

TwinERGY Integrated Data Management Platform-Release 2.00

D5.8

June 2023

Deliverable

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

DELIVERABLE REFERENCE NUMBER AND TITLE

D5.8 TwinERGY Integrated Data Management Platform – Release 2.00

Revision: 1

AUTHORS

Giorgos Papadopoulos	Marios Phinikettos
Suite5	Suite5



DISSEMINATION LEVEL

✓ **P** **Public**

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

Version History

REVISION	DATE	AUTHORS	ORGANISATION	DESCRIPTION
V0.95		Giorgos Papadopoulos, Marios Phinikettos	Suite5	1 st draft of the deliverable D5.8 - TwinERGY Integrated Data Management Platform-Release 2.00
.....				
.....				
.....				
.....				
.....				
V1.00 – Ready for Submission version		Giorgos Papadopoulos, Marios Phinikettos	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This deliverable contains original unpublished work except where clearly indicated otherwise.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 CINEA nor the European Commission is responsible for any use that may be made of the information contained therein.

© 2021 by TwinERGY Consortium

Executive Summary

Deliverable 5.8 presents the TwinERGY Integrated Platform Release 2.00 and provides an overview of the platform's services (Data Collection, Data Security, Data Storage, Platform Management). The purpose of this document is to inform the users of the TwinERGY Core Data Management Platform (CDMP) about the full range of functionalities that are available to them. This includes a description of all services and the associated functionalities, as well as functionality enhancements since the Platform's Release 1.00.

Deliverable D5.8 provides a more comprehensive overview of the TwinERGY CDMP Services that were previously covered in Deliverable D5.6 [1]. This latest deliverable incorporates updates and enhancements made to these services, including significant improvements in the data ingestion process. Notably, the introduction of the streaming Pub/sub mechanism expands the options for ingesting data, allowing users to stream their data directly to the platform in addition to using the traditional file upload and API retrieval methods.

Furthermore, the deliverable introduces the data anonymization handler as a data processing routine. This feature enables platform users to effectively manage any potentially sensitive information present in the data assets they upload to the CDMP, ensuring compliance with privacy regulations and safeguarding data confidentiality.

Finally, the extensions and updates made to the TwinERGY CIM with additional fields to ensure that both the pilot's available data and the module's data needs are captured successfully from the model, are presented in this deliverable.

Index

1. Introduction	8
1.1 Purpose of this deliverable.....	8
1.2 Scope of this deliverable.....	8
1.3 Structure of the document.....	8
1.4 Abbreviation List.....	9
2. TwinERGY Core Data Management Platform	10
3. Data Collection Service	12
3.1 Data Ingestion Job Creation.....	12
3.1.1 Data Ingester.....	13
3.1.2 Data Mapper.....	18
3.1.3 Data Curator.....	19
3.1.5 Data Anonymization Handler.....	19
3.1.6 Metadata Editor.....	21
4. Data Storage Service	23
4.1 Data Store.....	23
4.2 CIM Extension.....	23
5. Data Security Service	26
5.1 Access Policies Controller.....	26
6. Platform Management Service	27
6.1 Search & Retrieval Manager.....	27
6.2 Notifications Handler.....	28
7. Conclusions	30
REFERENCES	31
ANNEXES	26

List of Figures

Figure 1 TwinERGY Core Data Management Platform Conceptual Architecture	9
Figure 2 TwinERGY CDMP Landing Page	10
Figure 3 User registration	10
Figure 4 Data Ingestion Job Creation	12
Figure 5 Display of the data ingestion jobs page	12
Figure 6 File uploading in the TwinERGY CDMP	12
Figure 7 Overview of the sampling of the files uploaded	13
Figure 8 Process for appending data	13
Figure 9 Appending data in already stored datasets	14
Figure 10 Uploading bigger data files in a timely manner	14
Figure 11 API Data Ingestion	15
Figure 12 Introduction of new data ingestion methods	15
Figure 13 Streaming Pub/Sub Mechanism configuration	16
Figure 14 Setup of the streaming pub/sub connection	16
Figure 15 Streaming Data sample summary	17
Figure 16 Mapping data to the TwinERGY CIM	17
Figure 17 Curation Routines definition	18
Figure 18 Data Anonymization Handler Configuration	19
Figure 19 Data Anonymization type selection	19
Figure 20 Handling of quasi-identifier data fields	20
Figure 21 Data anonymization overview	20
Figure 22 Dataset's Metadata Editing	21
Figure 23 Provision of information on the granularity and coverage of the dataset	21
Figure 24 Data Store Configuration	22
Figure 25 Data Ingestion Jobs overview and completion status	22
Figure 26 Suggestion of new fields	23
Figure 27 Access Level Definition	25
Figure 28 Access Policies Definition	25
Figure 29 Retrieval of datasets	27
Figure 30 Notifications on ongoing/finished data ingestion jobs	28

List of Tables

Table 1 Additions of fields in the TwinERGY CIM

25

1. Introduction

1.1 Purpose of this deliverable

Deliverable 5.8 presents the final release of the TwinERGY Integrated Platform, release 2.00, along with its various services such as Data Collection, Data Security, Data Storage, and Platform Management. This deliverable builds upon the previous version, D5.6 - "TwinERGY Integrated Data Management Platform – Release 1.00", as well as on D5.7 – "Data Collection, Security, Storage & Management Services Bundles – Release 2.00" [2] providing detailed information on the services, their functionalities, and the updates and enhancements implemented in the transition from release 1.00 to the release 2.00 of the CDMP. The document utilizes intuitive and comprehensive screenshots to illustrate the process of setting up and utilizing the core functionalities of the platform.

1.2 Scope of this deliverable

Deliverable D5.8 - "TwinERGY Integrated Data Management Platform - Release 2.00," introduces the latest functionalities accessible to its end users. It provides updates on the implementation status of the TwinERGY Core Data Management Platform services and includes essential user routines, new features, and improvements since the previous release. The document also offers a user-friendly overview, complete with instructions and screenshots, to support platform understanding and usage.

1.3 Structure of the document

Deliverable D5.8 is structured as follows:

An overview of the TwinERGY Core Data Management Platform is provided in *Section 2. TwinERGY Core Data Management Platform* .

The Data Collection, Data Storage, Data Security, and Platform Management Services are fully described in *Sections 3. Data Collection Service through 6. Platform Management Service*. In order to inform and educate TwinERGY CDMP users on the features that are available on the TwinERGY Integrated Platform and the functionalities added in this version of the platform, appropriate screenshots are also provided for each functionality.

The deliverable is concluded in *Section 7. Conclusions*, with a quick summary of what has been covered 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

An "open", modular, and interoperable big data management platform serves as the foundation of the project and enables open standards-based data collection and management communication throughout the TwinERGY project's energy value chain. The TwinERGY Core Data Management Platform follows existing open energy standards, as listed in deliverable D5.1[3]. Additionally, the CDMP includes a homogenized Common Information Model (TwinERGY CIM) that guarantees semantic interoperability of the digitalized energy assets installed in the pilot sites as well as seamless integration, communication, and operation on top of any Energy Management System and Smart Home systems and devices. To guarantee the protection of end-user data and non-repudiation of DER assets, the TwinERGY Data Management Platform is outfitted with the necessary data security, privacy, authentication, and authorization mechanisms.

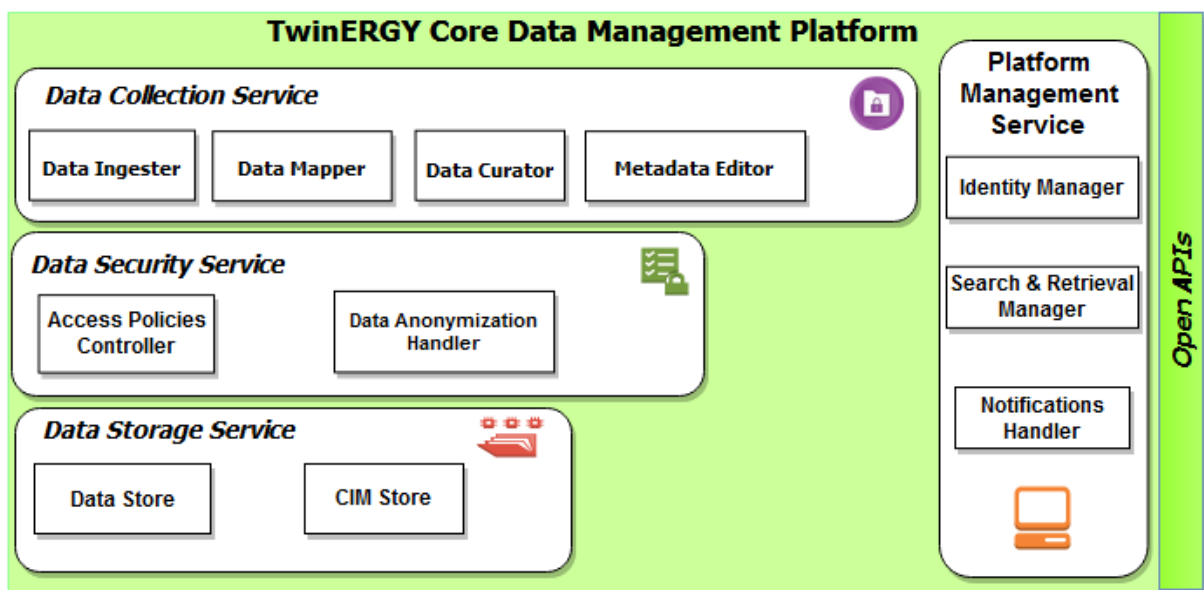


Figure 1 TwinERGY Core Data Management Platform Conceptual Architecture

The following services make up the TwinERGY Core Data Management Platform, as depicted in Figure 1 above:

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

The TwinERGY Core Data Management Platform, which can be found at <https://twinergy.s5labs.eu/>, is incorporated in this deliverable.

As shown in Figure 2, the landing page welcomes users to the TwinERGY CDMP and invites them to register or log in using their existing accounts.

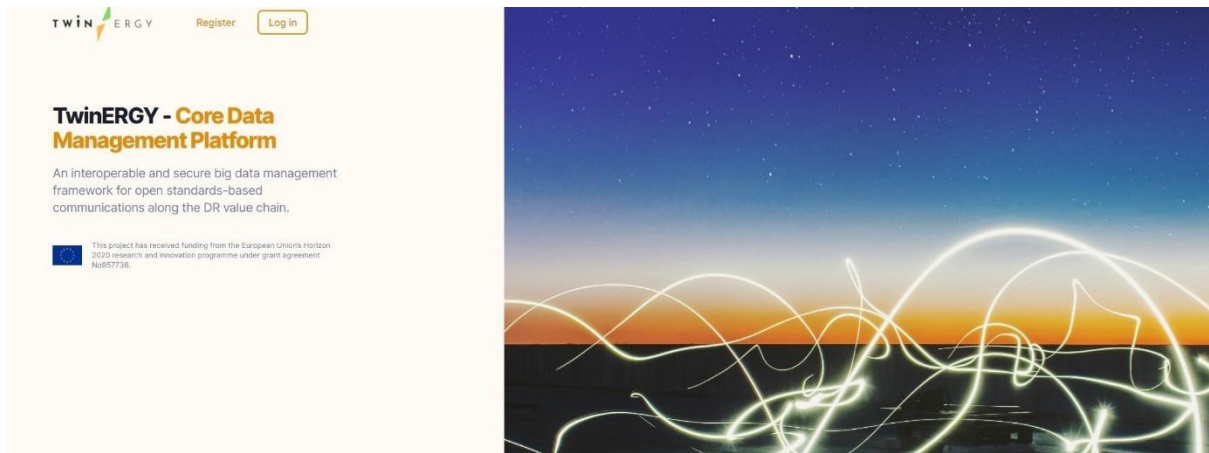


Figure 2 TwinERGY CDMP Landing Page

To begin the registration process on the TwinERGY CDMP, users are required to provide their first and last names, choose a username, enter their email address, and create a password that meets specific criteria. The password needs to be entered twice for verification, as shown in Figure 3.

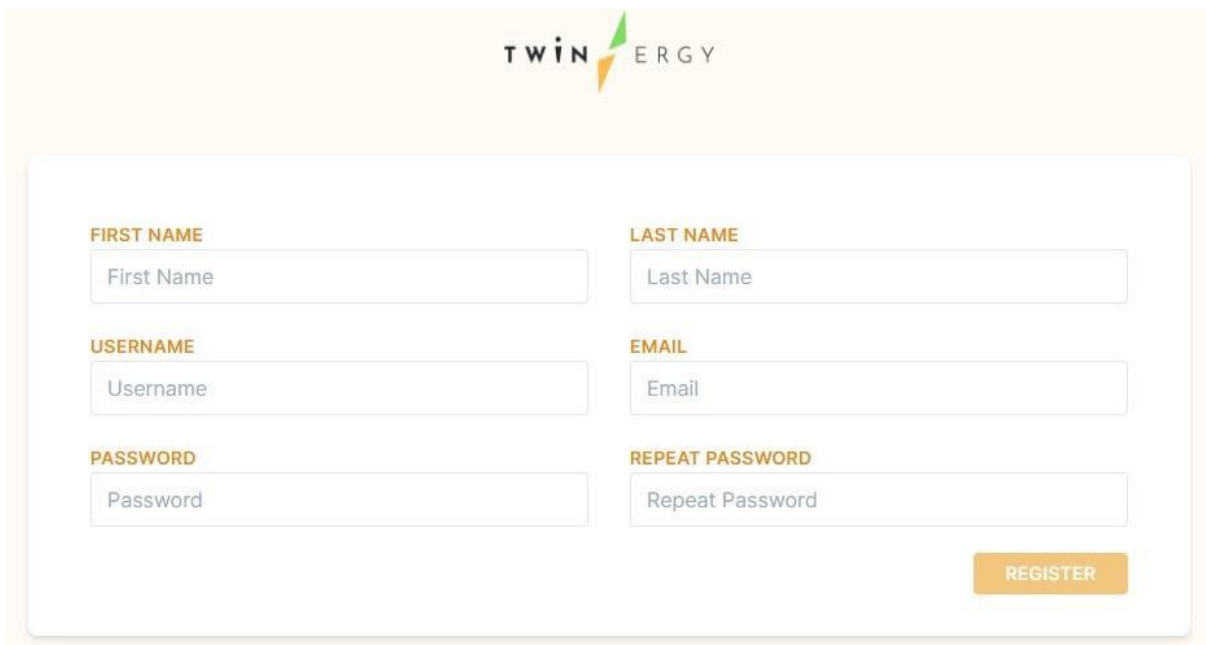


Figure 3 User registration

Upon successful completion of the registration process, users are directed to the login page, where they can enter their credentials and access the TwinERGY Core Data

Management Platform. These actions, which fall under the Platform Management Service, are designed to serve the Identity Manager functionality.

3. Data Collection Service

The TwinERGY Core Data Management Platform's data ingestion process is managed by the Data Collection Service, making it an essential component of the platform's overall configuration. The process includes ingesting data from numerous sources, semantically mapping the data once it is on the platform, data curation to ensure the accuracy and integrity of the data that will be analysed later, and metadata editing to add the necessary extra information to the dataset.

A number of functionalities for configuring the data ingestion process are provided by the Data Collection Service. Utilizing the Data Ingestor functionality, any CDMP user can upload data to the TwinERGY Integrated Platform infrastructure via dependable and secure methods, such as direct batch data uploading or real-time data uploading through APIs or KAFKA jobs. The TwinERGY Core Data Management Platform also saves the results of a data ingestion job as a dataset. The functionality of the Metadata Editor makes this process possible.

As part of the Data Mapper functionality, the user interface of the platform is designed and assembled to give platform users the necessary guidance so they can map all the fields of ingested data using the TwinERGY Common Information Model (CIM). Additionally, the user may describe the constraints and limitations that data ingested into the platform may have, as well as the precise actions that must be taken if any of these restrictions are breached. The high quality and value of the ingested data are ensured by the Data Curator functionality. When a validation option and a corrective action are combined to generate a curation routine for a particular field, the resultant curated data is applied to the converted data (obtained via the mapping process to the Common Information Model).

3.1 Data Ingestion Job Creation

Upon accessing the Core Data Management Platform, users are directed to the Data Collection page. To initiate the data ingestion process, they can click on the "Create" option in the top navigation bar. At this stage, users are required to select suitable processing routines and provide relevant parameters for the data ingestion job.

The process begins by creating a new data ingestion job and specifying the appropriate data processing routines. Users are prompted to provide essential details about the job, including a title, a brief description of the dataset to be ingested, and the data processing routines to be executed. While the data ingester, data mapper, and data store routines are pre-selected, users have the flexibility to choose whether to include data curation and data anonymization for specific fields, as the data curator and data anonymization handler routines are optional, as depicted in Figure 4.

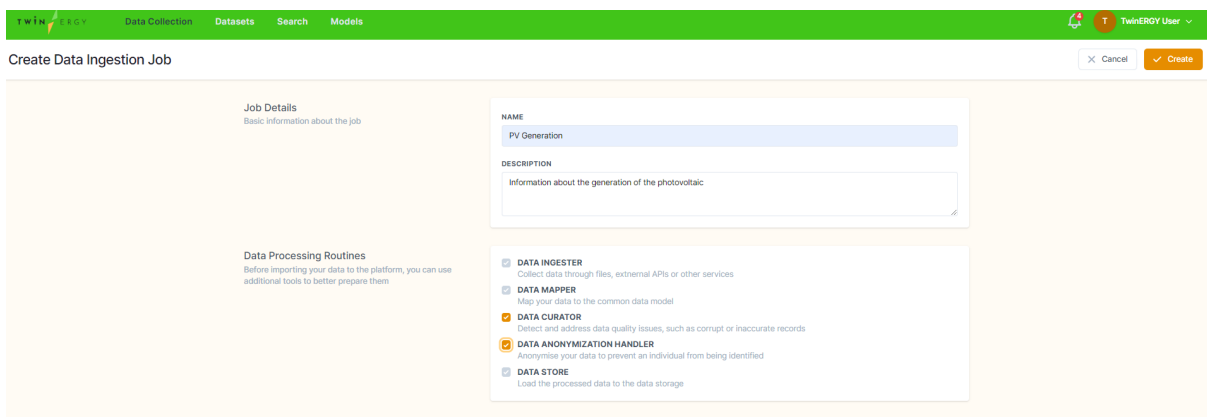


Figure 4 Data Ingestion Job Creation

After the required information is provided, the job is created and listed in the data ingestion jobs list, as shown in Figure 5.



Figure 5 Display of the data ingestion jobs page

3.1.1 Data Ingester

The user of the TwinERGY Core Data Management Platform can choose whether the data will be fed into the platform directly through file upload, APIs or KAFKA jobs. Users of the platform can also select the file format (csv, json) while using the file upload option. As shown in Figure 6, this process is completed by submitting a sample and the actual file to be uploaded.

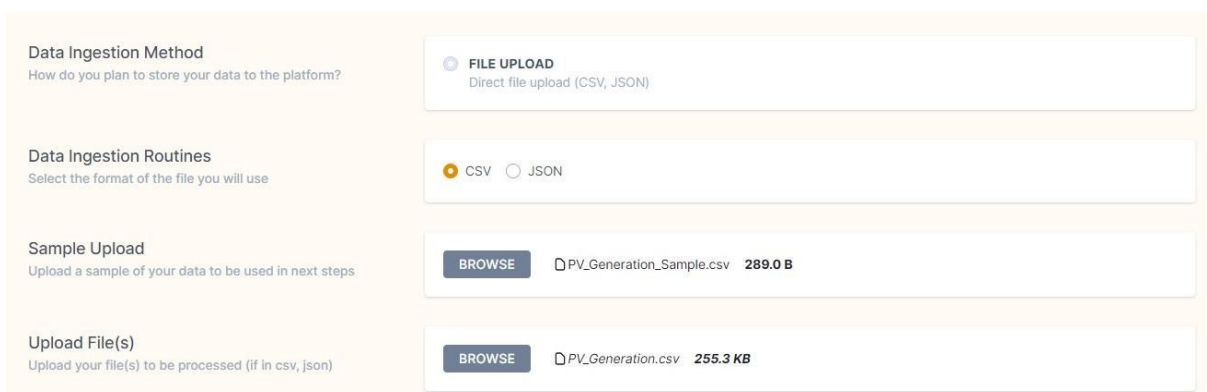
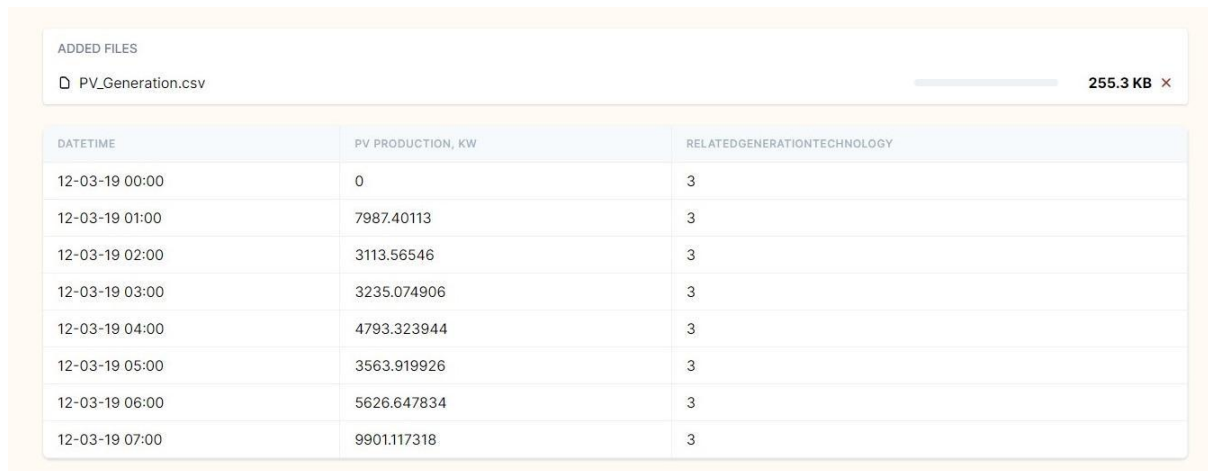


Figure 6 File uploading in the TwinERGY CDMP

Platform users receive a summary of the sample file they chose once all necessary information has been given, as depicted in Figure 7. Then they can initially Save and then Finalize the data ingester configuration.



DATETIME	PV PRODUCTION, KW	RELATEDGENERATIONTECHNOLOGY
12-03-19 00:00	0	3
12-03-19 01:00	7987.40113	3
12-03-19 02:00	3113.56546	3
12-03-19 03:00	3235.074906	3
12-03-19 04:00	4793.323944	3
12-03-19 05:00	3563.919926	3
12-03-19 06:00	5626.647834	3
12-03-19 07:00	9901.117318	3

Figure 7 Overview of the sampling of the files uploaded

As part of the improvements that took place since the initial implementation of the CDMP, significant enhancements were made to the data ingestion capabilities. Specifically, users are able select the "Append Data" option when utilizing the File Upload method for data ingestion. This functionality empowers users to incorporate new files into the CDMP that conform to the same format as the previously ingested data depicted in Figure 8.

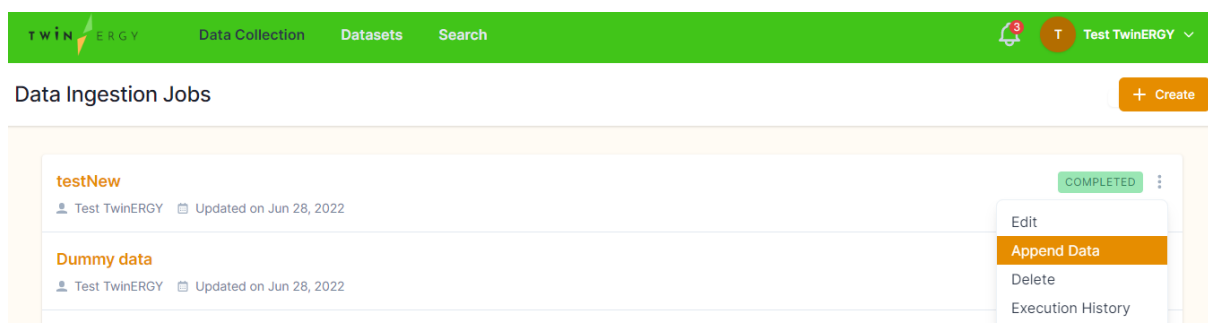


Figure 8 Process for appending data

This functionality allows the user whose data will be added, to append new files, in the same file format with the data that has already been ingested, as depicted in Figure 9.

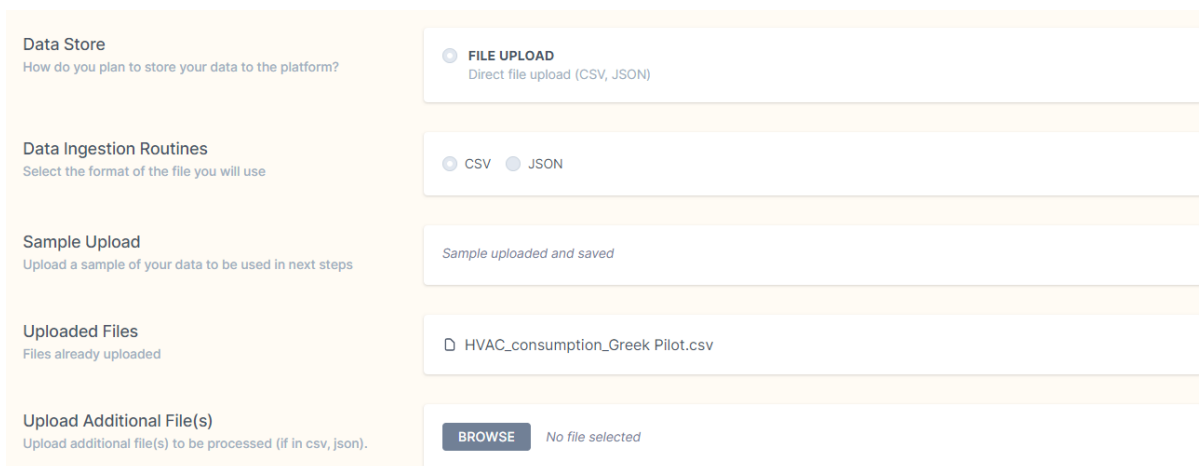


Figure 9 Appending data in already stored datasets

The TwinERGY Core Data Management Platform supports the ingestion of larger data volumes within a single dataset, as depicted in Figure 10, allowing for files up to approximately 500MB in size. Performance enhancements have also been implemented, enabling users to efficiently ingest files with complex or nested structures, specifically in JSON format. These enhancements have resulted in improved ingestion timeframes, typically around 5 minutes.

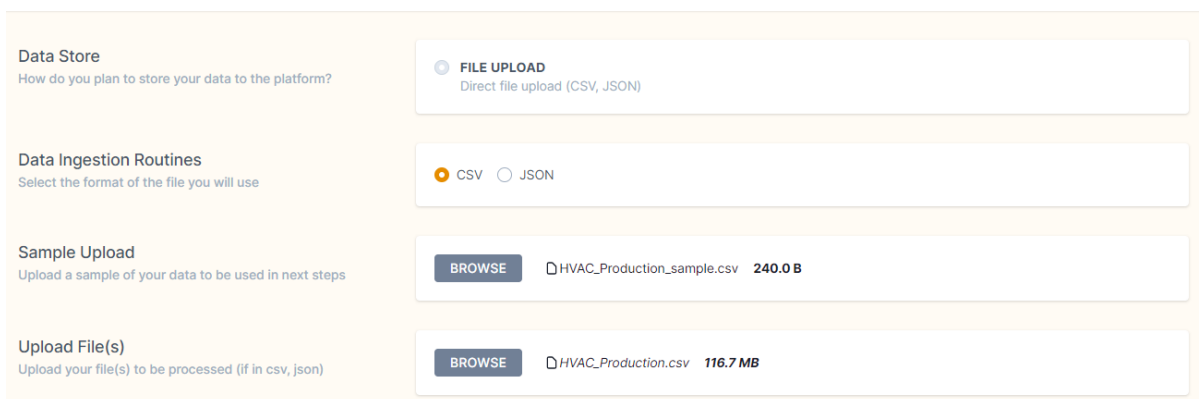
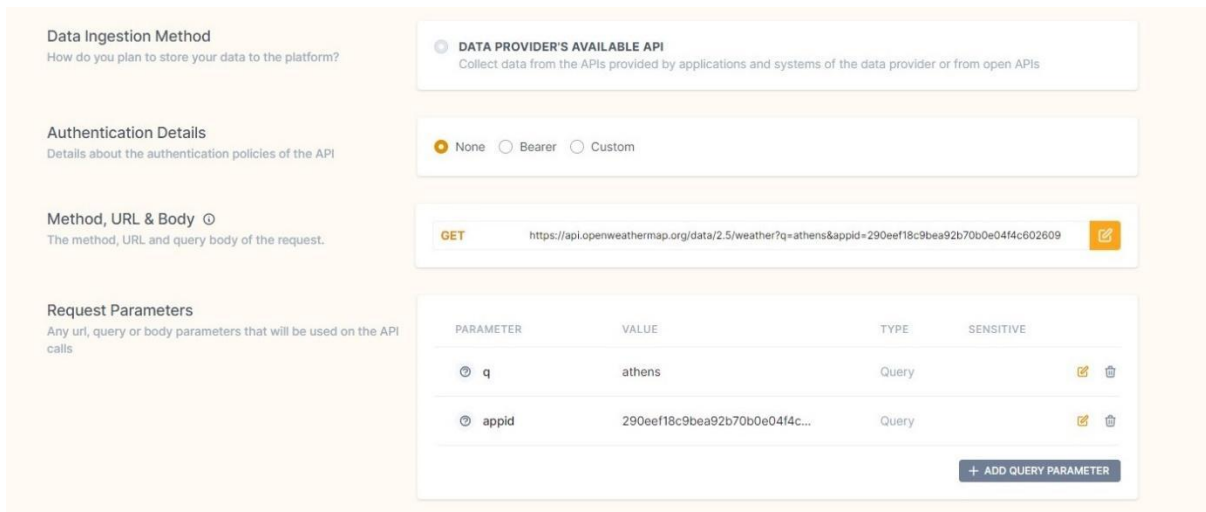


Figure 10 Uploading bigger data files in a timely manner

For API data ingestion, users who store their data within their organization's systems and expose it through APIs, can leverage the API retrieval functionality provided by the TwinERGY CDMP. To initiate the data ingestion process, users are presented with a configuration screen, as depicted in Figure 11, where they can specify the required details. This includes selecting the authentication type and providing the complete API URL, along with the desired method (e.g., GET) and any relevant request parameters.



Data Ingestion Method
How do you plan to store your data to the platform?

DATA PROVIDER'S AVAILABLE API
Collect data from the APIs provided by applications and systems of the data provider or from open APIs

Authentication Details
Details about the authentication policies of the API

None Bearer Custom

Method, URL & Body
The method, URL and query body of the request.

GET https://api.openweathermap.org/data/2.5/weather?q=athens&appid=290eef18c9bea92b70b0e04f4c602609

Request Parameters
Any url, query or body parameters that will be used on the API calls

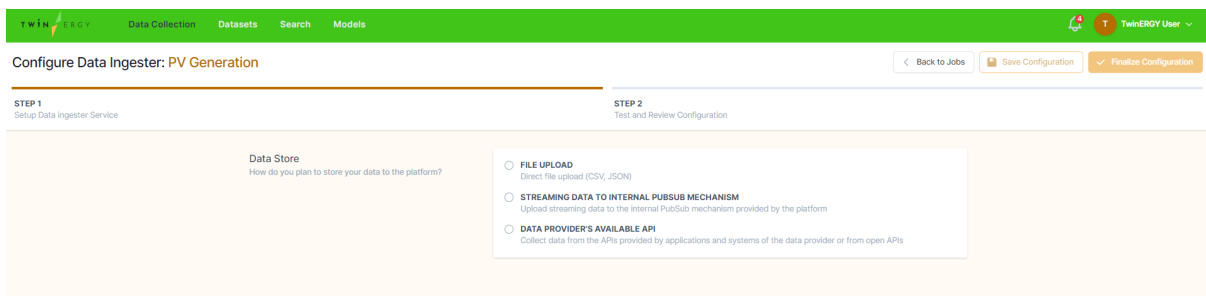
PARAMETER	VALUE	TYPE	SENSITIVE
q	athens	Query	<input checked="" type="checkbox"/>
appid	290eef18c9bea92b70b0e04f4c...	Query	<input checked="" type="checkbox"/>

+ ADD QUERY PARAMETER

Figure 11 API Data Ingestion

Users are also asked to select how frequently they want to retrieve data via the API and to specify appropriate scheduling intervals. To help the user quickly examine the data that will be saved on the platform, a summary of the structure that is captured, together with selected concepts, are provided.

Since the release of version 1.00, the TwinERGY Integrated Data Management Platform has undergone updates and introduced new features. One significant enhancement is the inclusion of the streaming Pub/sub mechanism, which allows users to stream their data to the platform in addition to the existing file upload and API retrieval methods.



Configure Data Ingestor: PV Generation

Back to Jobs Save Configuration Finalize Configuration

STEP 1
Setup Data Ingestor Service

STEP 2
Test and Review Configuration

Data Store
How do you plan to store your data to the platform?

- FILE UPLOAD**
Direct file upload (CSV, JSON)
- STREAMING DATA TO INTERNAL PUBSUB MECHANISM**
Upload streaming data to the internal PubSub mechanism provided by the platform
- DATA PROVIDER'S AVAILABLE API**
Collect data from the APIs provided by applications and systems of the data provider or from open APIs

Figure 12 Introduction of new data ingestion methods

With the streaming Pub/sub mechanism, users are provided with connection information, including the URL, port, SSL settings, topic, username, and password, as depicted in Figure 13.

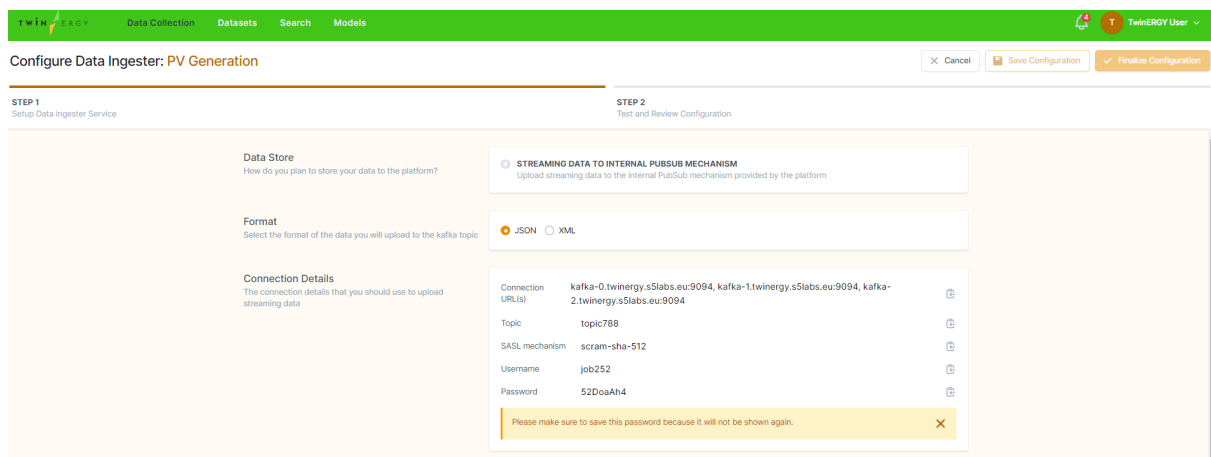


Figure 13 Streaming Pub/Sub Mechanism configuration

This enables them to establish a secure and reliable connection for streaming data. Furthermore, as shown in Figure 14, users have the flexibility to configure specific retrieval scheduling parameters based on their needs. As part of the setup process, users are also requested to upload a sample of the streaming data, facilitating efficient processing and handling of the incoming data stream.

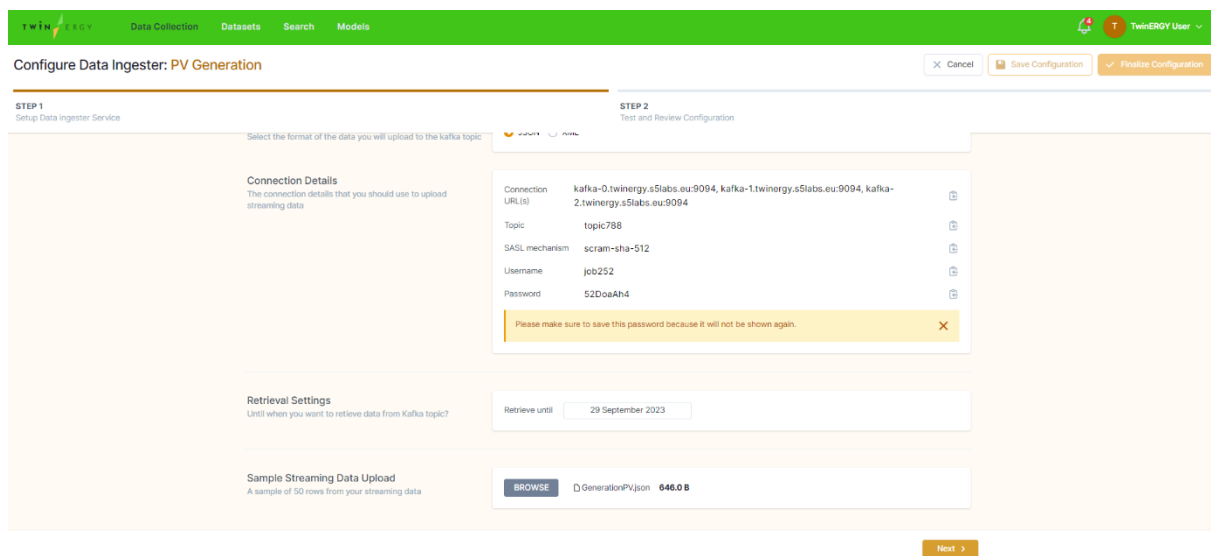


Figure 14 Setup of the streaming pub/sub connection

A summary of the uploaded data sample is provided to the user, allowing for a quick and convenient examination of the data that will be stored on the platform, as depicted in Figure 15.

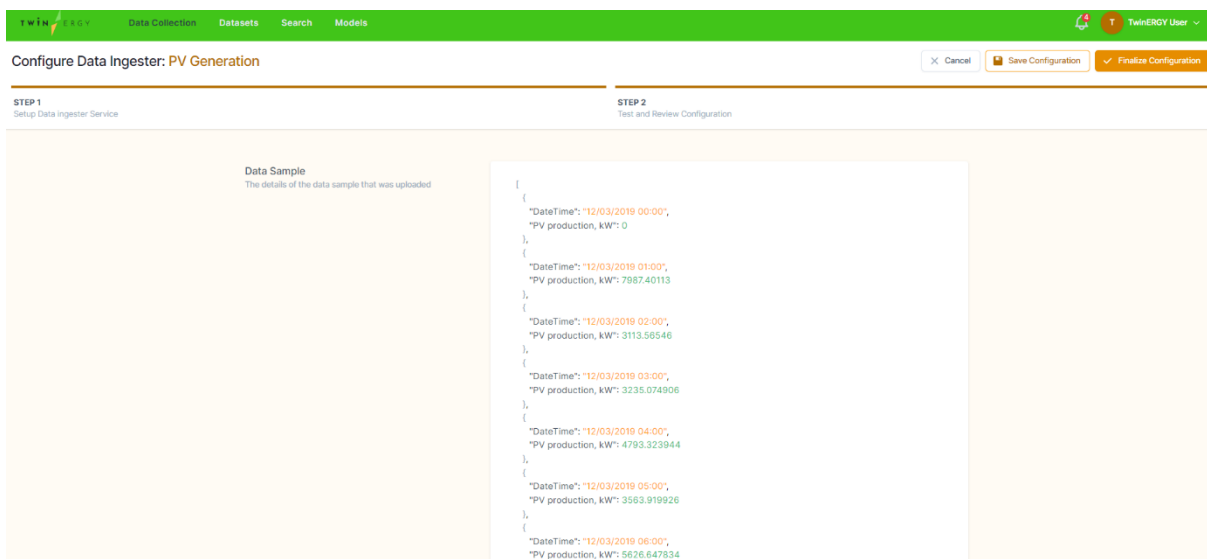


Figure 15 Streaming Data sample summary

3.1.2 Data Mapper

After finalizing the Data Ingestor settings, users can click on the Configure Data Mapper button to initiate the mapping process of the ingested data to the TwinERGY Common Information Model (CIM), as described in D5.1[3]. The platform presents the model's name and the standards used in developing the TwinERGY CIM. Users with knowledge of the specific standard to which their ingested data adheres can select it, and a list of TwinERGY CIM concepts is provided for selecting the entry point of the mapping process.

Then, users of the TwinERGY CDMP are directed to the mapping playground, where they can map the ingested data with the concepts and fields of the TwinERGY CIM, as shown in Figure 16.

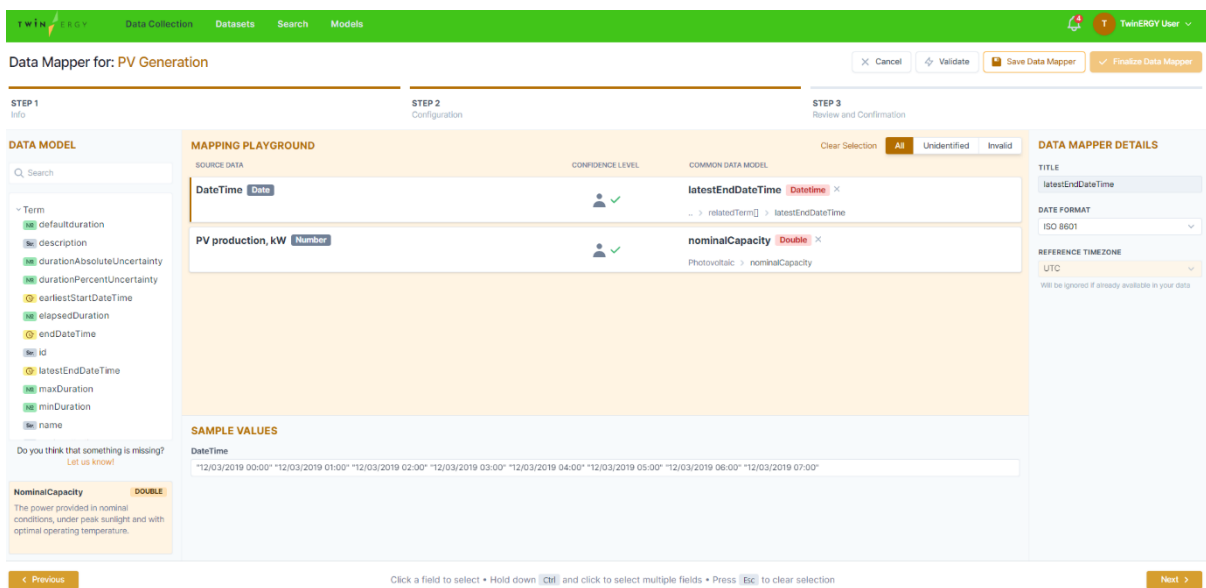


Figure 16 Mapping data to the TwinERGY CIM

Within the TwinERGY CDMP, users have the ability to search and explore the CIM to access the list of fields associated with the selected concept from the previous step. This list is displayed on the left side of the mapping playground page. In the center of the page, users can view the ingested data, while on the right side, they can specify the mapping details for each selected field.

The mapping playground provides various supportive features to simplify the mapping process. This includes mapping recommendations, confidence levels, search functionality, and drag-and-drop capabilities. Additionally, field descriptions and a data sample display are available to assist users. A validation option is provided to verify the accuracy of the mapping. If adjustments are needed, users can make them to ensure an effective mapping process.

Once the mapping is completed, users have the option to Save and Finish the process, and proceed to the Data Curator routine if it was selected as part of the configuration.

3.1.3 Data Curator

After the mapping process is complete, the users are redirected to the Data Curator configuration page where they can apply curation routines to the ingested data.

For each of the data fields, users have the option to define one or more curation routines and related outliers' criteria, as shown in Figure 17.

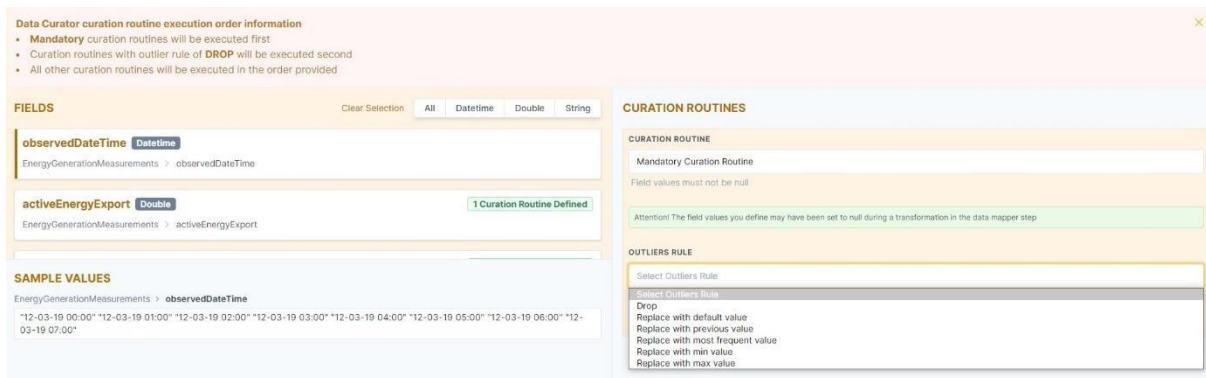


Figure 17 Curation Routines definition

After configuring the Data Curator in the TwinERGY CDM, users can review the defined curation routines for each data field and finalize the curation process. Once the Data Curator configuration is complete, users have the option to Save and Finish the process and then proceed to the Data Anonymization Handler routine, if it was included in their Data Ingestion Job configuration.

3.1.5 Data Anonymization Handler

Upon reaching this step, users are directed to the data anonymization handler configuration page, where they can apply anonymization rules to the ingested data, as depicted in Figure 18.

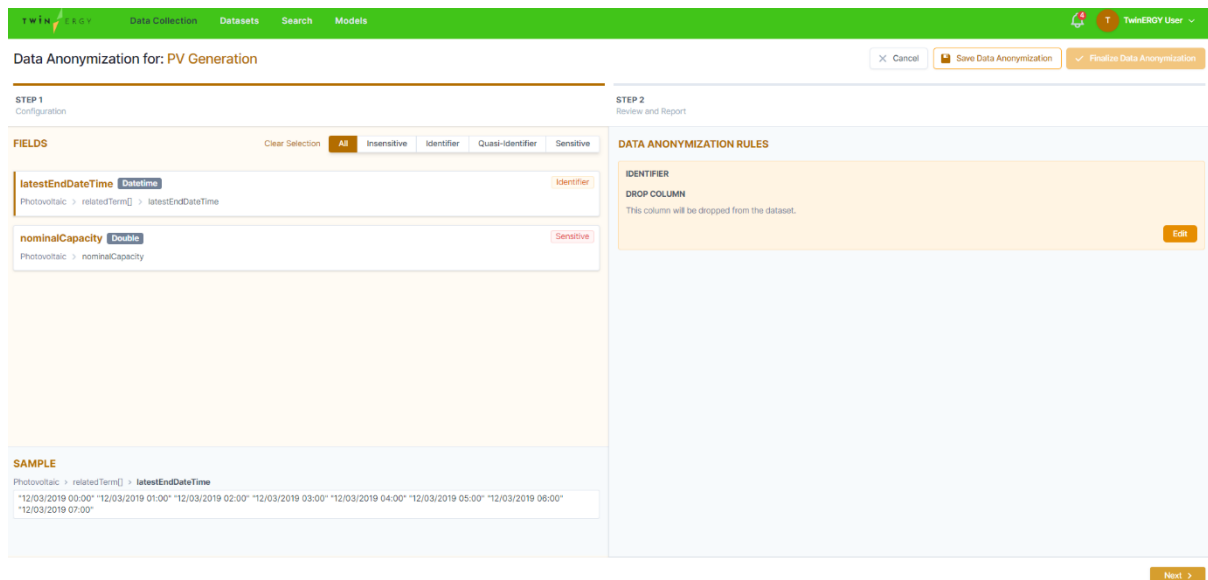


Figure 18 Data Anonymization Handler Configuration

Before making the data available on the TwinERGY Core Data Management platform, anonymization actions are performed on specific parts of the data, such as fields that may contain personally identifiable information. Users of the TwinERGY CDM have the option to anonymize their data and specify which fields they consider to be potentially containing sensitive information. The configuration allows users to define anonymization rules for "identifying," "quasi-identifying," and "sensitive" columns in the

data. If a user selects "identifier", as shown in Figure 19, as the anonymization type for a specific column, that column will be removed from the dataset.

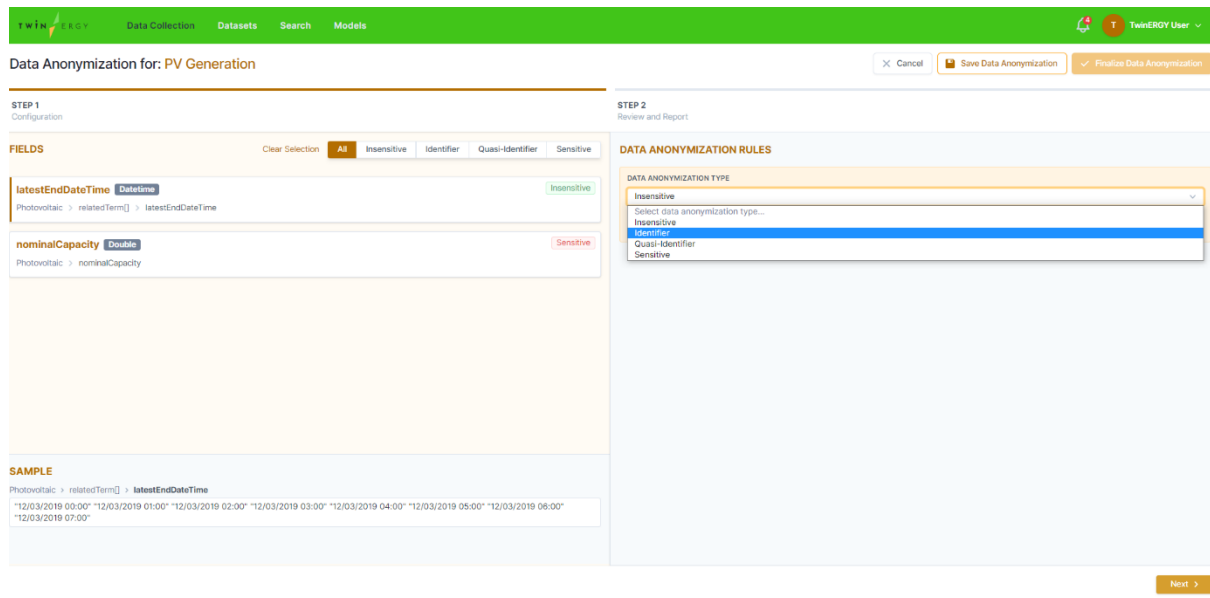


Figure 19 Data Anonymization type selection

The selection of quasi-identifiers, fields that may become identifiers when combined with other fields, is handled using generalization methods based on their specified type, ensuring the protection of private information, as shown in Figure 20.

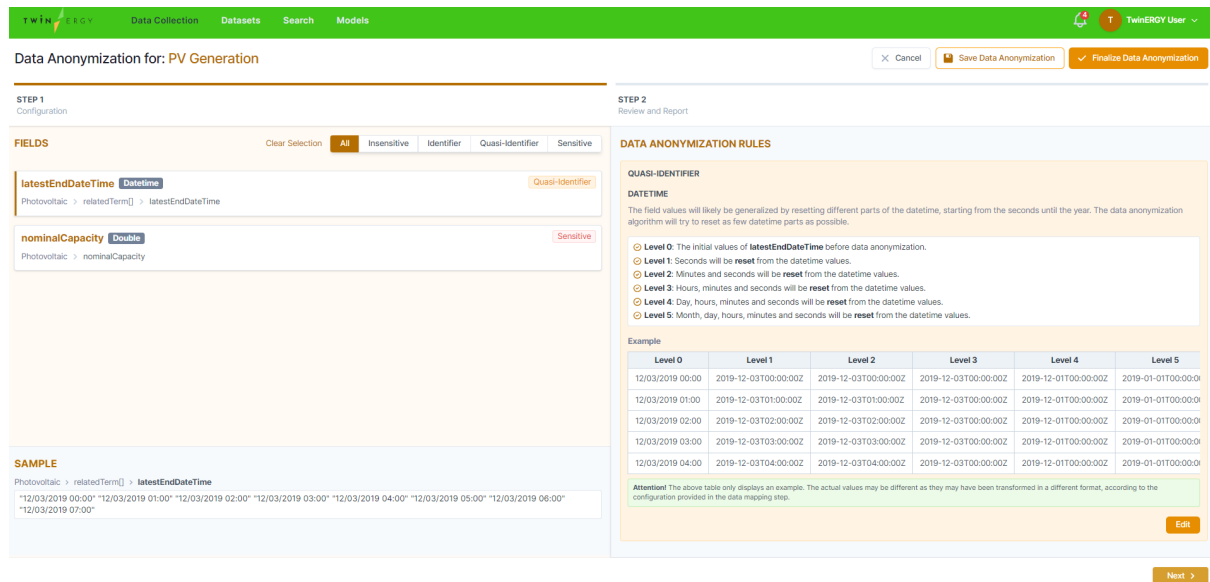


Figure 20 Handling of quasi-identifier data fields

Once the data anonymization handler configuration is finished, users can review the defined anonymization rules for each field of the dataset, as shown in Figure 21. They also have the option to set the acceptable information loss threshold and specify the parameters for the selected anonymization algorithm. This ensures that the

anonymization process is aligned with users' desired level of privacy protection while minimizing the loss of useful information.

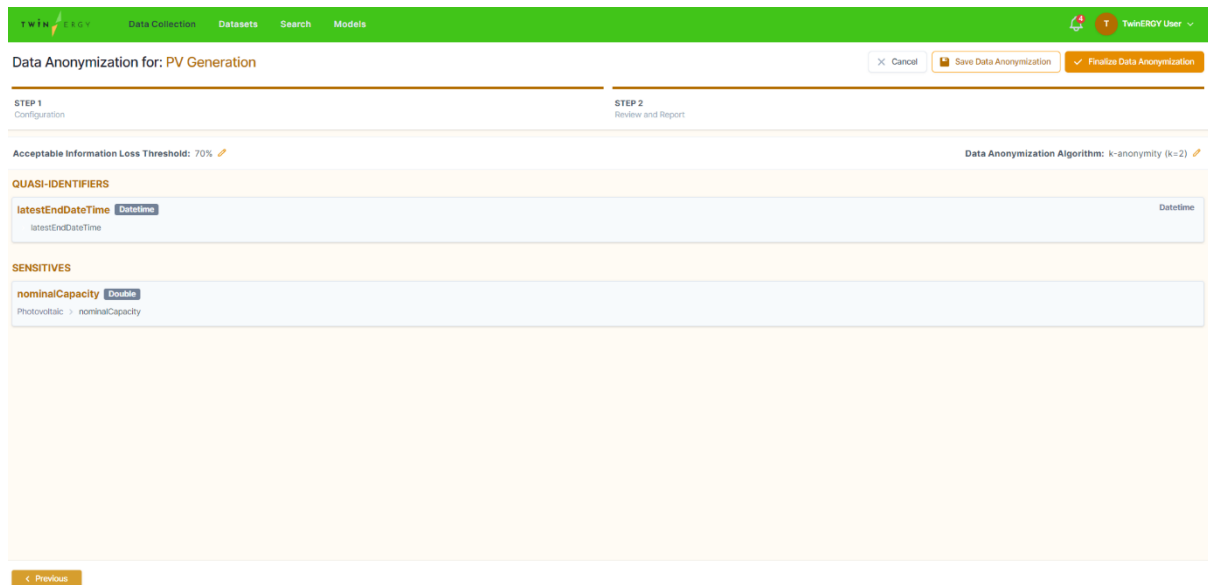


Figure 21 Data anonymization overview

3.1.6 Metadata Editor

When creating a dataset in the TwinERGY Core Data Management Platform, users are required to provide the dataset's full profile. By navigating to the Datasets tab and selecting it from the top navigation bar, users can enter a title and description for the dataset, specifying its contents, as depicted in Figure 22. Additionally, they have the option to define relevant tags to further categorize the dataset as part of completing the dataset profile.

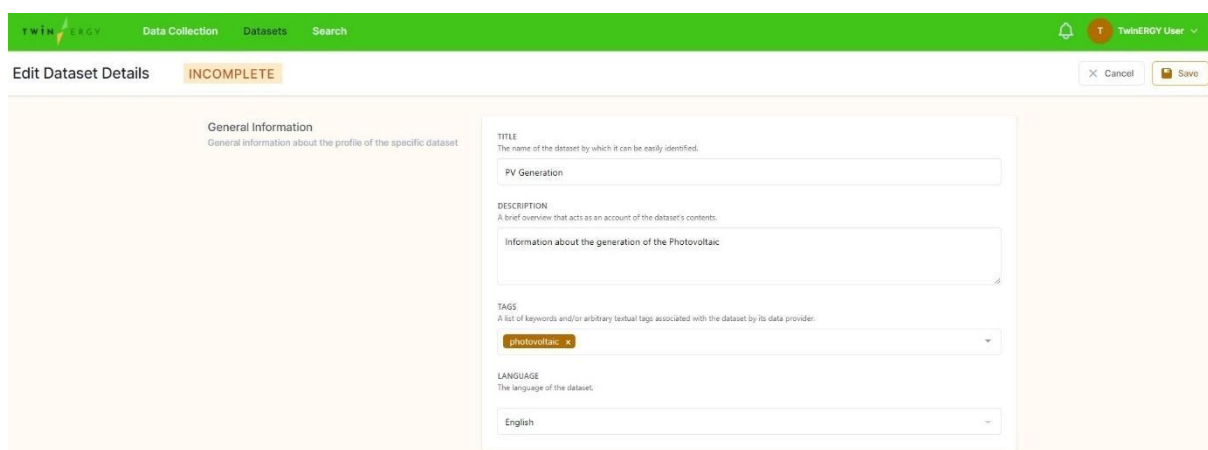


Figure 22 Dataset's Metadata Editing

The dataset's coverage and granularity, along with specific details about the data, are essential for providing comprehensive information. This includes identifying the type and format of the dataset, the language used, and the temporal and spatial coverage.

Additionally, specifying the temporal and spatial resolution units helps accurately characterize the dataset's temporal and spatial aspects, as depicted in Figure 23.

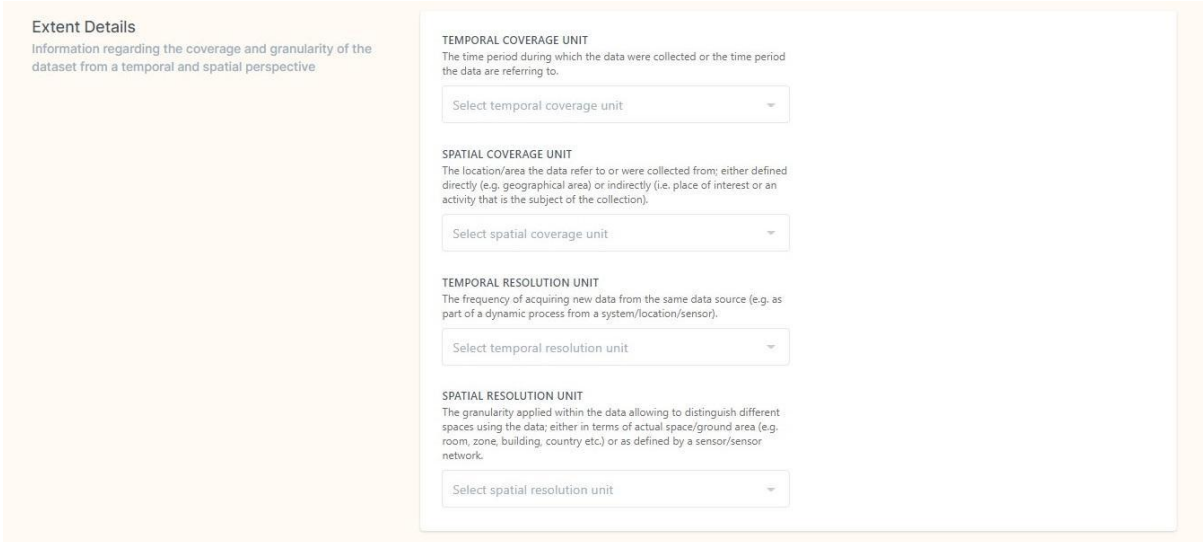


Figure 23 Provision of information on the granularity and coverage of the dataset

Users have the capability to apply specific access policies to individual datasets during the metadata updating process. This functionality, described in chapter 5. Data Security Service, falls under the Access Policies Controller section.

4. Data Storage Service

The Data Storage Service in the TwinERGY platform provides robust controls and diverse indexing methods to ensure reliable data storage. It efficiently handles the storage of a wide range of data types and their associated metadata. Different storage and indexing capabilities are available to accommodate various data storage requirements. The service also collects and maintains appropriate metadata that can be accessed by all modules. Furthermore, it logs relevant information for the smooth operation and administration of the TwinERGY platform, including user data and administrative details necessary to support the functionality of the CDMP.

4.1 Data Store

Upon completing a successful data ingestion job, the final stage is the Data Store, where TwinERGY CDMP users are prompted to provide a title and a brief description for the dataset, as illustrated in Figure 24. This information is required before the data can be saved and stored in the platform.

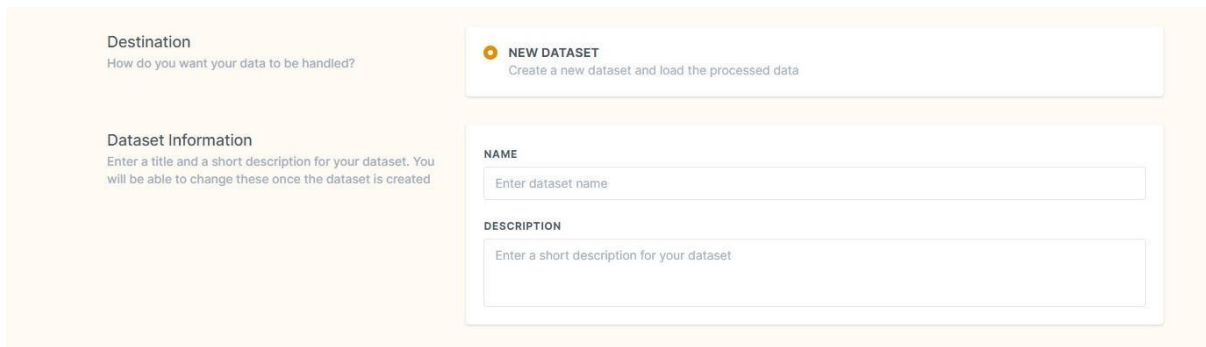


Figure 24 Data Store Configuration

Once all the required information about the dataset has been provided for storage in the TwinERGY Core Data Management Platform, users will receive a notification confirming the completion of the data ingestion job. They will then be directed to the data ingestion job page, where they can access an overview of all their initiated data ingestion jobs, including their current status, as depicted in Figure 25.



Data Ingestion Jobs		+ Create
PV Generation	COMPLETED	
TwinERGY User Updated on Mar 2, 2022		
Weather API Athens	CONFIGURATION, DATA MAPPER	
TwinERGY User Updated on Mar 2, 2022		

Figure 25 Data Ingestion Jobs overview and completion status

4.2 CIM Extension

The TwinERGY Common Information Model (CIM) is a dynamic model that undergoes continuous updates and expansion in response to project requirements, as outlined in Deliverable D5.1 - TwinERGY Common Information Model [3]. Users have the ability to propose changes or additions to the CIM, as illustrated in Figure 26. These suggestions are carefully evaluated and, if deemed appropriate, integrated into the evolving TwinERGY CIM.

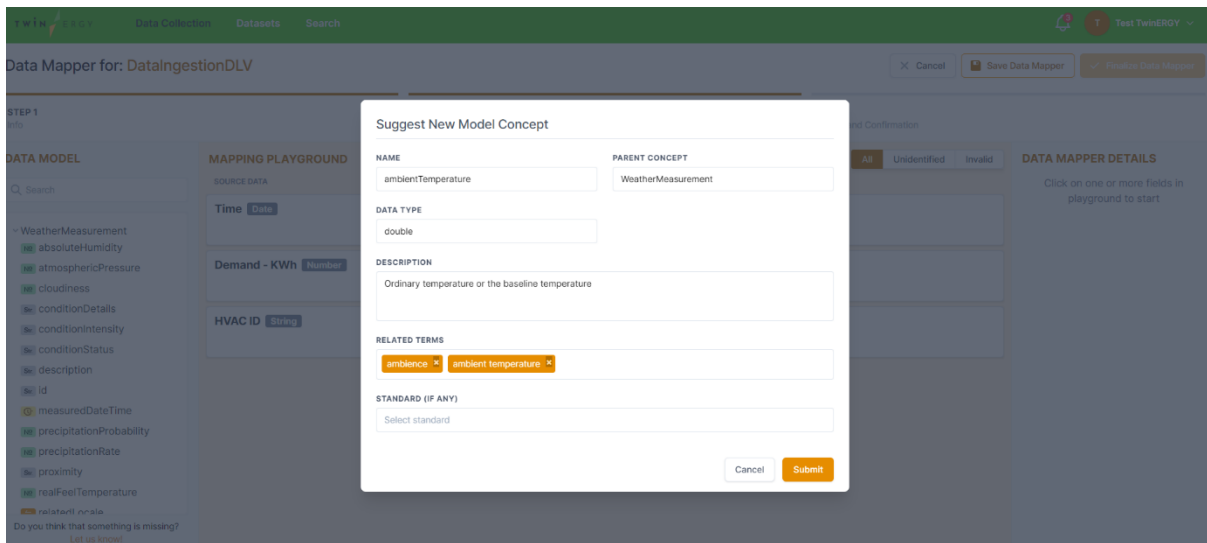


Figure 26 Suggestion of new fields

In the release 2.00 of the TwinERGY integrated platform, the Common Information Model (CIM) has been extended to ensure continuous updates and improvements, enhancing the reliability and long-term usability of the model. During the development of the previous release, module development partners encountered difficulties in mapping certain fields of their datasets to the existing fields and concepts of the TwinERGY CIM. To address these challenges, new fields were added or adjusted within specific concepts of the CIM, fulfilling the mapping requirements of the partners. Table 1 depicts the newly created fields (together with a brief description of them) since release 1.00, along with the corresponding TwinERGY CIM concept they can be found under:

Fields added to TwinERGY CIM	Field Description	TwinERGY Concept(s) found under
airTemperature	The air temperature within the indoor environment captured by the wrist worn wearable device	SensingMeasurement
globeTemperature	The globe temperature within the indoor environment captured by the wrist worn wearable device	SensingMeasurement
indoorHumidity	The relative humidity within the indoor environment captured by the wrist worn wearable device	SensingMeasurement

airVelocity	The air speed within the indoor environment captured by the wrist worn wearable device	SensingMeasurement
metabolicRate	The metabolic rate / activity level of the consumer as captured by the wrist worn wearable device	SensingMeasurement
clothingInsulation	An estimation for the level of thermal insulation provided from the garments worn by the consumer	Wellbeing
thermalComfort	The index of thermal comfort over the Fanger's thermal comfort scale	Wellbeing
thermalStatus	The corresponding literal explanation of the thermal comfort index (e.g., Neural, Slightly Cool, Warm)	Wellbeing

Table 1 Additions of fields in the TwinERGY CIM

5. Data Security Service

The Data Security Service plays a crucial role in addressing the data security and privacy concerns of users in the TwinERGY CDMP. Its primary objective is to build trust by providing robust mechanisms to protect and control access to the data stored and managed in the TwinERGY Core Data Management Platform.

One of the key features of the Data Security Service is the ability for users to design and implement access policies for their datasets. These access policies allow users to define specific rules and restrictions on who can access their data and how it can be accessed. The Access Policies Controller provides a graphical user interface that facilitates the creation of access policy routines, enabling users to easily configure complex access control mechanisms.

These access policy routines are stored within the platform, ensuring that the defined rules are consistently applied to the datasets. Users can also make modifications to these routines using the provided interface, allowing them to adapt the access policies as needed over time. This flexibility and control empower users to tailor the access permissions to their specific requirements and maintain the desired level of data security.

5.1 Access Policies Controller

Accordingly, the users of the platform can set the visibility levels and access policy routines for the respective datasets. They can also determine whether the dataset is for Private or General access, as displayed in Figure 27. The General access option establishes that no requirement for access policy routines to be completed is needed, thus the access to the specific dataset is free. On the contrary, the Private access option assures that the access to a specific dataset is only authorized if access policy routines are followed.

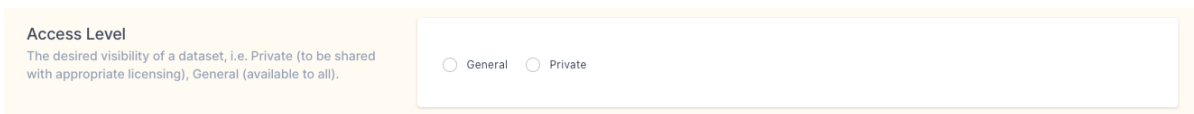


Figure 27 Access Level Definition

If the access level is set to Private, the required access policies, the applicable approach (Allow-all), and the inclusion of specific exceptions must be established, as shown in Figure 28. In order to add exceptions, the CDMP user need to define a user parameter, a condition (such as equal, not equal, etc.), and the parameter's value.

Access Policy
Define who has and doesn't have access to this dataset

Allow everyone to view this dataset

+ Add an exception

Field...

+ AND (add another condition)

You need to define at least one condition for policy to be added

Figure 28 Access Policies Definition

6. Platform Management Service

The Platform Management Service is crucial in TwinERGY Core Data Management Platform since it is responsible for the secure, reliable and unhindered registration of the users on the platform. Users are granted access to data that they are authorized to use via suitable authentication and security mechanisms. The Identity Manager functionality represents identity provider duties, such as establishing and managing identifying information for individuals permitted to access the TwinERGY CDMP, as well as providing authentication and authorization services to limit access to authorized users, as described in Section 2 (TwinERGY Core Data Management Platform) of this document.

Users of the platform can now search for data, using the Search & Retrieval Manager functionality, taking advantage of a user-friendly data search that supports a variety of search options. Additionally, while creating APIs for data retrieval, approved applications can use the TwinERGY Open APIs to configure data retrieval from a single dataset. The selection of particular data fields and filters, which are represented by API request parameters, are utilized to tailor the results.

Finally, the Notifications Handler functionality gives platform users instant updates on how a data ingestion job is progressing (successful or unsuccessful). Additionally, they can examine different notifications and if required, delete them.

6.1 Search & Retrieval Manager

The functionality is essential for TwinERGY's Core Data Management Platform because it enables users to define which of the available data is essential, find and define relevant data, and look into clear and comprehensive summaries of the results.

Users of the TwinERGY CDMP can view a list of the datasets they have by navigating to and selecting the relevant tab in the platform's top navigation bar. They are provided with the capabilities they need to filter the datasets based on the different data types. It's important to note that platform users have the option to select datasets that they or other users have published, as illustrated in Figure 29.

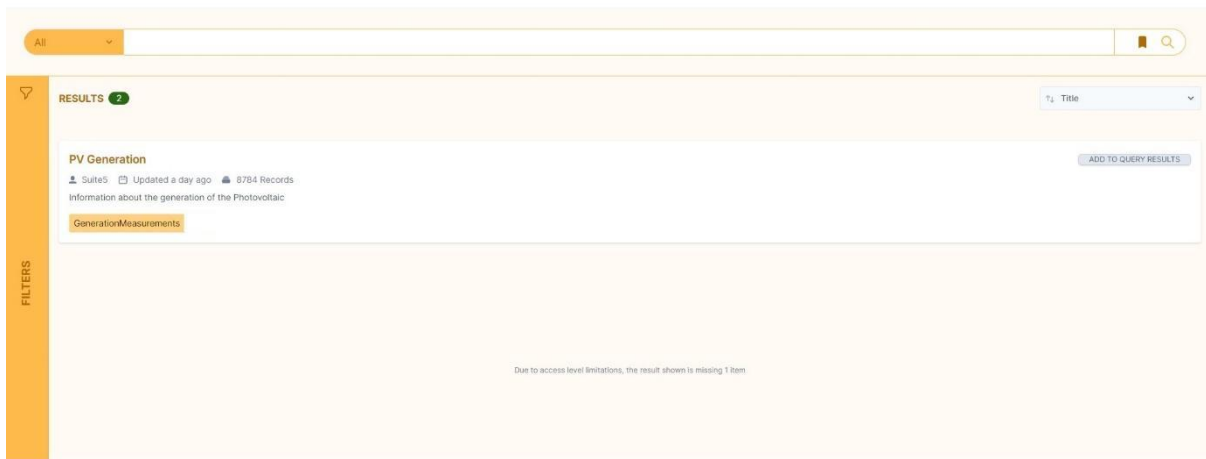


Figure 29 Retrieval of datasets

Users of the platform have more options for finding and searching for data within the TwinERGY Core Data Management Platform by utilizing the platform's powerful free-text search and data filtering capabilities. They can also search for data that is pertinent to their needs, and review the findings, to determine eligible candidates for retrieval. The options for retrieval must be defined through APIs, with the choice of the concepts that a user wishes to acquire from each dataset. The definition of concepts used as query parameters to filter query results is also necessary. Instructions on how to use the TwinERGY APIs to obtain the retrieval results are provided at the conclusion of the retrieval configuration. The endpoints are accompanied by authentication instructions for that purpose (i.e., for GET and POST methods).

6.2 Notifications Handler

The TwinERGY Core Data Management Platform provides users with appropriate updates regarding the status and progress of their data ingestion jobs, as depicted in Figure 30.

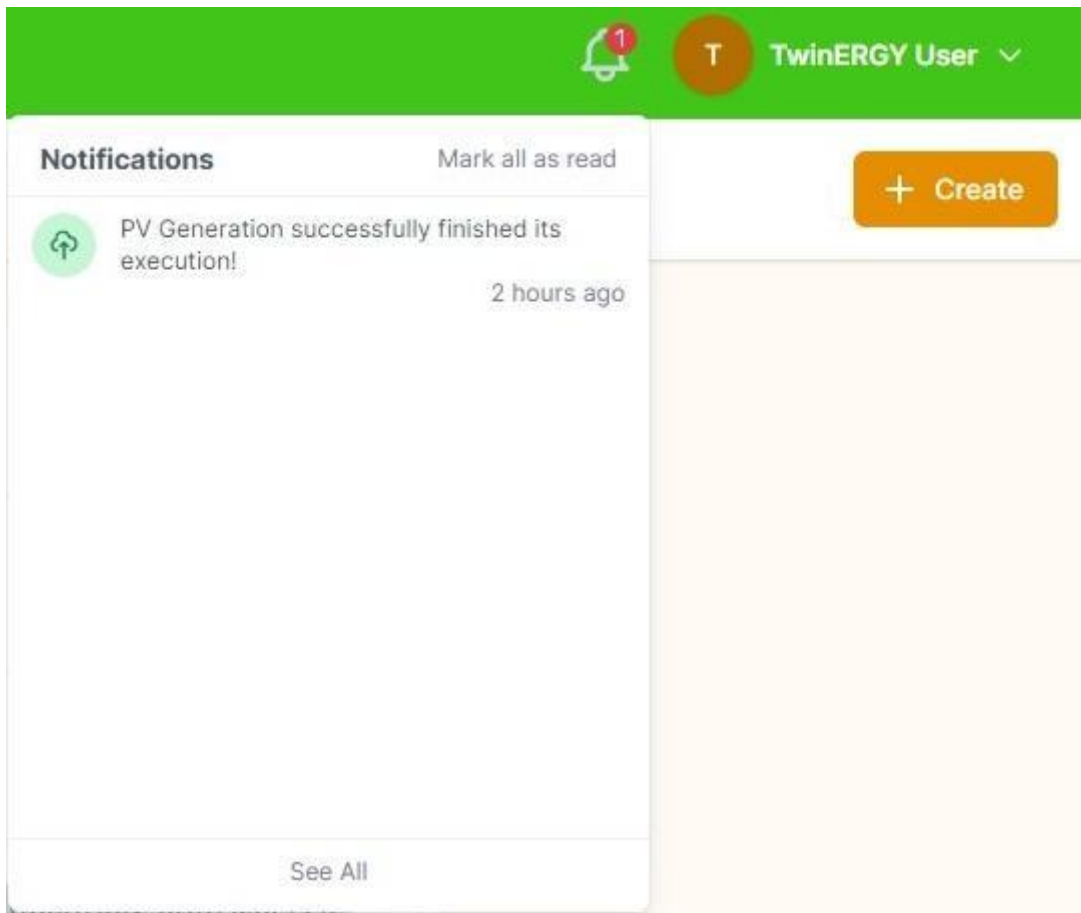


Figure 30 Notifications on ongoing/finished data ingestion jobs

7. Conclusions

Deliverable 5.8 marked the final release of the TwinERGY Integrated Platform, version 2.00. This release encompassed a range of services, including Data Collection, Data Security, Data Storage, and Platform Management. Building upon the previous version, D5.6 - "TwinERGY Integrated Data Management Platform – Release 1.00", and D5.7 - "Data Collection, Security, Storage & Management Services Bundles – Release 2.00", this deliverable provided comprehensive insights into the services' functionalities and the updates and enhancements introduced during the transition from release 1.00 to release 2.00 of the CDMP. In particular, this deliverable elaborated on:

- The significant updates made in the data ingestion process, introducing new features, notably with the addition of the streaming Pub/sub mechanism. This expanded the data ingestion options, allowing users to stream their data to the platform in addition to the existing file upload and API retrieval methods.
- The Data Storage service and the improvements made in the CIM Store functionality. These enhancements enabled the extension of the TwinERGY CIM by incorporating additional fields, thereby enhancing its capabilities beyond the previous version, presented in D5.6 - "TwinERGY Integrated Data Management Platform – Release 1.00".
- The vital role of the Data Security Service in ensuring the security and privacy of the data ingested by platform users. It focuses on addressing users' concerns and implementing measures to protect the confidentiality and integrity of their data. For that purpose, the functionality of the Data Anonymization Handler was introduced in the TwinERGY Integrated Platform, release 2.00 and elaborated in this deliverable.

The deliverable provided a comprehensive overview of these services, their functionalities, and the improvements made in the transition from release 1.00 to release 2.00. It demonstrated the platform's commitment to continuous enhancement and refinement to meet evolving user needs and ensure a robust and efficient data management solution.

REFERENCES

1. TwinERGY Consortium. (2020). TwinERGY D5.6 "TwinERGY Integrated Data Management Platform – Release 1.00"
2. TwinERGY Consortium. (2020). TwinERGY D5.7 "Data Collection, Security, Storage & Management Services Bundles – Release 2.00"
3. TwinERGY Consortium. (2020). TwinERGY D5.1 "TwinERGY Common Information Model"
4. TwinERGY Consortium. (2020). TwinERGY D5.5 "Data Collection, Security, Storage & Management Services Bundles – Release 1.00"

