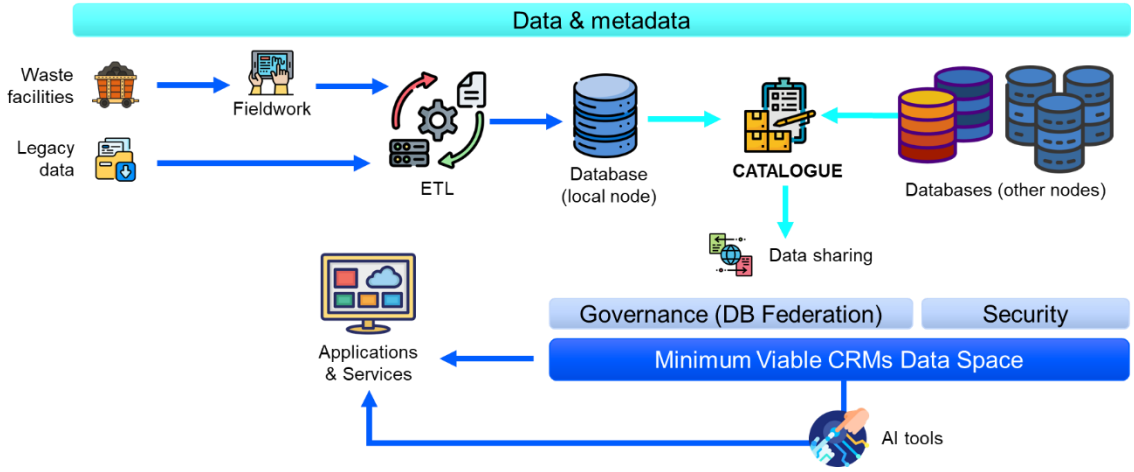


Figure 3-1. High-level data lifecycle diagram

The stages of this workflow are as follows:

1. **Data gathering** is the initial stage, focused on collecting information from various sources, which can be categorized into two main types:
    a. Legacy data, consisting of historical datasets preserved from earlier systems or processes. These provide valuable context, continuity and insights to inform and support the development of the data model. However, such data often exist in outdated formats or systems and may require transformation before reuse.
    b. New data collected through on-site surveys at waste facilities, using a purpose-built application developed for the project. This approach not only allows to generate additional information but also serves to test and refine the new data-gathering methodology aligned with the designed data model.

2. **ETL** is the process of extracting data from various sources, transforming it to meet quality standards and format requirements, and loading it into a centralized database. This step is particularly critical when dealing with legacy data, since they often require extensive cleaning and conversion due to differences in structure, quality or storage medium (paper records, spreadsheets, obsolete software systems, etc.).

3. **Data ingestion** is the formal upload of processed data into the target database or platform. Once ingested, data become accessible to authorized users, forming the basis for further use and analysis.

4. **Data cataloguing** stage involves the systematic documentation and labelling of datasets. A robust data catalogue ensures that users can discover, understand, and trust the data, thanks to well-maintained metadata, standardized descriptions and traceability of data origins and transformations.

5. **Data sharing** enables secure and governed access to federated datasets across different organizations or systems. Each data owner retains full control over their data, while robust governance, security protocols, and access policies ensure compliant, transparent and trustworthy collaboration.

6. **Data consumption** stage involves the responsible and policy-compliant use of shared data across the federated ecosystem. Consumers access data on demand, guided by usage rights, governance rules, and security standards, ensuring value-driven, ethical, and sustainable data use.

7. **AI-assisted data services**: AI may increase data value of data ecosystems by enabling smart data discovery, automated data classification, anomaly detection or real-time insights, among other features. When embedded within strong governance and security frameworks, AI tools help users navigate complex data landscapes efficiently, supporting privacy, access control and usage compliance while fostering scalable, intelligent and trusted collaboration.

3.2 Basic data model concepts

When designing a conceptual data model, it is essential to clearly establish the fundamental components that structure it. The concepts of phenomenon, property, attribute, inheritance, relation, restriction, observation, measurement and digital object provide the common vocabulary and semantic basis necessary to ensure the consistency of the model and its alignment with international standards such as the ISO 19100 family and INSPIRE.

These basic concepts are defined as follows:

- **Phenomenon**: any entity or process in the real world that can be observed, described, or modelled. It represents the object of interest in the application domain (*ISO 19101 Geographic information — Reference model*).
- **Property:** An abstract characteristic that describes an aspect of a geographical phenomenon or feature. It defines the type of information that can be associated

with a feature. Properties establish the semantic structure of the conceptual model. (*ISO 19109 Geographic information — General feature model and rules for application schema*).

- **Attribute:** representation of a property in the data model, usually expressed by a specific value (*ISO 19109 Geographic information — General feature model and rules for application schema*).
- **Inheritance:** mechanism by which one phenomenon or entity can derive from another, sharing its properties and attributes, and allowing for specialisation (*ISO 19109 Geographic information — General feature model and rules for application schema*).
- **Relation:** expression of how two or more phenomena are linked within the model. This can be dependency, association, spatial inclusion, temporal inclusion, or causal inclusion (*ISO 19101 Geographic information — Reference model*).
- **Restriction:** condition or rule that limits the possible values of an attribute or relationship, ensuring the consistency and quality of the model (*ISO 19109 Geographic information — General feature model and rules for application schema*).
- **Observation:** the act of collecting data on a phenomenon through a defined procedure (*ISO 19156 Geographic information — Observations, measurements and samples*).
- **Measurement:** value resulting from an observation, expressed on a specific scale or unit (*ISO 19156 Geographic information — Observations, measurements and samples*).
- **Digital object:** electronic representation of a phenomenon or set of observations/measurements, stored and managed in an information system (Directive 2007/2/EC – INSPIRE, Annex II).

3.3 International benchmarking

This section provides a preliminary analysis of international initiatives and projects related to the recovery of CRMs from coal waste, mine tailings, and coal combustion by-products. The analysis covers activities in the European Union, the United States, Canada, and Australia, highlighting the scope of work, the degree of technological maturity, and the relevance of CRMs within the broader circular economy agenda. The territories outside the EU have been selected because they share a strong mining heritage, where the legacies of coal exploitation and its associated residues represent both environmental challenges and opportunities for secondary resource recovery.

1. **European Union.** Within the European Union, the Critical Raw Materials Act (CRMA) establishes a list of 47 strategic projects designed to reinforce the EU's autonomy in securing CRMs supply. However, none of these initiatives

specifically address the recovery of CRMs from coal heaps, Tailing Storage Facilities (TSF) from coal mining, or coal combustion residues.

Beyond the CRMA, there are several research and development projects that provide important insights. One is REECOVERY (European Institute of Innovation and Technology - EIT Raw Materials), which focuses on recovering rare earth elements (REE), cobalt, zinc, and copper from mine waters and drainage systems in Los Frailes, Aznalcóllar, Spain. Its main contribution lies in validating technologies that transform polluted waters into valuable sources of critical materials, addressing both environmental risks and resource supply.

Another initiative is RAWMINA (Raw Materials Innovation for the Circular Economy: Sustainable Processing, Reuse, Recycling and Recovery Schemes, Horizon 2020), which develops a circular economy approach to mining waste management. The project includes pilot-scale demonstrations of technologies for metal recovery while also prioritizing the reduction of water and energy consumption. Its goal is to create scalable and sustainable processes that allow mining residues, including coal-related waste, to be valorised as a resource.

In Spain, The TI-RRES project (NextGenerationEU TI-RRES) carried out by the Spanish Geological and Mining Institute (Instituto Geológico y Minero de España, IGME-CSIC) aims to conduct a comprehensive study of the mining waste accumulated in the tailings and coal heaps of the Sn-Ta (tin–tantalum) mine in Golpejas, Salamanca. Its objective is to evaluate their potential for the recovery of rare earth elements (REE) and other CRMs, such as tantalum and niobium, as well as to determine the possibilities of utilizing industrial minerals. In this way, the project seeks to generate economic, environmental, and social benefits in the area by facilitating the restoration of land affected by mining waste facilities to a condition suitable for the establishment of a self-sustaining ecosystem over time. The initiative represents a clear example of Circular Economy and Ecological Restoration, both of which are pillars of the ecological transition demanded by the EU and society at large. By reintegrating the materials contained in mining waste into the production process, the project contributes to the generation of new products or secondary raw materials.

The SCIMIN-CRM project (Sustainable & Circular Production of Mineral Critical Raw Materials, 2024) is an EU-funded initiative under the Horizon Europe programme. Led by the National Association of Aggregates Manufacturers (ANEFA, Spain), its objective is to convert previously discarded mining waste facilities—such as tailings, stockpiles, and spoil heaps—into recoverable sources of CRMs. The project is designing a digital collaborative platform, implementing mobile crushing technologies, and developing four advanced processing methodologies. Its impact targets include increasing valorisation from near 0 % to 5 % of waste material and reducing evaluation times by 80 %. SCIMIN-CRM has pilot sites in Spain, Sweden, Austria, and Bosnia & Herzegovina, and aims to

promote sustainable mining practices, circular economy principles, and Europe's strategic autonomy in CRMs.

2. **United States of America**. The United States of America stand out as the most advanced region in terms of funded and production-level projects targeting CRMs recovery from coal waste. The DOE–NETL REE/CM Program (Department of Energy – National Energy Technology Laboratory Rare Earth Elements/Critical Materials, 2024) is a flagship initiative that has provided significant financial support to pilot plants, process scale-ups, and demonstration projects. The program's mission is to secure a domestic supply of rare earths and other critical materials by exploiting coal refuse, ash, overburden, and acid mine drainage (AMD). It not only promotes technological innovation but also strengthens national resilience against supply chain disruptions.
   Among the most prominent research and pilot projects is the work conducted at West Virginia University (WVU), through its Water Research Institute. Here, researchers have developed methods to recover REE from acid mine drainage, converting an environmental liability into a potential revenue stream. Supported by DOE and Environmental Protection Agency (EPA) funding, WVU has progressed from laboratory experiments to pre-commercial demonstration plants.
   The University of Kentucky's Center for Applied Energy Research (UK/CAER) has also played a critical role by developing pilot processes that extract rare earths from coal ash and coal-derived materials. Their work has successfully produced mixed rare earth oxides with purity levels of up to 98 percent, providing strong evidence for the economic feasibility of coal waste valorisation (UK/CAER - Rare-Earth Elements in US Coal-Based Resources: Sampling, Characterizations, and Round-Robin Inter-laboratory Study).
   In parallel, new DOE announcements for 2024–2025 under the Bipartisan Infrastructure Law provide fresh rounds of funding to expand this field, supporting projects that lower costs, improve environmental performance, and strengthen industrial supply chains for CRMs recovery from coal residues (DOE Funding Notice: Bipartisan Infrastructure Law, 2024).

3. **Canada**. In Canada, the recovery of CRMs from coal residues remains primarily at the research stage. Teams working with the Canadian Light Source and institutions in Alberta and Saskatchewan have carried out analytical campaigns to assess the concentrations of REE in coal ash. These studies employ advanced spectroscopic and mineralogical techniques to characterize the potential of these residues (Canadian Light Source, 2024).
   At the policy level, the Government of Canada's Critical Minerals Strategy explicitly acknowledges coal ash as a potential domestic resource for rare earths and scandium. Federal funding has been directed towards projects that

characterize these materials and explore suitable extraction technologies, although no large-scale pilots have yet been launched (Government of Canada, 2023).

Recent academic literature in Western Canada also points to the promising role of coal combustion by-products. Some studies combine geochemical assessments with machine learning methods to predict areas with higher REE recovery potential. While still at a low technology readiness level, this body of work establishes a knowledge base for future pilot-scale projects (Brendan et al, 2023).

4. **Australia**. Australia, like Canada, is currently focused on research and early-stage development. The Commonwealth Scientific and Industrial Research Organisation (CSIRO) has launched programs on "Rare Earths in Coal Waste," aimed at characterizing and testing methods for extracting REE from coal tailings and waste streams. A distinctive element of this initiative is its integration with existing coal wash plants, which could facilitate eventual industrial implementation (CSIRO, 2024a).

   In addition, the creation of national hubs for critical minerals R&D has expanded Australia's capacity to explore different supply pathways. Although much of the focus remains on clay-hosted rare earth deposits, these hubs also consider recovery from coal residues and urban mining as complementary strategies (CSIRO, 2024b).

   Australian universities, particularly Monash University and the Sustainable Minerals Institute at the University of Queensland, are investigating laboratory-scale processes for extracting rare earths from coal ash and tailings. These projects have demonstrated recovery efficiencies of over 90 percent, showcasing strong potential for scaling up (Monash University, 2025).

4 Definition of metadata profile

Metadata is the essential element that ensures discovery, location, understanding, and evaluation of datasets. Through structured and standardized descriptions, metadata allows users to identify available resources, learn about their main characteristics and determine their suitability for a specific use. Furthermore, they form the basis for ensuring technical and semantic interoperability between different data providers and consumers, which is critical in the context of European data spaces and the INSPIRE Directive.

The metadata profile should contribute to defining a sustainable metadata management framework, so that the data catalogue can be kept up to date, consistent, and aligned with international standards over time.

The following table summarizes the relevant metadata elements and their application to the Metadata Catalogue (adapted from INSPIRE/ISO 19115/DCAT-AP):

Table 4-1. Main metadata and their application to Data Catalogue

Metadata element (INSPIRE/ISO/DCAT-AP)	Application to Metadata Catalogue
Identification	Sheet with title, summary, normalized keywords
Geographical / Temporal extension	Bounding box, accepted temporal systems (ISO 8601/UTC), resolution
Data quality	Completeness degree, consistence and accuracy
Access and use	Licenses, legal restrictions and privacy
Conformity	Alignment degree with INSPIRE and Rulebook from T3.3.S1
Point of contact	Providing organization and technical manager
Access services	Links to WMS, WFS, APIs REST, downloads
Update and maintenance	Update frequency and responsible entity
Lineage	Traceability of the complete life cycle of the data, from its origin through all transformations and movements to its final destination

This set of elements ensures that the catalogue complies with INSPIRE's basic metadata requirements, while promoting interoperability, traceability and sustainability of information. It also enables the integration and reuse of data in INSPIRE- and DCAT-AP-compliant catalogues, contributing to the effective exchange of information within the European Data Spaces.

These metadata elements are defined in accordance with ISO 19115 and the INSPIRE Implementing Rules on Metadata (Regulation No 1205/2008) and are further aligned with the DCAT-AP profile to ensure cross-portal compatibility and harmonized discovery of datasets across European infrastructures.

The metadata profile will adopt ISO 19115 Geographic information – Metadata as a reference, aligned with INSPIRE implementation rules on Metadata (Regulation No 1205/2008) and the DCAT-AP profile, in which the specification of the Data Catalogue Vocabulary Application Profile for data portals in Europe is based so as to enable cross-data portal search for data sets and make public sector data better searchable across borders and sectors. This alignment ensures compatibility with European Spatial Data Infrastructures (SDI) and the correct publication of resources on the INSPIRE Geoportal.

INSPIRE defines a core metadata set that includes mandatory elements for spatial data sets and services, such as identification, conformity, access and use conditions, spatial and temporal extent, responsible organisation, and metadata date. The project's metadata profile will extend this basic set by incorporating additional elements from ISO 19115 and DCAT-AP, in order to ensure interoperability with other European data catalogues and facilitate integration into thematic data spaces.

5 High-Level Application Schema

The High-Level Application Scheme defines the structure and semantic organisation of the data model that will guide the description, exchange and interoperability of information within the project. It is developed in accordance with the principles established in *ISO 19109 Geographic information — General feature model and rules for application schema* and is aligned with the General Feature Model (GFM) defined by ISO and INSPIRE.

This scheme provides a bridge between the conceptual model (ISO 19101) and the data implementation level, identifying the main feature types, their properties, attributes, relationships and constraints, ensuring conceptual consistency and compliance with European spatial information standards.

The following figure shows the General Feature Model (GFM) according to INSPIRE / ISO 19109, which establishes the reference framework for object-based modelling in geographic information systems:

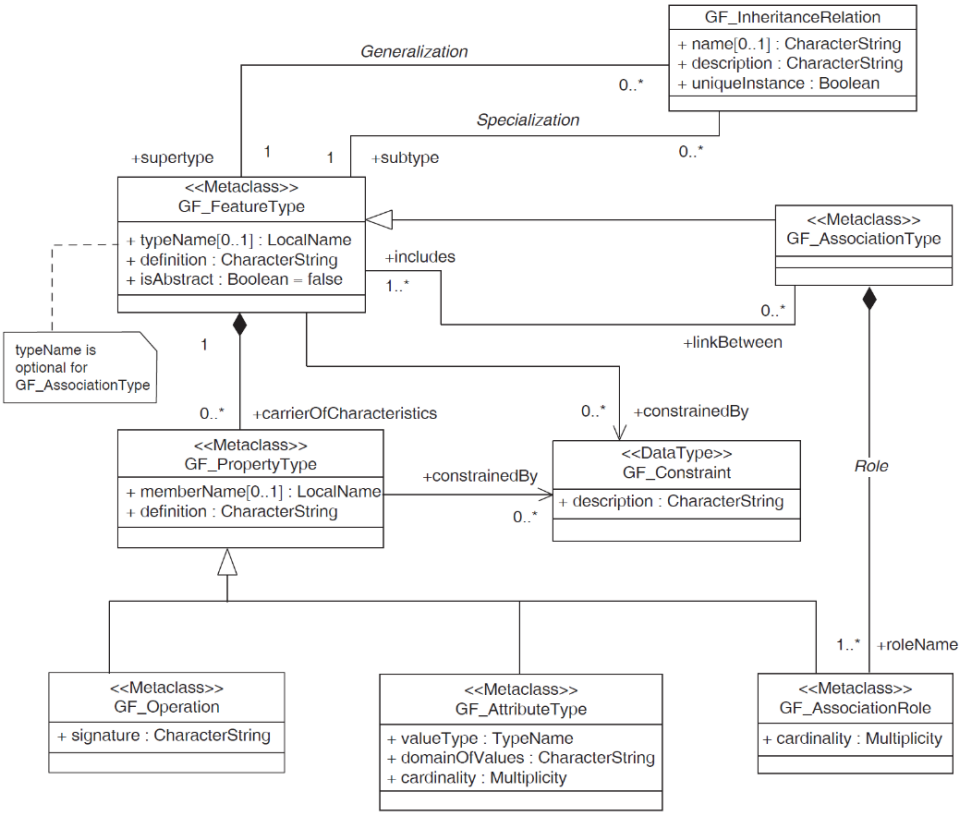


Figure 5-1. General Feature Model (INSPIRE/ISO 19109)

Based on this framework, the project incorporates the Observations and Measurements (O&M) structure defined by INSPIRE (D2.9 Guidelines for the use of Observations & Measurements) and ISO 19156, based on the Open Geospatial Consortium (OGC) standard. This conceptual framework describes how observations, properties and results relate to real-world phenomena.

The following figure shows the adaptation of INSPIRE's Specimen–Observation model to the domain of industrial waste, illustrating how waste is observed, measured and linked to analytical processes and their results, following the principles of O&M:

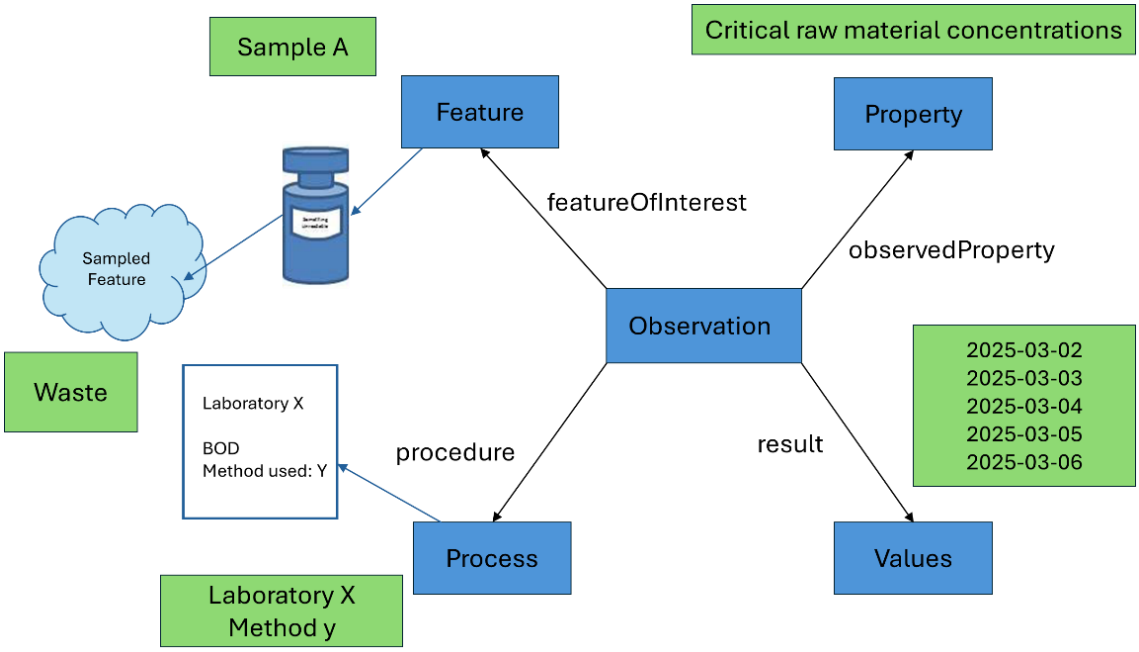


Figure 5-2. Adapted model of Observations and Measurements (based on INSPIRE D2.9 and ISO 19156, applied to the industrial waste domain)

In this adaptation:

- Feature of Interest (Waste) represents the waste material or site analysed.
- Sample (Specimen) is physical sample extracted from the waste for laboratory analysis.
- Property is a measured characteristic (e.g., concentration of critical materials, pH, metal content).
- Process is an analytical procedure (laboratory and method) used to obtain the observation.
- Observation is the act of measuring or recording the property.
- Values are measurement results, which may temporarily represent the observations.

This integration ensures that the project data model is consistent with ISO 19109, ISO 19156 and INSPIRE, and that it facilitates semantic interoperability with other data infrastructures based on the O&M model.

5.1 Design methodology

The design of the High-Level Application Schema follows the methodology established in *ISO 19103 Geographic information — Conceptual schema language*, specifically the guidelines described in Annex F.2 (Guidelines for information modelling). This Annex defines the phases and principles that must be applied to develop information models that comply with the ISO 19100 series of geographic information standards, using UML as the main conceptual representation tool.

According to Annex F.2 of ISO 19103, the modelling process is divided into the following phases:

- **Phase 0. Identify scope and context:** define the purpose, boundaries, and conceptual environment of the model.
- **Phase 1. Identify basic classes:** determine the key feature types relevant to the domain.
- **Phase 2. Specify relationships, attributes, and operations:** establish semantic connections between entities and define their properties and operations.
- **Phase 3. Complete constraints using text or OCL:** incorporate rules and constraints to ensure logical consistency.
- **Phase 4. Harmonise the model:** integrate submodels and external references into a coherent and interoperable schema.

The first three phases are essential for developing application schemes that comply with ISO standards for geographic information, while Phase 4 ensures the harmonisation and integration of sub-models into a coherent overall framework.

This structured methodology provides a clear roadmap for building an interoperable and semantically consistent model, aligned with INSPIRE, ISO 19109 and ISO 19156, and adaptable to the context of European Data Spaces.

5.2 Conceptual Data Model for Industrial Waste

The Conceptual Data Model for Industrial Waste is developed following the methodological principles established in ISO 19103 (Annex F.2) and ISO 19109, using an object-oriented approach that allows the main entities, properties and relationships within the industrial waste domain to be represented.

The model is structured around several thematic domains, each of which represents a key semantic component of the global framework.

These domains define the conceptual environment in which waste-related information is generated, managed and interpreted, ensuring interoperability and consistency with the ISO 19100 family of standards and INSPIRE data themes. The following figure shows the fundamental components surrounding the central concept of Industrial Waste:

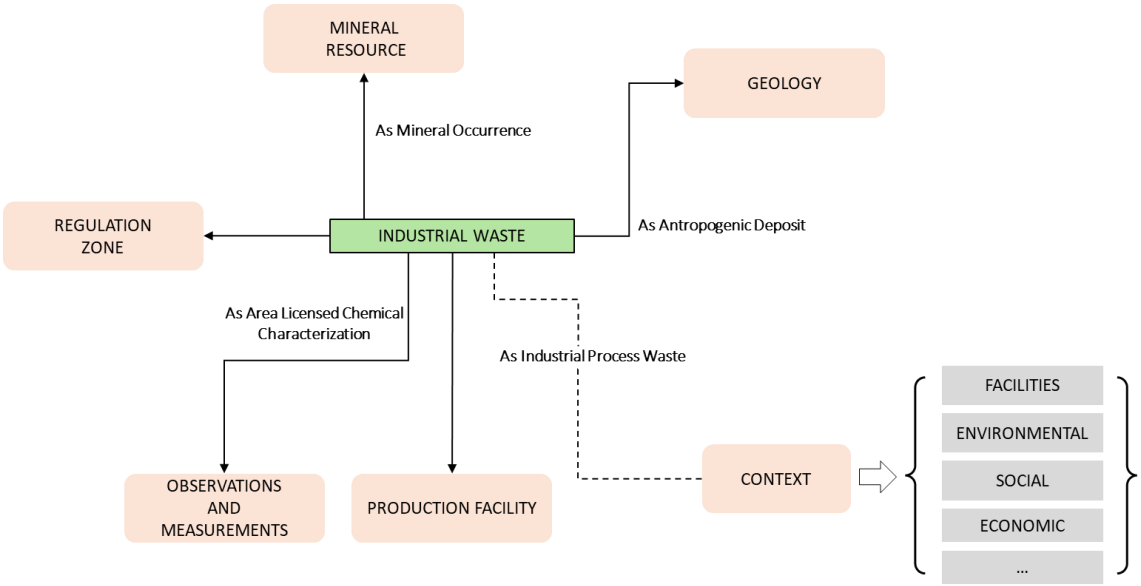


Figure 5-3. Main thematic domains of the data model for Industrial Waste

The main thematic domains and their functions are as follows:

- **Industrial Waste**: central domain of the model, representing the industrial waste dump or deposit. It is conceived as a spatial unit (geographical feature) associated with the accumulation, storage or disposal of waste materials from industrial processes (e.g. mining, metallurgy or energy generation).
- **Geology**: Describes the geological structure and composition of the environment in which industrial waste dumps or deposits are located. It includes entities from the GeoSciML model, which provide the physical-natural context where the waste originates or is located. This domain allows relationships to be established between the waste deposit and the underlying geological units, facilitating the interpretation of the physical-chemical behaviour of the waste material (e.g. acid drainage, permeability, mineral alteration).
- **Mineral Resource**: Represents mineral occurrences or potential resources identified from the characterisation of industrial waste. This domain is based on the INSPIRE conceptual model – Mineral Resources (D2.8.III.21) adapted to the context of industrial waste. This domain allows analytical data to be linked to the mineral characterisation of waste, assessing its resource recovery potential and

facilitating its integration into European data infrastructures on critical and recyclable raw materials.

- **Production Facility**: Represents the plants, infrastructure and industrial processes responsible for generating industrial waste. Includes information on the type of facility, processes and treatment methods in accordance with the INSPIRE Production and Industrial Facilities theme (D2.8.III.8).
- **Regulation Zone**: Defines the legal and administrative framework that regulates the management, environmental control, and supervision of waste dumps, including easement areas, protection limits, and regulatory constraints.
- **Context**: Groups cross-cutting dimensions such as environmental, social, economic and infrastructure conditions relevant to the management and assessment of waste deposits. This domain provides the context for integrating existing complementary information (e.g. environmental impact, land use, proximity to populations or logistics networks).

This domain-based conceptual structure establishes the basis for semantic integration and harmonisation with INSPIRE and GeoSciML models, ensuring compatibility with European spatial data infrastructures.

Based on the thematic framework defined above, the following figure presents the preliminary conceptual structure of the Industrial Waste data model, developed in accordance with ISO 19109 (General Feature Model) and harmonised with the INSPIRE, GeoSciML, EarthResourceML (ERML) and O&M (ISO 19156) standards.

This model illustrates how Industrial Waste, a landfill or industrial deposit, acts as the core of the system, connecting geological, mineralogical, production, analytical and regulatory information. The diagram integrates different levels of description, from the physical and chemical properties of the waste to observations derived from analysis and characterisation processes.

The model is designed to represent waste dumps with one or more industrial origins, so that the same facility may contain materials from different production processes or industrial plants.

This heterogeneity is managed in the model structure through the WastePart entity, which allows the total volume of waste to be broken down into parts or subunits, each of which can be associated with an industrial source, a production process or specific properties. The entities ProductionFacility and IndustrialProcess are linked, in turn, to WastePart, which allows for the documentation of the different sources of generation, the types of waste associated with them, and the intermediate transformations that may have occurred before their final disposal in the landfill.

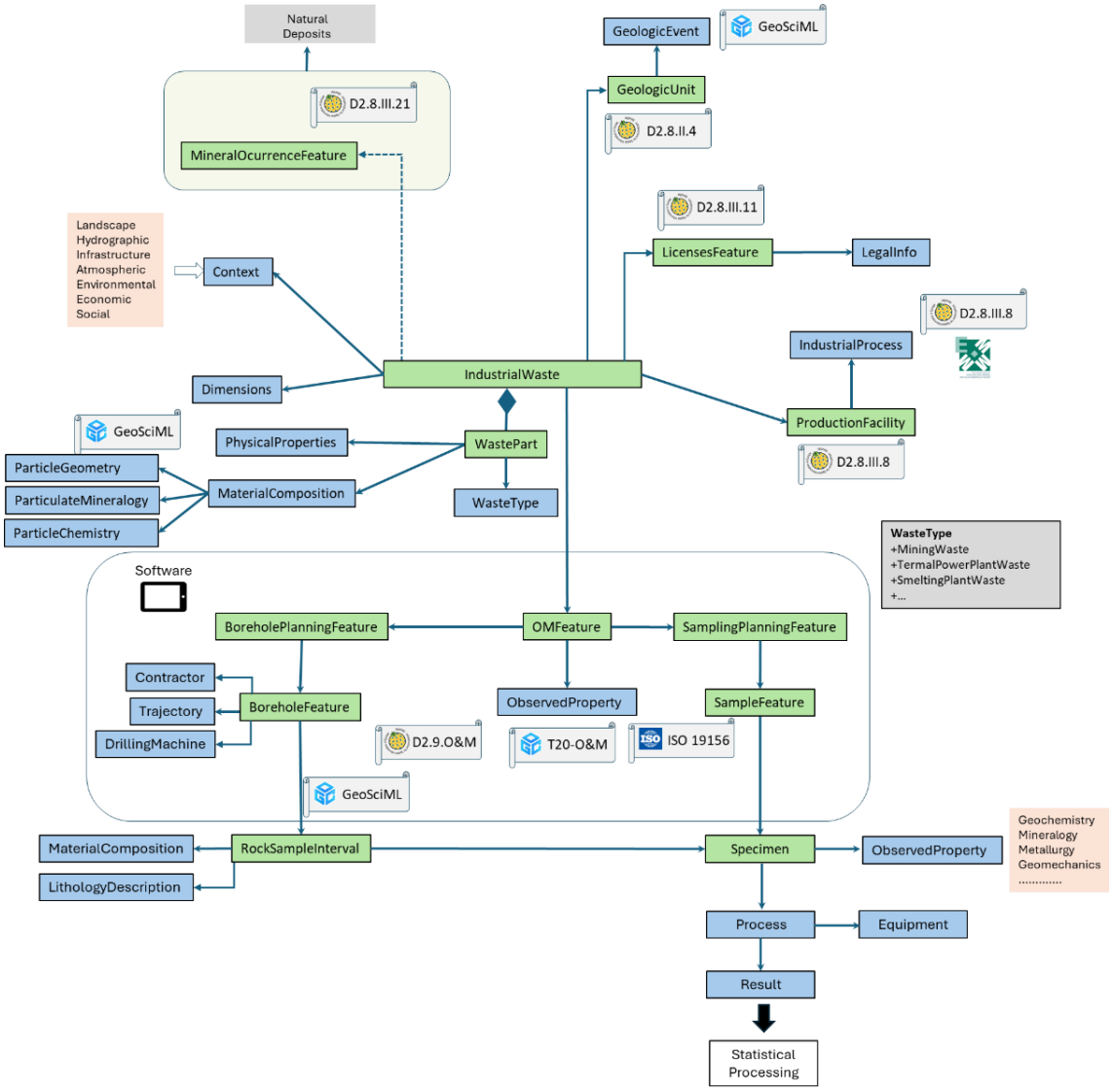


Figure 5-4. Preliminary conceptual model of Industrial Waste (based on ISO 19109, INSPIRE, GeoSciML and O&M)

The main components of the model are as follows:

- **Industrial Waste** (landfill or industrial deposit): Central entity representing the volume or facility where waste materials from one or more industrial processes accumulate. Includes identification attributes, geographical location, waste type (WasteType) and composition.
- **WastePart, PhysicalProperties**, **MaterialComposition**: These allow the internal structure and heterogeneous composition of the waste to be described.
  Each WastePart represents a part or fraction of the deposit associated with a specific industrial source or production process, with its own physical (density,

particle size, moisture content) and chemical (elements, compounds, proportions) properties.
Through these entities, the model can reflect the multiplicity of industrial sources and compositional variations within a single waste tip.

- **Geology (GeologicUnit):** Provides the link between the tailings and their geological context. Its integration into the model, based on GeoSciML (D2.8.II.4), allows the tailings information to be contextualised according to the geological substrate and terrain characteristics.
- **ProductionFacility and IndustrialProcess**: Document the facilities and processes where waste originates, related to WastePart entities to ensure the traceability of each material flow.
- **RegulationZone**: Defines the administrative or environmental framework under which the facility is managed, including licences, protection zones, usage restrictions, and regulatory measures.
- **OMFeature, ObservedProperty, SampleFeature, Specimen, Result**: Components of the O&M model (ISO 19156) that describe the observation process, from sampling planning and sampling to obtaining results. This block allows for the modelling of point analyses and chemical analysis.
- **BoreholeFeature and RockSampleInterval**: Derived from GeoSciML, these allow the representation of boreholes, sampling intervals, and lithological descriptions linked to both the waste and the underlying geological substrate.
- **Process and Equipment**: Describe the analytical procedures and equipment used in laboratory or on-site analyses, ensuring the reproducibility and traceability of results.
- **MineralOccurrenceFeature (Mineral Resource)**: Represents mineral occurrences or materials of interest identified in the waste through composition analysis.
  In this model, these occurrences are derived from the results of observations and measurements (O&M) and are described using the concepts of INSPIRE – Mineral Resources (D2.8.III.21) and ERML.
  Their inclusion allows the connection between the waste and the recoverable critical materials to be established, ensuring interoperability with European geoscientific standards.

In conjunction, the model allows the entire information chain to be traced, from the origin of the waste (multiple industrial processes) and its geological context, to the identification of valuable materials through analytical observations.

The MineralOccurrenceFeature entity, derived from the analysis results, is a key element in assessing the potential for reusing critical resources present in industrial waste.

In addition, the model is aligned with the interoperability principles of INSPIRE, GeoSciML, ERML and the future European Raw Materials Data Space, facilitating the integration of data from different sources, projects or national infrastructures.

This conceptual model is considered a preliminary version, which will serve as the basis for the development of the detailed UML schema and data catalogue in later phases of the project (tasks T3.2 and T3.3).

6 Governance and interoperability

Data interoperability is the ability of diverse systems, organizations, and technologies to exchange and use data coherently. Within the context of European Data Spaces, interoperability is a foundational principle enabling cross-domain collaboration, ensuring that data retains its meaning, structure, and trust regardless of where it originates or who uses it. Interoperability allows the seamless connection between participants, platforms, and services—supporting data sovereignty, innovation, and compliance with the FAIR principles (Findable, Accessible, Interoperable, Reusable). *ISO/IEC 19941:2017-5.2.1 – Information technology – Cloud computing – Interoperability and portability* will be the main guide to this end.

Achieving this requires not only compatible technical infrastructures but also a shared semantic foundation built upon controlled vocabularies and data ontologies. Controlled vocabularies standardize the terminology used to describe datasets, ensuring consistency and discoverability across repositories, while ontologies formally define the entities, attributes, and relationships within a domain, enabling semantic reasoning and machine-to-machine understanding.

In the context of European data spaces, these semantic tools are reinforced by interoperability frameworks such as Gaia-X, INSPIRE, and the Eclipse Dataspace Components (EDC):

- **Gaia-X** promotes interoperability through its Federation Services, Self-Descriptions, and Trust Framework, which define common schemas, identity mechanisms, and the use of the Dataspace Protocol (DSP) for interoperable data exchange across participants. *Gaia-X Compliance Document - 25.03 Release - Gaia-X Trust Framework* will be consulted.
- **INSPIRE** establishes a harmonized model for spatial data interoperability in the EU, using controlled vocabularies and common data specifications based on ISO 19100 standards to ensure that environmental and geospatial datasets can be seamlessly integrated. *A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructures*, published by EU in 2012, will be consulted.
- **Eclipse Dataspace Components**, an open-source implementation of the Dataspace Protocol, operationalizes interoperability at the technical layer, enabling compatible data sharing, contract negotiation, and transfer between connectors developed by different vendors (Eclipse Foundation, 2024).

Together, these frameworks illustrate that true interoperability in data spaces is multidimensional: it spans the semantic layer (vocabularies and ontologies), the

syntactic layer (shared data models and APIs), and the governance layer (trust, identity, and compliance). By aligning these elements, data spaces can ensure that data flows securely, meaningfully, and efficiently across organizational and national boundaries. Governance in data spaces will be addressed in detail in WP4.

6.1 Controlled vocabularies

A Controlled Vocabulary is "a set of terms and rules for their use in assigning material conditions to concepts for the purposes of indexing and information retrieval" (Council for Scientific Research – Consejo Superior de Investigaciones Científicas - CSIC, 2016) In general terms, a Controlled Vocabulary can range from a short list of clearly defined, mutually exclusive, and exhaustive terms that are the only options for use in a specific context, to something as complex as a thesaurus with thousands of terms and interrelationships. It usually includes preferred terms and variants. Controlled vocabularies are beneficial in the indexing process because cataloguers use the same term to refer to the same concept (for example, a person, a place, or a thing) in a consistent way. In the retrieval phase, controlled vocabularies guide users in their searches if they do not know the correct term for a given concept. There are different types of controlled vocabularies, such as: subject heading lists, authority files, taxonomies, alphanumeric classification schemes, thesauri, and ontologies.

6.1.1 Applications of controlled vocabularies

1. **To facilitate the indexing of content**: controlled vocabularies provide cataloguers and data managers with a fixed set of terms to describe resources. This avoids ambiguity when assigning keywords or subject headings. Instead of allowing variations like "coal waste," "coal residues," or "coal mining spoil," a controlled vocabulary ensures that all records use the preferred term "coal mine waste."
2. **To assist in searches and resource discovery**: users benefit from structured vocabularies because they can retrieve information even if they do not know the exact terminology. Synonyms, hierarchical structures, and cross-references guide users toward the relevant resources. Vocabulary mappings could assist a researcher by matching a query for "rare earth" to resources indexed as "REE" or "Rare Earth Elements".
3. **To improve the quality and consistency of bibliographic record metadata**: by standardizing the language used in metadata, repositories reduce errors, inconsistencies, and duplication. This improves data quality and ensures that records are easier to maintain and reuse. Ensuring that geographic names in the European Union follow an authority file such as the INSPIRE Data Specification on Geographical Names avoids inconsistencies.

4. **To semantically enrich resource descriptions**: controlled vocabularies often include hierarchical and associative relationships, which add semantic depth to resource descriptions. This enriches metadata beyond flat keywords and supports more meaningful connections between resources. For example, indexing a dataset on “coal mine tailings” also links it semantically to broader terms like “mining waste” and narrower terms like “acid mine drainage”.
5. **To promote interoperability with other platforms and enable data exchange**: using standard vocabularies aligned with international schemas (e.g., DCAT-AP, SKOS, ISO/IEC standards) ensures that data from one repository can be integrated or federated with data from others.
6. **To benefit from the opportunities of linked data**: controlled vocabularies are often published as Linked Open Data, meaning that terms are not just words but URIs that connect to broader knowledge graphs. This enables repositories to contribute to and benefit from the semantic web. For example, a record annotated with the URI for “Niobium (CHEBI:33344)” in the Chemical Entities of Biological Interest (ChEBI) ontology can be linked to datasets in chemistry, mining, and environmental sciences across platforms.

6.2 Data Ontology

Data ontology is a formal, explicit specification of concepts, entities, and their relationships within a particular domain of knowledge. Unlike controlled vocabularies or taxonomies (which list terms and categories), ontologies are semantic models that define not only what things are but also how they relate to each other in a way that can be processed by machines and understood consistently across systems. Ontologies are typically expressed in formal languages such as OWL (Web Ontology Language) or RDF (Resource Description Framework), enabling reasoning, interoperability, and linked data integration.

The DCMI is an example of ontology applied to data spaces. It is used for the description of resource metadata on the web. This ontology provides a set of terms and properties designed to facilitate the cataloguing and indexing of various types of online resources, such as documents, images, and videos. The use of DCMI enhances the accessibility and organization of information on the web, improving the management and retrieval of digital resources.

6.3 Interoperability in Data Spaces

Interoperability within data spaces is essential to ensure that heterogeneous systems, datasets, and services can exchange and interpret information consistently across organizational and national boundaries. In Europe, this principle is embodied through complementary frameworks that define both the technical and semantic foundations of

data exchange. Gaia-X establishes a federated and trusted data infrastructure where interoperability, achieved through a series of Framework definitions, ensures that different platforms and domains can interact seamlessly under shared governance and trust rules. The Data Spaces architecture developed by the Internation Data Spaces Association (IDSA), provides the reference model and interoperability standards—including usage control policies, connector interfaces, and the information model—that underpin Gaia-X and other European data ecosystems. Meanwhile, the EGDI exemplifies interoperability in practice by harmonizing geoscientific datasets across national geological surveys using INSPIRE-compliant data models, controlled vocabularies, and metadata catalogues. Together, these initiatives demonstrate how technical, semantic, and organizational interoperability form the backbone of the European data economy—linking domain-specific infrastructures such as EGDI with cross-domain federations like Gaia-X and IDS to enable secure, sovereign, and meaningful data sharing across sectors.

Unlike centralized data systems, Data Spaces operate in a federated and distributed environment—requiring robust mechanisms to maintain data integrity, ensure schema and metadata consistency, and preserve trust during sharing, transformation, and reuse:

- **Consistency** refers to the assurance that the representation of data, its metadata, and its associated policies remain coherent and aligned across systems. It implies that data updates, semantics, and access conditions are synchronized among federated nodes.
- **Integrity** focuses on the authenticity, completeness, and non-alteration of data, guaranteeing that datasets or metadata have not been tampered with, lost, or corrupted during transfer or storage processes.

6.3.1 Mechanisms for data consistency

Data consistency in federated environments is maintained through multiple layers of the data space architecture:

1. **Semantic and metadata consistency**. Achieved via shared vocabularies, ontologies, and schemas. In the European context, INSPIRE and DCAT-AP metadata standards ensure that resources are described uniformly, facilitating discoverability and interoperability. In Gaia-X, Self-Descriptions (SDs) provide a consistent metadata model for all participants, assets, and services, ensuring cross-domain harmonization (Gaia-X Architecture Document, 2025).
2. **Schema and data model alignment**. The International Data Spaces Information Model (IDSIM) defines standardized data structures and entity relationships (participants, connectors, contracts, etc.) across domains. This guarantees that

connectors interpret metadata and policies consistently, even across different vendors (International Data Spaces Information Model, 2022).

3. **Federated catalogue synchronization**. Federated catalogues, such as those implemented in EDC, ensure that participants maintain consistent and validated metadata about available assets. Each node synchronizes Self-Descriptions (JSON-LD) to reflect the current state of datasets and services in the network.
4. **Versioning and data lineage**. Version control mechanisms allow tracking of changes in datasets and metadata over time. Data lineage captures provenance information—who created or modified a dataset, when, and under what contractual conditions—ensuring consistency between data and its contextual documentation (An Overview of the PROV Family of Documents, 2013).

6.3.2 Mechanisms for data integrity

Maintaining integrity in Data Spaces requires verifiable and traceable mechanisms at both technical and governance levels.

1. **Cryptographic hashing and signatures**. Datasets, metadata, and policies can be hashed and digitally signed to verify authenticity and prevent tampering. In Gaia-X and IDS-based architectures, Verifiable Credentials (VC) and Verifiable Presentations (VP) (W3C standard) are used to authenticate participants and prove the integrity of Self-Descriptions and contractual claims (Verifiable Credentials Data Model v2.0, 2025).
2. **Secure transport and proof of possession**. Data transfers occur via mutual Transport Layer Security (mTLS) or equivalent cryptographic protocols ensuring end-to-end integrity and confidentiality. The Dataspace Protocol (DSP), as implemented in EDC, enforces integrity at the communication layer with proof-of-possession tokens and signature verification.
3. **Policy enforcement and usage control**. Integrity extends to compliance with usage rights: policies must be enforced consistently during and after data transfer. The IDS Usage Control Framework (based on ODRL) defines obligations, permissions, and prohibitions attached to each data asset, ensuring integrity of data use relative to agreed contracts (IDS Reference Architecture Model - Usage Control, 2023).
4. **Audit and traceability**. All transactions (catalogue queries, contract negotiations, data transfers) are logged with timestamps and cryptographic proofs. Gaia-X Federation Services and IDS Clearing House components provide secure audit mechanisms for verifying compliance and integrity.

In federated data spaces, these mechanisms work together through governance frameworks that define roles, validation procedures, and trust anchors. Gaia-X integrates semantic and trust consistency through its Trust Framework, which mandates validation of Self-Descriptions, issuance of Verifiable Credentials, and audit-based

verification of participant conformance. The IDSA provides the technical integrity model, specifying connector-to-connector trust, secure communication, and enforcement of contracts via Usage Control mechanisms, and the EDC operationalize these principles through open-source implementations, enabling cross-vendor interoperability and reproducible validation processes.

This alignment ensures that data integrity (authenticity, immutability, and traceability) and consistency (semantic coherence and synchronization) are enforced across every layer (from metadata to transport and governance) within the European federated data ecosystem.

See subsection “Use Cases” for further information on how these mechanisms are applied to data space interoperability.

7 Key success factors & Dependencies with other tasks

7.1 Awareness of the model as a work in progress

As mentioned above, it must be remembered that the data model conceptual design submitted at this stage is preliminary and may be subject to changes as the project progresses. As development advances and requirements evolve, adjustments may be necessary to ensure alignment with project needs, data availability and technical constraints. This model should therefore be considered a first version rather than a final specification.

7.2 Understanding the implications for standard adoption

Adopting standards when developing a data model for the characterization of CRMs in industrial waste deposits can offer clear benefits (interoperability, consistency, data reuse), but it also involves some significant challenges, especially in specific technical and operational contexts. The main issues to be considered are as follows:

- **Rigidity and lack of fit for specific contexts:** these standards are designed for broad interoperability, which can make them inflexible when dealing with local or niche needs (like industrial waste). In other words, they might lead to a data structure that does not align well with the type, resolution or detail of the data model to be defined. For instance, CRMs in waste may have chemical or textural properties not captured in the standards, requiring extensions that complicate interoperability.
  The availability of field data, both legacy information and data from surveys carried out in the context of this project, is once again one of the key requirements for the development of the most appropriate data model.

- **Lack of maturity for certain domains:** standards are usually more mature for certain domains (such as topography, hydrography, land use) but not necessarily for industrial waste or secondary raw materials. Semantic or structural gaps may arise, requiring development of application schemas or extensions, which further complicate adoption.

- **Mismatch between standards and operational reality:** standards usually evolve more slowly than data acquisition and analysis technologies (sensors, spectrometry, AI, etc.). The model may become outdated or fail to properly represent data generated by new characterization techniques.
  Therefore, a long-term vision that contemplates the need for evolving the data model in the future, especially after the conclusion of the project, is necessary.

- **Issues with real-world interoperability:** while standards aim for interoperability, in practice, different interpretations of the standards can lead to incompatible models. This may give the impression that interoperability is more of a bureaucratic burden than a functional advantage.

- **Technical complexity and learning curve:** these standards are technically complex and require advanced knowledge of geospatial modelling, UML, etc. For example, adopting the INSPIRE Geology data model requires understanding structures like GeologicUnit, MappedFeature, etc., which may not directly match how field data is collected.
  If the technical team does not have the necessary expertise, they may face a steep learning curve, delaying development or leading to implementation errors. In this case, providing some training for the technical team on the reference standards and related tools would be necessary.

7.3 Dependencies with other project tasks

A critical factor in the design and evolution of the data model is the nature of its dependencies on other project tasks, whether these are unidirectional or bidirectional.

The following figure provides a diagram of these dependencies:

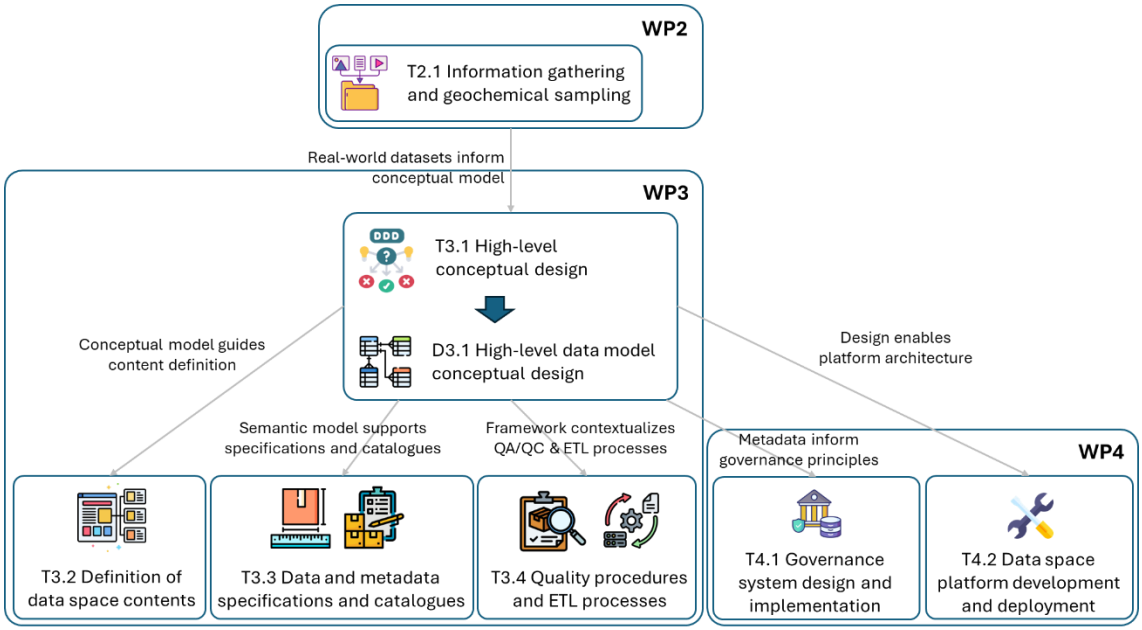


Figure 7-1. Dependencies of task T3.1 with other tasks

These dependencies are as follows:

- **T2.1 (Information gathering and geochemical sampling) → T3.1:**
  - ✓ Although T3.1 is primarily conceptual, it must be informed by T2.1 results to ensure the conceptual model accommodates real-world data collected during the development of T2.1.
  - ✓ T2.1 is included in *WP2 - Data acquisition from closed facilities* will produce raw georeferenced geochemical samples and gather legacy data regarding several phenomena and properties whose understanding is essential to ensure the model covers the right dimensions.

- **T3.1 → T3.2 (Definition of data space contents):**
  - ✓ T3.1 defines the fundamental components of the data model (phenomenon, property, attribute, relation, restriction, etc.) and establishes the metadata profile.
  - ✓ T3.2 requires the conceptual foundations from T3.1 to ensure data space contents are structured consistently. It depends on the high-level design to develop an Application Scheme that specifies how phenomena, attributes, relations, and restrictions are organized.

- **T3.1 → T3.3 (Data and metadata specifications and catalogues):**
  - ✓ T3.3 focuses on creating detailed data and metadata catalogues, semantic interoperability rules, and APIs. It cannot be effectively implemented without the conceptual framework of T3.1, as the latter establishes the semantic building blocks.
  - ✓ These catalogues must align with the conceptual model from T3.1 to ensure consistency and comparability.
  - ✓ One of the outcomes of T3.3 will be a rulebook for semantic interoperability, intended to provide a common guide within the Project ecosystem, minimize ambiguities and ensure the semantic consistency between different systems, and serve as a reference for the data harmonization and the development of interoperable tools. Obviously, the conceptual data model will be the first cornerstone in the creation of this rulebook.

- **T3.1 → T3.4 (Quality procedures and ETL processes):**
  - ✓ The ETL and QA/QC frameworks need to map to the conceptual definitions from T3.1 to automate quality enforcement correctly. T3.4 defines QA/QC policies and ETL processes to ensure integrity, quality, and consistency of captured data.
  - ✓ These processes must be aligned with the data model concepts and metadata rules established in T3.1.

- **T3.1 → T4.1 (Governance system design and implementation):**
  - ✓ Governance mechanisms must align with the conceptual definitions from T3.1 to ensure data sovereignty, FAIR compliance, and interoperability. T4.1 is included in *WP4 - Deploying a minimum viable data space* and establishes the governance framework for the data space.
  - ✓ Governance rules FAIR principles, compliance with European Data Strategy, etc.) rely on the metadata and data structures defined in T3.1.

- **T3.1 → T4.2 (Data space platform development and deployment):**
  - ✓ Without T3.1, the platform risks being inconsistent or misaligned with the agreed-upon conceptual data model.
  - ✓ The technical platform in T4.2 (included in *WP4 - Deploying a minimum viable data space*) will implement and operationalize the conceptual and semantic definitions from T3.1, ensuring scalability, interoperability and alignment with European standards (Gaia-X, DSP, EDC, etc.).

8 Expected benefits & Use cases

8.1 Expected benefits

The successful implementation of the CRMsDataSpace will generate significant benefits across economic, environmental, and governance domains. These benefits are interconnected, creating a virtuous cycle that supports a sustainable and competitive European industrial landscape.

8.1.1 Economic benefits

Among the potential economic benefits of the CRMsDataSpace initiative, the following must be highlighted:

- **Stimulating investment.** By providing clear, reliable, and comparable data on CRMs recovery potential, the data space de-risks initial prospecting and feasibility studies. This lowers the barrier to entry for private investment in resource recovery projects.
- **Enhancing market efficiency.** A centralised and transparent marketplace for data on secondary raw materials allows for more efficient allocation of resources, connecting potential project promoters with the most viable recovery sites.
- **Fostering innovation and new services.** The availability of high-quality, interoperable data will enable the creation of new data-driven services, such as advanced analytics for predictive modelling of recovery yields, logistics optimisation, and market analysis.
- **Reducing import dependency.** Facilitating the recovery of CRMs from domestic secondary sources directly contributes to the strategic autonomy of the EU, reducing reliance on volatile international supply chains.

8.1.2 Environmental and social benefits

CRMsDataSpace is expected to contribute to the following environmental and social benefits:

- **Promoting the circular economy.** The data space is a key enabler for transforming extractive waste from a liability into a valuable resource, embodying the core principles of the circular economy.
- **Supporting a just transition.** By identifying economic opportunities in former coal mining regions, the project directly supports the creation of new, high-skilled jobs in areas affected by deindustrialisation, ensuring that the green transition is socially equitable.

- **Land remediation and hazard reduction.** The recovery of materials from legacy waste sites can be coupled with environmental remediation activities, reducing potential long-term pollution risks and enabling the rehabilitation of land for new uses.

8.1.3 Governance and policy benefits

Finally, the CRMsDataSpace initiative is expected to deliver the following governance and policy benefits:

- **Streamlining regulatory compliance.** The data space provides a mechanism for Member States to fulfil their data provision obligations under the Critical Raw Materials Act in a standardised manner.
- **Informed policymaking.** Aggregated, cross-border data allows policymakers at the EU, national, and regional levels to develop more effective strategies for resource management, industrial policy, and regional development.
- **Enhanced cross-border collaboration.** A common data framework, built on standards like INSPIRE, removes linguistic and technical barriers, fostering collaboration between geological surveys, research institutions, and companies across Europe.

8.2 Use cases

To conceptualise the practical application of the CRMsDataSpace and illustrate the importance of some of the key concepts and features explained above, such as interoperability, consistency, integrity and security, among others, four key use cases are outlined in the following subsections.

These use cases describe how different user personas would interact with the system to achieve their objectives.

8.2.1 Prospecting and feasibility assessment for project promoters

A geologist or business development manager at a mining and minerals processing company wants to identify the top five most economically promising closed coal waste facilities in the Silesian Basin (Poland and Czechia) for the recovery of Germanium (Ge) and Gallium (Ga).

The following figure shows the event sequence related to this case:

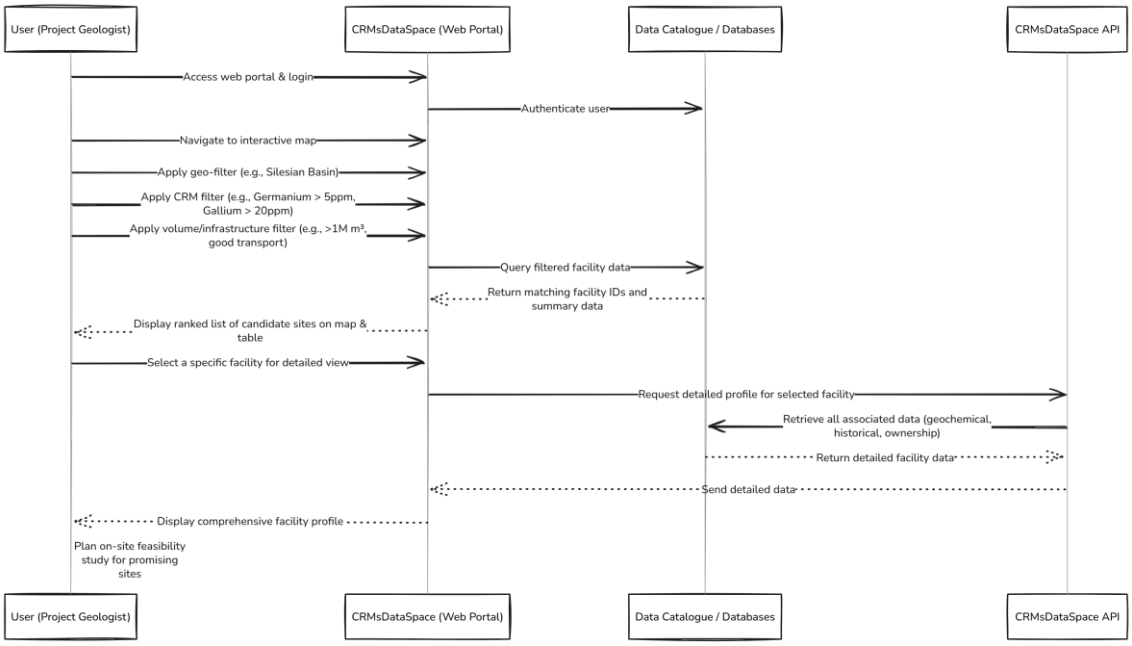


Figure 8-1. Sequencing diagram for prospecting and feasibility assessment use case

1. The user accesses the data space through a web portal with an interactive map interface.
2. They apply filters to query the database for facilities within the specified geographical area (NUTS 2 regions).
3. Further filters are applied based on material composition, specifically targeting sites with reported concentrations of Ge > 5 ppm and Ga > 20 ppm.
4. The user then narrows the search by querying for sites with volumetric estimates exceeding 1 million cubic meters and with direct access to road or rail infrastructure.
5. The system returns a ranked list of candidate sites, each with a detailed profile containing historical data, available geochemical analyses, ownership information, and links to original reports. This allows the company to focus its resources on conducting detailed, on-site feasibility studies for the most promising locations.

8.2.2 Regional development and just transition planning

A policy officer at a regional development agency in a former coal-mining region (e.g., Asturias, Spain or North Rhine-Westphalia, Germany) wants to develop a regional strategy to attract investment in the circular economy by quantifying the CRMs recovery potential and identifying infrastructure and skills gaps.

The following figure shows the event sequence related to this case:

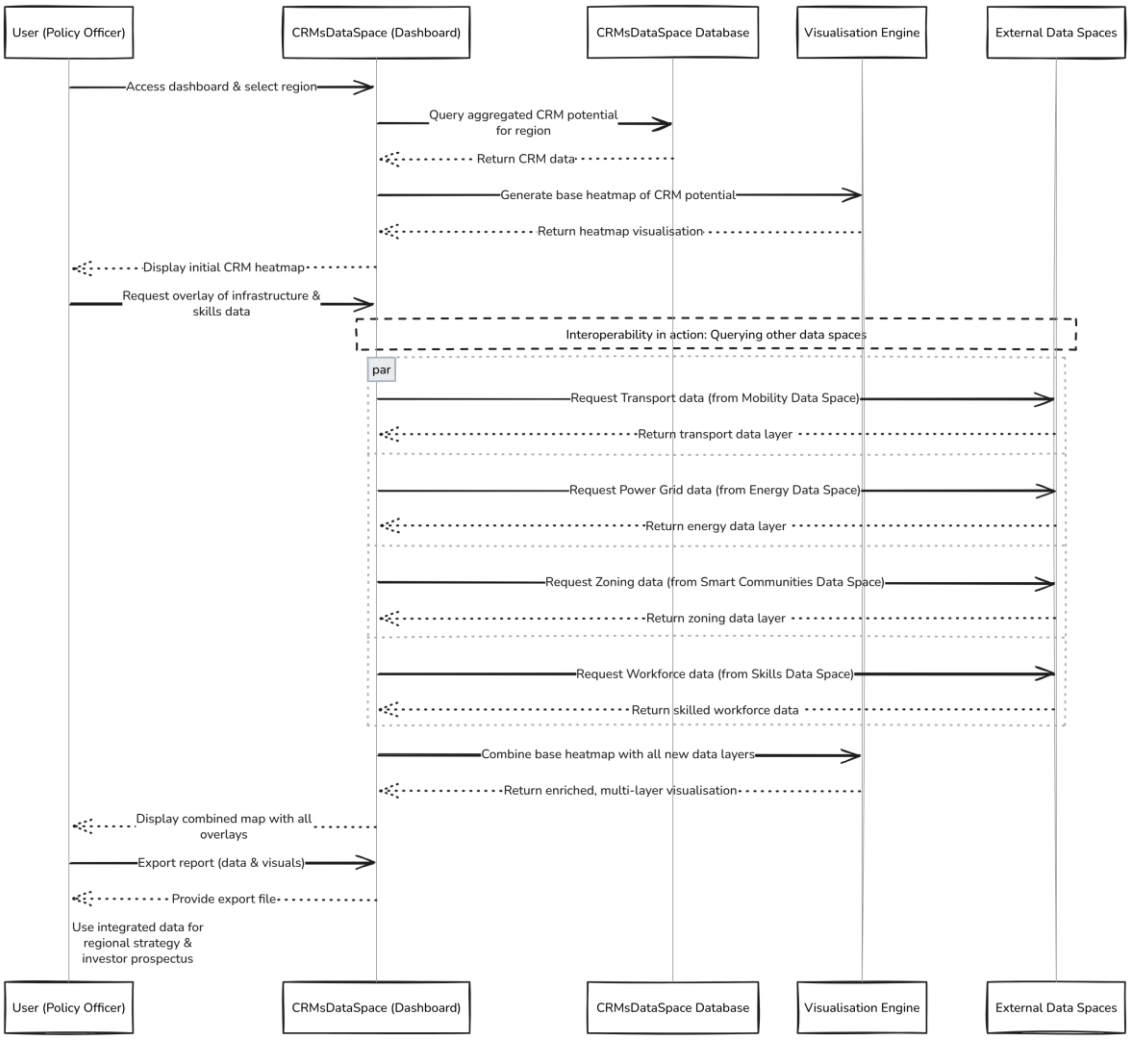


Figure 8-2. Sequencing diagram for regional development use case

1. The user accesses a dashboard view that provides aggregated analytics for their specific region.
2. They generate a heatmap showing the geographical concentration of high-potential waste facilities for specific CRMs relevant to battery or semiconductor value chains.
3. Crucially, the system then demonstrates the power of interoperability by dynamically querying other sectoral data spaces to enrich the analysis. The user overlays the CRMs map with vital logistical and socio-economic data layers, like:
    a. **Transport networks** (road, rail) are pulled from the **Common European Mobility Data Space** to assess site accessibility and logistical costs.
    b. **Power grid capacity** and major connection points are sourced from the **Common European Energy Data Space** to evaluate the energy supply needed for potential processing facilities.

c. **Zoning information for industrial parks** and water utility access are obtained via services federated within the **Smart Communities Data Space**.
d. **Data on the skilled workforce**, such as the number of available technicians and engineers in the local area, is accessed through the developing **Common European Skills Data Space**.

4. By seamlessly integrating these datasets, the user can identify not just the most resource-rich sites, but those with the highest overall project viability. This holistic view enables the creation of a powerful, data-driven prospectus for potential investors and a solid business case for targeted public investment in infrastructure upgrades or vocational training programs.

8.2.3 Academic research and environmental impact modelling

A university researcher specialising in geochemistry and environmental science is investigating the correlation between historical coal extraction/processing methods (e.g., washing techniques, combustion type) and the resulting mineralogical phase and leachability of CRMs and associated elements (e.g., Arsenic, Cadmium) in the waste.

The following figure shows the event sequence related to this case:

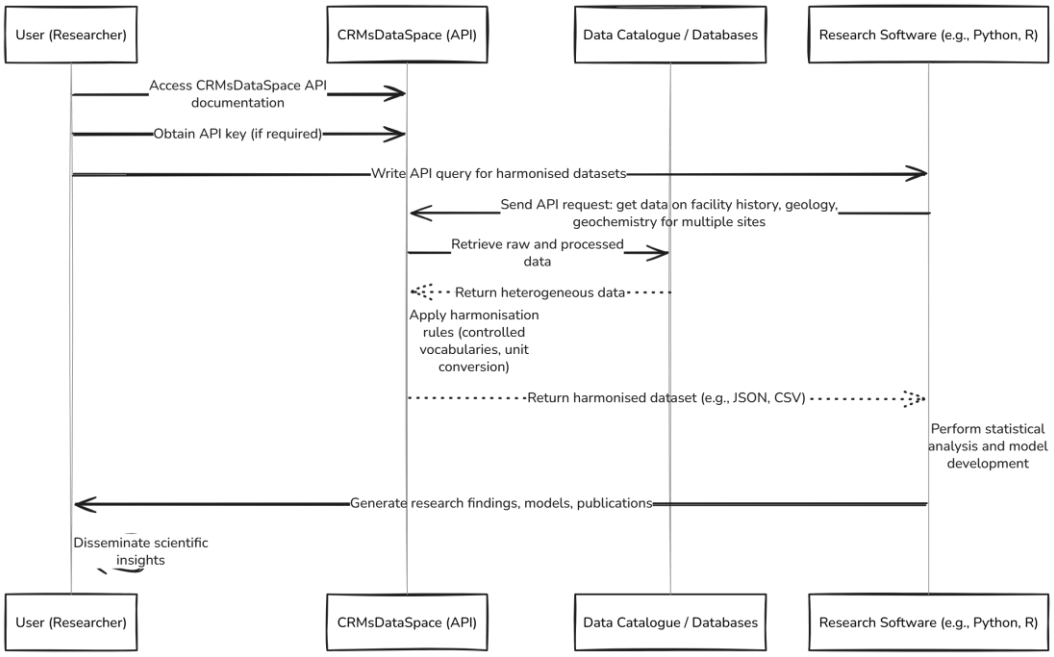


Figure 8-3. Sequencing diagram for environmental impact modelling use case

1. The researcher uses the data space's API (Application Programming Interface) to query and download harmonised datasets from hundreds of sites across multiple European countries.
2. The query combines data on the facility's operational history, the type of coal seam originally mined, and detailed geochemical assay data from various sampling campaigns.
3. The harmonised data model ensures that data from different national sources are comparable, using controlled vocabularies for geological terms and analytical methods.
4. This comprehensive dataset enables the researcher to perform large-scale statistical analysis and develop predictive models that would be impossible to create by manually collecting and cleaning heterogeneous data from individual sources, leading to new scientific publications and a better understanding of potential environmental risks.

8.2.4 AI-Powered predictive modelling for resource estimation

A data scientist at a geological services company, developing a predictive tool for project geologists and exploration firms to create and deploy a machine learning model that estimates the likely concentration of specific CRMs in under-sampled or un-sampled coal waste facilities, thereby enabling rapid, large-scale screening of assets. The following figure shows the event sequence related to this case:

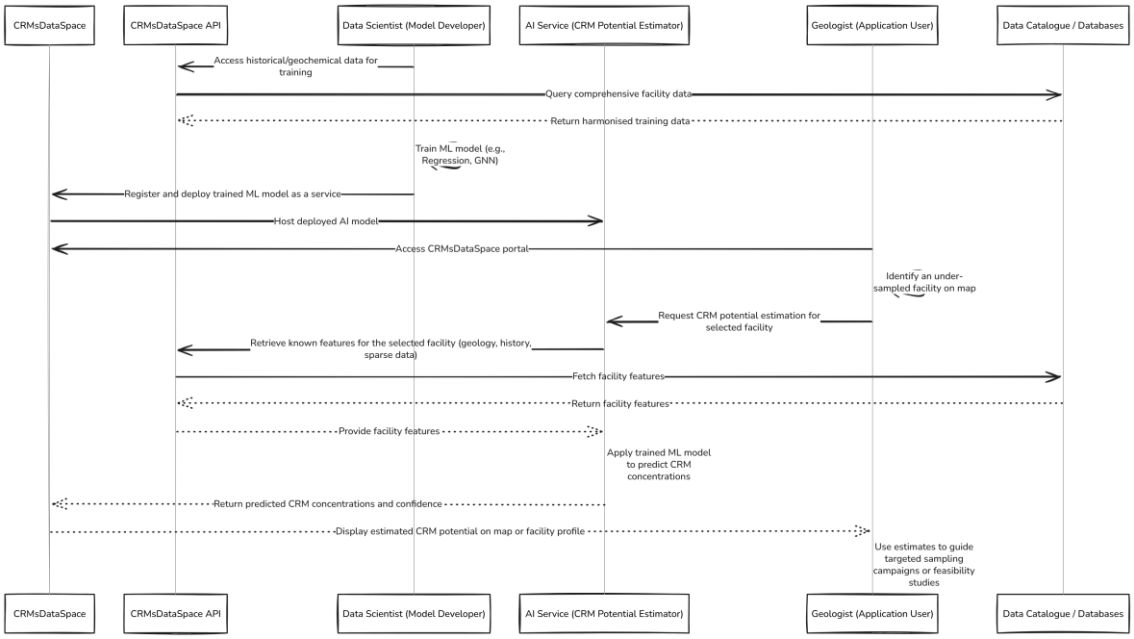


Figure 8-4. Sequencing diagram for AI-Powered predictive modelling use case

1. **Model Development (Training Phase):**
    a. The data scientist uses the CRMsDataSpace API to access a comprehensive, harmonised training dataset from all well-characterised facilities across Europe.
    b. The model's target variable is the known CRMs concentration (e.g., Gallium, Germanium).
    c. The predictor variables (features) include a wide range of data points available in the data space:
        i. Geological context (e.g., parent coal seam, host rock mineralogy).
        ii. Historical facility data (e.g., operational period, coal processing techniques used).
        iii. Available, albeit sparse, geochemical data.
        iv. Geospatial data (e.g., deposit morphology, proximity to geological faults).
        v. Proxy data from remote sensing (e.g., multispectral satellite imagery that can hint at specific mineral assemblages).
    d. Using this rich dataset, the data scientist trains a sophisticated model (e.g., a gradient boosting algorithm or a geospatial neural network) to learn the complex, non-linear relationships between these features and the actual CRMs concentrations.

2. **Application in Practice (Deployment Phase):**
    a. The trained model is deployed as a new, value-added service within the data space, perhaps called the "CRMs Potential Estimator".
    b. A project geologist evaluating a new region can select a facility with little to no public assay data.
    c. The application automatically gathers the known predictor variables for that site from the data space (its history, geology, etc.) and feeds them into the AI model.
    d. The model generates a prediction, providing an estimated concentration range for key CRMs (e.g., "Predicted Germanium: 8-15 ppm") along with a confidence score for the estimate.
    e. This allows the user to generate a regional "potential map" that fills in the gaps between known data points, highlighting previously overlooked areas that may warrant the expense of an on-site physical sampling campaign. This significantly reduces exploration risk and cost.

In conclusion, the use cases presented are not merely illustrative scenarios but concrete pathways to value creation. They demonstrate that CRMsDataSpace is envisioned as a dynamic, living ecosystem, not a static data archive. These examples underscore the critical importance of a robust, flexible, and interoperable conceptual design, as it is this foundation that will unlock the tangible benefits previously outlined.

9 Conclusions & Next steps

Deliverable 3.1 has established the **high-level conceptual design of the CRMsDataSpace data model**, which will serve as the cornerstone for developing a robust and interoperable data catalogue on CRMs from waste facilities. Several **conclusions** can be drawn from this initial stage:

1. **Standards as a foundation:** the adoption of established European and international standards (ISO 19100 family, INSPIRE, DCAT-AP, OGC frameworks) is essential to ensure semantic and technical interoperability across domains. Nevertheless, these standards may require extensions or contextual adaptations to properly capture the specificities of coal waste deposits and secondary raw materials.
2. **Balancing consistency and flexibility:** while standards guarantee comparability, strict adherence can create rigidity. Flexibility must therefore be incorporated into the model to accommodate evolving technologies, heterogeneous data, and local variations in waste characteristics.
3. **Dependencies across work packages:** the effectiveness of this deliverable depends heavily on the outcomes of related tasks (e.g., T2.1 sampling, T3.2 content definition, T3.3 catalogues, T4.1 governance, T4.2 platform deployment). This highlights the importance of coordination and iterative refinement throughout the project.
4. **Real-world data integration:** the conceptual model must also be grounded in real-world datasets, both from legacy and newly collected field data. This empirical evidence provides validation of abstract structures, reveals edge cases, and ensures the model's relevance in operational contexts.
5. **Work in progress:** this design is a first version, intended to evolve as requirements, standards, and available data mature. Stakeholder feedback and ongoing technical input will be crucial to align the model with project goals and external initiatives.

Building upon Deliverable 3.1, the following **next steps** are necessary to ensure successful progression:

- **Coordinate with subsequent tasks.** Ensure that outputs from this deliverable directly inform T3.2 (Definition of data space contents), T3.3 (Data and metadata specifications and catalogues), and T3.4 (Quality procedures and ETL processes). Early engagement with WP4 (*Deploying a minimum viable data space*, especially tasks T4.1 and T4.2 regarding governance and platform development) is essential to align the conceptual framework with technical and policy-level implementations.
- **Refine the conceptual model.** Incorporate lessons from field data (T2.1) and stakeholder feedback to adjust definitions of phenomena, attributes, relations,

and restrictions. Emphasis will be placed on harmonising coal waste-specific parameters with broader geoscientific standards.

- **Draft the rulebook for semantic interoperability.** Establish a project-wide reference for controlled vocabularies, units of measure, feature catalogues, and quality parameters to minimize ambiguity and guarantee semantic consistency.
- **Develop the metadata profile and application schema.** Translate the conceptual definitions into a structured metadata model (ISO 19115, DCAT-AP aligned) and a high-level application schema. These will provide the practical framework for cataloguing, discovery, and integration of datasets.
- **Pilot testing with sample datasets.** Apply the draft model to selected legacy and field datasets to validate its usability, identify structural gaps, and refine interoperability mechanisms. This stage will help anticipate integration challenges before large-scale implementation.
- **Stakeholder engagement and training.** Provide targeted workshops or training sessions to project partners and technical teams on the adopted standards and modelling approach. Building capacity will reduce the learning curve and foster consistent implementation.
- **Prepare for iterative updates.** Recognise the model as a living artifact. Mechanisms for continuous revision should be established, enabling adaptation to new analytical techniques, regulatory updates, or unforeseen technical requirements.

This deliverable represents a critical milestone in the CRMsDataSpace project. By laying the conceptual and semantic foundation, it prepares the ground for the next phases of detailed modelling, governance, and technical implementation. The ultimate goal remains clear: **to create a harmonised, interoperable, and practical data space that drives European resilience in CRMs** while supporting environmental sustainability and just transition objectives.

Some value creation pathways have been outlined as well via the defined use cases, which demonstrate the transformative potential of the CRMsDataSpace, not only for mining and industrial actors but also for regional policymakers, researchers, and the wider European economy. The conceptual design aims to deliver tangible **economic, social and environmental impacts**.

10 Glossary

AI – Artificial Intelligence

AMD – Acid Mine Drainage

ANEFA – Asociación Nacional de Empresas Fabricantes de Áridos

API – Application Programming Interface

ChEBI – Chemical Entities of Biological Interest

CM – Critical Materials

CRMA – Critical Raw Materials Act

CRMs – Critical Raw Materials

CSIRO – Commonwealth Scientific and Industrial Research Organisation

CSIC – Consejo Superior de Investigaciones Científicas

CSW – Catalogue Service for the Web

DCAT – Data Catalog Vocabulary

DACT-AP – Data Catalog Vocabulary-Application Profile

DCMI – Dublin Core Metadata Initiative

DOE – Department of Energy

DOE–NETL REE/CM – Department of Energy's National Energy Technology Laboratory program for Rare Earth Elements and Critical Materials

DSP – Data Space Protocol

EDC – Eclipse Dataspace Components

EIT – European Institute of Innovation and Technology

EGDI – European Geological Data Infrastructure

EPA – Environmental Protection Agency

EMMM – Exploration, Mining, Metals & Minerals Vertical

ERML – EarthResourceML

ETL – Extract, Transform, Load

EU – European Union

FAIR – Findable, Accessible, Interoperable and Reusable

GeoSciML – Geoscience Markup Language

GFM – General Feature Model

GIG-PIB – Główny Instytut Górnictwa - Państwowy Instytut Badawczy

GIS – Geographic Information System

GML – Geography Markup Language

IDSA – International Data Spaces Association

IDSIM – International Data Spaces Information Model

IGME – Instituto Geológico y Minero de España

INSEMEX – Institutul Național de Cercetare - Dezvoltare pentru Securitate Minieră și Protecție Antiexplozivă

INSPIRE – Infrastructure for Spatial Information in Europe

IEC – International Electrotechnical Commission

ISO – International Organization for Standardization

ISO/IEC – International Organization for Standardization / International Electrotechnical Commission

mTLS – Mutual Transport Layer Security

NETL – National Energy Technology Laboratory

NGO – Non-Governmental Organization

ODRL – Open Digital Rights Language

OGC – Open Geospatial Consortium

OWL – Web Ontology Language

Q/A-Q/C – Quality Assurance - Quality Control

RAWMINA – Raw Materials Innovation For The Circular Economy: Sustainable Processing, Reuse, Recycling And Recovery Schemes

R&D – Research & Development

RDF – Resource Description Framework

REE – Rare Earth Elements

RMIS – Raw Materials Information System

SATEC – Sistemas Avanzados de Tecnología, S.A.

SCIMIN-CRM - Sustainable & Circular Production of Mineral Critical Raw Materials

SDI – Spatial Data Infrastructure

SDs – Gaia-X Self-Descriptions

SKOS – Simple Knowledge Organization System

TI-RRES – Proyecto Tierra-Residuos

TSF – Tailing Storage Facilities

UK/CAER – University of Kentucky's Center for Applied Energy Research

UML – Unified Modeling Language

UNECE – United Nations Economic Commission for Europe

UNFC – United Nations Framework Classification for Resources

UNIOVI – Universidad de Oviedo

UK/CAER - University of Kentucky/Center for Applied Energy Research

URI – Uniform Resource Identifier

VC – Verifiable Credentials

VP – Verifiable Presentations

W3C – World Wide Web Consortium

WCS – Web Coverage Service

WFS – Web Feature Service

WMS – Web Map Service

WVU – West Virginia University

11 References

Brendan A. Bishop, Karthik Ramachandran Shivakumar, Daniel S. Alessi, Leslie J. Robbins (2023) Insights into the rare earth element potential of coal combustion by-products from western Canada. *Environmental Science Advances* 2.3, 529-542. https://doi.org/10.1039/d2va00310d

Canadian Light Source (2024). *Recovering rare earth elements from coal ash for clean energy technologies*. URL: https://www.lightsource.ca/public/news/2024-25-q2-jul-sept/recovering-rare-earth-elements-from-coal-ash-for-clean-energy-technologies.php (accessed 1 Jul. 2025)

Commonwealth Scientific and Industrial Research Organisation - CSIRO (2024a). *Rare earth elements in coal: doing more with coal waste*. URL: https://www.csiro.au/en/news/All/Articles/2024/February/RRE-in-coal-waste (accessed 1 Jul. 2025)

Commonwealth Scientific and Industrial Research Organisation - CSIRO (2024b). Accelerating Development of Australia's Rare Earth Resources (ADARER). URL: https://research.csiro.au/critical-minerals-hub/wp-content/uploads/sites/545/2024/06/Accelerating-Development-of-Australias-Rare-Earth-Resources-1.pdf (accessed 1 Jul. 2025)

Digital CSIC (2016). *Use of controlled vocabularies in repositories* (Spanish). URL: https://digital.csic.es/bitstream/10261/140742/1/Bernal_DIGITALCSIC_MECD.pdf (accessed 1 Jul. 2025)

Department of Energy (2024). *Funding Notice: Bipartisan Infrastructure Law: Advanced Processing of Critical Minerals and Materials for Industrial and Manufacturing Applications*. URL: https://www.energy.gov/fecm/funding-notice-bipartisan-infrastructure-law-advanced-processing-critical-minerals-and (accesses 1 Jul. 2025)

Department of Energy - National Energy Technology Laboratory. DOE–NETL REE/CM Program (2024). URL: https://www.netl.doe.gov/node/13953 (accessed 1 Jul. 2025)

Dublin Core Metadata Initiative – DCMI (2020). *DCMI Metadata Terms*. https://www.dublincore.org/specifications/dublin-core/ (accessed 1 Jul. 2025)

Eclipse Foundation (2024). *The Eclipse Dataspace Protocol Is Pioneering Interoperability and Standardisation in Dataspaces*. URL: https://newsroom.eclipse.org/eclipse-newsletter/2024/june/eclipse-dataspace-protocol-pioneering-interoperability-and (accessed 1 Jul. 2025)

EIT RM - 21033 Reecovery. *Circular economy to produce critical raw materials from water in the mining industry.* URL: https://www.reecovery.eu/ (accessed 1 Jul. 2025)

EuroGeoSurveys (2016). *European Geological Data Infrastructure (EGDI)*. URL: https://www.europe-geology.eu/ (accessed 1 Jul. 2025)

European Union (2006). *Directive 2006/21/EC of the European Parliament and of the Council of 15 March 2006 on the management of waste from extractive industries and amending Directive 2004/35/EC - Statement by the European Parliament, the Council and the Commission*. URL: https://eur-lex.europa.eu/eli/dir/2006/21/oj/eng (accessed 1 Jul. 2025)

European Union (2007). *Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*. URL: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32007L0002 (accessed 1 Jul. 2025)

European Union (2007). *INSPIRE Directive, establishing an infrastructure for spatial information in Europe to support Community environmental policies, and policies or activities which may have an impact on the environment*. URL: https://knowledge-base.inspire.ec.europa.eu/legislation/inspire-directive_en (accessed 1 Jul. 2025)

European Union (2008). *Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata*. URL: https://eur-lex.europa.eu/eli/reg/2008/1205/oj/eng (accessed 1 Jul. 2025)

European Union (2012). *A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructures*. URL: https://knowledge-base.inspire.ec.europa.eu/publications/conceptual-model-developing-interoperability-specifications-spatial-data-infrastructures_en (accessed 1 Jul. 2025)

European Union (2014). *Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC*. URL: https://eur-lex.europa.eu/eli/reg/2014/910/oj/eng (accessed 1 Jul. 2025)

European Union (2016). *Guidelines for the use of Observations & Measurements and Sensor Web Enablement-related standards in INSPIRE*. URL: https://knowledge-base.inspire.ec.europa.eu/publications/guidelines-use-observations-measurements-and-sensor-web-enablement-related-standards-inspire_en

(accessed 1 Jul. 2025)

European Union (2019). *Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)*. URL: https://eur-lex.europa.eu/eli/dir/2019/1024/oj/eng (accessed 1 Jul. 2025)

European Union (2020). *European Data Strategy*. URL: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en (accessed 1 Jul. 2025)

European Union (2024). *Regulation (EU) 2024/1252 of the European Parliament and of the Council of 11 April 2024 establishing a framework for ensuring a secure and sustainable supply of critical raw materials and amending Regulations (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1724 and (EU) 2019/1020 (Text with EEA relevance)*. URL: https://eur-lex.europa.eu/eli/reg/2024/1252/oj/eng (accessed 1 Jul. 2025)

European Union (2024). *INSPIRE Data Specification on Area Management/Restriction/Regulation Zones and Reporting Units – Technical Guidelines*. URL: https://knowledge-base.inspire.ec.europa.eu/publications/inspire-data-specification-area-managementrestrictionregulation-zones-and-reporting-units-technical_en (accessed 1 Jul. 2025)

European Union (2024). *INSPIRE Data Specification on Geology – Technical Guidelines*. URL: https://knowledge-base.inspire.ec.europa.eu/publications/inspire-data-specification-geology-technical-guidelines_en (accessed 1 Jul. 2025)

European Union (2024). *INSPIRE Data Specification on Mineral Resources – Technical Guidelines*. URL: https://knowledge-base.inspire.ec.europa.eu/publications/inspire-data-specification-mineral-resources-technical-guidelines_en (accessed 1 Jul. 2025)

European Union (2024). *INSPIRE Data Specification on Production and Industrial Facilities – Technical Guidelines*. URL: https://knowledge-base.inspire.ec.europa.eu/publications/inspire-data-specification-production-and-industrial-facilities-technical-guidelines_en (accessed 1 Jul. 2025)

European Union. *DCAT Application profile for data portals in Europe (DCAT-AP)*. URL: https://op.europa.eu/en/web/eu-vocabularies/dcat-ap (accessed 1 Jul. 2025)

Gaia-X (2024). *Gaia-X Compliance Document - 25.03 Release*. URL: https://docs.gaia-x.eu/policy-rules-committee/compliance-document/25.03/ (accessed 1 Jul. 2025)

Gaia-X (2025). *Gaia-X Architecture Document - 25.05 Release*. URL: https://docs.gaia-x.eu/technical-committee/architecture-document/25.05/ (accessed 1 Jul. 2025)

Government of Canada (2023). *Programs and funding for critical minerals projects – Federal Research and Development*. URL: https://www.canada.ca/en/campaign/critical-minerals-in-canada/federal-research-development.html (accessed 1 Jul. 2025)

Horizon 2020 – RAWMINA. *Raw Materials Innovation for the Circular Economy: Sustainable Processing, Reuse, Recycling and Recovery Schemes.* URL: https://rawmina.eu/ (accessed 1 Jul. 2025)

Horizon Europe - SCIMIN-CRM (2024). *Sustainable & Circular Production of Mineral Critical Raw Materials*. URL: https://scimin.eu/ (accessed 1 Jul. 2025)

International Data Spaces (2022). *International Data Spaces Information Model*. URL: https://international-data-spaces-association.github.io/InformationModel/docs/4.2.0/index.html (accessed 1 Jul. 2025)

International Data Spaces (2023). *Reference Architecture Model - Usage Control*. URL: *https://docs.internationaldataspaces.org/ids-knowledgebase/ids-ram-4/perspectives-of-the-reference-architecture-model/4_perspectives/4_1_security_perspective/4_1_6_usage_control* (accessed 1 Jul. 2025)

International Organization for Standardization – ISO (2014). *ISO 19101-1:2014 Geographic information — Reference model*. URL: https://www.iso.org/standard/59164.html (accessed 1 Jul. 2025)

International Organization for Standardization – ISO (2014). *ISO 19115-1:2014 Geographic information — Metadata*. URL: https://www.iso.org/standard/53798.html (accessed 1 Jul. 2025)

International Organization for Standardization – ISO (2015). *ISO 19109:2015 Geographic information — General feature model and rules for application schema*. URL: https://www.iso.org/standard/84700.html (accessed 1 Jul. 2025)

International Organization for Standardization – ISO (2023). *ISO 19156:2023 Geographic information — Observations, measurements and samples*. URL: https://www.iso.org/standard/82463.html (accessed 1 Jul. 2025)

International Organization for Standardization – ISO (2023). *ISO 19157-1:2023 Geographic information — Data quality*. URL:

https://www.iso.org/standard/78900.html (accessed 1 Jul. 2025)

International Organization for Standardization – ISO (2024). *ISO 19103:2024 Geographic information — Conceptual schema language*. URL: https://www.iso.org/es/contents/data/standard/08/34/83454.html (accessed 1 Jul. 2025)

International Organization for Standardization – ISO; International Electrotechnical Commission - IEC (2017). *ISO/IEC 19941:2017(E) Information technology — Cloud computing — Interoperability and portability.* URL: https://www.iso.org/standard/66639.html (accessed 1 Jul. 2025)

IUGS Commission for the Management and Application of Geoscience Information (CGI) (2008). *EarthResourceML (ERML)*. URL: https://earthresourceml.org/ (accessed 1 Jul. 2025)

Monash University (2025). *Urban mining: How coal waste could power the green revolution.* URL: https://lens.monash.edu/@technology/2025/08/13/1387740/urban-mining-how-coal-waste-could-power-the-green-revolution (accessed 1 Jul. 2025)

NextGenerationEU – TI-RRES. *Evaluación del potencial de recuperación de tierras raras (REE) y otros metales (Nb, Ta, Sn) a partir de residuos mineros de la mina de Golpejas y su restauración ecológica.* URL: https://www.igme.es/proyecto/proyecto-ti-rres/ (accessed 1 Jul. 2025)

Open Geospatial Consortium (2008). *OGC Geoscience Markup Language (GeoSciML)*. URL: https://www.ogc.org/standards/geosciml/ (accessed 1 Jul. 2025)

The Open Group (2008). *Exploration, Mining, Metals & Minerals Vertical (EMMM)*. URL: https://www.opengroup.org/content/exploration-mining-metals-minerals-emmm-forum-0 (accessed 1 Jul. 2025)

United Nations Economic Commission for Europe – UNECE (1997). *United Nations Framework Classification for Resources (UNFC)*. URL: https://unece.org/sustainable-energy/united-nations-framework-classification-resources-unfc (accessed 1 Jul. 2025)

University of Kentucky – Center for Applied Energy Research (UK/CAER). *Rare-Earth Elements in US Coal-Based Resources: Sampling, Characterizations, and Round-Robin Inter-laboratory Study*. URL: https://ees.as.uky.edu/uk-caer-receives-more-rare-earth-element-research-funding (accessed 1 Jul. 2025)

West Virginia University (2023). URL:

https://wvutoday.wvu.edu/stories/2023/04/05/wvu-researchers-earn-8m-for-rare-earth-extraction-facility-an-economic-and-environmental-game-changer (accessed 1 Jul. 2025)

World Wide Web Consortium – W3C (2013). *An Overview of the PROV Family of Documents.* URL: http://www.w3.org/TR/prov-overview/ (accessed 1 Jul. 2025)

World Wide Web Consortium – W3C (2024). *Data Catalog Vocabulary (DCAT) - Version 3*. URL: https://www.w3.org/TR/vocab-dcat-3/ (accessed 1 Jul. 2025)

World Wide Web Consortium – W3C (2025). *Verifiable Credentials Data Model v2.0*. https://www.w3.org/TR/vc-data-model-2.0/ (accessed 1 Jul. 2025)