RESPED

Enabling Energy Resilience through new energy flexible and affordable PED concepts

DELIVERABLE 4.2:
RESPED DATA MANAGEMENT PLATFORM

Date: 30/09/2026

AUTHOR(S)	ORGANISATION
Giorgos Papadopoulos	Suite5 Data Intelligence Solutions
Giorgos Bikas	Suite5 Data Intelligence Solutions

Co-funded by the European Union

Republic of Cyprus

RESEARCH & INNOVATION FOUNDATION

The project EP/DUT/0923/0006 is implemented under the programme of social cohesion "THALIA 2021-2027" co-funded by the European Union, through Research and Innovation Foundation.

Abstract

This deliverable reports the work carried out in Task 4.2 - Interoperable Data Collection and Management Platform of the RESPED project, and presents the resulting RESPED Data Management Platform. The document follows the work in the order it was carried out, from the data available at the demonstration sites to the delivered software. It starts with the data landscaping exercise that established which data assets are available at the demonstration sites and under which conditions they can be collected. It then presents the definition of the RESPED Common Information Model (RESPED-CIM) as the common semantic layer of the project, and the use case and requirements catalogue derived from both. It continues with the design of the platform services that answer those requirements and closes with the operational platform itself, illustrated through its main configuration and data management workflows. The platform supports file, third-party API and Kafka data collection, semantic harmonisation against the RESPED-CIM, data quality enhancement, anonymisation and configurable access control, and API endpoints through which harmonised datasets and results are exposed to the RESPED decision support dashboard developed in Task 4.3 and to other authorised data consumers.

Keywords

Data collection, data management platform, common information model, semantic interoperability, data harmonisation, data quality, anonymisation, access control, API, Positive Energy District

Revision	Date	Description	Author (Organisation)
0.9	28.09.2026	Draft for internal review	Giorgos Papadopoulos (Suite5)
0.91	30.09.2026	Adaptation based on feedback received from Czech Technical University	Giorgos Papadopoulos (Suite5)
0.92	30.09.2026	Adaptation based on feedback received from JOANNEUM RESEARCH	Giorgos Papadopoulos (Suite5)
1.0	30.09.2026	Ready to submit version	Giorgos Papadopoulos (Suite5)

EXECUTIVE SUMMARY

Under Task 4.2 of RESPED, the data layer of the project is delivered, namely the tools through which data flowing from building systems and distributed energy resources at the demonstration sites is collected, harmonised, quality-checked, protected and made available to the analytical services of the project. This deliverable reports what was built, why it was built that way, and how it is used.

The work started from the data itself. A data landscaping exercise was run with the demonstration site owners through a structured data collection template, covering the data assets that the sites can make available to the project. For each asset the exercise recorded format, velocity, temporal and spatial resolution, historical coverage, ownership, accessibility method, update strategy, privacy classification and the need for anonymisation. The picture that emerged is uneven, and that is what the platform had to absorb. One site exposes both a measured history and near real time measurements through its building management system, while the others contribute historical exports and modelled series. Each recorded attribute translated into a platform capability. The coexistence of file exports and live measurements defined the ingestion methods to be supported. The range of temporal resolutions defined how resolution is described in the dataset metadata. The dwelling level consumption series defined the anonymisation requirement. In addition, the privacy classifications declared by the data owners defined the access control model.

In parallel, the RESPED Common Information Model (RESPED-CIM) was defined as the semantic backbone of the platform. It was built for the project, from the landscaped data assets and the reporting needs of RESPED on one side, and from the standards and ontologies relevant to Positive Energy Districts on the other, namely SAREF with its SAREF4ENER and SAREF4BLDG extensions, the IEC Common Information Model, and IFC for building structure. The result is a model of thirty concepts and multiple fields per concept, covering demand, generation, flexibility, building structure, assets, weather, occupancy, incidents and KPI reporting. They are used by the platform as the target of every field mapping performed at ingestion.

Landscaping results and the CIM were then consolidated into a use case and requirements catalogue for the platform: eleven use cases and twenty-four functional requirements, cross-referenced to one another, and circulated to the consortium for review. Those requirements drove the design of five platform services, namely data collection, which covers ingestion, semantic harmonisation and data quality, security and storage, which covers the common information model, storage, anonymisation and access control, search and retrieval, notifications, and organisation management, and the platform was implemented against them.

The RESPED Data Management Platform is operational. It ingests data from files, from third-party REST APIs and from Kafka brokers, and maps every incoming field to the RESPED-CIM with unit normalisation applied at ingestion. It applies configurable cleaning rules per-field, as well as anonymisation rules for personally attributable data fields. The platform also enforces the access level set for each dataset by the user who created it, from visible to that user alone through to public. Harmonised datasets and results are exposed through authenticated, user-configured JSON endpoints. This document walks through these workflows as they appear in the platform.

With the platform being operational, the immediate next step is the connection of the Task 4.3 AI-enabled decision support dashboard to the live retrieval endpoints and the progressive extension of ingestion coverage as further data assets become available at the demonstration sites.

Table of contents

EXECUTIVE SUMMARY 2

TABLE OF CONTENTS 3

ACRONYMS AND TERMS 6

LIST OF FIGURES 7

LIST OF TABLES 7

INTRODUCTION 9

1.1. Purpose of the document 9

1.2. Approach and structure of the document 9

1.3. Relation to other project activities 10

2. DATA LANDSCAPING 11

2.1. Purpose and methodology 11

2.2. Attributes captured 11

2.3. Data assets and their sources 12

2.4. From data attributes to platform capabilities 13

3. THE RESPED COMMON INFORMATION MODEL 14

3.1. Purpose and methodology 14

3.2. Standards and data models considered 14

3.3. Design decisions 15

3.4. Structure of the model 16

3.5. EnergyDemandMeasurements as an example 19

3.6. Use of the model in the platform 20

4. USE CASES AND REQUIREMENTS 21

4.1. Methodology 21

4.2. Use case overview 21

4.3. Requirements overview 22

5. DESIGN OF THE RESPED DATA MANAGEMENT PLATFORM 24
5.1. Architecture overview 24
5.2. Data Collection Service 25
5.2.1 Data Ingester 25
5.2.2 Data Harmonisation 26
5.2.3 Data Quality 26
5.3. Security and Storage Service 27
5.3.1 Common Information Model 27
5.3.2 Secure Data and Metadata Storage 27
5.3.3 Anonymisation 27
5.3.4 Access Control 28
5.4. Search and Retrieval Service 28
5.4.1 Search and Query Creation 28
5.4.2 Open APIs 28
5.5. Notification Service 28
5.6. Organisation Management Service 29
5.7. Technology stack and deployment 29

6. THE RESPED DATA MANAGEMENT PLATFORM IN OPERATION 30
6.1. Platform dashboard 30
6.2. Configuring data ingestion 30
6.3. Data Harmonisation 33
6.4. Data quality and cleaning 34
6.5. Anonymisation 35
6.6. Data catalogue and access control 36
6.7. API configuration and data retrieval 37
6.8. Organisation and user management 38

7. CONCLUSIONS AND NEXT STEPS 39

8. REFERENCES 40

ANNEX I – T4.2 USE CASE CATALOGUE 41

ANNEX II – REQUIREMENTS COMPLIANCE MATRIX 47

ACRONYMS AND TERMS

Acronym	Full name
AI	Artificial Intelligence
API	Application Programming Interface
CIM	Common Information Model
CSV	Comma-Separated Values
DCP	Data Collection Platform (identifier prefix used for T4.2 use cases and requirements)
DUT	Driving Urban Transitions Partnership
Dx.y	Deliverable x.y
ETSI	European Telecommunications Standards Institute
EV	Electric Vehicle
GDPR	General Data Protection Regulation
HTTP(S)	Hypertext Transfer Protocol (Secure)
IEC	International Electrotechnical Commission
IFC	Industry Foundation Classes
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
OCPP	Open Charge Point Protocol
OpenADR	Open Automated Demand Response
OSCP	Open Smart Charging Protocol
PED	Positive Energy District
PV	Photovoltaic
REQ	Requirement
RESPED-CIM	RESPED Common Information Model
REST	Representational State Transfer
SAREF	Smart Applications REFerence ontology
SAREF4BLDG	SAREF extension for Buildings
SAREF4ENER	SAREF extension for Energy

Acronym	Full name
Tx.y	Task x.y
UC	Use Case
UI	User Interface
ULL	Urban Living Lab
URL	Uniform Resource Locator
USEF	Universal Smart Energy Framework
WP	Work Package

LIST OF FIGURES

Figure 1: Overview of the RESPED-CIM concepts by domain group 16
Figure 2: Architecture of the RESPED Data Management Platform 24
Figure 3: Platform dashboard with summary counters and recent ingestion executions 30
Figure 4: Ingestion jobs of the user 31
Figure 5: Creation of a new ingestion job 32
Figure 6: Setup of a file ingestion job 33
Figure 7: Linking of source columns to the RESPED-CIM 34
Figure 8: Configuration of data quality rules per field 35
Figure 9: Classification of fields for anonymisation 36
Figure 10: Data catalogue with the assets available to the user 37
Figure 11: Selection of assets and fields for a retrieval query 38

LIST OF TABLES

Table 1: Relation of Task 4.2 to other RESPED activities 10
Table 2: Extract of the RESPED data collection template 11
Table 3: Overview of the data landscape across the demonstration sites 12
Table 4: Translation of data asset attributes into platform capabilities 13
Table 5: RESPED-CIM design decisions and their rationale 15
Table 6: Concepts of the RESPED Common Information Model 19
Table 7: Extract of the EnergyDemandMeasurements concept (18 of 42 fields) 20
Table 8: T4.2 use case overview 22
Table 9: T4.2 requirements overview 23
Table 10: Source types supported by the Data Collection Service 26
Table 11: Data quality and cleaning rules 26
Table 12: Anonymisation strategies and the field classifications they apply to 27
Table 13: Platform technology stack 29

Introduction

1.1. Purpose of the document

This deliverable reports the work performed in Task 4.2 - Interoperable Data Collection and Management Platform of Work Package 4 of the RESPED project, and presents the software delivered by that task, the RESPED Data Management Platform.

Task 4.2 is led by Suite5 with the participation of all partners and runs from M10 to M21 [1]. It covers the collection of data as batch files, through third-party APIs and through streaming mechanisms, the definition of the RESPED data model and of the mapping mechanisms that apply it, data quality enhancement, data anonymisation and access control, and the standards based API endpoints through which the collected data is retrieved. Its purpose is to give RESPED a single, controlled point through which data produced at the demonstration sites is collected, understood, cleaned, protected and shared. Without it, every analytical activity in the project would have to negotiate its own access to raw site data, in whatever format the site happens to produce, with no common vocabulary and no consistent handling of personal data. The platform removes that problem: data enters once, is mapped once to a common model, and is served from there to everything downstream.

The chapters of the document are sequential. The deliverable starts with the analysis of the available data at the pilot sites, continues with the tasks that the platform has to do as a result, the way the platform was designed to execute them, and how the platform looks in operation, with each chapter providing the necessary information to the next one.

1.2. Approach and structure of the document

The scope of this document is Task 4.2 and the platform it delivers. The contractual scope describes the capabilities the platform should offer, while the specific needs of the demonstration sites become clear only once their data is examined. The work was therefore carried out in four steps, which also give the structure of this document.

- Data landscaping. A structured template was circulated to the data-owning partners, capturing for every candidate data asset its format, velocity, resolution, coverage, ownership, accessibility, update strategy and privacy classification. This established the scope of what the platform must actually handle, and is reported in Chapter 2.
- Common Information Model definition. The RESPED-CIM was defined from the landscaped data assets and the reporting needs of the project, constructed against the relevant standards and ontologies so that its concepts and fields carry established semantics. It is reported in Chapter 3.
- Use case and requirements definition. The landscaping outcomes and the CIM were consolidated into eleven use cases and twenty-four functional requirements, circulated to the consortium for review. They are reported in Chapter 4.
- Design and implementation. The requirements were allocated to platform services, the services were designed and implemented, and the platform was deployed and is ready to be populated with data from the demonstration sites. The design is reported in Chapter 5 and the platform in operation in Chapter 6.
- Chapter 7 concludes, presents what was delivered and states the next steps.
  Annex I contains the use case catalogue and Annex II the requirements compliance matrix.

1.3. Relation to other project activities

Task 4.2 analyses the data related pilot level activities of the project and provides feedback to the analytical ones.

Activity	Relation to Task 4.2
T1.4 - Integrated methodological framework and KPIs	Defines the project KPI framework [2]. The platform stores data to be used for the visualisation of KPIs at the RESPED dashboard.
T4.3 - AI enabled decision support dashboard	Primary consumer of the platform. Retrieves harmonised datasets and results exclusively through the platform's API endpoints, with no direct access to site systems.
T3.1 - Energy Modelling	Provider of modelled outputs that can be ingested into the platform as datasets
T2.4 - Evaluation and experiences of Living Lab PEDs	Uses the data held by the platform for the monitoring and evaluation of the Living Lab PEDs.

Table 1: Relation of Task 4.2 to other RESPED activities

2. Data Landscaping

2.1. Purpose and methodology

Data landscaping is the activity of establishing, before any development starts, what data exists, who owns it, in what form it can be obtained, and under what conditions it may be used. In the context of D4.2 it had a second purpose: to convert those findings into platform requirements, so that the platform is specified against data that is known to be available rather than against data that is assumed to be.

The activity was run in four steps:

- A data collection template was defined by Suite5 and circulated to the data-owning partners, with a worked example for each column so that entries would be comparable across partners.
- Partners completed the template for every data asset they can make available to the project, whether already available or planned.
- Suite5 reviewed each entry with the provider, resolved ambiguities in resolution, units and coverage, and obtained access to the source systems to verify the declarations against the actual data.
- The verified entries were analysed attribute by attribute and translated into platform capabilities, which then entered the requirements catalogue.

The template was maintained as a living document throughout the task.

Basic information	Data asset features				Availability	Data governance	
Data asset	Format	Velocity	Temporal resolution	Temporal coverage	Owner	Privacy	Anonymisation
Energy imported from the grid (kWh)	CSV	Near real-time	5 min	From 1.1.2024	NOHO	Private	Yes
PV generation (kWh)	CSV	Near real-time	15 min	From 1.1.2024	NOHO	Private	Yes
Heat pump heating consumption (GJ)	CSV	Near real-time	5 min	From 1.1.2024	NOHO	Private	Yes
Gas boiler heating consumption (GJ)	CSV	Near real-time	5 min	From 1.1.2024	NOHO	Private	Yes
Electricity consumption, flats 1-48 (kWh)	CSV	Near real-time	5 min	From 1.1.2024	NOHO	Private	Yes
Cold water consumption, flats 1-48 (m^3)	CSV	Near real-time	5 min	From 1.1.2024	NOHO	Private	Yes
Electricity consumption, EV charging points 1-20 (kWh)	CSV	Near real-time	5 min	From 1.1.2024	NOHO	Private	Yes
Space heating, cooling, DHW and heat pump demand (W)	CSV	Batch	15 min	From 1.8.2024	JR	Private	No
PV generation, internal gains, indoor and outdoor conditions	CSV	Batch	15 min	From 1.8.2024	JR	Private	No
Thermal energy demand of the PED buildings	CSV	Batch	Hourly	1 year	UNICAL	Proprietary	No

Table 2: Extract of the RESPED data collection template

2.2. Attributes captured

The template captures four groups of attributes for each data asset. Basic information includes the title of the asset and a description, together with the corresponding column in the exported files and,

where an interface exists, the corresponding data point in the source system API. Data asset features include volume, format, velocity, historical availability, temporal resolution, temporal coverage and spatial coverage. Data asset availability captures the owner, whether the asset originates from a third party, the provider, the accessibility method, the frequency of updates and the update strategy. Data governance includes the privacy classification, the need for anonymisation and the types of analysis currently performed on the data.

Not every attribute could be completed for every asset, since some depend on how the data is delivered and others on decisions that remain with the data owners. Table 2 shows an extract of the template for a selection of assets from the three demonstration sites. It is limited to the attributes that were declared consistently across them, namely the title of the asset, its format, velocity, temporal resolution and temporal coverage, its owner, its privacy classification and the need for anonymisation.

Each of these groups exists for a reason, and each drives a specific part of the platform.

2.3. Data assets and their sources

The landscaping covered the three demonstration sites of the project with available data. Table 3 summarises the outcome per site. The complete asset level record, with the attribute values declared for each entry, is maintained in the data collection template referenced in Section 2.1.

Demonstration site	Data categories provided	Availability and resolution	Collection mode
Czech site (Cihlovka 2, Kukleny, Hradec Kralove)	District level grid import and export, PV generation, thermal production from heat pumps and gas boilers for space heating, cooling and domestic hot water, ventilation heating, common area and dwelling level electricity, dwelling level heating, cooling, EV charging	Measured data. Historical series from 2024 onwards and near real time, at five and fifteen minute resolution	File export and third-party API
Austrian site	Building and district level energy consumption and generation measurements	Measured data at 15 minute resolution. Historical series	File export
Italian site	Thermal energy demand of the PED buildings	Modelled. Hourly series covering a reference year	File export
All sites, contextual data	Weather observations and forecasts for the site locations	Measured and forecast, continuous	Third-party API

Table 3: Overview of the data landscape across the demonstration sites

On the Czech site, measurements are produced and stored by the Mervis building management system from which historical files can be downloaded. The Mervis system offers additionally an API, which serves the same measurements programmatically. The remaining sites deliver exported files, which were inspected directly.

Data inspection established what the ingestion mechanism has to support across sites, namely source specific naming conventions, the use of kWh, MWh and GJ across assets measuring comparable quantities, cumulative meter readings alongside interval values, and gaps in the series where meters were offline. The API analysis established the corresponding data points for each asset, the authentication scheme, the query parameters available for time range selection, and the response structure.

The outcome of this analysis was recorded back into the data collection template, which links the file based assets to the columns in which they appear in the exported files.

2.4. From data attributes to platform capabilities

Each attribute captured in the template answers a design question. Table 4 sets out that relation.

Data attribute	Platform capability it determines
Method of data collection	Support for the specific ingestion methods required: file upload, scheduled API pull, Kafka subscription
Format of data files	Support for the file formats encountered, CSV and JSON, with the number and date formats of the source declared at harmonisation
Temporal and spatial coverage and resolution	Metadata definition
Accessibility method	Support for retrieval variations and for the authentication schemes of the source systems
Update strategy	Support for append behaviour on subsequent executions of the same pipeline
Need for anonymisation	Per-field anonymisation, based on the role of each field in identifying a person
Privacy classification	Dataset level access set by the creating user, from a single user through to public
Units used at source	Unit normalisation applied at ingestion time as part of field mapping
Historical availability and gaps	Completeness indicators computed per execution, and gaps handled in the preparation of the results served to the analytical services

Table 4: Translation of data asset attributes into platform capabilities

3. The RESPED Common Information Model

3.1. Purpose and methodology

The platform ingests data from systems that were not designed to talk to each other, and serves it to analytical components that were not designed around any of those systems. A common information model is what makes that possible: a single vocabulary, defined once, and used to map every incoming field. Without it, each consumer would need to understand each source, and every new source would multiply the integration effort.

Suite5 defined the RESPED Common Information Model, RESPED-CIM, to serve that role. Three principles governed its definition.

- Standards first. Every concept and field is defined against established models and ontologies, so that RESPED terminology carries semantics that are recognised outside the project. Terms were introduced by the project only where the standards reviewed offer no equivalent.
- Fit to the demonstration sites. The model covers what RESPED measures and reports. Concepts with no counterpart in the project were left out rather than added for completeness, because an oversized model makes mapping harder and adds no value.
- Available in the platform. The model is not a paper artefact. It is loaded into the platform and used as the target of every mapping, which imposes practical constraints on naming, typing, and the metadata carried by each field.

The definition started from the data assets established by the landscaping activity and the reporting needs of the project, and continued with the analysis of the relevant standards and ontologies. The concepts, fields, types, units and relationships were then defined, and the model was uploaded to the platform and used.

3.2. Standards and data models considered

The review covered the models that are relevant to the domains RESPED touches: the PED definition, energy assets and devices, measurements, and flexibility and resilience events.

SAREF, the Smart Applications REFerence ontology [3], is the ETSI-standardised reference ontology for smart appliances and the de facto starting point for device-level semantic interoperability in the energy domain. Its core abstraction, devices that perform functions, expose states and produce measurements of properties with explicit units, matches the structure of the data produced at the RESPED sites, where every asset is a device reporting measured properties at intervals.

SAREF4ENER [4] extends the core ontology towards energy management, adding the notions of power and energy profiles, flexibility, and the interaction between a customer energy manager and controllable devices. It is the most direct source of alignment for the energy asset concepts of RESPED-CIM, namely heat pumps, boilers, EV charging stations and photovoltaic systems. The measurement concepts follow the measurement semantics of the core SAREF ontology.

SAREF4BLDG [5] extends the core ontology towards buildings, importing the building topology of IFC, building, storey, space, zone, and the equipment installed in them. It is the source of alignment for the spatial and built environment concepts of RESPED-CIM, including occupancy, which matter in RESPED because measurements have to be attributed to a flat, a storey, a building and a district, and aggregated along that hierarchy.

The IEC Common Information Model, standardised in the IEC 61970 [6] and IEC 61968 [7] series, is the reference model for electricity network information and for the exchange of operational data between utilities. Its relevance to RESPED lies in disruption events. Outages, incidents and their logs have mature definitions in IEC CIM, and using them keeps the project aligned with how distribution system operators describe the same facts. Grid network and market level entities, such as substations, transformers and system operators, were not modelled as concepts, since they fall outside the boundary of a Positive Energy District and cannot be measured from the data available at the sites.

IFC, the Industry Foundation Classes [8], is the open standard for building information modelling and the origin of the spatial hierarchy adopted through SAREF4BLDG. RESPED does not carry full building geometry, but it does need the spatial containment relationships and a stable identifier scheme for spaces, so that a measurement recorded for a flat can be resolved to its storey, its building and the district.

Further initiatives were reviewed and not adopted as structural sources, either because their scope is narrower than RESPED needs or because the relevant content is already reachable through the models above: OCPP and OSCP for electric vehicle charging, USEF and OpenADR for flexibility market interactions, and the various smart city KPI vocabularies. Where these define terminology that the project uses, that terminology is reflected in the related terms attached to RESPED-CIM fields rather than in additional concepts.

3.3. Design decisions

The model was built for RESPED and the decisions taken during its creation are recorded here because they explain why the model looks the way it does.

Decision	Rationale
Limit the model concepts to the necessary ones	Only concepts with a counterpart in the demonstration sites or in the reporting needs of the project were defined. A compact model makes the mapping step faster and less error prone for platform users.
Cover electricity, thermal, gas and water demand within a single demand concept	The sites report heating, cooling, domestic hot water and ventilation demand alongside electricity, and report grid import and export at the district boundary. Carrying them in one concept keeps cross vector comparison straightforward, while the energy vector remains explicit on every record.
Adopt the IFC-derived identifier scheme for BuildingSpace	Aligns space identifiers with the way the sites describe their own dwellings, which removes an entire class of mapping errors at ingestion.
Add PositiveEnergyDistrict as an explicit concept	The PED is the unit of analysis of the project and the level at which most KPIs are reported. It needed to exist as a first-class concept rather than being implied by a grouping of buildings.
Include KeyPerformanceIndicator and KeyPerformanceIndicatorValue	Allows KPI definitions and their computed values to be carried in the same model as the measurements they derive from, so that reported results remain traceable to their inputs.

Table 5: RESPED-CIM design decisions and their rationale

3.4. Structure of the model

RESPED-CIM consists of 30 concepts and 781 fields. The concepts fall into six domain groups, namely spatial and built environment, energy assets and devices, measurements, flexibility and resilience, evaluation and KPIs, and PED core and temporal context. Figure 1 gives an overview of the concepts in each group, together with the standards used during their definition.

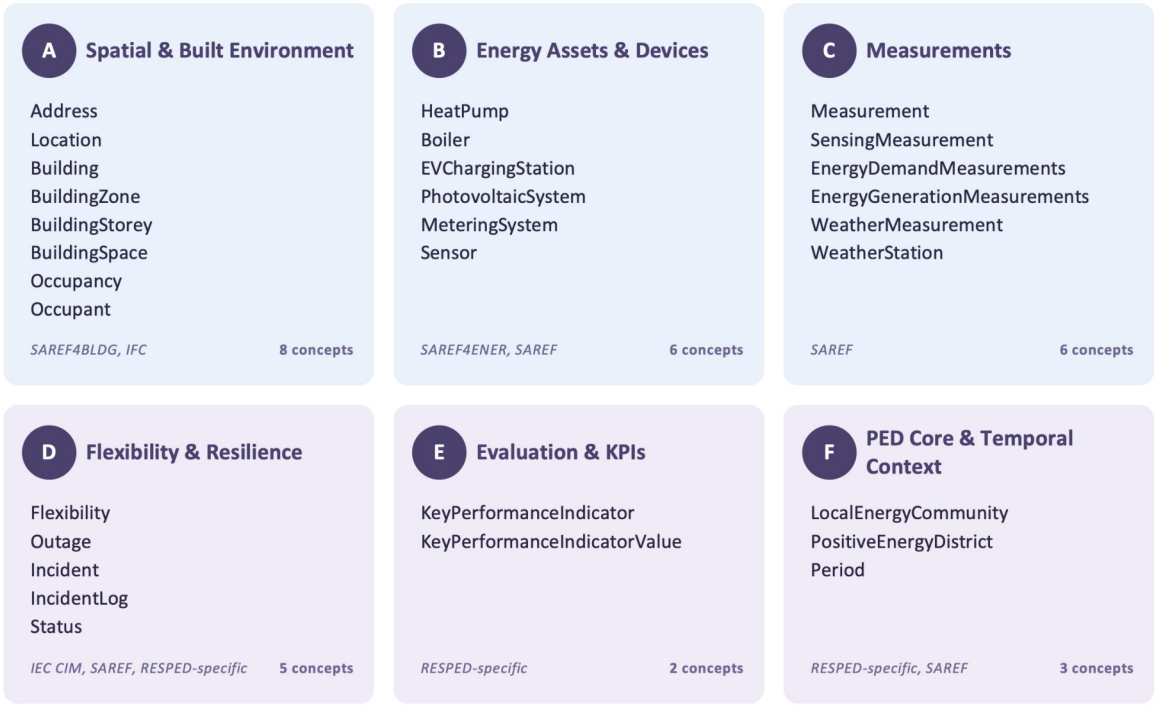


Figure 1: Overview of the RESPED-CIM concepts by domain group

Table 6 lists the concepts, the number of fields each carries, its role in the model and the standard that affected their definition.

Group	Concept	Fields	Role in the model	**Standard**
A. Spatial & Built Environment	Address	24	Postal address of a stakeholder, building or asset	SAREF4BLDG
	Location	36	Geographic location, coordinates and spatial reference	SAREF4BLDG, IFC
	Building	33	Building characteristics, construction, envelope and installed systems	SAREF4BLDG, IFC

Group	Concept	Fields	Role in the model	**Standard**
	BuildingZone	34	Aggregation of spaces or partial spaces sharing a common characteristic	IFC
	BuildingStorey	24	Storey of a building, with elevation and the spaces it contains	IFC
	BuildingSpace	49	Individual space within a building, in RESPED a residential flat	SAREF4BLDG, IFC
	Occupancy	18	Occupancy of a building, storey or space over time	SAREF4BLDG
	Occupant	16	Person present in a building at a given time	SAREF4BLDG
B. Energy Assets & Devices	HeatPump	25	Heat pump asset characteristics and operating parameters	SAREF4ENER
	Boiler	11	Gas boiler asset providing space heating and domestic hot water	SAREF4ENER
	EVChargingStation	18	Electric vehicle charging point, its connectors and charging sessions	SAREF4ENER
	PhotovoltaicSystem	44	Photovoltaic system characteristics, orientation, capacity and inverter information	SAREF4ENER
	MeteringSystem	19	Traditional or smart metering system and its registers	SAREF
	Sensor	24	Sensing device, its type, location and measured properties	SAREF
C. Measurements	Measurement	30	Generic measurement of a property, with value, unit, timestamp and quality attributes	SAREF

Group	Concept	Fields	Role in the model	**Standard**
	SensingMeasurement	41	Measurements produced by sensing devices, including environmental and comfort parameters	SAREF
	EnergyDemandMeasurements	42	Electricity, heating, cooling, domestic hot water and ventilation demand, and grid import and export at the district boundary	SAREF
	EnergyGenerationMeasurements	23	Generation measurements, in RESPED primarily photovoltaic production at site and system level	SAREF
	WeatherMeasurement	34	Observed and forecast atmospheric conditions at a location	SAREF
	WeatherStation	16	Weather station or weather data source providing the measurements	SAREF
D. Flexibility & Resilience	Flexibility	32	Flexibility capacity and forecast for controllable assets, covering both upward and downward availability	RESPED-specific
	Outage	26	Grid or energy disruption event recorded at a demonstration site	IEC CIM
	Incident	26	Problem occurring in the network or in a device that may affect operation	IEC CIM
	IncidentLog	17	Log report associated with an incident event	IEC CIM
	Status	19	State of a system or device, such as on/off or open/closed	SAREF

Group	Concept	Fields	Role in the model	**Standard**
E. Evaluation & KPIs	KeyPerformanceIndicator	14	Definition of a KPI, including its formula, unit and reporting frequency	RESPED-specific
	KeyPerformanceIndicatorValue	13	Computed value of a KPI for a given period or timestamp	RESPED-specific
F. PED Core & Temporal Context	LocalEnergyCommunity	26	Local energy community, used to represent the collective energy arrangement of a district	RESPED-specific
	PositiveEnergyDistrict	9	The Positive Energy District as the unit of analysis of the project	RESPED-specific
	Period	38	Time period or interval to which a record refers	SAREF

Table 6: Concepts of the RESPED Common Information Model

3.5. EnergyDemandMeasurements as an example

EnergyDemandMeasurements is the most heavily used concept of the model, since every consumption series collected from the demonstration sites maps to it. It carries 42 fields, and Table 7 shows an extract of them. Each field declares its type, and where applicable its measurement unit, which is what allows the platform to apply unit normalisation automatically at ingestion.

Field	Type	Measurement unit
totalEnergyConsumption	double	Kilowatt hour
activePower	double	Kilowatt
spaceHeatingDeviceLoad	double	Kilowatt
boilerLoad	double	Kilowatt
realTimeDemand	double	Kilowatt
airConditioningSystemHeatingConsumption	double	Kilowatt hour
electricVehicleLoad	double	Kilowatt hour

Field	Type	Measurement unit
peakLoad	double	Gigawatt
airConditioningSystemLoad	double	Kilowatt
gridLoad	double	Gigawatt
forecastLoad	double	Kilowatt
forecastDateTime	datetime	-
name	string	-
id	string	-
dayAheadDemand	double	Kilowatt
updatedDateTime	datetime	-
createdDateTime	datetime	-
description	string	-

Table 7: Extract of the EnergyDemandMeasurements concept (18 of 42 fields)

3.6. Use of the model in the platform

RESPED-CIM is loaded into the platform and exposed to the user at the harmonisation step of every ingestion pipeline. When a source is configured, the platform detects the incoming schema and presents each source field along with the concepts and fields of the RESPED-CIM. The user maps or excludes each field, the platform records the mapping together with any unit transformation implied by the difference between the source unit and the model unit, and applies it automatically on every subsequent execution of that pipeline.

4. Use cases and requirements

4.1. Methodology

The use case and requirements catalogue is the bridge between the findings of Chapters 2 and 3 and the platform design of Chapter 5. It was produced in four steps. Use cases were drafted from the landscaped data assets, the contractual scope and the needs expressed by the analytical tasks. Functional requirements were then derived from each use case. The result was consolidated and overlapping entries were removed. Finally, the catalogue was reviewed by the consortium and adapted.

Use cases describe what the platform does from the point of view of an actor. Requirements state what the platform must provide for that to be possible. Requirements using shall are mandatory, while requirements using should are recommended. Every requirement is traceable to at least one use case, and every use case is covered by at least one requirement.

The catalogue contains 11 use cases and 24 functional requirements. Use case identifiers follow the convention RESPED-UC-DCP-XX and requirement identifiers follow RESPED-REQ-DCP-XX, where DCP stands for the data collection platform.

4.2. Use case overview

The eleven use cases cover the full lifecycle of data inside the platform, from the configuration of a source to the consumption of the resulting dataset by an analytical service. The table below summarises them. The complete catalogue, with actors, triggers, pre-conditions and post-conditions for each, is given in Annex I.

UC ID	Title	Priority	Requirements (REQ-DCP-)
UC-01	Ingest PED-related data from heterogeneous sources	Mandatory	01, 02, 03
UC-02	Harmonise ingested data using RESPED semantics	Mandatory	04, 05, 06
UC-03	Validate and improve data quality	Mandatory	07, 08
UC-04	Manage dataset metadata for discovery and retrieval	Mandatory	09, 10
UC-05	Control access to PED data	Mandatory	11, 12
UC-06	Anonymise sensitive PED data	Mandatory	13, 14
UC-07	Expose data to authorised services via standardised APIs	Mandatory	15, 16, 17
UC-08	Track data provenance and processing history	Recommended	18
UC-09	Notify users of data ingestion and quality events	Recommended	19, 20

UC ID	Title	Priority	Requirements (REQ-DCP-)
UC-10	Expose results for retrieval by analytical services	Mandatory	21, 22, 23
UC-11	Manage platform users and organisations	Mandatory	24

Table 8: T4.2 use case overview

Use cases 01 to 03 describe the ingestion configuration workflow, which is a single multi-step process in the platform: the user configures the source, maps its fields to the RESPED-CIM, and optionally configures quality rules and anonymisation before activating the pipeline. Use cases 04 and 05 describe what happens to the dataset once it exists - how it is described, discovered and protected. Use case 06 covers anonymisation, which applies only to datasets that are complete at ingestion. Use cases 07 and 10 describe the output side of the platform. Use cases 08, 09 and 11 cover provenance, notifications and administration.

4.3. Requirements overview

The twenty-four functional requirements derived from the use cases are listed below, each with the use case it comes from and whether it is mandatory or recommended. They cover data collection with its harmonisation and quality steps, security and storage, search and retrieval, notifications, and organisation management.

Twenty of the twenty-four requirements are mandatory. The four recommended ones concern provenance reporting and notifications, which improve the operation of the platform but are not preconditions for the data flows the project depends on.

Requirement	Summary	Type	Use case
REQ-DCP-01	Three source types supported: File, External API, Kafka	Mandatory	UC-01
REQ-DCP-02	Multi-step ingestion configuration workflow	Mandatory	UC-01
REQ-DCP-03	Operational monitoring view of active pipelines	Mandatory	UC-01
REQ-DCP-04	Field mapping to RESPED-CIM with exclusion option	Mandatory	UC-02
REQ-DCP-05	Inferred types, sample values and unit normalisation labels	Mandatory	UC-02
REQ-DCP-06	Multi-energy vector semantics aligned with SAREF and SAREF4ENER	Mandatory	UC-02
REQ-DCP-07	Per-field quality and cleaning rule configuration	Mandatory	UC-03
REQ-DCP-08	Quality indicators computed and attached per execution	Mandatory	UC-03

Requirement	Summary	Type	Use case
REQ-DCP-09	User-defined and system-managed dataset metadata	Mandatory	UC-04
REQ-DCP-10	Searchable and filterable data catalogue	Mandatory	UC-04
REQ-DCP-11	Four level dataset access, from Only me to Public	Mandatory	UC-05
REQ-DCP-12	Access policy editable by the data owner at any time	Mandatory	UC-05
REQ-DCP-13	Per-field anonymisation configuration for complete datasets	Mandatory	UC-06
REQ-DCP-14	Anonymisation applied before the dataset is catalogued	Mandatory	UC-06
REQ-DCP-15	User configured retrieval endpoints over datasets and results	Mandatory	UC-07
REQ-DCP-16	Retrieval query wizard with access token authentication	Mandatory	UC-07
REQ-DCP-17	JSON responses on all outbound endpoints	Mandatory	UC-07
REQ-DCP-18	Processing summary recorded for every execution	Recommended	UC-08
REQ-DCP-19	Notification of pipeline and quality events	Recommended	UC-09
REQ-DCP-20	User-configurable notification triggers and thresholds	Recommended	UC-09
REQ-DCP-21	Results available for retrieval across demand, generation, flexibility and grid vectors	Mandatory	UC-10
REQ-DCP-22	Configurable gap filling and imputation	Mandatory	UC-10
REQ-DCP-23	Temporal aggregation and resampling at configurable resolutions	Mandatory	UC-10
REQ-DCP-24	User and organisation administration with role assignment	Mandatory	UC-11

Table 9: T4.2 requirements overview

5. Design of the RESPED Data Management Platform

5.1. Architecture overview

The platform is organised as a set of services, each owning one stage of the data lifecycle and each answering a defined group of requirements. Data moves through them in one direction, collected, harmonised, quality-checked, anonymised, stored and catalogued, then retrieved, with configuration and monitoring available at every stage. Configuration and execution are deliberately kept apart. The user sets up each stage separately and the platform then runs them as one pass from ingestion to the stored dataset, reporting the outcome of each stage as it goes. Retrieval is a later and independent step over the stored dataset. Figure 2 shows how the services fit together.

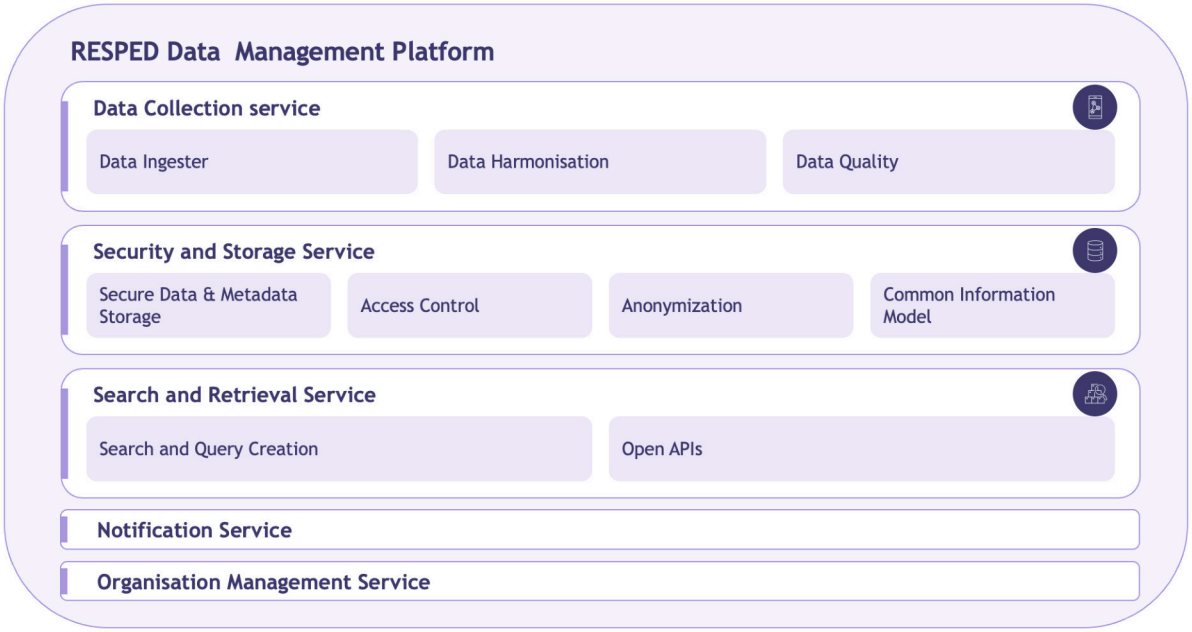


Figure 2: Architecture of the RESPED Data Management Platform

Five services compose the platform. The Data Collection Service carries data from its source to the point at which it is ready to be stored. The Security and Storage Service holds it, protects it, and holds the semantic model against which it was mapped. The Search and Retrieval Service is how data is identified and prepared to be retrieved from the platform. The Notification Service and the Organisation Management Service are cross-cutting: they own no stage of the data lifecycle but serve every stage of it.

- Data Collection Service, which acquires data and prepares it for storage. It comprises the Data Ingester, which manages the connection to each source and the scheduling and execution of ingestion pipelines, the Data Harmonisation component, which maps source fields to the RESPED-CIM and applies unit normalisation and transformation rules, and the Data Quality component, which applies per-field validation and cleaning rules and computes quality indicators.
- Security and Storage Service, which persists harmonised time series and their metadata and governs what may be done with them. It comprises Secure Data and Metadata Storage, Access Control, Anonymisation, and the Common Information Model, held here as the registry against which harmonisation maps and retrieval describes.

- Search and Retrieval Service, through which data leaves the platform. It comprises Search and Query Creation, through which a user identifies the datasets and results they need and defines the selection over them, and Open APIs, through which that selection is served to authorised consumers.
- Notification Service, which informs users of pipeline events.
- Organisation Management Service, which registers organisations, administers their membership and assigns the roles of their members.

The services group capabilities by ownership, not by order of execution, and two consequences of that are worth stating before the sections that follow. Anonymisation sits with storage because both concern what may be held and by whom, but it is invoked by the ingestion pipeline and applied to the data before it is persisted, so that no identifiable form of a dataset ever reaches the catalogue. The Common Information Model sits with storage for the same reason, as the registry the platform holds, but it is consumed by the Data Harmonisation component at ingestion and by the Search and Retrieval Service at the point of serving.

The unit of work in the platform is the ingestion pipeline. A pipeline binds a source, a mapping, an optional set of quality rules, an optional anonymisation configuration and a schedule into a single configured object that produces and maintains exactly one dataset. Every subsequent execution of the pipeline applies the same configuration, which is what makes repeated collection from the same source possible.

5.2. Data Collection Service

The Data Collection Service answers requirements REQ-DCP-01 to 08. It brings data into the platform and leaves it harmonised, validated and ready to be stored. Its three components run as a single pass over each file, response or message: the Data Ingester acquires the data, the Data Harmonisation component gives it RESPED semantics, and the Data Quality component applies the configured cleaning rules. Each is configured separately, which is what allows a pipeline to be corrected one stage at a time rather than rebuilt.

5.2.1 Data Ingester

The Data Ingester answers requirements REQ-DCP-01 to 03. It supports three source types, each configured through the same workflow but with its own connection parameters.

Source type	Used in RESPED for	Configuration parameters
File	Historical exports delivered by the demonstration sites and modelled datasets produced by project partners, in CSV and JSON	Sample file and full dataset file, in CSV or JSON. Append is not available when anonymisation is enabled
External API	Scheduled collection of measurements from site systems that expose a programmatic interface, and from third-party sources such as weather services	Endpoint URL, HTTP method, authentication scheme and credentials, custom headers, query parameters, pagination behaviour, and collection schedule
Kafka	Streaming ingestion where a site publishes measurements as they are produced	Broker address, topic, consumer group, security protocol and credentials, and message schema

Table 10: Source types supported by the Data Collection Service

Common to all three is the job configuration: the name and description of the pipeline, the pre-processing steps it includes, the access level of the job, and the schedule where one applies. The service maintains an operational view of every active pipeline showing its current status, the time of its last execution and its outcome.

Execution is asynchronous. A pipeline runs immediately on creation, according to its schedule, or in response to an incoming message where the source is a Kafka topic. When an execution fails, the user is notified and the failure is recorded in the execution history of the pipeline. The configuration itself is not invalidated, so once the cause has been corrected the pipeline can run again.

5.2.2 Data Harmonisation

The Data Harmonisation component answers requirements REQ-DCP-04 to 06. It is the step at which data acquires meaning in RESPED terms.

On first configuration of a pipeline, the service inspects the incoming data and detects the source schema: the column names and their data types, inferred from sample values. It presents each detected field alongside a selector containing the RESPED-CIM concepts and fields, and the user maps the field to its target or leaves it unlinked, in which case it is not stored.

Where the unit declared by the target field differs from the unit of the source, the service displays the normalisation that will be applied, for example kWh to MWh, or GJ to kWh, and applies it at ingestion time on every execution. This matters in RESPED because the same physical quantity arrives in different units across assets, and comparability across the district depends on resolving that consistently rather than at the point of analysis.

5.2.3 Data Quality

The Data Quality component answers requirements REQ-DCP-07 and 08. Quality configuration is an optional step of the ingestion workflow, enabled per pipeline and configured per field.

Rule type	Options and behaviour
Null handling	Drop the record, fill with the field mean, fill with zero, or fill with a configured constant
Range validation	Minimum and maximum acceptable values per field, with a configurable action on violation, replacing the value or dropping the record
Duplicate handling	Keep first, keep last, or drop all duplicate records for a given timestamp and identifier combination
Quality indicators	Completeness score and related indicators computed at every execution and attached to the dataset as metadata

Table 11: Data quality and cleaning rules

The rules are applied on every execution, not only on the first, so quality handling is a property of the pipeline rather than a one-off correction. Records affected by each rule are counted and reported in the processing summary of the execution.

5.3. Security and Storage Service

Once data has been collected and harmonised it has to be stored and protected, which is the role of the Security and Storage Service. It covers the common information model against which everything held is described, the storage of the data and its metadata, the anonymisation rules of personally attributable fields before they are stored, and the access control enforced on everything the platform holds. It answers requirements REQ-DCP-09, 11 to 14 and 18.

5.3.1 Common Information Model

RESPED-CIM, defined in Chapter 3, is not a document the platform consults but a model loaded into it. The Security and Storage Service holds it as a registry: the thirty concepts, the fields of each, the declared type of every field and its measurement unit. The registry is read by the Data Harmonisation component, which presents it to the user as the set of available mapping targets and derives from the declared units the normalisation to apply.

5.3.2 Secure Data and Metadata Storage

Harmonised measurements are persisted as time series. The data is high-frequency and append-heavy, and is consumed by time range and by identifier. Metadata is maintained alongside in two layers. User-defined attributes are set by the platform user: the title and description of the asset, and its access level. System-managed attributes are maintained by the platform: the number of records, the timestamp and outcome of the most recent execution, the anonymisation flag and the quality indicators. The extent of the data, meaning its temporal and spatial coverage and resolution, can be set by the user or derived by the platform.

5.3.3 Anonymisation

A substantial part of the landscaped data is recorded at dwelling level, and dwelling level electricity, heating, cooling and water series are personally attributable by construction: a consumption profile for a single dwelling describes the behaviour of its occupants, and is therefore personal data under the GDPR [9]. Anonymisation is therefore applied inside the ingestion pipeline, before the dataset reaches the catalogue, so that identifiable data never becomes discoverable or retrievable in the first place.

Strategy	Behaviour
Remove	The field is excluded from the stored dataset. It applies to fields classified as identifiers, which point to a person directly.
Generalise	Numeric values are grouped into intervals or sets of values, only as far as needed for the level of protection set for the dataset. It applies to quasi-identifiers.
Date truncation	Datetime values are reduced to the hour, the day, the month or the year. It applies to quasi-identifiers.
Keep	The field is stored unchanged. It applies to insensitive fields, and to sensitive fields, whose protection comes from the generalisation of the quasi-identifiers around them.

Table 12: Anonymisation strategies and the field classifications they apply to

Each field is classified by the role it plays in identifying a person, as insensitive, identifier, quasi-identifier or sensitive, and the strategy follows from that classification and from the type of the field. The level of protection and the acceptable loss of information are set once for the dataset, so values are generalised only as far as needed. Anonymisation rules are computed over the dataset as a whole rather than record by record and are therefore offered only for datasets that are complete when they enter the platform. Where a live source carries personally attributable fields, those fields are excluded at the harmonisation step instead.

5.3.4 Access Control

Access is controlled at dataset level, and the level is chosen by the user who created the dataset rather than centrally. Four levels are available. Only me keeps the dataset visible to its creator alone. Restricted access opens it to selected users within the same organisation. Organisation level access opens it to every user of that organisation. Public access opens it to all platform users. The level is set when the dataset is created, can be changed at any time afterwards, and is enforced on every request.

5.4. Search and Retrieval Service

The Search and Retrieval Service answers requirements REQ-DCP-10, 15 to 17 and 21 to 23. It is the interface through which data is obtained from the platform, and it works in two stages: the user first identifies the datasets and results they need, then receives them through an endpoint configured over that selection.

5.4.1 Search and Query Creation

The catalogue presents every dataset and every result the user is entitled to see. It can be searched by name and filtered by status and by asset type. Each entry shows its main metadata, the number of records it holds, how it can be accessed and its access level. The selection over the chosen assets, meaning the fields a query returns and the indexed fields that can filter it, is defined when the retrieval query is created.

5.4.2 Open APIs

Retrieval endpoints serve the harmonised pilot datasets on one hand and results on the other, results being processed data packages that the analytical and forecasting components of Task 4.3 work with. In both cases the user selects one or more assets from the catalogue and the fields to return, and the indexed fields can then be used as query parameters to filter the data. The platform generates the endpoints of the query, together with instructions for authentication, pagination, sorting and the measurement units of the returned fields. All endpoints return JSON and are authenticated by an access token issued by the user. That user passes the endpoint URL, its parameters and the token to the application that needs the data, and the token is what authorises the calls.

5.5. Notification Service

The Notification Service answers requirements REQ-DCP-19 and 20. It surfaces three classes of event in the platform interface: successful pipeline execution, pipeline failure such as a connection timeout or a malformed file, and detection of data quality issues exceeding configured thresholds. Users configure which event types they are notified of and the threshold values that trigger quality alerts, so that an operator running twenty pipelines is not notified twenty times a day for normal operation.

5.6. Organisation Management Service

The Organisation Management Service answers requirement REQ-DCP-24. The platform is organised around organisations rather than around individual accounts. Membership is then administered by the organisation itself. Each organisation has an owner with full control over it, managers, who can invite new members and promote or demote them between member and manager, and members, who work with the data of the organisation.

5.7. Technology stack and deployment

The platform is implemented on a stack chosen for the characteristics of the data and for operational simplicity in a research project.

Layer	Technology and rationale
Backend services	TypeScript with NestJS on Node.js. Modular services with REST and GraphQL APIs for data, workflows and platform configuration
Data storage	PostgreSQL, MongoDB, MinIO with S3-compatible interfaces.
Asynchronous execution	Python, PyArrow
User interface	Vue.js with TypeScript, Vuex and Tailwind CSS.
Packaging and deployment	Docker and Kubernetes.

Table 13: Platform technology stack

The platform is deployed as a containerised stack and is accessible to consortium partners over authenticated HTTPS. Configuration is externalised, credentials for source systems are held in the platform secret store rather than in pipeline definitions.

6. The RESPED Data Management Platform in Operation

This chapter presents the platform as it is used. It follows the order a user follows: entering the platform and seeing the state of the data, configuring a source, mapping it to the RESPED-CIM, applying quality rules, applying anonymisation, finding the resulting dataset in the catalogue, configuring the endpoint through which it is served, and managing the organisation and its members. Each section states what the screen does, what the user controls on it, and which requirements it satisfies.

The platform is deployed at https://resped.s5projects.eu. Access requires an account, and users sign in from the login page, after which the dashboard is displayed.

6.1. Platform dashboard

The dashboard is the landing view of the platform and shows at a glance whether data is arriving and whether its processing succeeds. Three counters give the total number of datasets, ingestion jobs and retrieval queries. Below them, the most recent ingestion executions of the user's organisation are listed, each with its status, the name of the job, the outcome message and the time it ran. A failed execution shows the error returned, so the user can open the job, correct the cause and run it again. The view can be refreshed on demand, and the time of the last update is displayed. From the side menu the user moves to data ingestion, the data catalogue or the API configuration.

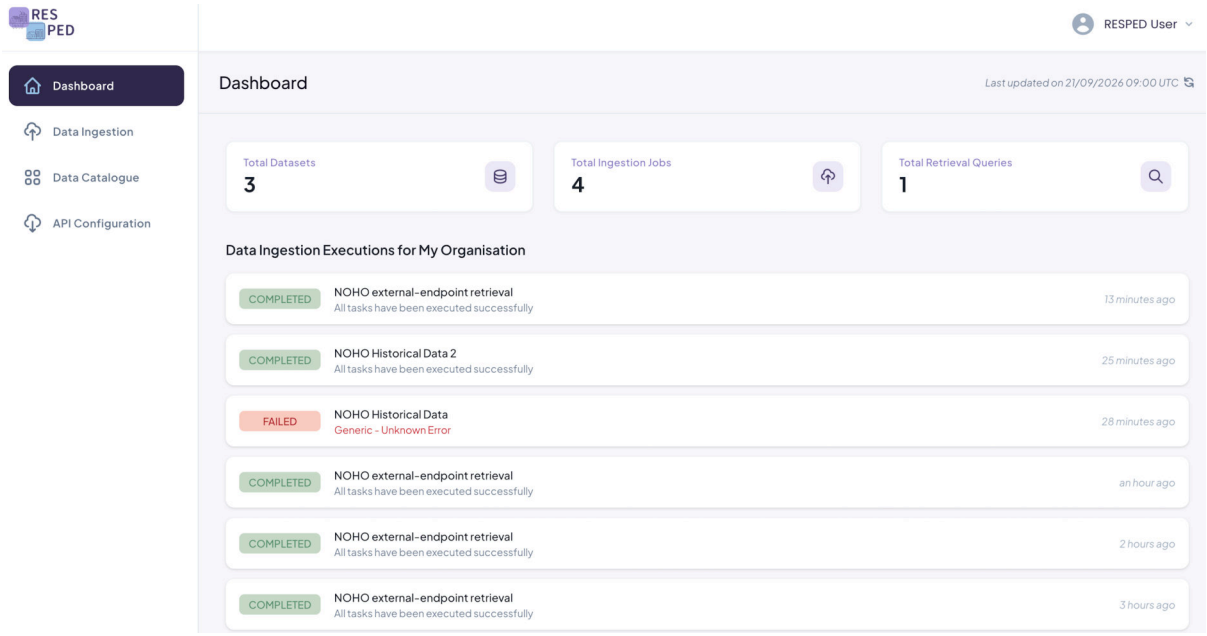


Figure 3: Platform dashboard with summary counters and recent ingestion executions

6.2. Configuring data ingestion

The Data Ingestion area lists the ingestion jobs of the user. Jobs can be searched by name, filtered by status and by ingestion source, and sorted by the time of their last update, status or name. Each job is shown with its name, its source type and format, and a short description. It also shows the processing stages configured for that job, such as ingestion, harmonisation, data quality, anonymisation and storage, each marked with the outcome of the latest execution. The creator of the job, the date of the last execution and its overall status complete the view. Where an execution fails, the stage at which it stopped is marked, so the user sees immediately where the correction is needed.

New jobs are created from the same view, which also gives access to the incidents reported for the user's jobs.

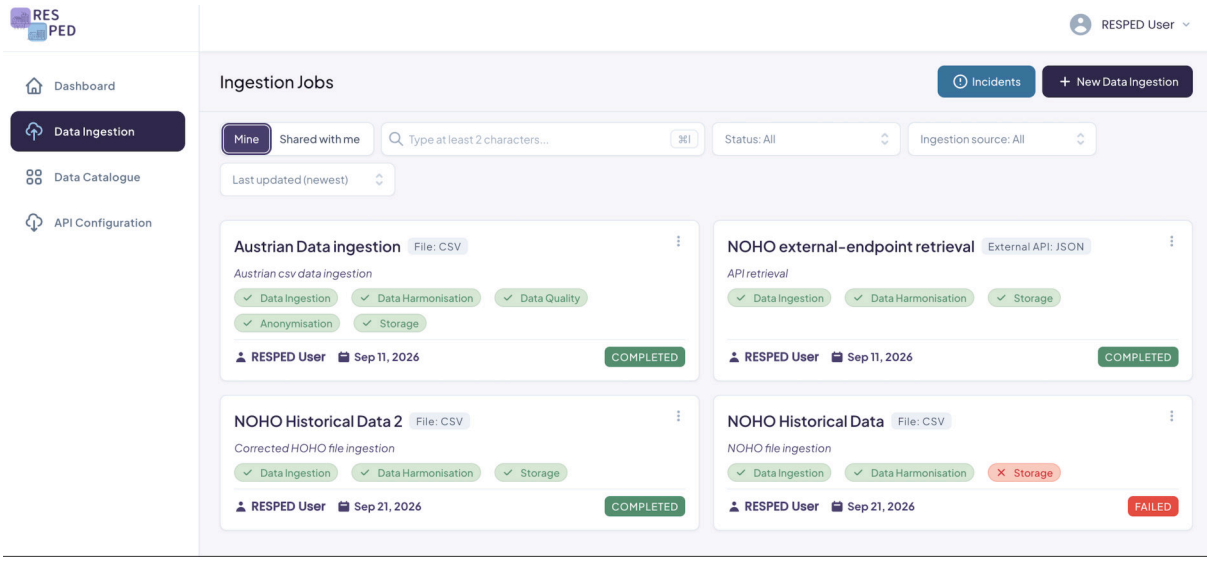


Figure 4: Ingestion jobs of the user

A new ingestion job is created from a single form. The user gives the pipeline a name and a short description, and then selects the ingestion source, which is a file upload in CSV or JSON, an external Kafka service or an external API. Once a source is selected, the pre-processing steps of the job are chosen. Data harmonisation, which maps the data to the RESPED-CIM, and storage are always part of the job. Data quality, which detects and addresses corrupt or inaccurate records, and anonymisation, which prevents individuals from being identified, are added only where the data requires them.

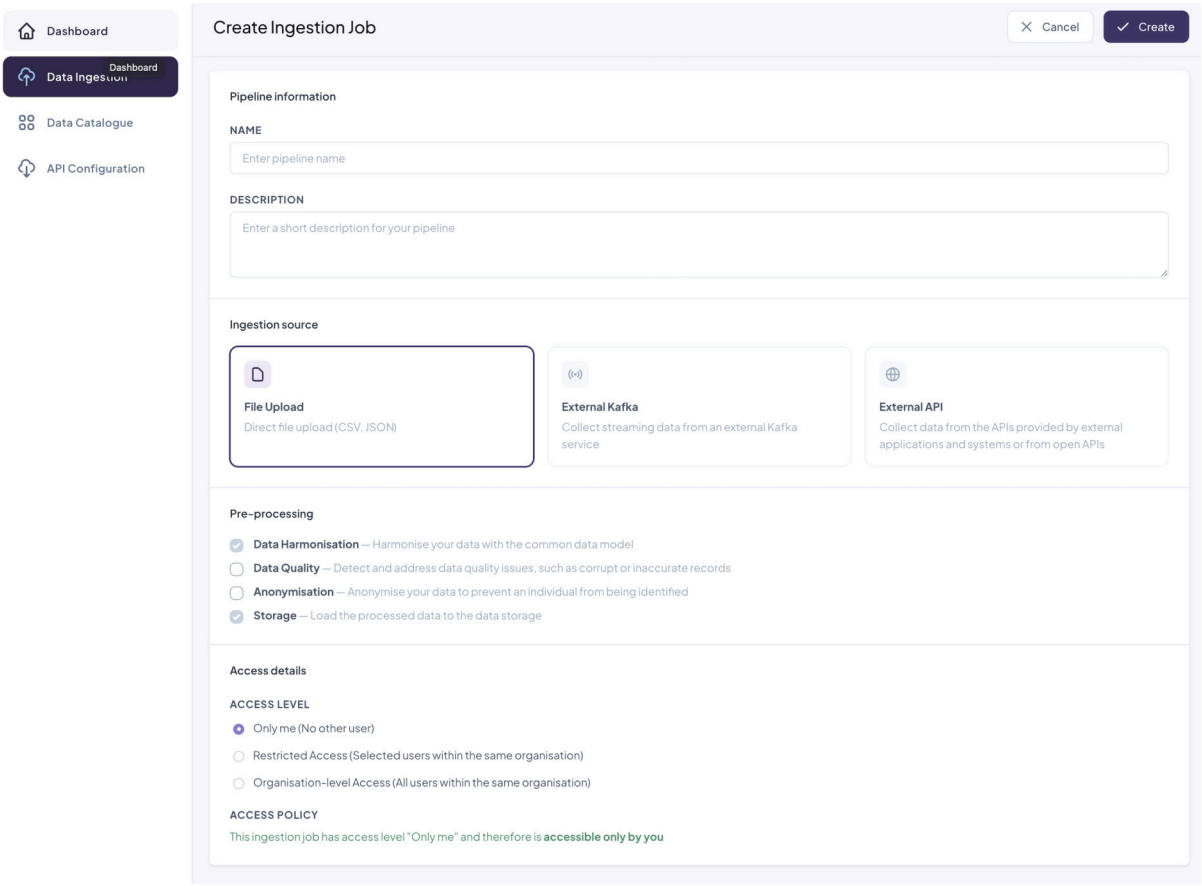


Figure 5: Creation of a new ingestion job

File ingestion is used for the historical exports delivered by the demonstration sites and for the modelled datasets produced by project partners. Once the job is created, its configuration continues in two steps. In the first, the user selects the file format, CSV or JSON, and provides two files. The sample file is a small extract of the data, which the platform uses to present the data and to prepare the configuration that follows. The full dataset file is the complete data to be ingested. Keeping the two apart allows the configuration to be prepared and checked on a light file before the complete dataset is processed.

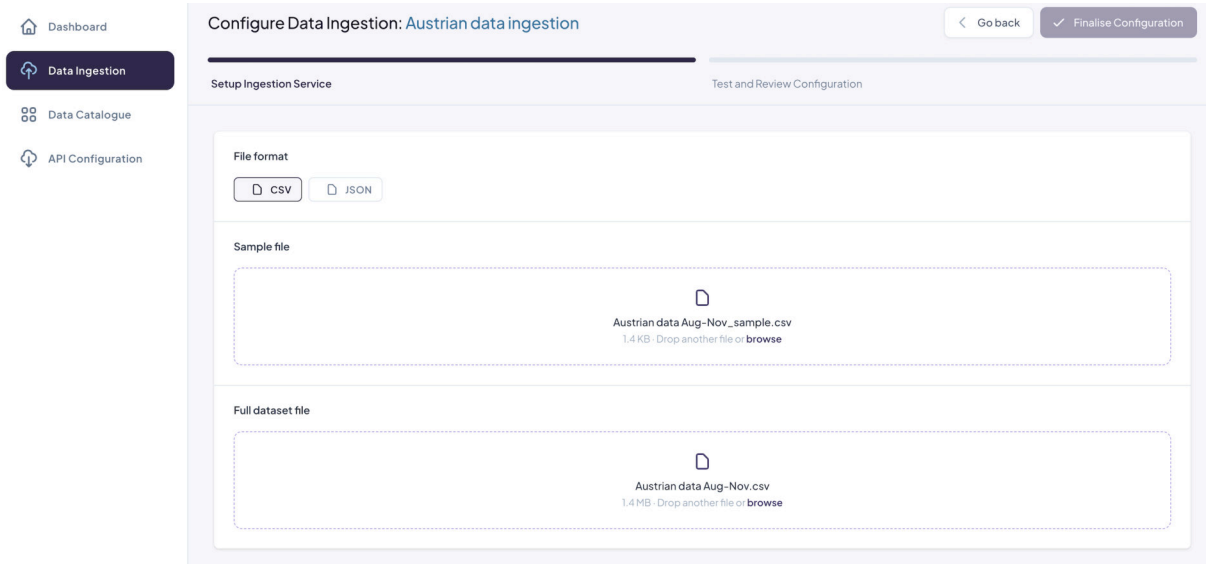


Figure 6: Setup of a file ingestion job

In the second step the platform lists the files to be uploaded with their size and presents a sample of the data as a table, with the source columns as they appear in the file. The user confirms that the file has been read correctly before completing the configuration.

External API ingestion is used for the scheduled collection of measurements from site systems that expose a programmatic interface, such as the building management system at the Czech site, and from third-party sources such as weather services. The user supplies the endpoint URL and HTTP method, selects the authentication scheme and enters the corresponding credentials, adds any custom headers, and defines the query parameters used to request a time range. The schedule is set on the same screen.

Kafka ingestion covers the case where a site publishes measurements as they are produced rather than exposing them for collection. The user supplies the broker address, the topic and consumer group, and the security protocol and credentials. Once activated, the pipeline consumes messages as they arrive and applies the same harmonisation, quality and storage steps as any other pipeline.

6.3. Data Harmonisation

Harmonisation is the step where the ingested data is linked to the RESPED-CIM. The screen is organised in three parts. On the left are the concepts and fields of the RESPED-CIM, which can be searched and are shown with their data type. Selecting a field displays its description and, where relevant, the unit in which it is measured. In the middle are the columns of the ingested data, as they appear in the source, each with its detected type. The user selects a column and links it to a field of the model, which is then shown next to the column. On the right are the details of the selected column, which adapt to its type. For a number, the user states how it is written in the source, such as the thousands separator and the decimal point, and the unit in which it is measured. The platform then shows the unit the values will be converted into to match the model. For a date, the user states its format, which the platform can also infer from the data, the order of day, month and year, and the time zone, unless the data already carries one. Below the columns, sample values of the selected column are displayed. They help the user see what a column actually contains before linking it. The linking has to be done for every column that should be stored in the dataset. Columns that are not linked are ignored and are not stored.

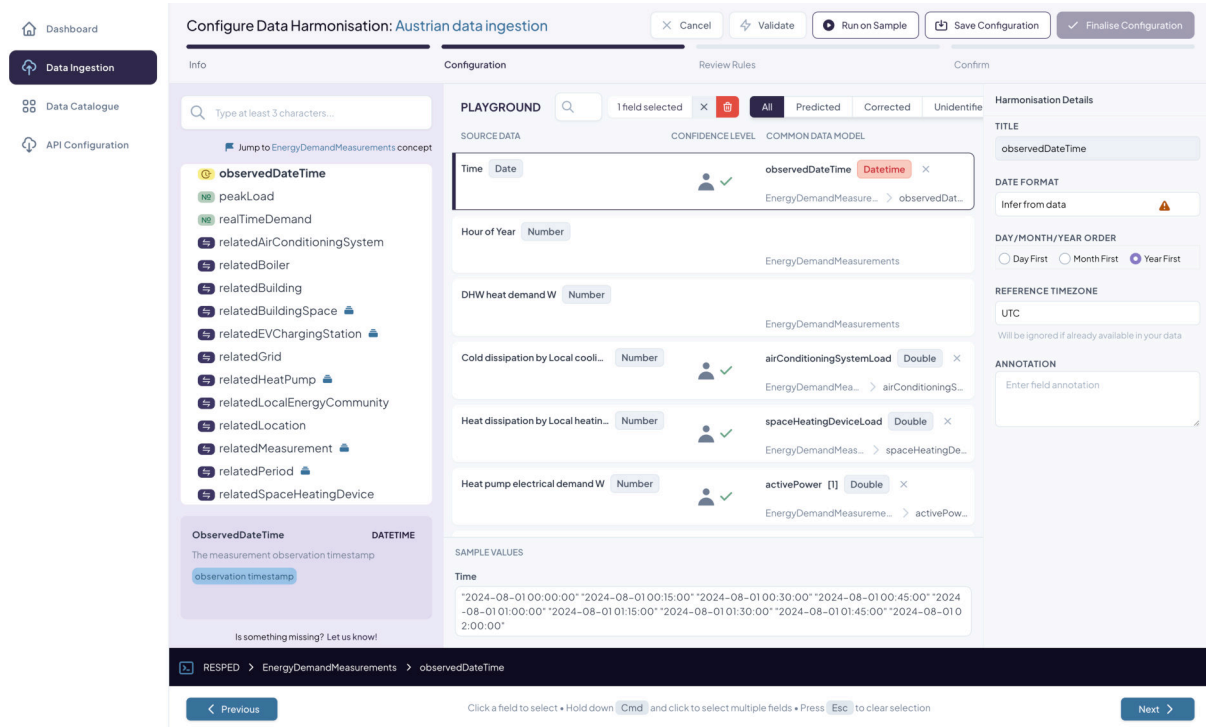


Figure 7: Linking of source columns to the RESPED-CIM

A confirmation step then lists the source columns linked to the model, each next to the concept and field it maps to, and separately the columns left unidentified, which will not be stored. The harmonisation is completed from there and is applied on every execution of the job.

6.4. Data quality and cleaning

Where data quality is selected as a pre-processing step, the quality rules are configured on the fields that were linked to the RESPED-CIM during harmonisation. The fields are listed with their type and can be filtered by it. The sample values of the selected field are shown below, so each rule can be set with the actual data in view. For each field, the user defines how missing values are handled, the range of acceptable values together with the action to take when a value falls outside it, and how duplicate records are treated. Each field shows how many rules have been defined for it. The rules can be tried on the sample data and saved before they are reviewed and confirmed. They are applied on every execution of the job. The completeness of the data and related quality indicators are computed at each execution and attached to the dataset as metadata.

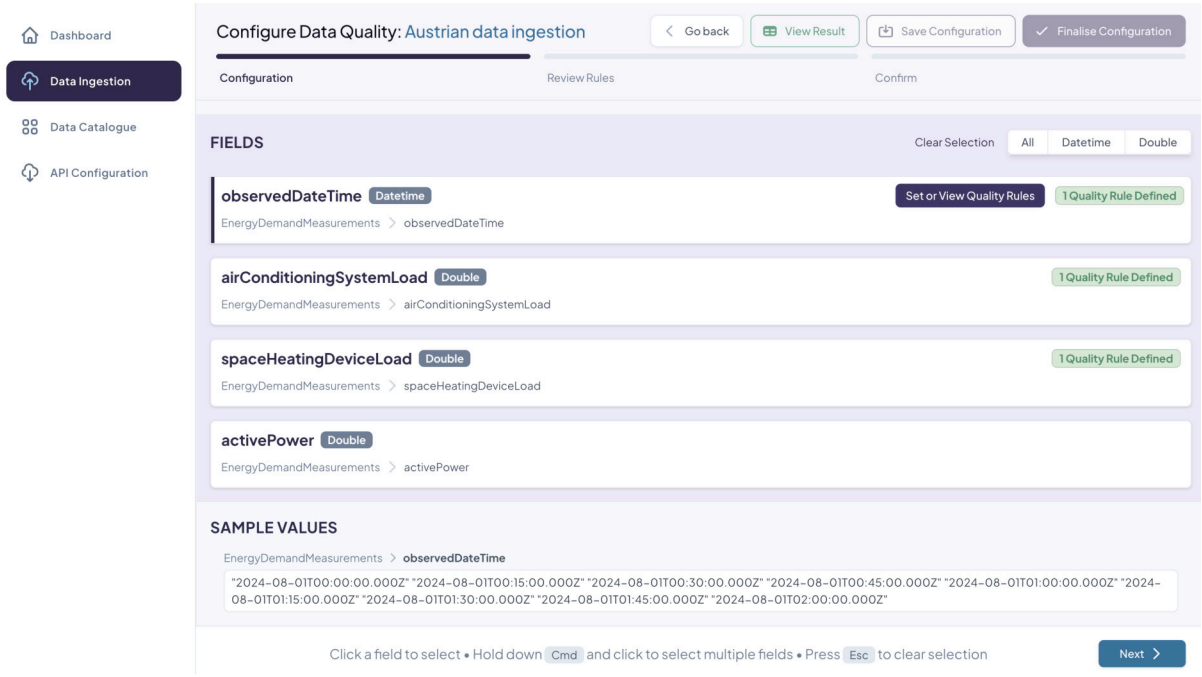


Figure 8: Configuration of data quality rules per field

A confirmation step then states each rule in plain language, with the condition that triggers it and the action taken, and lists the fields left without rules.

6.5. Anonymisation

Where anonymisation is selected as a pre-processing step, it is configured field by field on the fields linked to the RESPED-CIM during harmonisation. For each field the user states the role it plays in identifying a person. An insensitive field is kept as it is. An identifier, which points to a person directly, is removed. A quasi-identifier, which could identify a person when combined with other fields, is generalised. Numbers are grouped into intervals or sets of values, and dates are truncated to the hour, the day, the month or the year. A field can also be marked as sensitive, meaning its content is private, such as a consumption profile, even though it does not identify anyone on its own. Its values remain unchanged. The protection comes from generalising the quasi-identifiers, so that a sensitive value can no longer be linked to an individual. Sample values of the selected field are shown below, so each choice is made with the actual data in view. The level of protection to be achieved and the acceptable loss of information are set once for the dataset, so the platform generalises values only as far as needed.

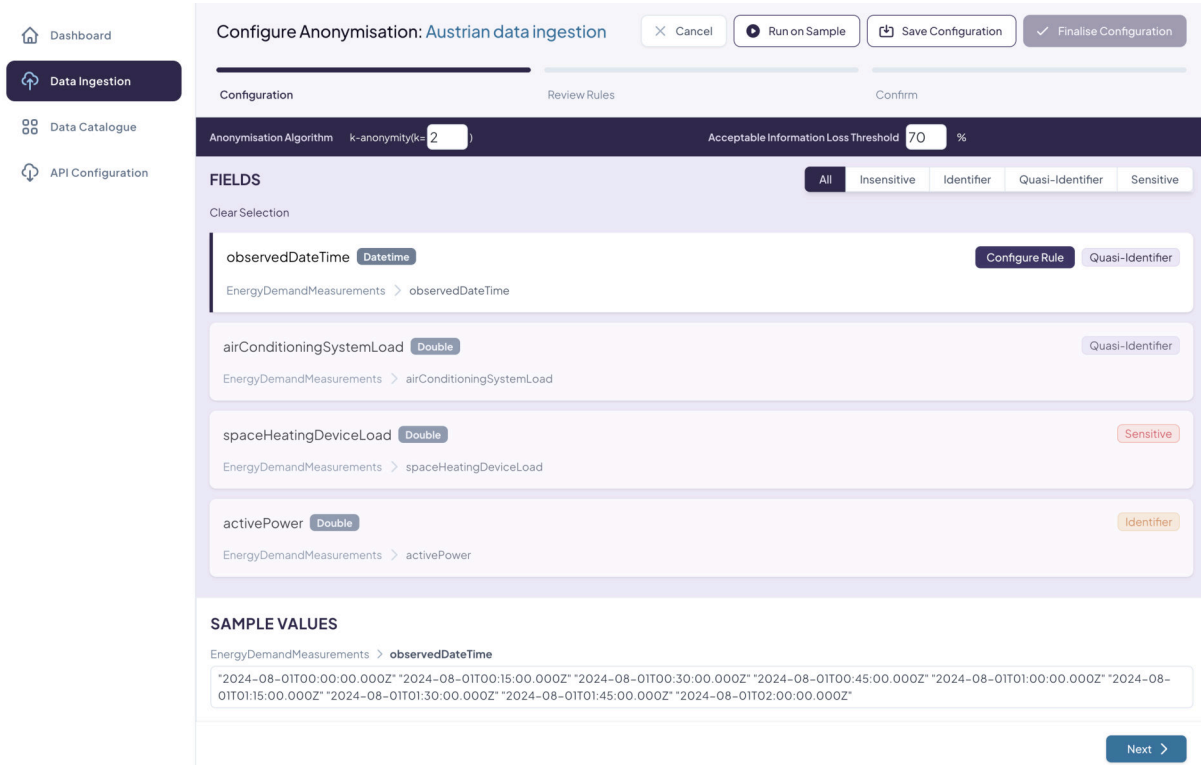


Figure 9: Classification of fields for anonymisation

A confirmation step then groups the fields by their classification and lists, for each generalised field, the successive levels of generalisation the platform may apply.

6.6. Data catalogue and access control

Once an ingestion job has run, the resulting dataset appears in the Data Catalogue. The catalogue lists the assets available to the user. They can be searched by name, filtered by status and by asset type, and sorted by the time of their last update, name and status. Each asset is shown with its name, its type and a short description, together with its main metadata: the number of records it holds, how it can be accessed and its access level. The creator, the date of the last update and the status of the asset complete the view. The catalogue shows only the assets that the access level of each one allows the user to see.

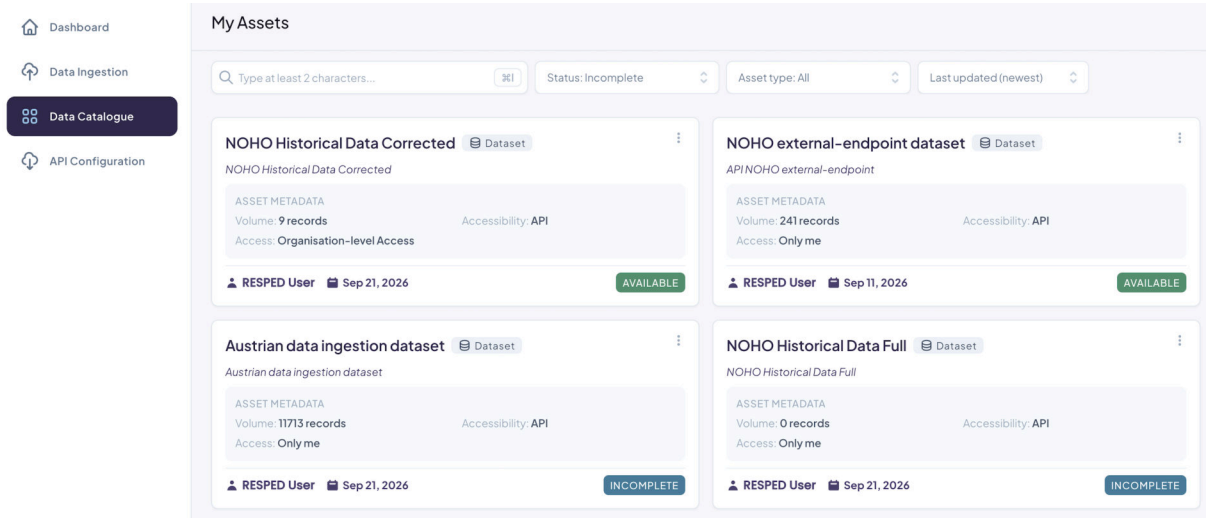


Figure 10: Data catalogue with the assets available to the user

Each asset in the catalogue can be edited. Its general information, a title and a description, is set there. The same view holds the access level of the asset. The asset can be kept to its creator, opened to selected users of the same organisation, opened to the whole organisation, or made public to all users of the platform. Metadata such as the number of records and the time of the last update is maintained by the platform on every execution, so the user only provides what cannot be derived from the data itself. An asset is marked as incomplete until its extent has been described, and as available once it has.

6.7. API configuration and data retrieval

Data leaves the platform through retrieval queries that users configure themselves. The API Configuration area lists the retrieval queries of the user. Each is shown with its name, the assets it draws on, a short description, the dates of its creation and last update, and whether the underlying assets are available. Queries can be searched, filtered by status and sorted, and new ones are created from the same view.

A retrieval query is configured in three steps. In the first, the user selects one or more assets they are entitled to access, datasets or results, and then the fields of each asset the query should return. Indexed fields are marked, since only these can be used as query parameters to filter the data. Fields that were generalised during anonymisation appear as ranges rather than single values.

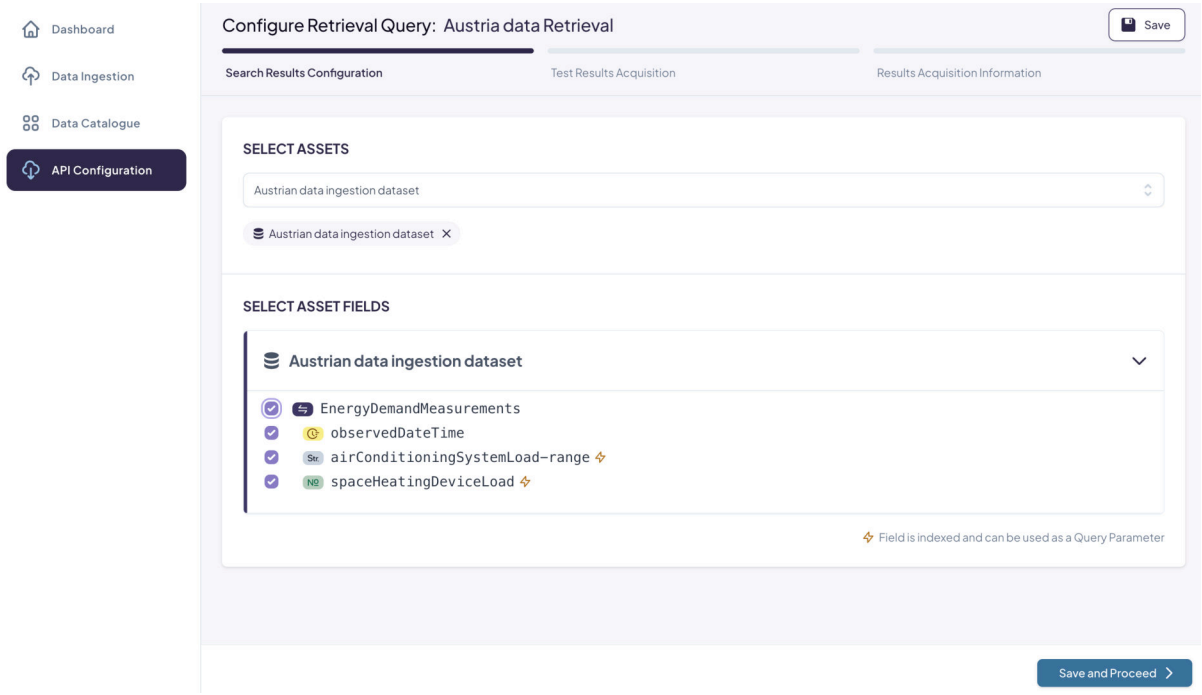


Figure 11: Selection of assets and fields for a retrieval query

In the second step, the user tests the query by running it with sample parameters and inspecting the results the platform returns. In the third, the platform provides everything needed to use the query from an external application. This covers the endpoints to call, how requests are authenticated with an access token issued by the user, how larger result sets are retrieved page by page and how results can be sorted. The user then passes the endpoint, its parameters and the token to the application that needs the data. All responses are returned in JSON.

6.8. Organisation and user management

Access to the platform is granted through an organisation. An organisation is registered with its general information, its type, which is what it is in the project, and its location. The organisation then manages its own membership. Its members are listed with their name, contact address, the roles they hold within the organisation and their current status. The owner holds full control. Managers invite new members and promote or demote them between member and manager. Members work with the data of the organisation and decide, dataset by dataset, who else sees what they create. Administration is kept inside the organisation, since each partner knows who in its own team needs access.

7. Conclusions and next steps

Task 4.2 delivers the data layer of RESPED. The work established what data exists at the demonstration sites, defined what it means, set out what the platform has to do with it, and built the platform itself.

The work started with a data landscaping exercise covering the data assets of the demonstration sites, recording for each its format, velocity, resolution, coverage, ownership, accessibility, update strategy and privacy classification, and maintained as a living record that also links the file-based assets to the columns of the exported files. It continued with the RESPED Common Information Model, 30 concepts and 781 fields, defined for the project against SAREF, SAREF4ENER, SAREF4BLDG, IEC CIM and IFC, loaded into the platform and used as the target of every field mapping. It was followed by a use case and requirements catalogue of eleven use cases and twenty-four functional requirements, reviewed by the consortium.

The result is the RESPED Data Management Platform: five services covering collection from files, third-party APIs and Kafka brokers with semantic harmonisation and per-field data quality enhancement, secure storage of the data and its metadata with anonymisation, dataset-level access control and provenance, search over the catalogue and standards-based retrieval of harmonised datasets and results.

The platform is operational and ready to be populated with data from the demonstration sites while the work continues in three directions. The first is the connection of the Task 4.3 decision support dashboard to the live retrieval endpoints, which is the integration step that turns two components into one system and is reported in D4.3 at M24. The second is the progressive extension of ingestion coverage as further data assets become available at the demonstration sites, which the platform absorbs as additional pipelines. The third is the operational monitoring of the platform through the remainder of the project, including the quality indicators it produces, which feed the evaluation of the Living Lab PEDs in Task 2.4.

8. References

[1] RESPED, Enabling Energy Resilience through new energy flexible and affordable PED concepts - Full Proposal, DUT Partnership, 2024.

[2] RESPED, D1.2 - Integrated methodological framework and KPIs, 2026.

[3] ETSI TS 103 264, SmartM2M; Smart Applications; Reference Ontology and oneM2M Mapping (SAREF).

[4] ETSI TS 103 410-1, SmartM2M; Extension to SAREF; Part 1: Energy Domain (SAREF4ENER).

[5] ETSI TS 103 410-3, SmartM2M; Extension to SAREF; Part 3: Building Domain (SAREF4BLDG).

[6] IEC 61970 series, Energy management system application program interface (EMS-API) - Common Information Model (CIM).

[7] IEC 61968 series, Application integration at electric utilities - System interfaces for distribution management.

[8] ISO 16739-1, Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries.

[9] Regulation (EU) 2016/679, General Data Protection Regulation.

Annex I – T4.2 use case catalogue

This annex contains the complete use case catalogue for Task 4.2, version 1.0, as reviewed by the consortium. Eleven use cases are documented, each with its description, actors, trigger, pre-conditions, post-conditions, priority and the requirements derived from it.

RESPED-UC-DCP-01 | Ingest PED-related data from heterogeneous sources

UC ID	RESPED-UC-DCP-01
Title	Ingest PED-related data from heterogeneous sources
Description	The platform ingests PED-related data originating from heterogeneous sources and providers through a dedicated ingestion configuration workflow. Three source types are supported: File (batch upload of CSV or JSON files, e.g., NOHO energy import/export, PV generation, heat pump consumption, EV charging, and per-flat metering data), External API (scheduled pull from third-party REST endpoints, with configurable authentication, headers, and query parameters), and Kafka (streaming ingestion via Kafka brokers, with configurable topic, consumer group, and security protocol). For each source, the platform user configures the name and description of the ingestion job, the pre-processing steps it includes, the access level of the job and, where applicable, the schedule. Harmonisation and storage are always part of the job, while data quality rules and anonymisation are optional steps configured before the pipeline is activated. Once activated, the pipeline executes immediately, or according to its schedule or in response to incoming data events.
Actors	Data Provider (e.g., NOHO, UNICAL); Platform User
Trigger	Platform User initiates a new ingestion pipeline via the ingestion configuration workflow; a scheduled pipeline trigger fires; an incoming data push is received from a connected Kafka source
Pre-conditions	Platform is operational; user account is active; source credentials and connection parameters are available
Post-conditions	Ingestion pipeline is configured and activated; first execution is completed or scheduled; an ingestion log entry is created with timestamp and status
Priority	1 - Mandatory
Related Reqs	RESPED-REQ-DCP-01, RESPED-REQ-DCP-02, RESPED-REQ-DCP-03

RESPED-UC-DCP-02 | Harmonise ingested data using RESPED semantics

UC ID	RESPED-UC-DCP-02
Title	Harmonise ingested data using RESPED semantics
Description	As part of the ingestion configuration workflow, the platform user maps each source data field to the corresponding RESPED Common Information Model (CIM) field via a dedicated harmonisation step. The platform lists all detected source columns with their inferred data types and sample values. For each column, the user selects the target RESPED CIM field or marks it for exclusion. The platform displays unit normalisation labels per field (e.g., kWh ↔ MWh), indicating the transformation that will be applied at ingestion time. Multi-energy vector data is handled through distinct semantic classes in the RESPED CIM, aligned with SAREF and SAREF4ENER ontologies. The confirmed mapping is stored for the pipeline and applied automatically on every subsequent execution.
Actors	Platform User; Platform (CIM mapping engine)

Trigger	Platform User advances to the harmonisation step within the ingestion configuration workflow following source configuration (UC-DCP-01)
Pre-conditions	Source configuration is complete; RESPED CIM field definitions are available in the platform; source schema has been detected or manually defined
Post-conditions	Field mapping configuration is saved; unit normalisation and further mapping rules are registered for the pipeline; harmonised data will be produced on every pipeline execution according to the confirmed mapping
Priority	1 - Mandatory
Related Reqs	RESPED-REQ-DCP-04, RESPED-REQ-DCP-05, RESPED-REQ-DCP-06

RESPED-UC-DCP-03 | Validate and improve data quality

UC ID	RESPED-UC-DCP-03
Title	Validate and improve data quality
Description	As an optional step within the ingestion configuration workflow, the platform user configures per-field data quality and cleaning rules that will be applied automatically on every pipeline execution. The step is selected as a pre-processing step when the job is created and is presented after the harmonisation step. For each mapped field, the user may configure the handling of missing values, range validation with a configurable action on violation, and duplicate handling. Quality indicators, including a completeness score, are computed at each execution and attached to the dataset as metadata. Records that match configured rules are handled according to the defined strategy.
Actors	Platform User; Platform (quality engine)
Trigger	Platform User selects data quality as a pre-processing step when creating the job and advances to the quality configuration step within the ingestion workflow
Pre-conditions	Harmonisation step (UC-DCP-02) is complete, and data quality is selected as a pre-processing step
Post-conditions	Per-field quality rules are saved for the pipeline; quality indicators are computed and stored as metadata on each execution; non-conforming records are handled according to the configured strategy
Priority	1 - Mandatory
Related Reqs	RESPED-REQ-DCP-07, RESPED-REQ-DCP-08

RESPED-UC-DCP-04 | Manage dataset metadata for discovery and retrieval

UC ID	RESPED-UC-DCP-04
Title	Manage dataset metadata for discovery and retrieval
Description	Metadata for every dataset is composed of two layers: user-defined attributes set by the platform user, and system-managed attributes automatically maintained by the platform, collectively supporting data discovery, search and retrieval through a searchable data catalogue for authorised users. User-defined attributes include the title and description of the dataset, its extent, meaning its temporal and spatial coverage and resolution where these are not derived from the data, and its access level. System-managed attributes include the number of records, the time of the last update, the status of the dataset, the

	anonymisation flag and the quality indicators. Metadata is set after the first pipeline execution and updated automatically on subsequent executions.
Actors	Platform User; Platform (metadata engine)
Trigger	A pipeline executes successfully for the first time, registering a new dataset; metadata partly configured by the platform user and partly from the system; a subsequent execution updates an existing dataset's metadata; a platform user queries the data catalogue
Pre-conditions	At least one ingestion pipeline has been configured and executed; metadata schema is defined
Post-conditions	Metadata is stored and indexed; datasets are discoverable and filterable via the catalogue; search results are returned to the user
Priority	1 - Mandatory
Related Reqs	RESPED-REQ-DCP-09, RESPED-REQ-DCP-10

RESPED-UC-DCP-05 | Control access to PED data

UC ID	RESPED-UC-DCP-05
Title	Control access to PED data
Description	The platform enforces dataset-level access control based on an access level set by the user who created the dataset. Four access levels are supported. Only me keeps the dataset visible to its creator alone. Restricted opens it to selected users within the creator's own organisation. Organisation opens it to every user of that organisation. Public opens it to all authenticated platform users. The access level is managed by the creator and can be updated at any time after registration. The platform enforces the active level on every data access request, regardless of the requesting user's role.
Actors	Platform User (Data Owner); Platform User (Data Consumer); Platform (access control engine)
Trigger	Data owner sets or updates an access policy for a dataset; a platform user or service requests access to a dataset
Pre-conditions	Dataset is registered in the platform; data owner account is active; requesting user is authenticated
Post-conditions	Access is granted or denied based on the active access policy
Priority	1 - Mandatory
Related Reqs	RESPED-REQ-DCP-11, RESPED-REQ-DCP-12

RESPED-UC-DCP-06 | Anonymise sensitive PED data

UC ID	RESPED-UC-DCP-06
Title	Anonymise sensitive PED data
Description	Where a dataset contains personally attributable data, such as dwelling level energy, heating, cooling and water consumption, anonymisation is applied as an optional step within the ingestion configuration workflow, before data reaches the catalogue. Anonymisation is restricted to datasets that are complete at ingestion, since the rules are computed over the dataset as a whole. It is configured per field at pipeline setup time, by classifying each field according to its role in identifying a person. Insensitive fields are kept as they are, identifiers are removed, and quasi-identifiers are generalised, with numbers grouped into

	intervals or sets of values and dates truncated to the hour, day, month or year. Sensitive fields remain unchanged and are protected by the generalisation of the quasi-identifiers. The level of protection and the acceptable loss of information are set once for the dataset.
Actors	Platform User; Platform (anonymisation engine)
Trigger	Platform User selects anonymisation as a pre-processing step when creating a job over a complete dataset and advances to the anonymisation step within the ingestion workflow
Pre-conditions	The dataset is complete at ingestion, the harmonisation step (UC-DCP-02) is complete, and anonymisation is selected as a pre-processing step
Post-conditions	Per-field anonymisation rules are applied to the data before it is stored, and the anonymised dataset is made available according to its access level
Priority	1 - Mandatory
Related Reqs	RESPED-REQ-DCP-13, RESPED-REQ-DCP-14

RESPED-UC-DCP-07 | Expose data to authorised services via standardised APIs

UC ID	RESPED-UC-DCP-07
Title	Expose data to authorised services via standardised APIs
Description	The platform provides standards-based API endpoints through which authorised internal services and external consumers can retrieve harmonised datasets and results, results being pre-processed data packages for the analytical and forecasting modules of T4.3. Retrieval queries are created by the platform user via a dedicated configuration wizard. The user selects one or more assets from the catalogue and the fields to return, tests the query, and receives auto-generated endpoint URLs together with instructions for pagination, sorting and measurement units. Indexed fields can be used as query parameters. All endpoints are authenticated with an access token issued by the user and return responses in JSON format. Retrieval queries are managed via the platform's API configuration module.
Actors	Platform User; T4.3 Dashboard Service; Platform (API gateway)
Trigger	A platform user creates or updates a retrieval query over assets they have access to, or an authorised consumer or service issues an API request against a configured endpoint
Pre-conditions	User is authenticated and authorised to access the selected assets, the assets are available and harmonised, and the configuration wizard is accessible
Post-conditions	Endpoint URLs are generated and accessible to the user, and requested data is returned in JSON format upon a valid API call
Priority	1 - Mandatory
Related Reqs	RESPED-REQ-DCP-15, RESPED-REQ-DCP-16, RESPED-REQ-DCP-17

RESPED-UC-DCP-08 | Track data provenance and processing history

UC ID	RESPED-UC-DCP-08
Title	Track data provenance and processing history

Description	The platform records provenance and processing history for every pipeline execution. Provenance captures the origin of each dataset: the data source, the data owner, and the timestamp of ingestion. Processing history captures the outcome of each configured step applied during execution: the number of records ingested, records dropped or modified during harmonisation, records affected by quality rules (e.g., filled, clipped, or dropped), and fields transformed by anonymisation rules (e.g., generalised or removed). No prior versions of the dataset are retained. When a dataset is updated, e.g., when NOHO appends new CSV files or UNICAL provides a revised batch of modelled PED energy consumption, a new execution record is created, and dataset metadata is updated accordingly.
Actors	Platform (processing engine); Platform User
Trigger	A pipeline execution completes; a platform user inspects the processing history of a dataset
Pre-conditions	At least one ingestion pipeline is active and has executed successfully
Post-conditions	A processing summary is recorded for the execution; dataset metadata is updated to reflect the latest execution
Priority	2 - Recommended
Related Reqs	RESPED-REQ-DCP-18

RESPED-UC-DCP-09 | Notify users of data ingestion and quality events

UC ID	RESPED-UC-DCP-09
Title	Notify users of data ingestion and quality events
Description	The platform generates and delivers notifications to relevant platform users upon occurrence of significant data lifecycle events. Notification triggers include: successful pipeline execution, pipeline failure (e.g., connection timeout, malformed file), and detection of data quality issues exceeding configured thresholds. Notifications are surfaced via the platform's user interface. Users can configure which event types they are notified of and the quality threshold values that trigger alerts.
Actors	Platform (notification engine); Platform User
Trigger	A monitored pipeline or quality event occurs
Pre-conditions	At least one ingestion pipeline is active; notification preferences are configured
Post-conditions	Notification is delivered to the relevant users via the platform UI
Priority	2 - Recommended
Related Reqs	RESPED-REQ-DCP-19, RESPED-REQ-DCP-20

RESPED-UC-DCP-10 | Expose results for retrieval by analytical services

UC ID	RESPED-UC-DCP-10
Title	Expose results for retrieval by analytical services
Description	The platform exposes results, structured and pre-processed data packages designed to feed the AI models and analytical algorithms of T4.3. Results are assembled from quality-validated, harmonised and aggregated time series covering energy demand (electricity, heating, cooling), generation (PV), flexibility

	proxies (EV charging, heat pump) and grid interaction (import/export). Gaps are filled using configurable imputation strategies, and time series are aggregated at configurable resolutions. Results are listed in the data catalogue alongside the datasets and are retrievable in JSON through user-generated endpoint URLs, as described in UC-DCP-07.
Actors	Platform (data preparation engine); T4.3 Dashboard Service; Platform User
Trigger	A platform user or service issues a request via a retrieval endpoint configured over a result
Pre-conditions	Quality-validated, harmonised and aggregated data is available, a retrieval endpoint over the result is configured and active, and the requesting service is authenticated
Post-conditions	The result is returned via the endpoint in JSON format
Priority	1 - Mandatory
Related Reqs	RESPED-REQ-DCP-21, RESPED-REQ-DCP-22, RESPED-REQ-DCP-23

RESPED-UC-DCP-11 | Manage platform users and organisations

UC ID	RESPED-UC-DCP-11
Title	Manage platform users and organisations
Description	The platform organises access around organisations. Organisations are registered with their identifying and location details and with an organisation type that states their role in the project, such as energy application owner, data provider or data consumer. Membership is administered by each organisation itself. Each organisation has an owner with full control over it, managers, who invite new members and promote or demote them between member and manager, and members. Every member carries a status, so participation can be withdrawn without removing the account.
Actors	Platform Administrator; Organisation Owner; Organisation Manager; Platform
Trigger	A new organisation is registered, a manager invites a new member, or the role or status of a member requires updating
Pre-conditions	Platform is operational, and the organisation and its owner or manager accounts are active
Post-conditions	The organisation is registered, or the member account is created, updated or its role changed
Priority	1 - Mandatory
Related Reqs	RESPED-REQ-DCP-24

Annex II – Requirements compliance matrix

This annex states, for each of the twenty-four functional requirements defined for Task 4.2, how the delivered platform meets it and where in this document the corresponding capability is described. All requirements are met.

Requirement	Requirement text	Status	How it is met
REQ-DCP-01	The platform shall support ingestion of PED-related data via three source types: File (CSV, JSON), External API (REST endpoints with configurable authentication and query parameters), and Kafka (brokers with configurable topic, consumer group and security protocol).	Met	Three source types implemented in the Data Ingester - Sections 5.2.1, 6.2
REQ-DCP-02	The platform shall provide a dedicated multi-step ingestion configuration workflow through which the platform user configures the name and description of the ingestion job, the source type, the pre-processing steps to include, the access level of the job and, where applicable, the schedule.	Met	Ingestion job creation and configuration workflow - Sections 5.2.1, 6.2
REQ-DCP-03	The platform shall provide an operational monitoring view of all active ingestion pipelines, displaying current status, last execution time, and success/failure outcome, accessible to all platform users.	Met	Pipeline status in the dashboard and in the ingestion jobs list - Sections 5.2.1, 6.1, 6.2
REQ-DCP-04	The platform shall map each source data field to the corresponding RESPED CIM field via a dedicated harmonisation step within the ingestion workflow, allowing the platform user to select the target CIM field or mark a field for exclusion.	Met	Harmonisation step with RESPED-CIM field selector, unlinked columns not stored - Sections 5.2.2, 6.3
REQ-DCP-05	The platform shall display inferred data types, sample values, and unit normalisation labels per field during the harmonisation step, indicating the transformation that will be applied at ingestion time.	Met	Detected types, sample values and unit conversion shown at mapping - Sections 5.2.2, 6.3
REQ-DCP-06	The platform shall handle multi-energy vector data through distinct semantic classes in the RESPED CIM, aligned with SAREF and SAREF4ENER ontologies.	Met	Distinct semantic classes per energy vector in RESPED-CIM - Chapter 3, Section 5.2.2
REQ-DCP-07	The platform shall support per-field data quality and cleaning rule configuration as an optional step within the ingestion workflow, covering null handling (e.g., drop row, fill with mean, fill with zero, fill with constant), range validation with configurable action on violation, and duplicate handling.	Met	Per-field rules for missing values, ranges and duplicates - Sections 5.2.3, 6.4
REQ-DCP-08	The platform shall compute quality indicators including completeness score at each pipeline execution and attach them to the dataset as metadata.	Met	Completeness and related indicators computed per execution and stored as metadata - Sections 5.2.3, 6.4
REQ-DCP-09	The platform shall maintain dataset-level metadata composed of user-defined attributes (e.g., title, description, extent and access level) and system-managed attributes (e.g., number of records, time of last	Met	User-defined and system-managed metadata, edited from the asset - Sections 5.3.2, 6.6

Requirement	Requirement text	Status	How it is met
	update, quality indicators), automatically created on first pipeline execution and updated on every subsequent execution.		
REQ-DCP-10	The platform shall provide a searchable data catalogue allowing platform users to discover datasets and results by name and by filtering on metadata attributes such as status and asset type.	Met	Catalogue searchable by name, filterable by status and asset type, filtered by entitlement - Sections 5.4.1, 6.6
REQ-DCP-11	The platform shall enforce dataset-level access control based on an access level set by the user who created the dataset, supporting four access levels: Only me (creator alone), Restricted (selected users within the creator's organisation), Organisation (all users of that organisation), and Public (all authenticated users).	Met	Four dataset access levels enforced on every access path - Sections 5.3.4, 6.6
REQ-DCP-12	The platform shall allow the creator of a dataset to set and update its access level at any time after registration.	Met	Access level editable by the creator from the asset - Sections 5.3.4, 6.6
REQ-DCP-13	The platform shall support per-field anonymisation configuration as an optional step within the ingestion workflow, restricted to datasets that are complete at ingestion, in which each field is classified as insensitive, identifier, quasi-identifier or sensitive, identifiers are removed and quasi-identifiers are generalised, with numbers grouped and dates truncated according to the field type.	Met	Per-field classification with removal and generalisation, for complete datasets - Sections 5.3.3, 6.5
REQ-DCP-14	The platform shall apply confirmed anonymisation rules to the data prior to storing the dataset in the catalogue.	Met	Anonymisation applied inside the ingestion pipeline before storage - Sections 5.3.3, 6.5
REQ-DCP-15	The platform shall support outbound retrieval endpoints serving both harmonised datasets and results, the latter being pre-processed data packages for the analytical services.	Met	Retrieval endpoints over datasets and results - Sections 5.4.2, 6.7
REQ-DCP-16	The platform shall allow platform users to create outbound API endpoint URLs via a configuration wizard, by selecting one or more assets and their fields, with auto-generated endpoint URLs whose calls are authenticated with an access token issued by the user.	Met	Retrieval query wizard with generated endpoints and access token authentication - Sections 5.4.2, 6.7
REQ-DCP-17	All outbound API endpoints shall return responses in JSON format.	Met	All outbound endpoints return JSON - Sections 5.4.2, 6.7
REQ-DCP-18	The platform should record a processing summary for every pipeline execution, capturing: data source, data owner, timestamp, number of records ingested, and the outcome of each configured processing step (harmonisation, quality, anonymisation).	Met	Processing history recorded per execution - Section 5.3.2
REQ-DCP-19	The platform should notify platform users of significant pipeline events including successful execution, pipeline failure, and detection of data quality issues exceeding	Met	Execution, failure and quality notifications in the platform interface - Section 5.5

Requirement	Requirement text	Status	How it is met
	configured thresholds, delivered via the platform user interface.		
REQ-DCP-20	The platform should allow users to configure which event types trigger notifications and the quality threshold values that activate quality alerts.	Met	User-configurable notification event types and quality thresholds - Section 5.5
REQ-DCP-21	The platform shall prepare and expose results, as pre-processed data packages assembled from quality-validated, harmonised and aggregated time series covering energy demand, generation, flexibility proxies and grid interaction vectors, formatted in JSON.	Met	Results listed in the catalogue and retrievable through retrieval endpoints - Sections 5.4.2, 6.7
REQ-DCP-22	The platform shall apply configurable gap-filling and imputation strategies to time series when preparing results.	Met	Configurable gap filling and imputation applied when results are prepared, reported in result metadata - Section 5.4.2
REQ-DCP-23	The platform shall support temporal aggregation and resampling of time series at configurable resolutions (15-minute, 60-minute, daily) using configurable aggregation methods (sum, mean, min, max) when preparing results.	Met	Results prepared and stored at 15-minute, 60-minute and daily resolution with configurable aggregation method - Section 5.4.2
REQ-DCP-24	The platform shall support the registration of organisations with their type, and the administration of their membership by the organisation itself, with owner, manager and member roles, where managers invite new members and promote or demote them between member and manager.	Met	Organisation registration and membership administration with owner, manager and member roles - Sections 5.6, 6.8