



TwinERGY Integrated Data Management Platform-Release 1.00

D5.6

October 2022

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

D5.6 TwinERGY Integrated Data Management Platform – Release 1.00

Revision: v1.0

AUTHORS

Anastasios Tsitsanis	Marios Phinikettos
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
v0.1	26/10/2022	Anastasios Tsitsanis Marios Phinikettos	Suite5	1 st draft of the deliverable D5.6 - TwinERGY Integrated Data Management Platform-Release 1.00
v0.2	31/10/2022	Moisés Antón García Ana Isabel Martínez García	ETRA	Peer review of the document. Minor comments/Corrections
v0.3	31/10/2022	Fynn Christian Bollhöfer Axel Balke	TH-OWL	Peer review of the document. Minor corrections
v0.4	31/10/2022	Apostolos Kapetanios Konstantinos Kotsalos	ED	Peer review of the document. Minor corrections
v0.5	31/10/2022	Anastasios Tsitsanis Marios Phinikettos	Suite5	Comments/corrections made during the review process are addressed.
v0.6	31/10/2022	Anastasios Tsitsanis Marios Phinikettos	Suite5	Draft sent to PC for submission
v1.0	31/10/2022	Athanasios Chassiakos, Stylianos Karatzas, Vasiliki Lazari, Anthony Papamanolis	UoP	Draft submitted to EC by the PC

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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

© 2022 by TwinERGY Consortium

Executive Summary

Deliverable 5.6 presents the TwinERGY Integrated Platform Release 1.00 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 Beta Release.

Deliverable D5.6 goes into further detail about the TwinERGY CDMP Services that were covered in Deliverable D5.4 [1]. Updates to certain of these services have been incorporated into this deliverable, including Data Ingestor improvements to address the updating of data by appending new rows in existing datasets in the TwinERGY Platform, performance improvements for the more timely upload of data as well as the TwinERGY CIM extension with additional fields to ensure that both the pilot's available data and the module's data needs are captured successfully from the model.

Additional enhancements and optimizations will be considered as the platform development activities continue and will be introduced in the TwinERGY Integrated Platform - Release 2.00 on M32.

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
3.1.1.1 Update of data – appending of new rows	15
3.1.1.2 Efficient file upload	15
3.1.2 Data Mapper	16
3.1.3 Data Curator	17
3.1.4 Metadata Editor	17
4. Data Storage Service	19
4.1 Data Store	19
4.2 CIM Extension.....	19
5. Data Security Service	21
5.1 Access Policies Controller	21
6. Platform Management Service	22
6.1 Search & Retrieval Manager	22
6.2 Notifications Handler	23
7. Conclusions	25
REFERENCES	26
ANNEXES	26

List of Figures

Figure 1 TwinERGY Core Data Management Platform Conceptual Architecture	10
Figure 2 TwinERGY CDMP Landing Page	11
Figure 3 User registration	11
Figure 4 Data Ingestion Job Creation	13
Figure 5 Display of the data ingestion jobs page	13
Figure 6 File uploading in the TwinERGY CDMP	13
Figure 7 Overview of the sampling of the files uploaded	14
Figure 8 API Data Ingestion.....	14
Figure 9 Process for appending data	15
Figure 10 Appending data in already stored datasets.....	15
Figure 11 Uploading bigger data files in a timely manner	16
Figure 12 Mapping data to the TwinERGY CIM.....	16
Figure 13 Curation Routines definition.....	17
Figure 14 Dataset's Metadata Editing	18
Figure 15 Provision of information on the granularity and coverage of the dataset	18
Figure 16 Data Store Configuration	19
Figure 17 Data Ingestion Jobs overview and completion status.....	19
Figure 18 Suggestion of new fields	20
Figure 19 Proposed fields be added in the TwinERGY CIM.....	20
Figure 20 Access Level Definition	21
Figure 21 Access Policies Definition	21
Figure 22 Retrieval of datasets	23
Figure 23 Notifications on ongoing/finished data ingestion jobs.....	24

1. Introduction

1.1 Purpose of this deliverable

The TwinERGY Integrated Platform's release 1.00 is presented in deliverable 5.6, together with the platform's services, particularly the Data Collection, Data Security, Data Storage, and Platform Management services. This deliverable is an updated version of D5.4, "TwinERGY Integrated Platform - Beta Release," and it details all services and the functionalities connected with them, as well as the updates and enhancements made from the Beta Release to the release 1.00 of the CDMP. Intuitive and comprehensive screenshots are utilized to demonstrate the process for establishing and using the core platform functionalities.

1.2 Scope of this deliverable

The newest functionalities made accessible to its intended customers are described in Deliverable D5.6, "TwinERGY Integrated Platform - Release 1.00," which also provides information on the implementation status of the TwinERGY Core Data Management Platform. This release 1.00 of the platform includes the most crucial user routines for carrying out basic platform processes, numerous functionalities incorporated, as well as the improvements made since the platform's previous version. An end-to-end usage overview is given with the goal of achieving it, along with detailed instructions and pertinent screenshots.

1.3 Structure of the document

Deliverable D5.6 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[2]. 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.

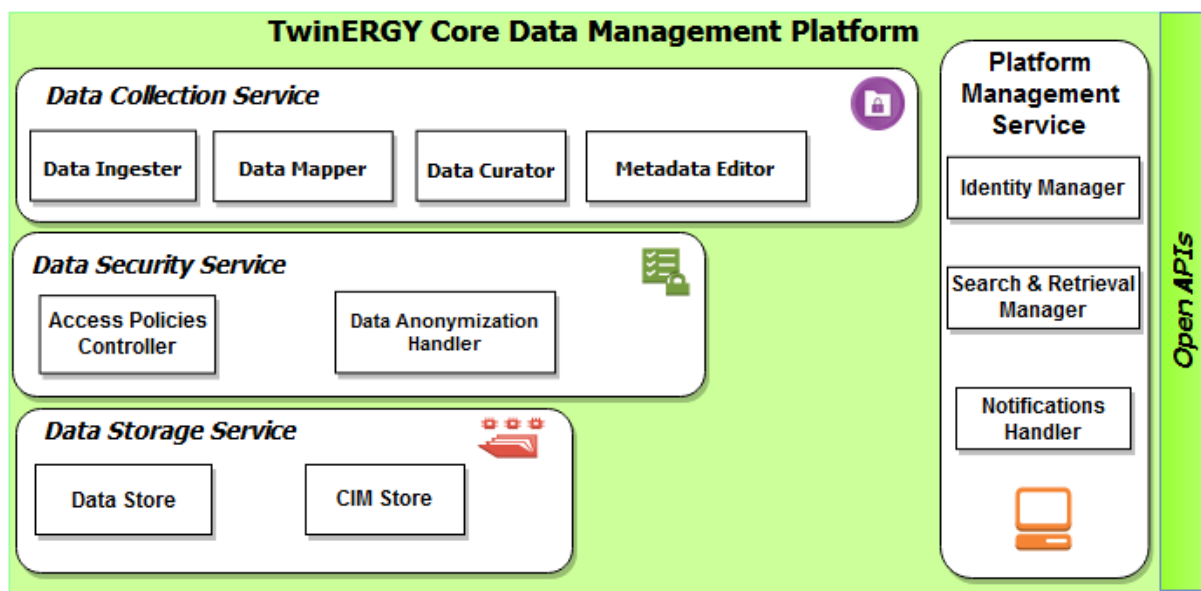


Figure 1 TwinERGY Core Data Management Platform Conceptual Architecture

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 included 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.

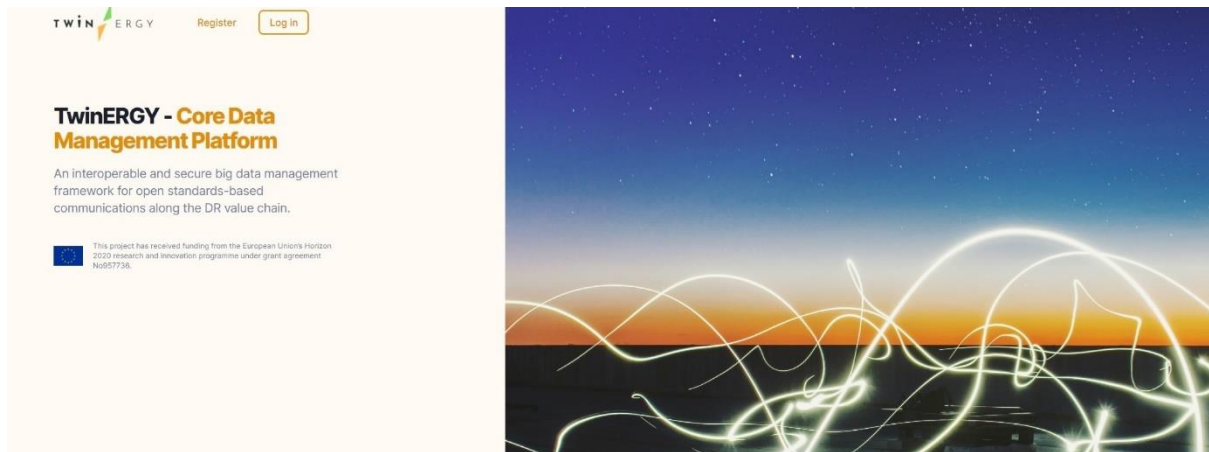


Figure 2 TwinERGY CDMP Landing Page

Users of the TwinERGY CDMP can commence the registration process by providing their first and last names, a username, an e-mail address, and a password that must meet specific requirements and be submitted twice, as shown in Figure 3.

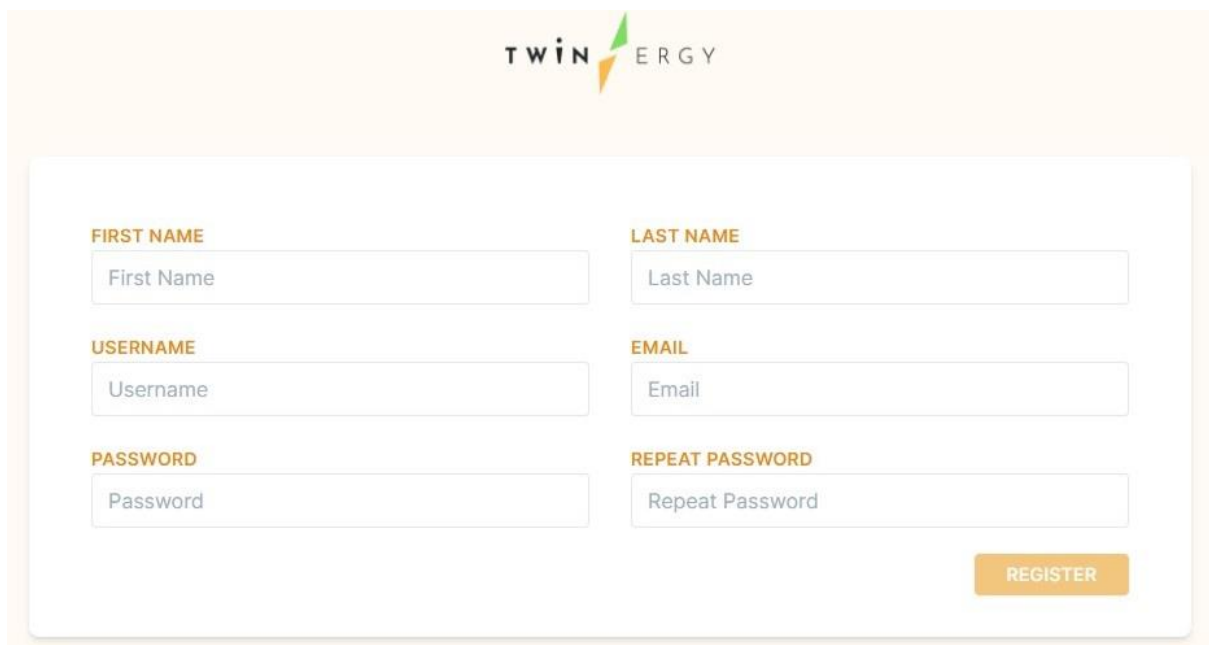


Figure 3 User registration

After completing the registration process, users are routed to the login page, where they can submit their credentials and log in to the TwinERGY Core Data Management Platform. Both actions (registration and login) are part of the Platform Management Service and serve the purpose of 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. 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

When users initially access the Core Data Management Platform, they are presented with the Data Collection page. The data ingestion job is created after selecting Create from the top navigation bar. Additionally, platform users must choose appropriate processing routines and provide data ingestion job parameters.

The procedure starts by creating a fresh data ingestion job and specifying the proper data processing routines. The user is prompted to submit some basic details regarding the data ingestion job, including the title, a brief description of the dataset to be ingested, and the data processing routines that will be performed. Although the data ingester, data mapper, and data store data processing routines are pre-selected, the platform users can decide whether or not to curate the dataset since the data curator routine is 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 or through API. 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.

ADDED FILES		
D PV_Generation.csv		255.3 KB ×
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

If platform users keep data in their organization systems and expose it through APIs, they can also use the TwinERGY platform's API retrieval functionality. To start the process of data ingestion, an appropriate configuration screen is presented to them, as illustrated in Figure 8. Following the selection of the authentication type, the platform user is prompted to submit the entire API URL, as well as the method to be used (e.g., GET), in conjunction with any potential request parameters.

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

Authentication Details
Details about the authentication policies of the API

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

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

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

None
 Bearer
 Custom

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

PARAMETER	VALUE	TYPE	SENSITIVE
⊙ q	athens	Query	✎ ✖
⊙ appid	290eef18c9bea92b70b0e04f4c...	Query	✎ ✖

+ ADD QUERY PARAMETER

Figure 8 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.

3.1.1.1 Update of data – appending of new rows

This functionality is part of the release 1.00, extending the functionalities described in the beta release. When the data ingestion method is the File Upload, the users of the platform can choose "Append Data" from the options at the right side of each data ingestion job in the Data Ingestion Jobs overview, as depicted in Figure 9.

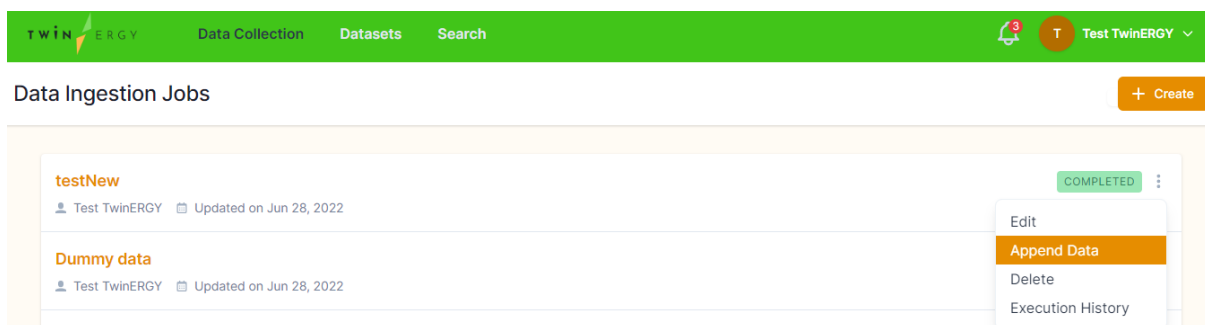


Figure 9 Process for appending data

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 10.

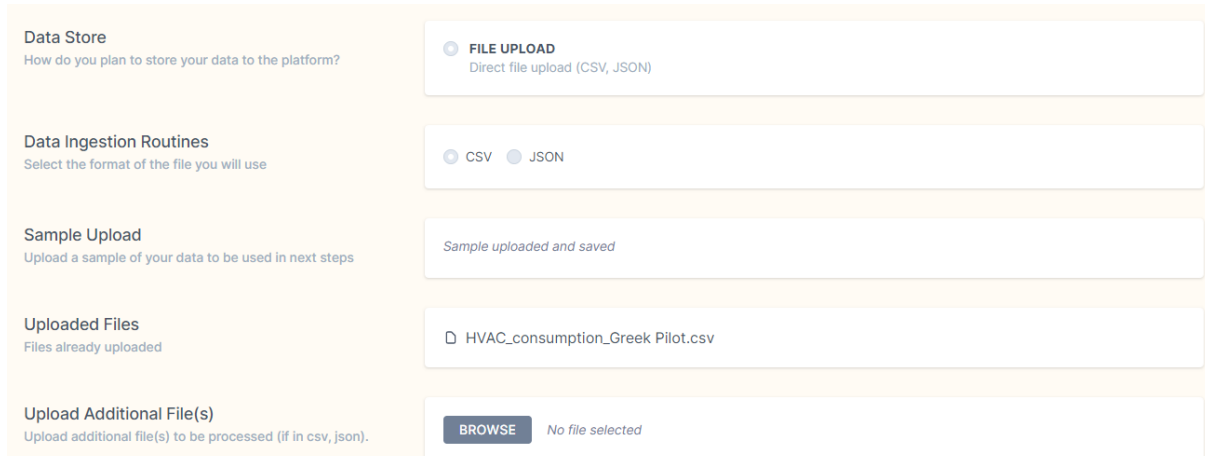
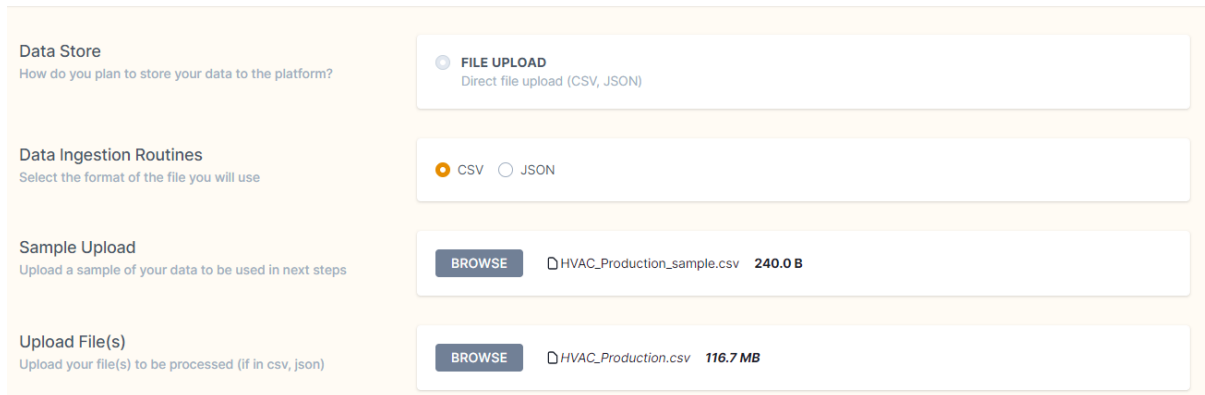


Figure 10 Appending data in already stored datasets

3.1.1.2 Efficient file upload

These enhancements are part of release 1.00 and build upon the beta release's functionalities. All functionalities of the TwinERGY Core Data Management Platform beta release were carefully examined and updated as necessary based on integration testing in order to be prepared for cases that users will attempt to upload files with considerable size or structure complexity. In the TwinERGY CDMP, larger data volumes can be ingested in a single dataset (with a threshold at approx. 500MB). Additionally, the various performance improvements that have been performed enable users to ingest files with

complex/nested structure (in JSON format), as shown in Figure 11, in acceptable timeframes (~5 minutes).



The screenshot shows a user interface for uploading data. It is divided into four sections:

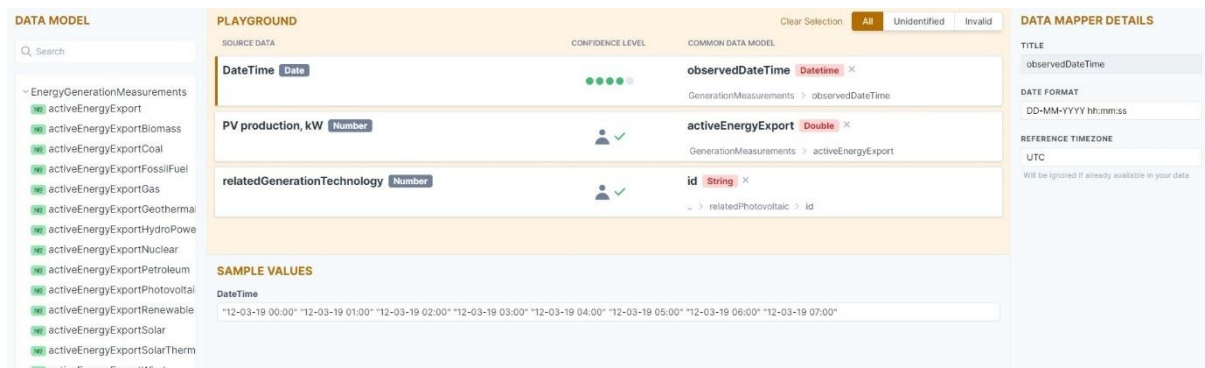
- Data Store:** A question "How do you plan to store your data to the platform?" with a radio button selected for "FILE UPLOAD" (Direct file upload (CSV, JSON)).
- Data Ingestion Routines:** A question "Select the format of the file you will use" with radio buttons for "CSV" (selected) and "JSON".
- Sample Upload:** A question "Upload a sample of your data to be used in next steps" with a "BROWSE" button and a file selection showing "HVAC_Production_sample.csv" (240.0 B).
- Upload File(s):** A question "Upload your file(s) to be processed (if in csv, json)" with a "BROWSE" button and a file selection showing "HVAC_Production.csv" (116.7 MB).

Figure 11 Uploading bigger data files in a timely manner

3.1.2 Data Mapper

By clicking on the Configure Data Mapper button, after completing the Data Ingestor settings, the process of mapping the ingested data to the TwinERGY CIM defined concepts, as described in D5.1[2], begins. The platform initially displays the model's name, the standards considered in the development of the TwinERGY Common Information Model, which can be selected if the user is knowledgeable of the corresponding standard to which the ingested data adheres, and a list of the TwinERGY CIM concepts defined to select the mapping process's entry point.

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 12.



The screenshot shows the "Data Mapper" interface. It is divided into four main sections:

- DATA MODEL:** A search bar and a list of data models including "EnergyGenerationMeasurements", "activeEnergyExport", "activeEnergyExportBiomass", "activeEnergyExportCoal", "activeEnergyExportFossilFuel", "activeEnergyExportGas", "activeEnergyExportGeothermal", "activeEnergyExportHydroPower", "activeEnergyExportNuclear", "activeEnergyExportPetroleum", "activeEnergyExportPhotovoltaic", "activeEnergyExportRenewable", "activeEnergyExportSolar", "activeEnergyExportSolarTherm", and "activeEnergyExportWind".
- PLAYGROUND:** A table for mapping source data to common data model concepts.

SOURCE DATA	CONFIDENCE LEVEL	COMMON DATA MODEL
DateTime [Date]	●●●●●	observedDateTime [Datetime] × GenerationMeasurements > observedDateTime
PV production, kW [Number]	●●●●●	activeEnergyExport [Double] × GenerationMeasurements > activeEnergyExport
relatedGenerationTechnology [Number]	●●●●●	id [String] × > relatedPhotovoltaic > id
- SAMPLE VALUES:** A section showing "DateTime" with a list of sample values: "12-03-19 00:00" "12-03-19 01:00" "12-03-19 02:00" "12-03-19 03:00" "12-03-19 04:00" "12-03-19 05:00" "12-03-19 06:00" "12-03-19 07:00".
- DATA MAPPER DETAILS:** A section showing "TITLE" (observedDateTime), "DATE FORMAT" (DD-MM-YYYY hh:mm:ss), and "REFERENCE TIMEZONE" (UTC). A note says "Will be ignored if already available in your data".

Figure 12 Mapping data to the TwinERGY CIM

The users of the TwinERGY CDMP can search through the CIM and discover the list of fields for the concept that was selected in the previous step, on the left side of the mapping playground page. The user can also see the data that has been ingested in the centre of the playground page, as well as specify the mapping details for each selected field on the right side of the page. Suitable support and mapping recommendations are

provided, together with pertinent confidence levels, search and drag-and-drop functionality, field descriptions, and a data sample display, to make the mapping process simpler. There is also a validation option that informs the user whether the mapping completed is accurate. If not, the user can always make the necessary adjustments so to complete the process in an effective manner. Users can then choose to Save and Finish the process and carry on with the Data Curator (if selected) routine.

3.1.3 Data Curator

After the mapping process is complete, the users are sent 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 13.

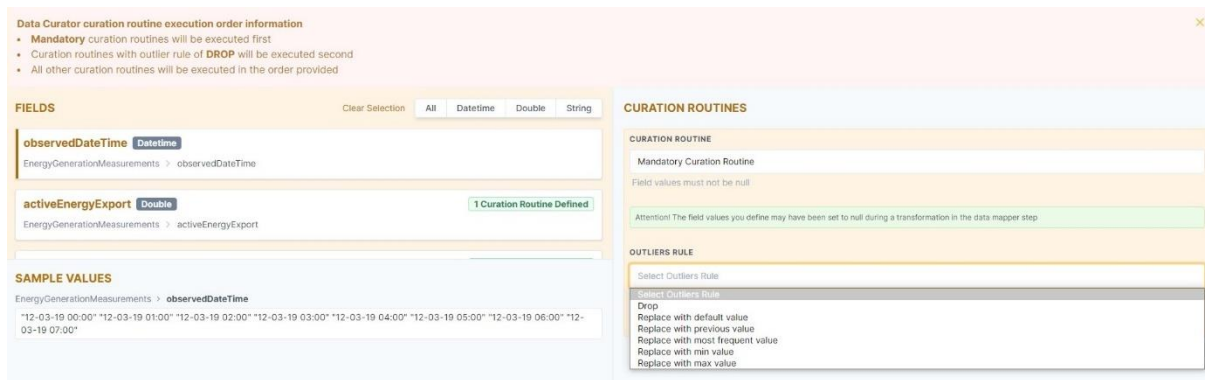


Figure 13 Curation Routines definition

Once the Data Curator configuration is complete, the TwinERGY CDMP users can examine the curation routines defined for every single data field and then finalize the curation process.

3.1.4 Metadata Editor

Any result of a data ingestion job is preserved as a dataset in the TwinERGY Core Data Management Platform; therefore, users must specify the dataset's full profile, as depicted in Figure 14. After selecting the Datasets tab from the top navigation bar, they are requested to provide a title for the dataset and a description of what it contains, as well as define any appropriate tags while finishing the dataset profile.

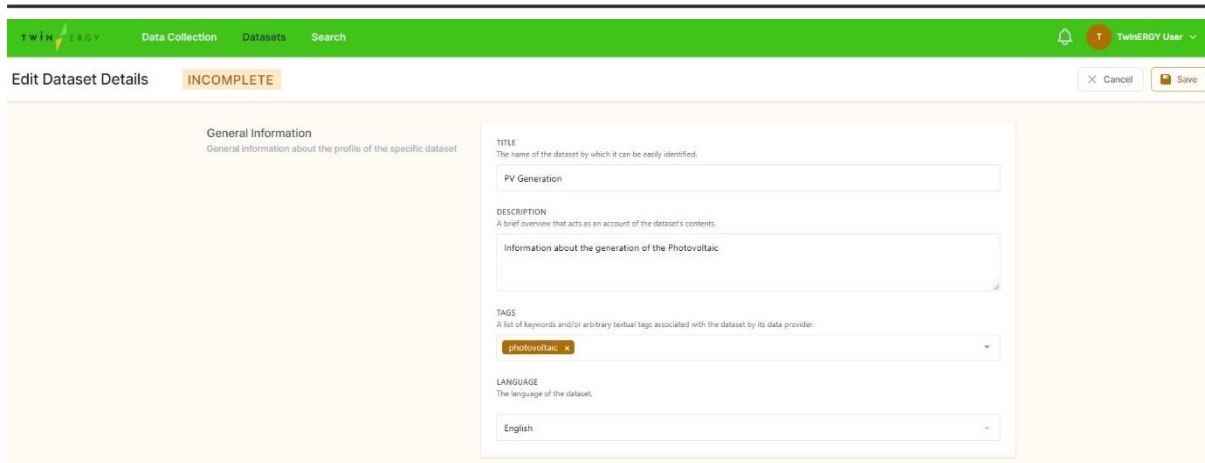


Figure 14 Dataset's Metadata Editing

Information about the dataset's coverage and granularity as well as specifics about these aspects of the data are also deemed necessary. This way, the type and the format of the dataset, the language of the dataset, and the dataset's temporal and spatial coverage as well as temporal and spatial resolution units, can be identified, as depicted in Figure 15.

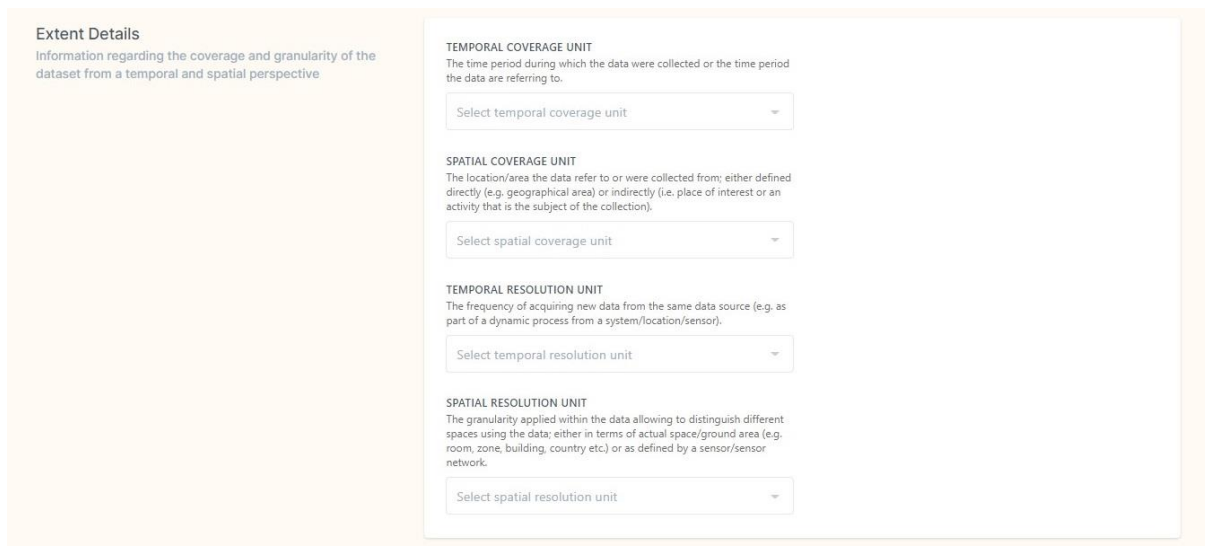


Figure 15 Provision of information on the granularity and coverage of the dataset

Specific access policies per dataset can be applied as part of the metadata updating process; this functionality is discussed in chapter 5. Data Security Service, under the Access Policies Controller section.

4. Data Storage Service

The Data Storage Service offers solid controls and a variety of indexing approaches to meet the requirements for reliable data storage and indexing. A large variety of data, as well as the associated metadata, must be stored efficiently and reliably via the Data Storage Service. Different storage and indexing capabilities are intended to satisfy different needs, depending on the type of data stored in the CDMP and how it is accessible when it comes to data storage. At the same time, appropriate metadata is gathered so that it is accessible to all modules. Additionally, this service keeps log-related information for the TwinERGY platform's use and operation, including user information and any administrative information needed to maintain the CDMP's functionality.

4.1 Data Store

A successful data ingestion job ends with the Data Store stage. TwinERGY CDMP users are requested to give the dataset a title and a concise description, as shown in Figure 16, before the data can be saved in the platform.

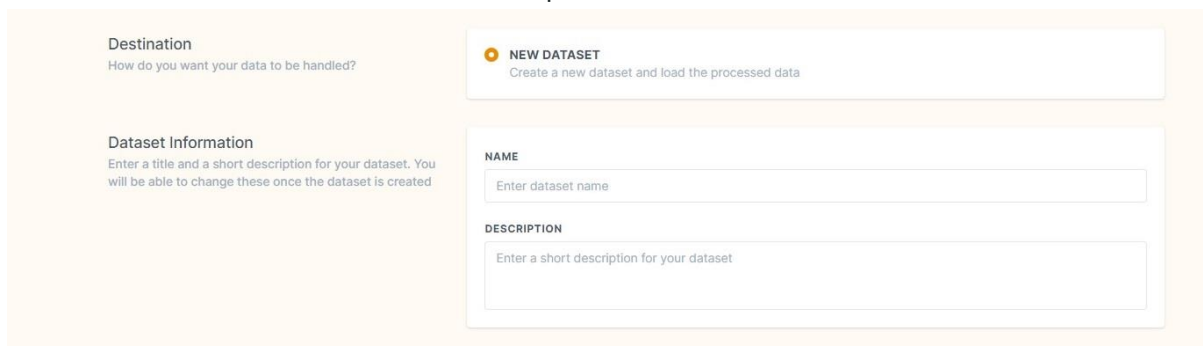


Figure 16 Data Store Configuration

After providing all necessary details about the dataset to be stored in the TwinERGY Core Data Management Platform, the platform users are informed that the data ingestion job is completed and are then redirected to the data ingestion job page, where they can view all the initiated data ingestion jobs as well as their current status, as depicted in Figure 17.



Figure 17 Data Ingestion Jobs overview and completion status

4.2 CIM Extension

The TwinERGY CIM is essentially seen as a "living" model that will be regularly changed and grow relying on the demands of the project, as defined in Deliverable D5.1 (TwinERGY

Common Information Model). Changes or additions to the Common Information Model can be suggested, as shown in Figure 18, so to be further assessed and accordingly be incorporated in the TwinERGY CIM.

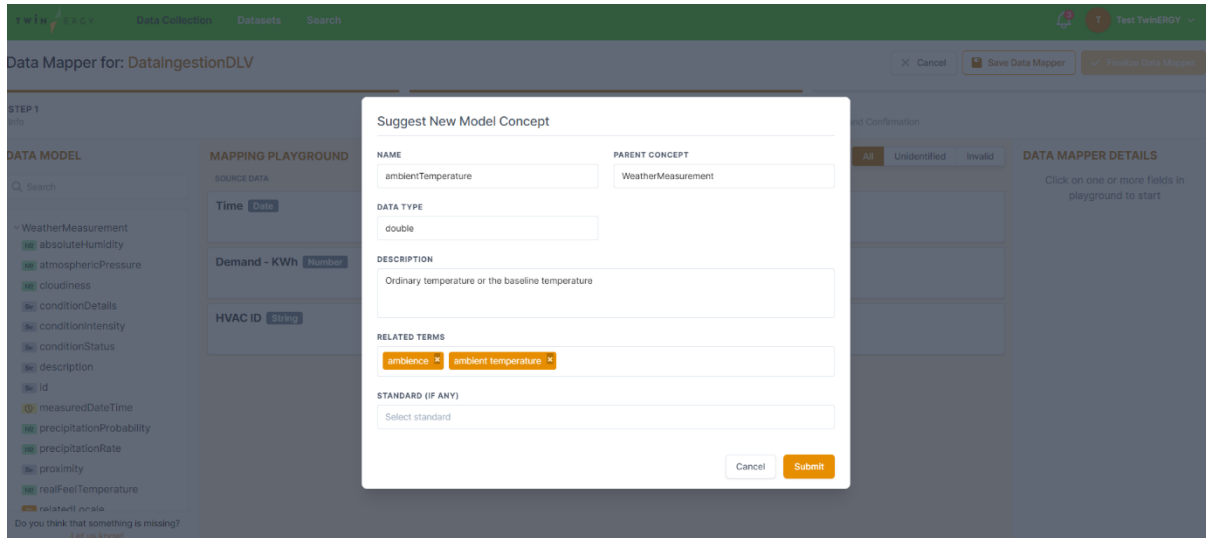


Figure 18 Suggestion of new fields

This CIM extension, which is a part of release 1.00, makes sure that the model is constantly updated and improved so that its use is reliable and long-lasting. After a careful assessment of each specific piece of information required, as described in the previous step, the suggested fields are inserted under the appropriate concept and are ready to be used during the mapping process, as shown in Figure 19.

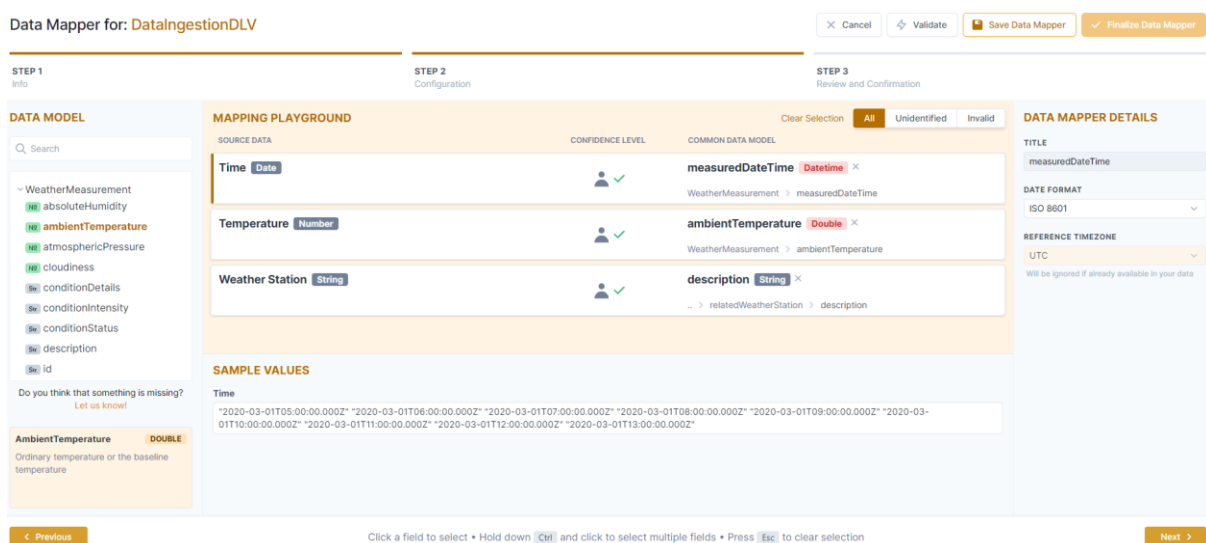


Figure 19 Proposed fields be added in the TwinERGY CIM

5. Data Security Service

The Data Security Service addresses the data security and privacy concerns of TwinERGY CDMP users with regard to the data that will be ingested and handled in the TwinERGY Core Data Management Platform. The Data Security Service helps users in developing trust in the TwinERGY CDMP by enabling them to design simple and flexible access policy routines, that limit access requests to their data on the platform. To enhance efficiency, access policy routines for each dataset can be created utilizing the functionality of the Access Policies Controller. Through a graphical user interface, users can create complex access policy routines for their datasets. The routines are retained, and by utilizing the provided interface, the users can easily make changes to them.

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 20. 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.

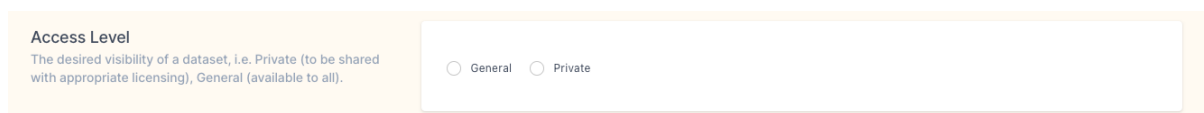


Figure 20 Access Level Definition

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 21. 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.

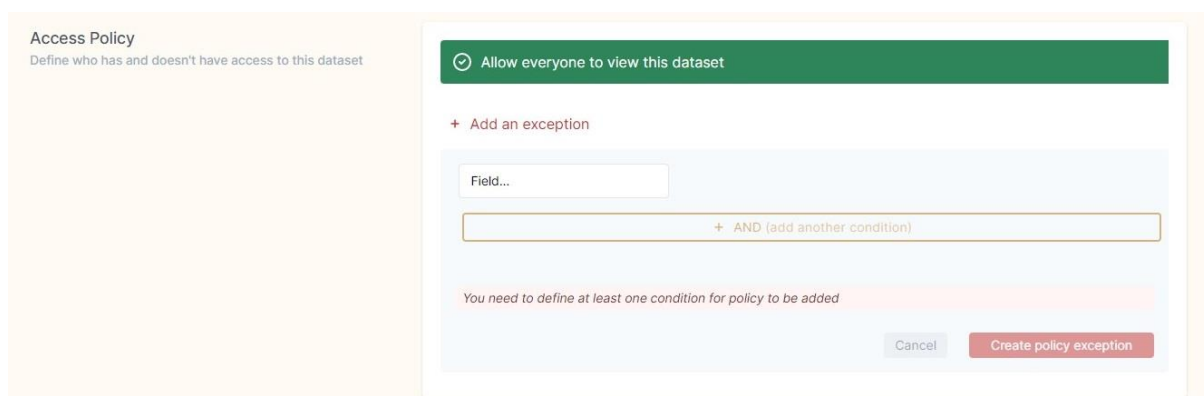


Figure 21 Access Policies Definition

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 22.

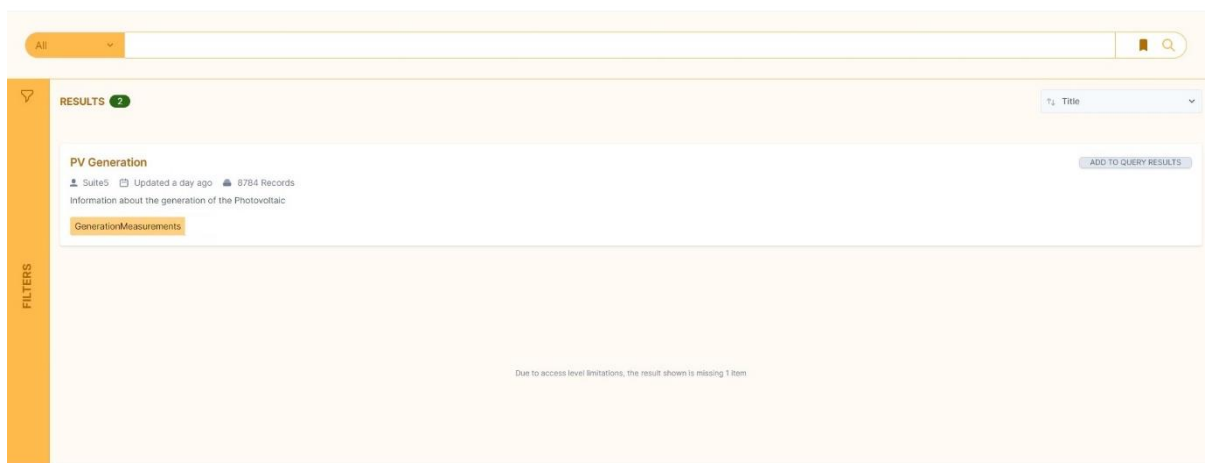


Figure 22 Retrieval of datasets

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 23.

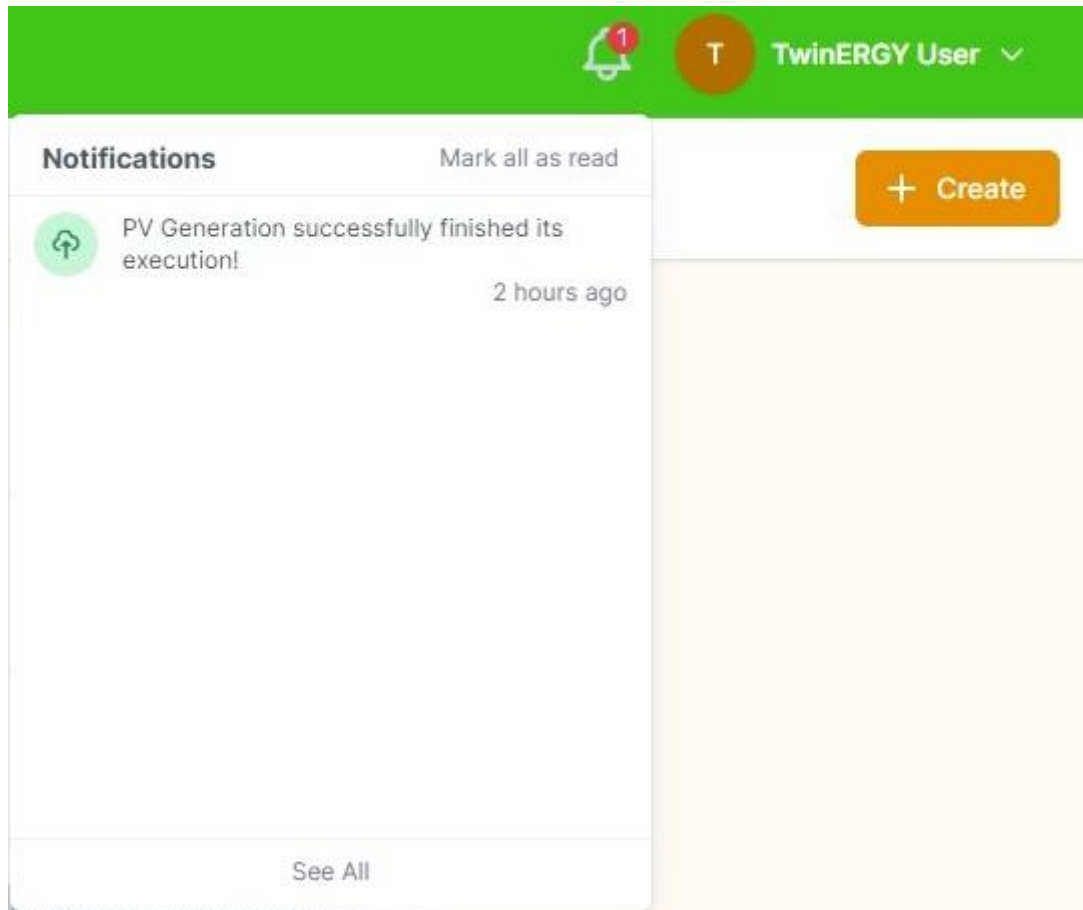


Figure 23 Notifications on ongoing/finished data ingestion jobs

7. Conclusions

Deliverable 5.6 includes the TwinERGY Integrated Platform's release 1.00. The document comprises an improved version of the functionalities described in Deliverable 5.4, as well as the platform's services (Data Collection, Data Security, Data Storage, and Platform Management). In particular, this deliverable elaborated on:

- The platform's data ingestion process that incorporates improvements to the Data Ingestor functionality, which allowed for larger and more complex file uploads to the TwinERGY CDMP as well as updates to data by appending new rows.
- The Data Storage Service, which incorporated improvements to CIM Store functionality, which allowed for TwinERGY CIM extension with additional fields than the original version that was presented in D5.1 – TwinERGY Common Information Model.
- 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 platform management functionalities that allow users to edit and modify the information in their personal profiles, search for data that they are permitted to use in accordance with the access policies that have been established and receive notifications regarding the success or failure of data ingestion jobs.

In comparison to Deliverable D5.5 [3] "Data Collection, Security, Storage & Management Services Bundles - Release 1.00," D5.7 [4] "Data Collection, Security, Storage & Management Services Bundles - Release 2.00" will present the final version of the TwinERGY Platform core services for data collection, security, storage, governance and management, offering a more developed and enhanced version of the services, while D5.8 [5] "TwinERGY Integrated Platform- Release 2.00" will present the final version of the Integrated TwinERGY Data Management Platform.

REFERENCES

1. TwinERGY Consortium. (2020). TwinERGY D5.4 "TwinERGY Integrated Platform – Beta Release"
2. TwinERGY Consortium. (2020). TwinERGY D5.1 "TwinERGY Common Information Model"
3. TwinERGY Consortium. (2020). TwinERGY D5.5 "Data Collection, Security, Storage & Management Services Bundles – Release 1.00"
4. TwinERGY Consortium. (2020). TwinERGY D5.7 "Data Collection, Security, Storage & Management Services Bundles – Release 2.00"
5. TwinERGY Consortium. (2020). TwinERGY D5.8 "TwinERGY Integrated Data Management Platform– Release 2.00"

ANNEXES

TwinERGY services and implementation dependencies with other projects

Taking into consideration that TwinERGY is an Innovation Action project, it is anticipated that the development activities build on different background technologies, either open-source or owned by the technical partners. The following table presents in detail the background components that have been leveraged in the Data Collection Service, the Data Security Service and the Platform Management Service and the updates that have been already introduced in the current release of the TwinERGY Core Data Management Platform. The Data Storage Service has been developed from scratch based on open-source technologies (as described in D5.5 [3]).

Service	Built from scratch in TwinERGY project	Dependency from another project	Enhancements made in comparison to previous implementation
Data Collection	Data Curator, Metadata Editor: Developed from scratch (leveraging the experiences from past big data projects, especially ICARUS, and open-source technologies)	Data Ingestor: Built on top of the BIMERR Data Ingestor & Fetcher Data Mapper: Built on top of the BIMERR Model Mapper	Data Ingestor <ul style="list-style-type: none"> • Full refactoring for improving performance • Alignment with the TwinERGY technology stack For Data Mapper: <ul style="list-style-type: none"> • Full refactoring for improving accuracy and performance • Alignment with the TwinERGY technology stack • Enhancements for effectively handling the TwinERGY CIM
Data Security	Access Policies Controller: Developed from scratch (leveraging the experiences from past big data projects, especially ICARUS, and open-source technologies)	-	-
Platform Management	Identity Manager: Developed from scratch (leveraging	Search & Retrieval Manager: Built on top of the BIMERR Building	<ul style="list-style-type: none"> • Partial refactoring for more efficient search

Service	Built from scratch in TwinERGY project	Dependency from another project	Enhancements made in comparison to previous implementation
	<p>open-source technologies) Notifications Handler: Developed from scratch (leveraging open-source technologies)</p>	<p>Information Query Builder</p>	<ul style="list-style-type: none"> • New retrieval layer from the TwinERGY Storage Service • Adapted metadata schema