



TwinERGY Integrated Data Management Platform - Alpha, Mockups Release

D5.3

December 2021

Deliverable

PROJECT ACRONYM	GRANT AGREEMENT #	PROJECT TITLE
TWINERGY	957736	Intelligent interconnection of prosumers in positive energy communities with twins of things for digital energy markets

DELIVERABLE REFERENCE NUMBER AND TITLE

D5.3 TwinERGY Integrated Data Management Platform – Alpha, Mockups Release

Revision: 1

AUTHORS

Tasos Tsitsanis	Giorgos Papadopoulos
Suite5	Suite5



Funded by the Horizon 2020 programme of the European Union
Grant Agreement No 957736

DISSEMINATION LEVEL

✓ P Public

C Confidential, only for members of the consortium and the Commission Services

Version History

REVISION	DATE	AUTHORS	ORGANISATION	DESCRIPTION	
V1.00	15/12/2021	Tasos Georgios Papadopoulos	Tsitsanis, Suite5	Ready for review version	
	21/12/2021	Ana Isabel García, Moisés Antón García	Martínez Antón	ETRA	Suggestions for table formatting/ Minor Changes
	21/12/2021	Konstantinos Kotsalos, Kapetanios	Apostolos	ED	Minor Corrections/Changes and Suggestions
	23/12/2021	Fynn Bollhöfer, Axel Balke	Christian	TH-OWL	Minor text changes
V1.01 Consolidated Version	– 27/12/2021	Tasos Georgios Papadopoulos	Tsitsanis, Suite5		
V1.02 – Ready for Submission	29/12/2021	Tasos Georgios Papadopoulos	Tsitsanis, Suite5		Ready for submission

Statement of Originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Legal Disclaimer

The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any particular purpose. The above referenced authors will have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law. The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the European Union. Neither the INEA nor the European Commission is responsible for any use that may be made of the information contained therein.

© 2021 by TwinERGY Consortium

Executive Summary

The alpha, mock-ups release of the TwinERGY Integrated Data Management Platform is presented in Deliverable 5.3, as described in the TwinERGY DoA [1] , as well as the services that are delivered within the platform (Data Collection, Data Security, Data Storage, Platform Management). All services and their functionalities are described in depth in this document, to show the range of functionalities that the TwinERGY CDMP users, including data providers and all involved actors that will exploit the data available, would be able to access.

D5.3 describes the services that platform users will have access to, in order to appropriately configure and carry out platform functions (data collection, data security and data storage), as well as Platform Management functions (Identity Manager, Search & Retrieval Manager and Notifications Handler). To this end, ten (10) mock-ups have been created to explain and thoroughly clarify the most important functionalities delivered in the three core services, as well as to present the Platform Management functionalities that will be available to TwinERGY users. The services are summarized as follows: (a) the data collection service, which outlines the process that the platform users utilize to upload data; (b) the data security service, which outlines the functionalities for effectively addressing users' data security and privacy concerns regarding the data that will be ingested and handled in the TwinERGY Core Data Management Platform; and (c) the data storage service, which provides appropriate strategies and indexing methods to fulfil the need for robust data storage and indexing.

Through the use of appropriate mock-ups, this deliverable gradually introduces and describes all of the fundamental functionalities, that will serve as a foundation for the TwinERGY development activities of the Integrated Data Management Platform Beta Release on M16 (February 2022), which will be documented in the TwinERGY Deliverable D5.4. Any adjustments and improvements that may occur as the project progresses, as well as any new requirements that might arise, will be considered for integration into the future editions of the TwinERGY Core Data Management Platform.

Index

1. Introduction	8
1.1 Purpose of this deliverable	8
1.2 Scope of this deliverable	8
1.3 Structure of the document	9
1.4 Abbreviation List	9
2. TwinERGY Core Data Management Platform	11
3. Data Collection Service	19
3.1 Data Ingester	19
3.2 Data Mapper	21
3.3 Data Curator	22
3.4 Metadata Editor	23
4. Data Security Service	25
4.1 Access Policies Controller	25
4.2 Data Anonymization Handler	26
5. Data Storage Service	28
5.1 Data Store	28
6. Platform Management Service	29
6.1 Identity Manager	29
6.2 Search & Retrieval Manager	30
6.3 Notifications Handler	31
7. Conclusions	33
REFERENCES	34

List of Figures

Figure 1 TwinERGY Core Data Management Platform Conceptual Architecture	18
Figure 2 Ingestion of files in the TwinERGY CDMP	20
Figure 3 Ingestion of API data in the TwinERGY CDMP	21
Figure 4 Mapping data to the TwinERGY CIM	22
Figure 5 Curation Routines definition	23
Figure 6 Metadata Editor	24
Figure 7 Establishing access policies	26
Figure 8 Data Anonymization Configuration	27
Figure 9 User profile creation	30
Figure 10 Retrieval of datasets	31
Figure 11 Notifications on ongoing/finished data ingestions	32

List of Tables

Table 1 TwinERGY Pilot sites data availability	12
Table 2 Data requirements of the TwinERGY modules	14
Table 3 Analysis of the TwinERGY Use Cases	16
Table 4 TwinERGY CDMP Functional Requirements	18

1. Introduction

1.1 Purpose of this deliverable

Deliverable 5.3 describes the alpha, mock-ups release of the TwinERGY Integrated Data Management Platform, covering the services that are introduced within the platform, namely the Data Collection, the Data Security, the Data Storage, and the Platform Management service. By identifying all services and their associated functionalities, this deliverable defines the extent of the functionalities that the users of the TwinERGY Core Data Management Platform will be able to access.

Platform functions per service are introduced in this deliverable. They cover the needs for data ingestion, data mapping, data curation and metadata editing which are part of the Data Collection Service. Also, access policy control and data anonymization are part of the Data Security Service. Data store, combined with data backup and recovery mechanisms, along with the CIM store are included in the Data Storage Service. Finally, identity management, search and data retrieval and notifications management are part of the Platform Management Service. They are serving as the basis for further elaboration in D5.2 "Data Collection, Security, Storage & Management Services Bundles – Beta Release"[5] , in which the rapid prototyping and beta release of the services will be offered, as well as in D5.4 "TwinERGY Integrated Platform– Beta Release", enriching the initial functionality offered in the current deliverable.

1.2 Scope of this deliverable

D5.3 attempts to define and offer suitable guidance on the services that TwinERGY CDMP users will be able to access. This document's specific scope is described as follows:

- Define the main services, as well as the numerous functionalities that the platform's users will be able to access; in the sections that follow, each service and its associated functionalities are thoroughly described.
- Utilize this alpha version of the TwinERGY Integrated Data Management Platform as the foundation for the platform's beta release of the various Data Services Bundles' integration activities, offering the connections between the various services presented, which will be further analysed and extensively described in the respective deliverables (TwinERGY Deliverables 5.2 (M14), 5.4 (M16), 5.5 (M20), 5.6 (M24), 5.7 (M28), 5.8 (M32)).

1.3 Structure of the document

D5.3 is structured as follows:

Section 2 presents the activities undertaken towards the definition of the functional requirements of the TwinERGY Core Data Management Platform and the introduction of the conceptual architecture of the platform, by identifying the data requirements of both the pilot sites of TwinERGY project, as well as those of the modules to be developed. In addition, an analysis of the project's nine (9) use cases delineating the interactions of the TwinERGY CDMP with the different modules, systems and devices is performed.

Section 3 through Section 6 present a thorough description of the Data Collection, Data Security, Data Storage and Platform Management Services, and their relevant functionalities, along with appropriate mock-ups, respectively.

Section 7 concludes the deliverable by presenting a brief overview of what has been described within the document.

1.4 Abbreviation List

Acronym	Full Name
API	Application Programming Interface
CA	Consortium Agreement
CDMP	Core Data Management Platform
CIM	Common Information Model
CSV	Comma Separated Values
D	Deliverable
DER	Distributed Energy Resources
EC	European Commission
EMS	Energy Management System
ICT	Information and Communication Technology
GA	Grant Agreement
H2020	Horizon 2020 The EU Framework Programme for Research and Innovation
JSON	JavaScript Object Notation
PubSub	Publish-Subscribe
XML	Extensible Markup Language
WP	Work package

2. TwinERGY Core Data Management Platform

The project's backbone is an "open", modular, and interoperable big data management platform [2], that will enable open standards-based data collection and management communication throughout the project's related energy value chain. The TwinERGY Data Management Platform will adhere to existing open standards in the energy arena (e.g., OpenADR, USEF, etc.), while also incorporating a homogenized Common Information Model that ensures semantic interoperability of the digitalized energy assets involved in the pilot sites, as well as seamless integration, communication, and operation on top of any Energy Management System, as well as Smart Home systems and devices. To assure end-user data security and non-repudiation of DER assets, the TwinERGY Data Management Platform will be equipped with appropriate data security, privacy, authentication and authorization methods.

The following activities were carried out in order to ensure that the TwinERGY Platform will capture both the project's pilot sites, as well as the modules data needs, combined with the project use cases' technical characteristics:

- Data collection actions in the TwinERGY pilot sites are aimed at capturing useful information on present or future dataset availability, data format, and velocity in the four TwinERGY pilot sites.

Pilot	Data	Format	Velocity
Bristol City (England)	Measuring and storing power grid condition data such as busbar min/max, min/max and averages voltages, RMS current and the voltage phase angle, active and reactive power per feeder	csv, json	Near Real-time
Municipality of Benetutti (Italy)	Data related to the energy imported from the grid to cover prosumers' demand needs, along with smart metering data on residential and commercial prosumers' installations	csv, json	Batch, Near Real-time
Athens (Greece)	Data related to the energy imported from the grid and sensing measurements related to Ambience and Occupancy, such as Temperature, Humidity, Lighting and Presence	xslx	Batch

Hagedorn (Germany)	Transformer station per line power, voltage, power factor measurements, solar radiation measurements at the transformer station. Smart metering data on residential and commercial prosumers' installations and power, voltage, and current measurements. Battery storage assets data spanning from battery limits to storage capacity measurements. Weather data. Public Charging Station data, including power limit measurements, the total number of charging points and connectivity nodes. RES asset data, including PV, batteries and EVs data	csv, json	Batch, Real-time
--------------------	---	-----------	------------------

Table 1 TwinERGY Pilot sites data availability

- Identification and analysis of the requirements for the TwinERGY modules in order to capture the data requirements for the modules and the Digital Twins functionality implementation.

Module	Data Needs	Format	Velocity	Accessibility Method
M1 Consumer well-being module	Basic user identification, Environmental, Energy consumption and production, Energy prices, User comfort	json	N/A	HTTP Request/ API/ DB Query
M2 & M3 Consumer and Neighbourhood demand flexibility profiling Module	Electricity pricing, building total electricity and natural gas consumption, consumption deriving from different energy sources, PV generation, energy storage in the building, battery capacity, capacity to charge EV, battery storage capacity within the community, generation in case of community PV array, capabilities of the BMS, base building loads, flexible loads, thermostat, humidity and	json, csv	Batch	API

	CO2 Readings, Forecasted weather data			
M4 Home & Tertiary real time Energy Monitoring Module	Energy consumption of single electric appliances and single power lines, Indoor environment monitoring	csv/json	near real-time/ batch	As downloadable file/API
M5 RES integration and DER management Module	Local weather forecasting, Smart meter, Local consumers, Grid Storage, Community flexibility, Local RES asset, Grid/node limits	N/A	N/A	N/A
M6 Risk Management and event handling Module	Energy consumption of single electric appliances and single power lines	csv/json	near real-time/ batch	As downloadable file/API
M7 Electric Mobility as a Service Module	Accounts of owners of electric vehicle, electric vehicle details, EV user's payment details, charging station details	json	Real-time	N/A
M8 Customer Deployment and Social Engagement Module	Energy consumption, Energy production, Energy production /storage/forecast for community PV/ BESS/EVCP	csv	near real-time/ batch	As downloadable file
M9 Transactive Energy Module	Identification data, transaction data, balances and block chain related data	json	Real-time	API

Table 2 Data requirements of the TwinERGY modules

- Analysis of the nine TwinERGY project use cases to determine the scope of each, the probable actors involved, and the data requirements for their implementation, aiming also to identify the interactions of the TwinERGY CDMP with the different modules, systems and devices, as identified in Deliverable D4.4 "System's Architecture"[3].

High Level Use Cases Number	Title	Goal	Input	Output
-----------------------------	-------	------	-------	--------

1	Home Energy Management	Maximize self-consumption and self-sufficiency	HEMS gateway; PV, storage, Env. Sensors.	Home & Tertiary Module
2	RES generation in domestic and tertiary buildings	Minimizing energy costs and overall carbon emissions produced by end users and the community	iSCAN, Community battery Storage gateway; Raspberry Pi - community battery storage, HEMS; Raspberry Pi - PV home battery, Asset.	RES integration DER management Module, Consumer Demand Flexibility Profiling Module
3	Grid capacity enhancement utilizing e-mobility	Offer ancillary services to the DSO, the development of a smart charging scheme for EV owners and the participation of EV batteries in flexible energy markets	HEMS; Raspberry Pi - Asset.	Home & Tertiary Module
4	Prosumer's empowerment in local energy trading markets	Enable grid decentralization and democratization; balancing the grid as well as increasing the quality and reliability of it	iSCAN, Community battery Storage gateway; Raspberry Pi - community battery storage, HEMS; Raspberry Pi - PV home battery, Asset.	RES integration DER management Module, Consumer Demand Flexibility Profiling Module

5	Enhance grid flexibility through DER Management	Measure combined network data in order to calculate the network status in real time	HEMS gateway; Raspberry Pi - PV home battery, Community battery Storage Gateway; Raspberry Pi - community battery storage; Grid Gateway; RTU	RES integration DER management Module, TwinEV Module
6	Consumers' engagement in Demand Side Management Programs utilizing feedback mechanisms	Increase residential demand flexibility and decrease residential energy use	Grid Gateway; RTU; HEMS Gateway; Raspberry Pi - Asset.	Home & Tertiary Module, RES integration DER management Module
7	Consumer's engagement in demand response programs utilizing a socio-economic context	Influence energy exchanges between households, and consumer attitudes towards benefit and comfort	HEMS Gateway; Raspberry Pi - Asset.	Comfort & Wellbeing Module, Social Network Module
8	Consumer's engagement in demand response programs utilizing personalized comfort/health-oriented services	Accurately depict consumers' comfort / wellbeing level; the consumers become predictable energy wise; feedback notifications can be provided to them	HEMS Gateway; Raspberry Pi - Sensor, Physiological Sensor Controller; Home & Tertiary Module, Neighborhood Demand Flexibility Profiling Module,	Comfort & Wellbeing Module

9	Consumer Engagement in Demand Response Programs Utilizing Digital Twin Prediction Capabilities for Dynamic VPPs	Dashboards will be created for each of the pilot sites by utilizing the digital twins of the TwinERGY pilot sites	HEMS; Raspberry Pi - Asset.	Home & Tertiary Module, Neighborhood Demand Flexibility Profiling Module, Consumer Demand Flexibility Profiling Module
---	---	---	-----------------------------	--

Table 3 Analysis of the TwinERGY Use Cases

Following the completion of the aforementioned activities and evaluation of the findings and feedback obtained, a list of functional requirements is produced below, expressed in the form of "The platform shall perform an action" when a desired/recommended functionality is represented.

Requirement_ID	Description
Req_001	The platform shall have access to DER management system data.
Req_002	The platform shall provide the possibility for the ingestion of real-time datasets.
Req_003	The platform shall have access to Energy Management Systems data.
Req_004	The platform shall have access to Smart Home Systems data.
Req_005	The platform shall provide the possibility to communicate with the DERs to receive data.
Req_006	The platform shall provide the possibility to communicate with the relevant modules to send data.
Req_007	The platform might have access to grid data.
Req_008	The platform shall have access to sensor data.
Req_009	The platform shall have the capability to consume data provided from gateways.
Req_010	The platform shall allow the mapping and storage of all available datasets under a common information model (CIM).
Req_011	The platform should provide the possibility to a user to upload data files (csv, json).

Req_012	The platform shall provide the definition of metadata of available datasets.
Req_013	The platform should allow the user to define data anonymization routines.
Req_014	The platform should allow the user to define data curation routines on the data that the user owns.
Req_015	The platform shall formulate and enforce a data access control decision based on the associated data access policies.
Req_016	The platform should enable the periodic data retrieval from Open Data APIs.
Req_017	The platform should ensure an intuitive, easy-to-use user interface.
Req_018	The platform should ensure a user-interface with minimum latency.
Req_019	The platform should be able to handle big volumes of data.

Table 4 TwinERGY CDMP Functional Requirements

In order to address the requirements listed above, the following architecture of the CDMP is introduced.

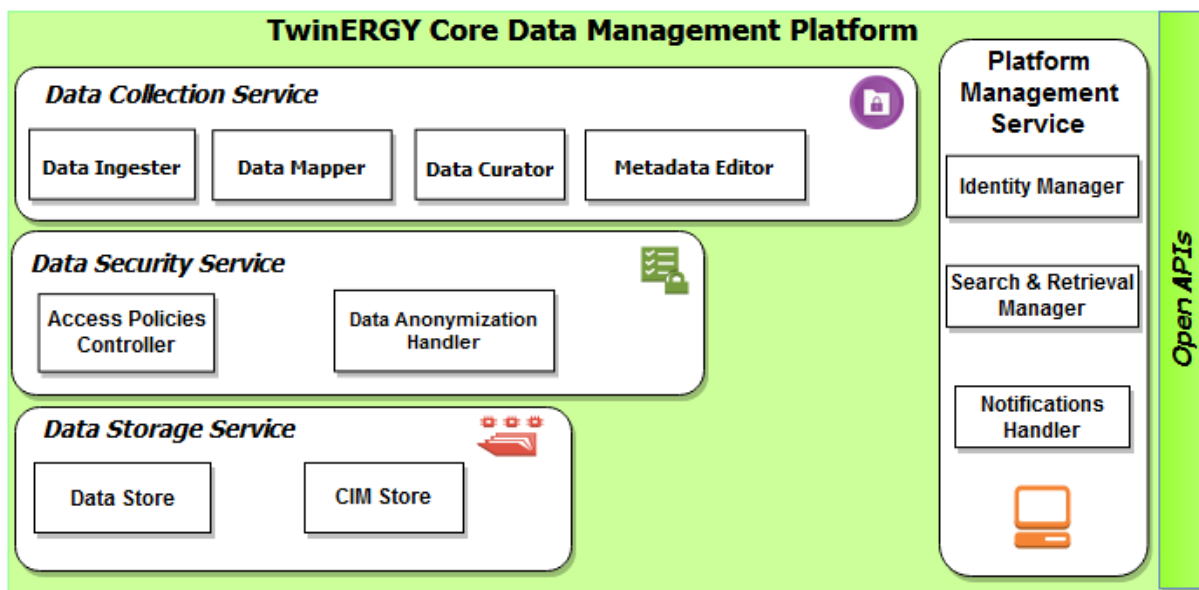


Figure 1 TwinERGY Core Data Management Platform Conceptual Architecture

The TwinERGY Core Big Data platform is made up of the following essential services, as shown in Figure 1 above:

- Data Collection Service
- Data Security Service
- Data Storage Service
- Platform Management Service

Each one of these services is analysed in the chapters below.

3. Data Collection Service

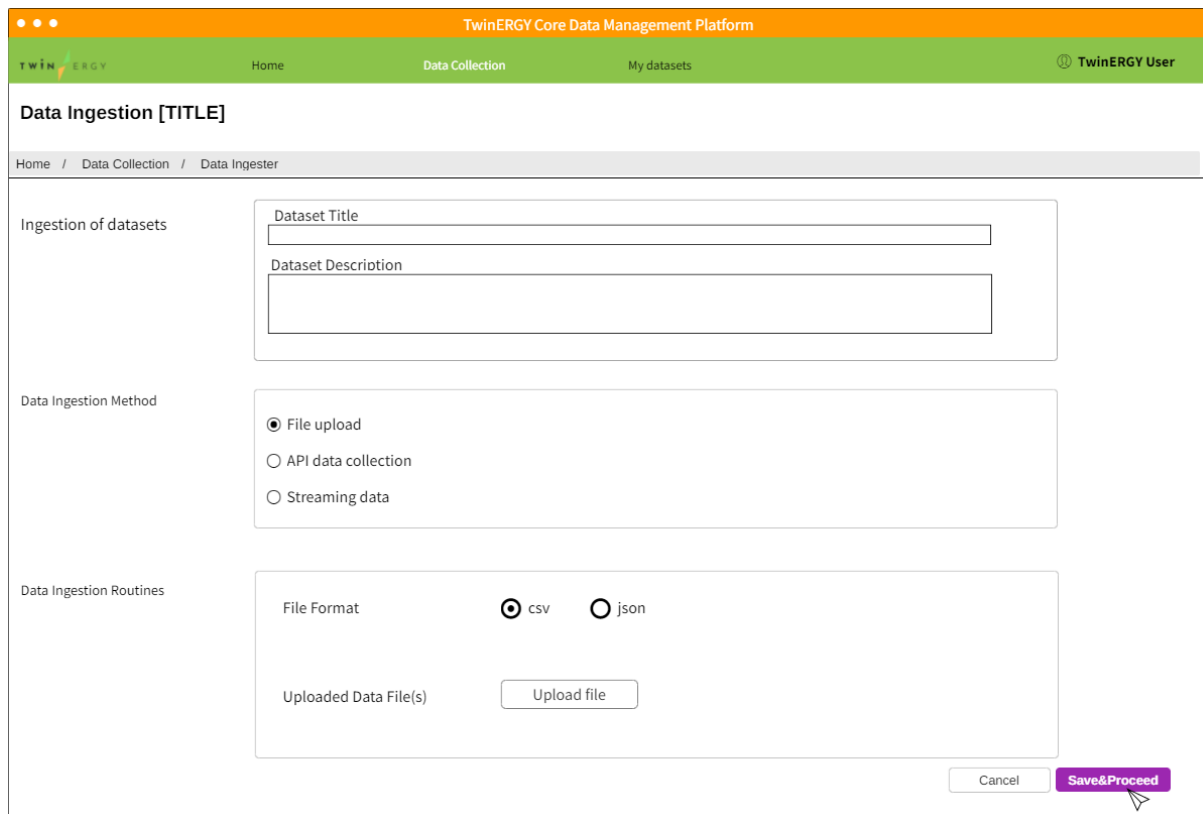
The Data Collection service, which is responsible for the data ingestion process in the TwinERGY Core Data Management Platform, is a critical part of the overall Platform configuration because it provides a wide range of functionalities. It includes data ingestion from various sources, semantic mapping of the data ingested in the platform, data curation to ensure the accuracy and integrity of the data to be processed later and metadata editing to complement the dataset with the necessary additional information. The Data Collection Service's key functionalities are described in the subsequent sections.

3.1 Data Ingestor

The role of the Data Ingestor is to simplify the data ingestion from various sources into the TwinERGY CDMP. The users uploading data will use the platform in order to:

- Select the data ingestion method to be utilized for the dataset to be ingested into the platform
 - Ingestion data from files: enables retrieval of data from files in formats that can be processed (e.g., csv, json).
 - Ingestion data from APIs: enables retrieval of data via pilot systems APIs as well as Open Data APIs (e.g., weather data, other local sources)
 - Ingestion of streaming data: enables the asynchronous service-to-service communication via PubSub mechanisms
- Provide the required data ingestion routines of the method selected, including the user preferences.

Figure 2 and 3 below, offer respective mock-ups of the data ingestion page to be presented to the platform user. In Figure 2, the user is asked to enter the title and the description of the dataset to be ingested, and after having selected the file upload method, in the data ingestion routines, the format of the file, as well as the actual file to be ingested can be defined.



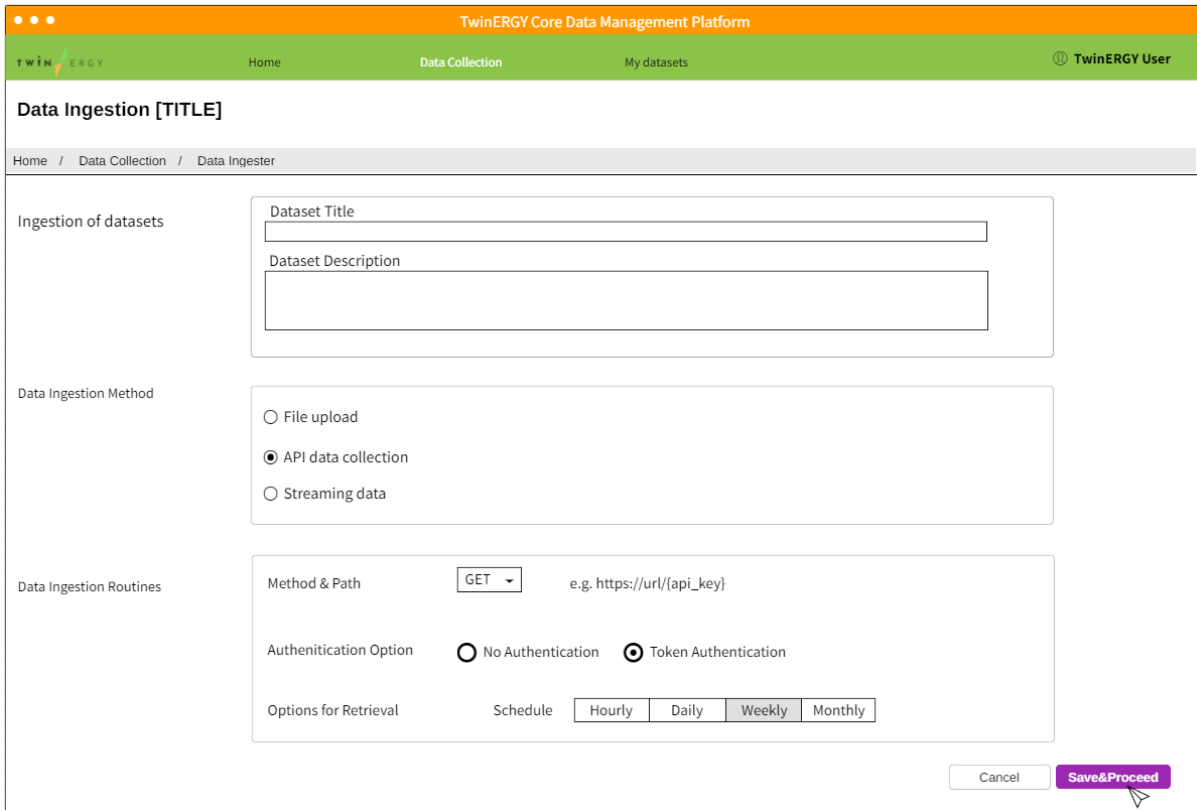
The screenshot shows the 'Data Ingestion [TITLE]' page in the TwinERGY Core Data Management Platform. The page has a green header with the TwinERGY logo and navigation links for 'Home', 'Data Collection', and 'My datasets'. The user is logged in as 'TwinERGY User'. The breadcrumb trail is 'Home / Data Collection / Data Ingester'. The main content area is divided into three sections:

- Ingestion of datasets:** Contains two text input fields for 'Dataset Title' and 'Dataset Description'.
- Data Ingestion Method:** Contains three radio button options: 'File upload' (selected), 'API data collection', and 'Streaming data'.
- Data Ingestion Routines:** Contains a 'File Format' section with radio buttons for 'csv' (selected) and 'json', and an 'Uploaded Data File(s)' section with an 'Upload file' button.

At the bottom right of the form, there are two buttons: 'Cancel' and 'Save&Proceed'.

Figure 2 Ingestion of files in the TwinERGY CDMF

Similarly, in Figure 3, the platform user has to enter the title and the description of the dataset to be ingested, but in this case, the API data collection method is chosen. In the data ingestion routines, the method and URL of the API service, combined with the authentication options to be used and the retrieval options need to be defined, in order to execute the ingestion method according to the user needs.



The screenshot shows the 'Data Ingestion [TITLE]' form in the TwinERGY Core Data Management Platform. The form is divided into three main sections:

- Ingestion of datasets:** Contains two text input fields for 'Dataset Title' and 'Dataset Description'.
- Data Ingestion Method:** Features three radio buttons: 'File upload', 'API data collection' (which is selected), and 'Streaming data'.
- Data Ingestion Routines:** Includes a 'Method & Path' dropdown menu set to 'GET', with an example URL 'e.g. https://url/{api_key}'. Below this are 'Authentication Option' radio buttons for 'No Authentication' and 'Token Authentication' (selected). At the bottom, there are 'Options for Retrieval' and a 'Schedule' section with buttons for 'Hourly', 'Daily', 'Weekly', and 'Monthly'. The 'Save&Proceed' button is highlighted in purple.

Figure 3 Ingestion of API data in the TwinERGY CDMP

3.2 Data Mapper

The users of the TwinERGY CDMP will be able to undertake necessary mapping activities, to ensure that all fields of ingested data are mapped to TwinERGY Common Information Model (CIM), as defined in D5.1 "TwinERGY Common Information Model"[4]. Semantic mapping is an integral part of the data handling process and serves as an enabler for other Data Management Platform components' functionalities. This functionality enables platform users to:

- Match the ingested data to the TwinERGY CIM by making the necessary adjustments to the data fields to correspond to the CIM concept names.
- Undertake the relevant mapping activities, in order to comply with the TwinERGY CIM, on the ingested data. For instance, the data values are matched to the CIM measurement units and the associated data types, using this functionality.

Figure 4 demonstrates the mapping playground to be offered to the platform user, in which the concept mapping and the transformation details can be defined.

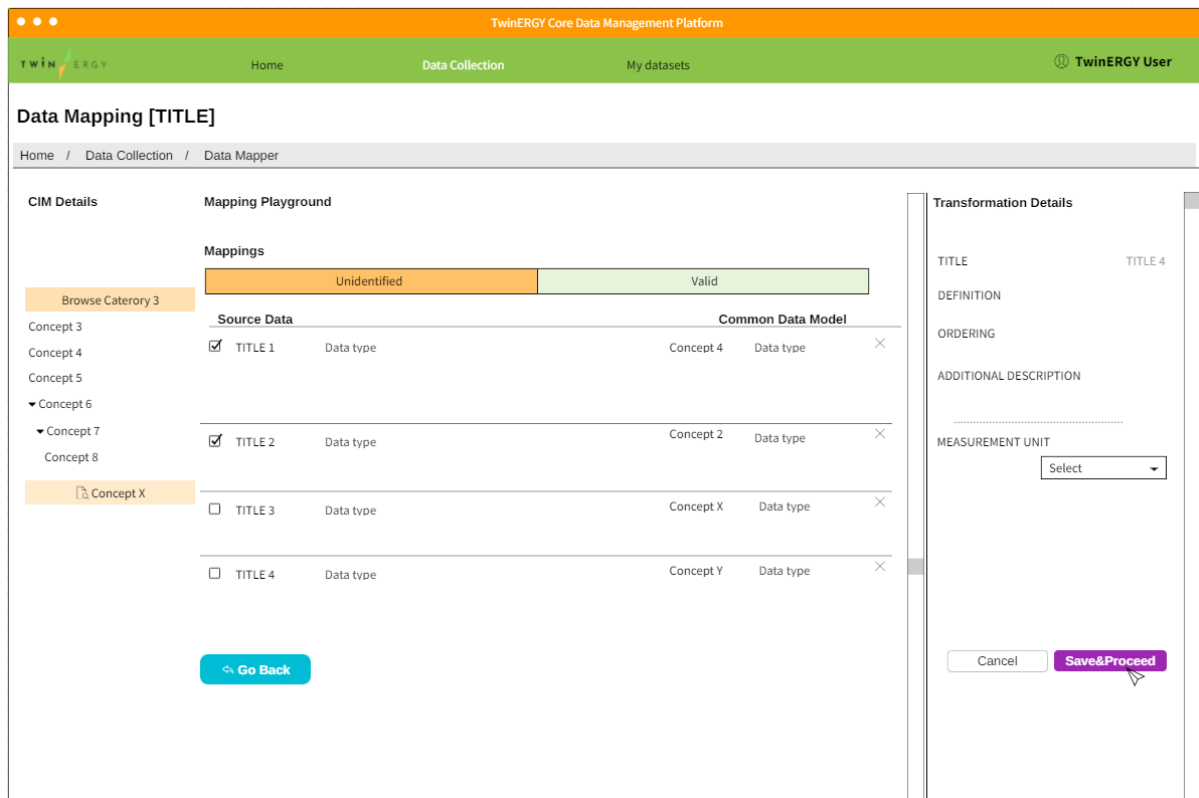


Figure 4 Mapping data to the TwinERGY CIM

3.3 Data Curator

Through the development of the Data Curator, the Data Collection Service guarantees that data ingested into the Core Data Management Platform is correct and complete. By eliminating or correcting incomplete and inconsistent data, data quality and reusability are improved. In this sense, the Data Curator will implement data curation techniques such as simple value substitutions, reformatting, and duplication removal, as well as more advanced tasks such as outlier identification and substitution. The Data Curator will allow data curation routines to be executed, determined by the platform user as constraints and limitations that the data ingested may have. Precise actions that must be performed if any of these constraints are breached, are also defined during the curation configuration process. A number of data validation processes are available depending on the data type of each column/field and are complemented by two types of corrective measures: discarding entries and modifying values. A curation rule is generated by combining a validation option with a corrective action for a specific field, and all curation routines are applied to the modified data (which comes from the semantic mapping to the TwinERGY CIM), resulting in curated data.

Figure 5 demonstrates the columns selection process and examples of the curation routines and constraints that can be defined from the platform user.

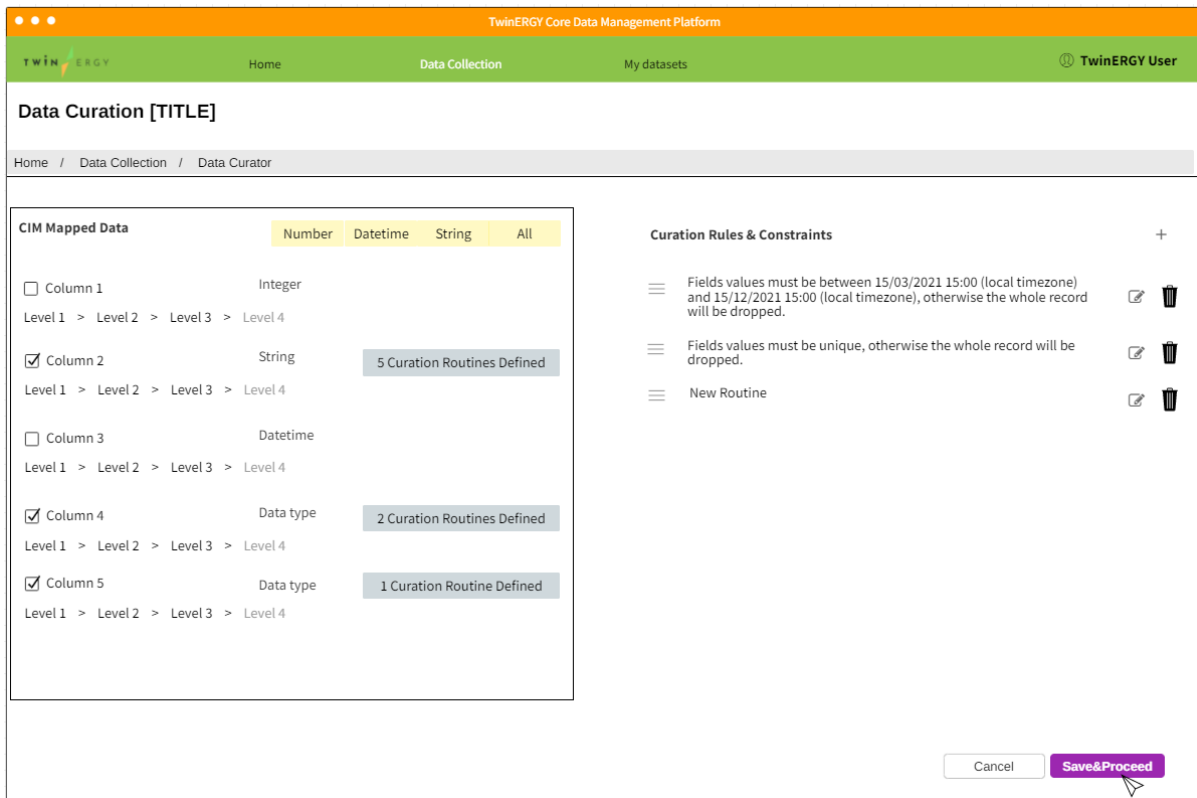


Figure 5 Curation Routines definition

3.4 Metadata Editor

Any result of the data ingestion is saved as a dataset in the TwinERGY Core Data Management Platform; therefore, users uploading data must define a complete profile of the dataset. Users of the platform will be allowed to issue the dataset title and define what is included in the dataset after going to My Datasets on the top navigation bar. Additionally, platform users will be able to specify the type of dataset, the format in which the data will be available, the language of the dataset, as well as specifics about the temporal coverage, spatial coverage, temporal resolution, and spatial resolution units of the data via a series of drop-down lists, as presented in Figure 6 below.

TwinERGY Core Data Management Platform

TWIN ENERGY
Home
Data Collection
My datasets
TwinERGY User

Edit Dataset Metadata

Home / Data Collection / Metadata Editor

General Information

Dataset Title

Dataset Description

Distribution

Type Format Language

Extent

Temporal Coverage Unit Temporal Resolution Unit

Spatial Coverage Unit Spatial Resolution Unit

Figure 6 Metadata Editor

4. Data Security Service

The Data Security Service is responsible for addressing the TwinERGY CDMP users' data security and privacy concerns regarding data that will be ingested and handled in the TwinERGY Core Data Management Platform. As a result, it provides numerous functionalities, such as the Access Policies Controller, and the Data Anonymization Handler, that may be intuitively configured in the TwinERGY Platform. The Data Security Service's key functionalities are described in the subsequent sections.

4.1 Access Policies Controller

By allowing users to create access routines that govern access requests on their data within the TwinERGY Core Data Platform in a flexible and clear manner, Access Policies Controller functionality strengthens users' trust in the whole TwinERGY Core Data Platform. Permission policies can be created for each dataset to achieve the required behaviour. The Access Policies Controller effectively separates issues between policy development and policy enforcement, ensuring that data may only be seen and accessed by eligible platform users. In particular, this functionality enables platform users to:

- Define, configure, and modify access policies: Using a graphical user interface, users may create sophisticated access policy routines for their datasets. These criteria could be based on both data fields and the characteristics of the requestor. Policies can be used to delineate under which circumstances access is allowed, and they can be linked together to create complex routines using Boolean logic. The routines are saved, and the user that uploads data can easily modify them using the same interface.
- Enforce data access policies: The access policies that are defined and configured, are applied when a request is made to access data that has been ingested into the TwinERGY Core Data Management Platform.

The users of the TwinERGY platform will need to set the access levels for the related datasets. The users can choose whether the dataset is for:

- i) Private access (access to the dataset is allowed only if access policies are met) or
- ii) General access (Free access to the dataset, no requirement for access policies to be met).

If the access level is set to private, the applicable access policies, as well as the relevant approach (Allow-all) and the inclusion of specific exceptions, must be defined. A user that uploads data must choose a user parameter (e.g., organization country, organization kind, etc.), a condition (e.g., equal, etc.), and the value of the parameter to add exceptions, as demonstrated in Figure 7.

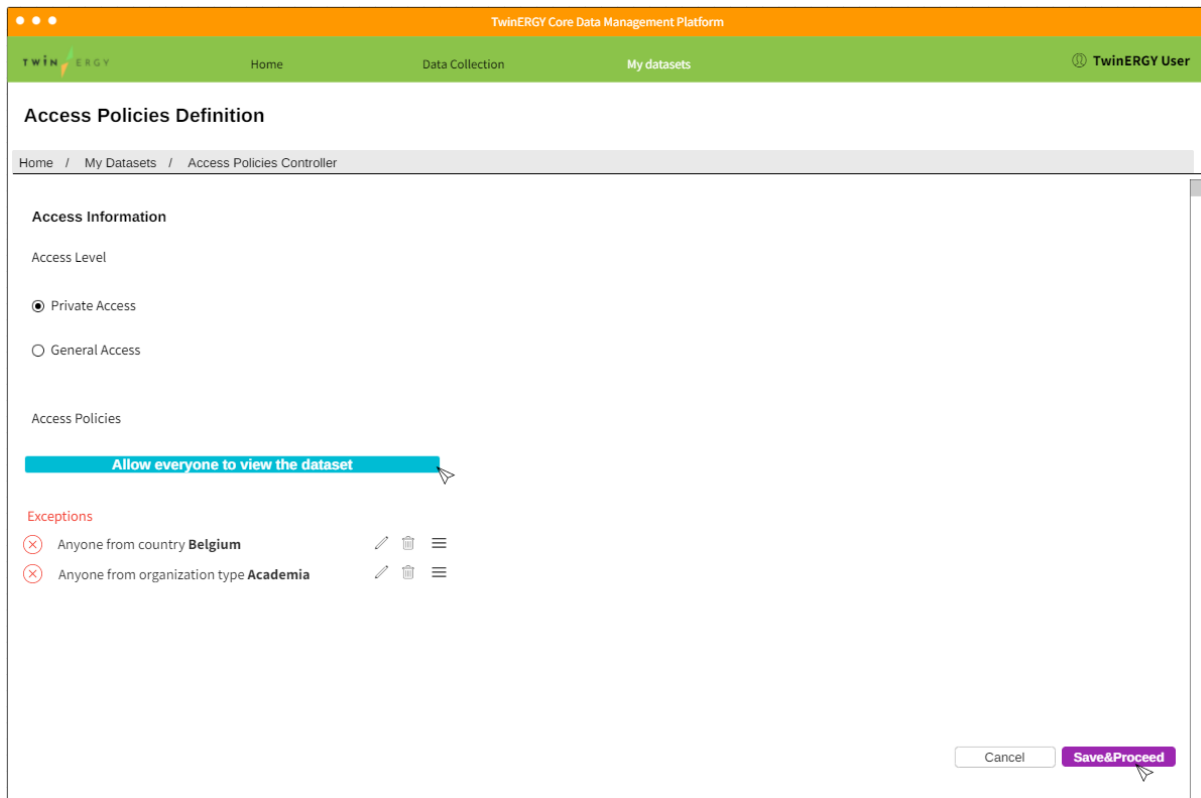


Figure 7 Establishing access policies

4.2 Data Anonymization Handler

The Data Anonymization Handler safeguards data from unintended disclosure of personal or corporate information. Anonymization actions that reflect into defined parts of the data (i.e., which fields) are conducted prior to making the data available on the TwinERGY platform. Users of the TwinERGY CDMP are provided with adequate tools, should they need to anonymize their data, as well as any fields in their data that potentially personal identifying information is contained. Additionally, it is important that platform users will have to determine an admissible threshold for information loss as well as the exact parameters of the chosen anonymization method. This functionality allows for:

- Field anonymization: The Data Anonymization Handler allows users to identify any data fields that include sensitive or identifying personal information, in accordance with the GDPR regulation.
- Data anonymization routines execution: The Data Anonymization Handler allows the platform user to choose the anonymization method that should be implemented in an easy manner based on the type of characteristic that the user defines (personal, identifying, quasi-identifying) and the data type. Such methods include generalization approaches to create arithmetic intervals or categories for numeric fields, and masking approaches for string fields.

Figure 8 offers an overview of the data anonymization configuration page, allowing users to define which columns pose sensitive of identifying information, as well as the admissible threshold for information loss.

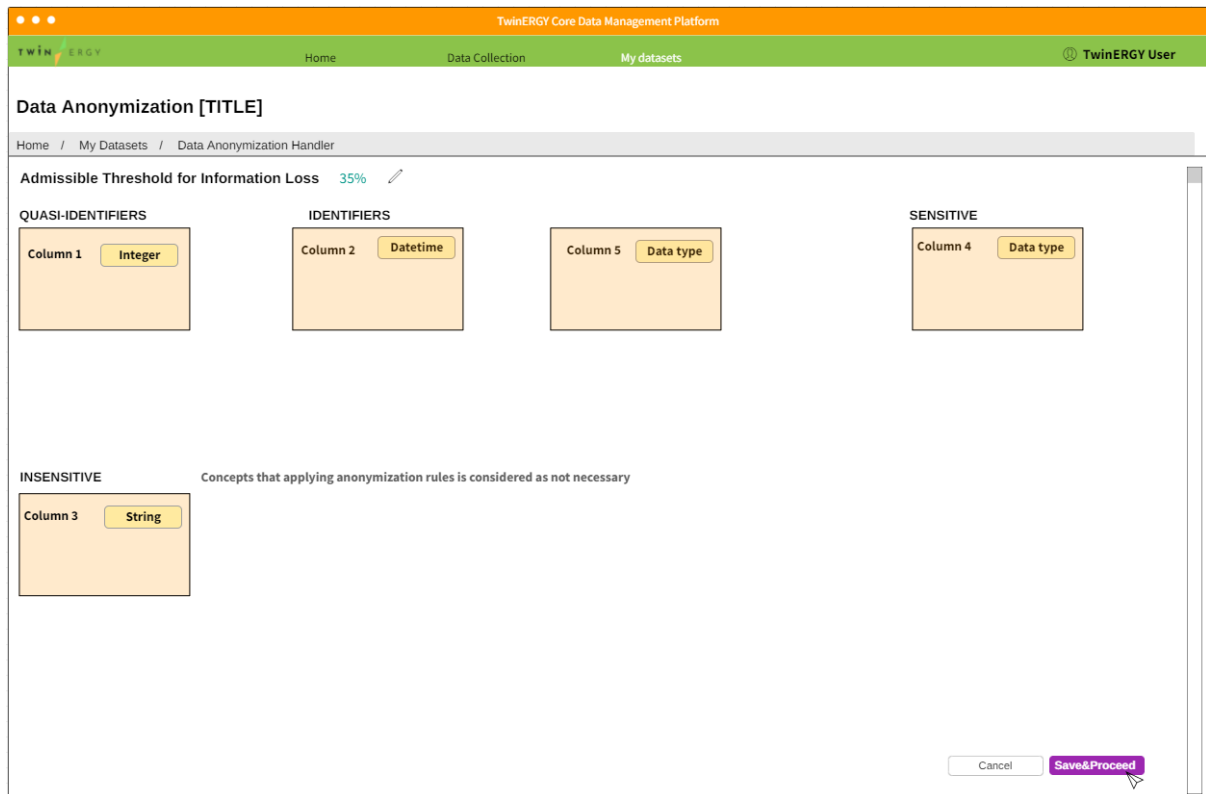


Figure 8 Data Anonymization Configuration

5. Data Storage Service

Through the provision of resilient measures and multiple indexing approaches, this service addresses the demand for reliable data storage and indexing. Different storage and indexing methods are planned to satisfy diverse needs depending on the type of information to be stored in the TwinERGY Platform and how it will be retrieved.

5.1 Data Store

The Data Storage Service is responsible for storing a wide range of data, as well as their related metadata, in a secure and dependable manner. All storing and indexing tools utilized in the context of TwinERGY are taken into account, with a focus on the following:

- Data storage: storing various data (i.e., datasets) as well as associated metadata in order to make it available to all TwinERGY platform modules and services.
- Log data storage: storing log-related information for the TwinERGY Core Data Management platform's operation and usage, as well as data from users and organizations, as well as any administrative information required for the TwinERGY Platform's smooth operation.
- CIM storage: storing the TwinERGY Common Information Model in its different versions, as well as the concepts and fields relevant to it.

An extra layer of data protection will be ensured with the utilization of appropriate Backup and Recovery Services, for applications, files, and servers that are created, administered, monitored, and supported. Creating backups of available data, protecting them in the event of security breaches or natural disasters that disrupt regular operations, and restore the data quickly enough to minimize the business damage is today regarded important for any organisation.

The safety and security of the data made available in the CDMP will be ensured by choosing the suitable backup and recovery solution for big data. This solution will also be customisable to meet the demands of the TwinERGY project.

6. Platform Management Service

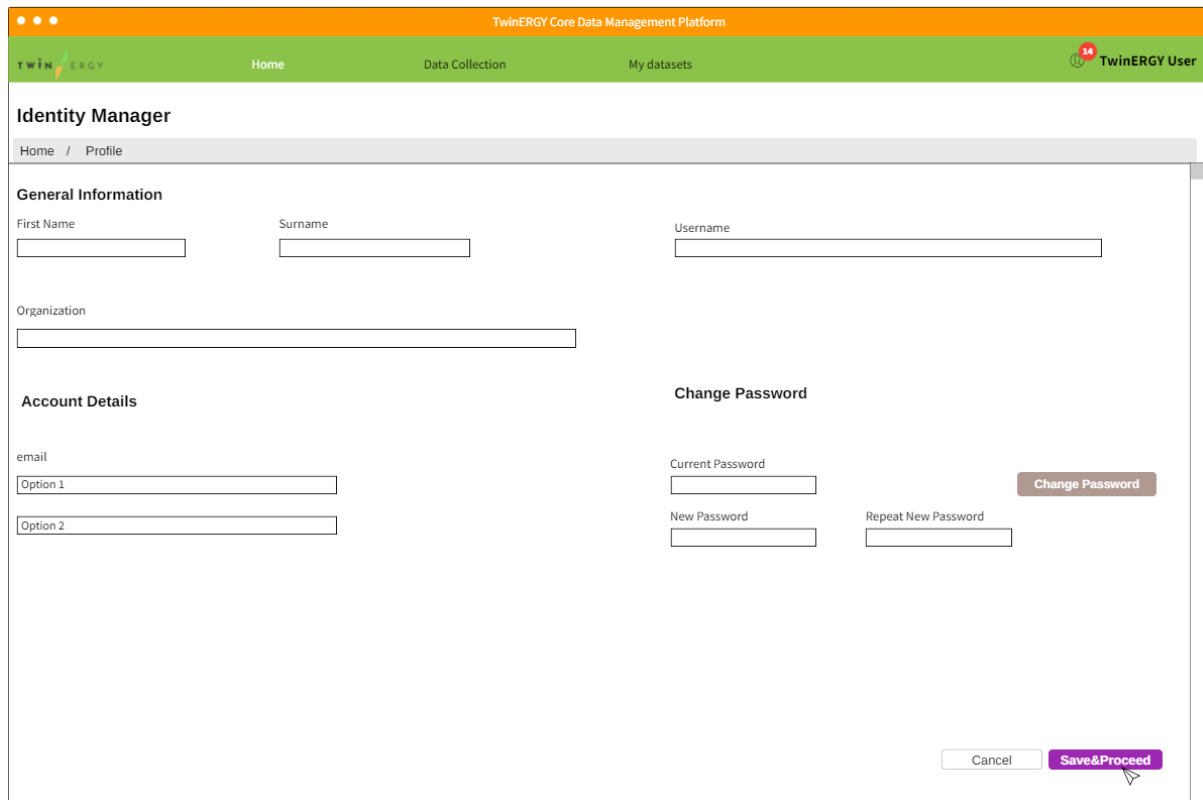
The Platform Management Service in TwinERGY's Core Data Management Platform is critical because it establishes the methods and processes for users to register securely and reliably on the platform. Users are provided access to the data they are eligible to search and retrieve, combined with proper authentication and authorization methods. Users will also be notified via platform notifications if data ingestion occurs on the platform, based on their preferences.

6.1 Identity Manager

Since it includes various features and layers of security, such as identity checks and reliable registration and authentication of platform users, the User Management functionality is an important part of the Platform Management Service. According to the data access policy routines stated in 4.2 Data Anonymization Handler, authenticated users are only granted access to data that they are permitted to use. This functionality allows for:

- Investigation of identity information for TwinERGY Core Data Management Platform users: The Identity Manager functionality acts as an identity provider, generating and managing identity information for TwinERGY platform users.
- Provision of authentication and authorization for modules and services: The Identity Manager functionality provides authentication and authorization services to limit access to authorized users based on data access policy routines.

Figure 9 presents a mock-up for the user profile creation, in which first, last and username need to be defined, combined with organization, e-mails and password details.



The screenshot shows the 'Identity Manager' interface within the 'TwinERGY Core Data Management Platform'. The user is logged in as 'TwinERGY User'. The page is titled 'Identity Manager' and has a breadcrumb trail 'Home / Profile'. The form is divided into two main sections: 'General Information' and 'Account Details'.

General Information:

- First Name:
- Surname:
- Username:
- Organization:

Account Details:

- email:
-

Change Password:

- Current Password:
- New Password:
- Repeat New Password:
-

At the bottom right, there are two buttons: 'Cancel' and 'Save&Proceed'.

Figure 9 User profile creation

6.2 Search & Retrieval Manager

Search & Retrieval Manager is a crucial part of TwinERGY's Core Data Management Platform, allowing users to search for and discover data that can be proven valuable, identify and define which of these available data are relevant to their needs, and finally, have a clear and comprehensive view of the provided results. Also, because it operates as a single point of entry for apps (both TwinERGY and third-party) to retrieve data from the CDMP that they're allowed to access, the setting of APIs for data retrieval is a major aspect of this functionality. The Search & Retrieval Manager functionality accepts all API calls before retrieving and aggregating data from the various services required to respond to the call, before eventually returning the appropriate result.

In Figure 10, it is illustrated how the platform user will be able to select the datasets of interest, the desired fields under each concept and the query parameters to be used for the creation of the API URL, that will allow the retrieval of the selected data from the platform.

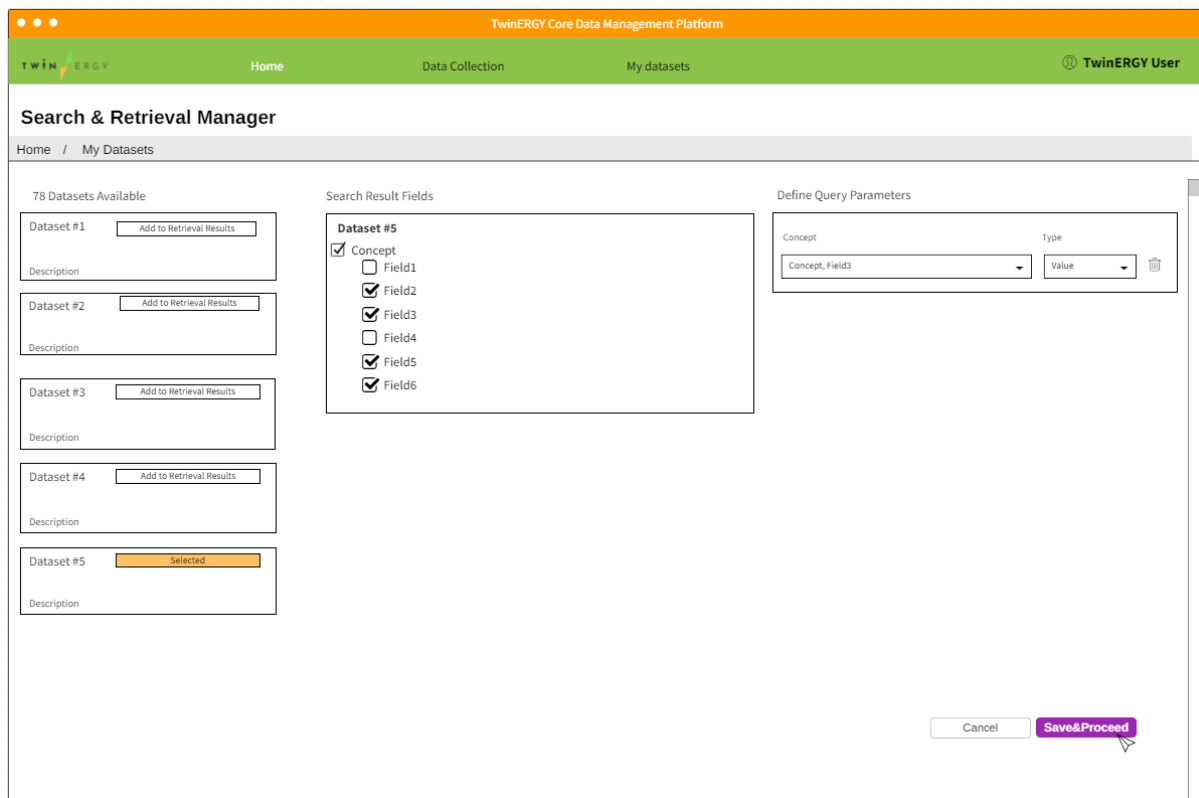


Figure 10 Retrieval of datasets

6.3 Notifications Handler

The users of the TwinERGY Core Data Management Platform will be able to receive notifications about certain events that occur on the platform and are relevant to their data jobs. Real-time information on the progress of ongoing data ingestion will be updated at significant milestones. Within the TwinERGY platform, notifications will be sent to ensure that essential information reaches the appropriate stakeholders as quickly and easily as possible. This functionality will allow for:

- Notifications delivery on various occurrences in the TwinERGY CDMP: The Notifications Handler functionality generates and sends out pertinent notifications about the data ingestion process (regarding successful completion each time it is executed, or details for any failure).
- Notifications management: Users can study and act on various notifications, as well as delete them as needed. Users can also personalize which notifications they receive based on their particular preferences.

Figure 11 presents the envisioned Notifications Manager page, that will display the notifications with their respective statuses.

TwinERGY Core Data Management Platform

TWIN ENERGY
Home
Data Collection
My datasets
14 TwinERGY User

Notifications Handler

Home / Notifications

Data Ingestion_TITLE 1_ Short Description	Status ✔ Successfully executed	Details Delete
Data Ingestion_TITLE 2_ Short Description	Status ✘ Failed to Execute	Details Delete
Data Ingestion_TITLE 3_ Short Description	Status ✘ Failed to Execute	Details Delete

[Go Back](#)

«
1
2
...
34
35
»

Figure 11 Notifications on ongoing/finished data ingestions

7. Conclusions

In Deliverable D5.3 ("TwinERGY Integrated Data Management Platform – Alpha, Mockups Release"), a detailed description of the TwinERGY Core Data Management Platform Services for configuring and executing the core platform functionalities (i.e., data ingestion, data security, and data storage), as well as different Platform Management functionalities (i.e., Identity Manager, Search & Retrieval Manager, Notifications Handler), was provided. This deliverable specified the alpha version of the TwinERGY Platform, offering a first introduction to the envisioned user interface via relevant mock-ups for the main functionalities of the defined services. In particular, this deliverable elaborated on:

- The Data Collection service, which is responsible for the data ingestion process in the TwinERGY Core Data Management Platform.
- The Data Security Service, which is responsible for addressing the platform users' data security and privacy concerns regarding the data that are ingested in the TwinERGY CDMP.
- The Data Storage Service, which is responsible for addressing the demand of the platform users for reliable data storage and indexing.
- The Platform Management Service, which is responsible for establishing the methods and processes for users to register securely and reliably on the platform.

Further elaboration of those services will be offered in D5.2 "Data Collection, Security, Storage & Management Services Bundles – Beta Release", in which the rapid prototyping and beta release of the services is planned, while D5.4 "TwinERGY Integrated Platform– Beta Release", will present an enriched version of the initial functionalities described in the current deliverable.

REFERENCES

[1] DoA Part A, TwinERGY Consortium Agreement No 957736

[2] DoA Part B, TwinERGY Consortium Agreement No 957736

[3] TwinERGY Consortium. (2020). TwinERGY D4.4 "System's architecture"

[4] TwinERGY Consortium. (2020). TwinERGY D5.1 "TwinERGY Common Information Model"

[5] TwinERGY Consortium. (2020). TwinERGY D5.2 "Data Collection, Security, Storage & Management Services Bundles – Beta Release"