# TWIN ERGY

# Data Collection, Security, Storage & Management Services Bundles - Release 2.00

D5.7

February 2023

# 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 |

## D5.7
## Data Collection, Security, Storage & Management
## Services Bundles – Release 2.00

**Revision: v1.0**

*AUTHORS*

| Alexandros Tsitsanis | Marios Phinikettos | Chrysanthos Michael |
|---|---|---|
| Suite5 | 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 |
|----------|------|---------|--------------|-------------|
| V0.1 | 17.02.2023 | Alexandros Tsitsanis Marios Phinikettos Chrysanthos Michael | Suite5 | 1st draft of the deliverable D5.7 - Data Collection, Security, Storage & Management Services Bundles – Release 2.00 |
| V0.2 | 22.02.2023 | Ana Isabel Martínez, Moisés Antón | ETRA | Feedback from Reviewer |
| V0.3 | 22.02.2023 | Axel Balke, Fynn Christian Bollhöfer | TH-OWL | Feedback from Reviewer |
| V0.4 | 22.02.2023 | Konstantinos Kotsalos, Apostolos Kapetanios | ED | Feedback from Reviewer |
| V0.5 | 23.02.2023 | Alexandros Tsitsanis Marios Phinikettos Chrysanthos Michael | Suite5 | Consolidating feedback received |
| V0.6 | 28.02.2023 | Alexandros Tsitsanis Marios Phinikettos Chrysanthos Michael | Suite5 | Consolidated Version – Submitted to PC |
| V1.0 | 28.02.2023 | Athanasios Chassiakos, Stylianos Karatzas, Vasiliki Lazari, Antonis Papamanolis | UoP | Final draft submitted to EC by the PC |

TWIN ERGY

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

# 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.

TWIN ERGY

Twinergy has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 957736

# Executive Summary

1. Deliverable D5.7 " Data Collection, Security, Storage & Management Services Bundles – Release 2.00" summarizes three (3) tasks, namely Task 5.2 "Data Management Platform Backbone Infrastructure," Task 5.3 "Core Data Ingestion, Curation, and Management Services," and Task 5.4 "Data security, encryption, and privacy mechanisms" of WP5 " Data Collection and Communication Platform", as depicted in the TwinERGY DoA DoA Part A, TwinERGY Consortium Agreement No 957736
2. TwinERGY Consortium. (2020). TwinERGY D5.5 "Data Collection, Security, Storage & Management Services Bundles – Release 1.00"
3. TwinERGY Consortium. (2020). TwinERGY D5.8 "TwinERGY Integrated Data Management Platform – Release 2.00"
4. DoA Part B, TwinERGY Consortium Agreement No 957736
5. Data Catalog Vocabulary (DCAT) - Version 2 (w3.org) and expands on D5.5 "Data Collection, Security, Storage & Management Services Bundles – Release 1.00, DoA Part A, TwinERGY Consortium Agreement No 957736
6. TwinERGY Consortium. (2020). TwinERGY D5.5 "Data Collection, Security, Storage & Management Services Bundles – Release 1.00"
7. TwinERGY Consortium. (2020). TwinERGY D5.8 "TwinERGY Integrated Data Management Platform – Release 2.00"
8. DoA Part B, TwinERGY Consortium Agreement No 957736

Data Catalog Vocabulary (DCAT) - Version 2 (w3.org)". The current deliverable provides the Release 2.00 and features some further functionalities for the intended users of the TwinERGY Core Data Management Platform, in relation to D5.5. These functionalities include the Data Anonymization Handler functionality, which provides platform users with appropriate data interventions in the circumstance that they need to anonymize fields in their data that might contain personally identifiable information, as well as the streaming data to internal Publish/Subscribe [14] mechanism, a newly introduced method for data collection.

This deliverable dives into the development details of the services, covering the following aspects:

- The current status of various functionalities provided from each service, expanding the functionalities introduced in D5.5 "Data Collection, Security, Storage & Management Services Bundles – Release 1.00".
- The frameworks and libraries used to implement each service with their corresponding versions.

- The interface specifications that were created, in order to document the technical details of each service's APIs.

The current deliverable is the 2.00 version of TwinERGY Core Data Management Platform's services on M28. The delivered services are going to be represented with appropriate screenshots in D5.8 "TwinERGY Integrated Data Management Platform– Release 2.00", which will describe the final (2.00) release of the TwinERGY Integrated Data Management Platform on M32 (June 2023).

# Index

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

# List of Figures

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

# List of Tables

# 1. Introduction

## 1.1 Purpose of this deliverable

The backbone of the project is an "open," modular, and interoperable big data management platform, as captured in the TwinERGY DoA [12], which enables open standards-based data collection and management communication throughout the project's energy value chain. The TwinERGY Data Management Platform complies with existing open energy standards (like OpenADR, USEF, etc.) and incorporates a homogenized Common Information Model (CIM) that ensures semantic interoperability of the digitalized energy assets used 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. The TwinERGY Data Management Platform has the required data security, privacy, authentication, and authorization processes in place to safeguard end-user data confidentiality and the non-repudiation of DER assets.

To define the functional requirements for the development of the TwinERGY CDMP and the various services provided by the platform, a thorough analysis of the data availability in the pilot sites, the data requirements of the TwinERGY modules, and the way that the TwinERGY use cases are addressed, was conducted. The various functionalities introduced to the Data Collection, Data Security, Data Storage, and Platform Management services, as well as those implemented during the release 1.00, are thoroughly detailed in this deliverable. As an enhanced version of the TwinERGY Core Data Management Platform Services Bundles, D5.7 is a more developed and feature-rich version of D5.5, detailing the upgrades and improvements made in comparison to the previous release. These functionalities include the Data Anonymization Handler functionality, which provides platform users with appropriate data interventions in cases that they need to anonymize fields in their data that might contain personally identifiable information, as well as the ingestion of streaming data via Pub/sub mechanism, a newly introduced method for data collection.

## 1.2 Scope of this deliverable

The major scope of D5.7 is to provide a comprehensive overview of the TwinERGY Core Data Management Platform Services version 2.00. The deliverable goes through the TwinERGY CDMP's services in greater detail, as well as their functionalities. Data Collection, Data Security, Data Storage, and Platform Management services, are all described in detail by:

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

- Reporting on the implemented functionalities for each service, as well as the functionalities introduced in the release 1.00.
- Describing the libraries and frameworks used to implement each service.
- Documenting the development of relevant APIs that make communication and information sharing across services easier.

## 1.3 Structure of the document

The structure of the document is as follows:

Section 2 provides a brief description of the TwinERGY Core Data Management Platform conceptual architecture, the services and their functionalities, setting the basis for the next section. It also provides an overview of the platform's services and shows how the functionalities of each service are developed. Additionally, the technological infrastructure that was used to establish each service and the specifications of the various interfaces that have been created, together with the appropriate licensing and access methods are reported.

Deliverable D5.7 "Data Collection, Security, Storage, and Management Services Bundles – Release 2.00" is concluded with Section 3, which provides a brief synopsis of what has been covered in the various sections of the deliverable.

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

## 1.4 Abbreviation List

| Acronym | Full Name |
|---------|-----------|
| API | Application Programming Interface |
| CDMP | Core Data Management Platform |
| CIM | Common Information Model |
| CSV | Comma-separated values |
| D | Deliverable |
| DCAT | Data Catalog Vocabulary |
| DoA | Description of Action |
| DER | Distributed Energy Resources |
| DMP | Data Management Platform |
| GDPR | General Data Protection Regulation |
| JSON | JavaScript Object Notation |
| M | Month |
| Pub/sub | Publish/Subscribe |
| UI | User Interface |
| WP | Work Package |
| XML | Extensible Mark-up Language |

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

# 2. TwinERGY Core Data Management Platform and Services Overview

## 2.1 TwinERGY Core Data Management Platform

Data Collection, Data Security, Data Storage, and Platform Management are all part of the TwinERGY Core Data Management Platform, as depicted in Figure 1. The platform's services and their respective functionalities are detailed and further examined in the sections that follow throughout this deliverable.
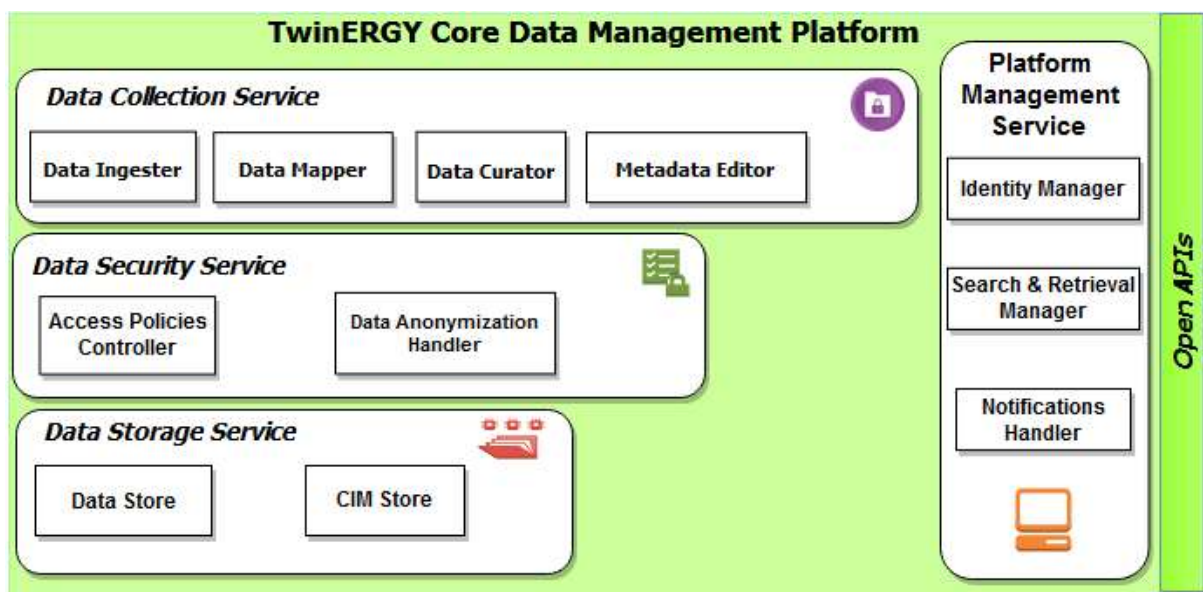


*Figure 1 The TwinERGY Core Data Management Platform conceptual architecture*

The above architecture of the CDMP was introduced, in order to address the requirements that were identified, after the completion of the activities and evaluation of the findings and feedback obtained from the TwinERGY Use Cases elaboration, the TwinERGY pilot sites data availability activities and the identification of the various data requirements of the different modules to be developed in the context of the TwinERGY project.

In the following sections, a detailed overview of the services and of their functionalities is provided, along with the state of development of each service's functionalities and the interfaces specifications created to describe the technical details of each service's APIs.

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

## 2.2 Services updates and overview

### 2.2.1 Outline of services updates

Between release 1.00 of the Data Collection, Security, Storage & Management Services Bundles and release 2.00, a number of enhancements were made to the CDMP, its services, and their respective functionalities. By simultaneously addressing the needs that emerged from the elaboration of the TwinERGY Use Cases, the availability of data in the pilot sites, and the identification of the various data requirements of the modules, these enhancements were meant to increase the Core Data Management Platform's overall performance. More specifically, these improvements refer to:

1. Introduction of new ingestion methods, enabling platform users to stream their data to the platform using the streaming Pub/sub mechanism. The users are provided with the necessary connection information (url, port, SSL, topic, username and password), they can define specific retrieval parameters while configuring the streaming Pub/sub mechanism, and they are also requested to upload a sample of the streaming data. The introduction of the new data collection method complements the already available data collection mechanisms, that are addressing Req_001 to Req_009 of the TwinERGY CDMP Functional Requirements, as shown below in the respective rows of the requirements list. The entire list of requirements can be found in the

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

2. Annex section.

*Table 1 Requirements 001 to 009, that led to the introduction of new data ingestion methods*

| 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. |

3. <u>Data Anonymization Handler</u>, a functionality necessary to safeguard potentially sensitive information by anonymizing parts of data (i.e., fields), in order to maintain privacy and confidentiality. This functionality addresses requirement Req_013 of the TwinERGY CDMP Functional Requirements, as shown below in the corresponding segment of the requirements list.

*Table 2 Requirement 013, that led to the implementation of the Data Anonymization Handler functionality*

| Req_013 | The platform should allow the user to define data anonymization routines. |
|---------|---------------------------------------------------------------------------|

In the following sections, an updated detailed overview of the services and of their functionalities is provided, along with the state of development of each service's functionalities and the interfaces specifications created to describe the technical details of each service's APIs.

## 2.2.2 Data Collection Service

The Data Collection Service introduces a variety of functionalities that are relevant to the data collection process, including:

i.  Data Ingester, which introduces proper procedures and methods for ingesting data into the CDMP;
ii. Data Mapper, which allows ingested datasets to be aligned with the CIM;

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

iii.    Data Curator, which ensures the datasets quality; and

iv.    Metadata Editor, which enables the users of the platform to define specific aspects of the ingested dataset, in order to provide a detailed profile of the given dataset.

The appropriate handling of data management procedures enables data collection setup, allowing users to specify the data ingestion method, according to which the data is uploaded to the TwinERGY CDMP. The platform integrates data from a variety of sources in a straightforward manner. Users uploading data use the platform to clarify the way and the type of data they own and are ready to share, as well as to define the parameters of the data ingestion process, outline their needs and preferences on the way that the data is inserted in the platform (file uploading, users APIs, or Data Streaming), and perform proper mapping actions on all fields of the data, all of which are based on the TwinERGY Common Information Model.

The Data Mapper functionality determines how the fields of a dataset are aligned with the concepts of the TwinERGY CIM. When the CIM and the dataset are matched, a consistent domain-specific understanding of the data is established, making it easier to use for data-related services in the CDMP. To guarantee adherence with the CIM, the application of appropriate mapping and transformation routines on datasets allows for the changing of field names and the conversion of measurement units.

Additionally, during the curation configuration process, the Data Curator functionality allows the user to specify the restrictions and limitations that the data ingested into the platform may have, as well as the precise activities undertaken if any of these constraints are not satisfied. This configuration refers to a set of curation routines that are performed by carrying out the required data curation processes.

Any result of data ingestion is saved as a dataset in the TwinERGY Core Data Management Platform and with the Metadata Editor functionality, the users of the platform can specify a comprehensive dataset profile, by determining the dataset title, description and additional metadata in alignment with the DCAT recommendations.

### 2.2.2.1 State of Development

The Data Collection Service release, as specified in this deliverable, introduces a range of functionalities that are described below.

The functionalities that were part of release 1.00 of the Data Collection Service are depicted as follows:

- Data ingestion process configuration: describes in detail the needs and preferences on the way that the data are ingested into the TwinERGY Core Data Management Platform (file uploading, API, streaming data).
- Ingestion of data from files: allows data retrieval from files in popular file formats (e.g., csv).

TWIN ERGY

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

- API Data Ingestion: This functionality allows data to be retrieved from both pilot system APIs and Open Data APIs (e.g., weather data, other local sources).
- Reliable and safe data upload: Data is uploaded to the TwinERGY Core Data Management Platform using reliable and secure processes.
- Mapping activities on ingested data: This functionality ensures that the data values are matched to the CIM provisions (e.g., data types, measurement units) by performing the necessary mapping actions on the ingested data.
- Dataset & metadata creation: Any outcome of a data ingestion is recorded as a dataset in the Data Storage Service of the TwinERGY Core Data Management Platform with this functionality. Users of the platform can give a dataset a title and associated metadata, as well as describe what is contained inside it.
- Appending data in data files: With this functionality, the TwinERGY CDMP users are able to append new data in already stored datasets within the platform.
- Large file uploading: This functionality ensures that bigger volumes of data can be ingested in the TwinERGY CDMP, in a single dataset.
- Faster uploading: This functionality ensures that data are ingested faster in the Core Data Management Platform, due to the refactoring performed.

The functionalities that are part of the 2.00 release, extending the functionalities of the release 1.00, of the Data Collection Service are depicted as follows:

- Streaming data to internal Pub/sub mechanism: This functionality enables the users of the Core Data Management Platform to stream their data to the internal Pub/Sub mechanism of the platform, which facilitates the ingestion of real-time datasets.

## 2.2.2.2 Technological Infrastructure

The back - end of the Data Collection Service is developed with **Python 3.10**[1], using also a range of open-source technologies, namely **Django 4.0**[2], **Pandas 1.3.0**[3] and **Docker 20.10.11**[4].

The front-end of the Data Collection Service is developed with **Vue.js V2**[5].

## 2.2.2.3 Interface Specification

*Table 3 Start a new data ingestion job interface specification*

| Name | Start a new data ingestion job |
|------|-------------------------------|
| Description | Initiates the data ingestion process |

---

[1] https://www.python.org/downloads/release/python-3100/

[2] https://www.djangoproject.com/weblog/2021/dec/07/django-40-released/

[3] https://pandas.pydata.org/docs/whatsnew/index.html

[4] https://docs.docker.com/engine/release-notes/

[5] https://vuejs.org/v2/guide/

| Reference Code | DCS01 | | |
|---|---|---|---|
| Service | Data Collection | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |
| | POST | N/A | 200 OK<br>201 Created<br>401 Unauthorized<br>403 Forbidden<br>404 Not Found |
| Endpoint URI | /twinergy-api/data-collection/ | | |
| Interaction with Services | Data Collection Service | | |

*Table 4 Retrieve a data ingestion job interface specification*

| Name | Retrieve a data ingestion job | | |
|---|---|---|---|
| Description | Initiate the data ingestion retrieval process | | |
| Reference Code | DCS02 | | |
| Service | Data Collection | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |
| | GET | id: The identifier of the data collection | id (string)<br>title (string)<br>user (string)<br>configuration (json)<br>created (datetime)<br>status (string) |
| Endpoint URI | /twinergy-api/data-collection/{id} | | |
| Interaction with Services | Data Collection Service | | |

*Table 5 Delete a data ingestion job interface specification*

| Name | Delete a data ingestion job | | |
|---|---|---|---|
| Description | Deletes a data ingestion job | | |
| Reference Code | DCS03 | | |
| Service | Data Collection | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |
| | DELETE | id: The identifier of the data collection | 200 OK<br>401 Unauthorized<br>403 Forbidden<br>404 Not Found |

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

| Endpoint URI | /twinergy-api/data-collection/{id} |
|---|---|
| Interaction with components | Data Collection Service |

*Table 6 Ingester Configuration interface specification*

| Name | Ingester Configuration | | |
|---|---|---|---|
| Description | Creates the data ingester settings | | |
| Reference Code | DCS04 | | |
| Service | Data Collection | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |
| | POST | id: The identifier of the data ingester configuration | 200 OK 201 Created 401 Unauthorized 403 Forbidden 404 Not Found |
| Endpoint URI | /twinergy-api/data-collection/data-ingester-config/{id} | | |
| Interaction with Services | Data Collection Service | | |

*Table 7 Ingester Retrieval interface specification*

| Name | Ingester Retrieval | | |
|---|---|---|---|
| Description | Retrieves the data ingester settings | | |
| Reference Code | DCS05 | | |
| Service | Data Collection | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |
| | GET | id: The identifier of the data ingester retrieval | id (string) step-id (string) title (string) configuration (json) created (datetime) status (string) |
| Endpoint URI | /twinergy-api/data-collection/data-ingester-config/{id} | | |
| Interaction with Services | Data Collection Service | | |

*Table 8 Mapper Configuration interface specification*

| Name | Mapper Configuration |
|---|---|
| Description | Creates the data mapper settings |

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

| Reference Code | DCS06 | | |
|---|---|---|---|
| Service | Data Collection | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |
| | POST | id: The identifier of the data mapper configuration | 200 OK 201 Created 401 Unauthorized 403 Forbidden 404 Not Found |
| Endpoint URI | /twinergy-api/data-collection/ data-mapper-config /{id} | | |
| Interaction with Services | Data Collection Service | | |

*Table 9 Mapper Retrieval interface specification*

| Name | Mapper Retrieval | | |
|---|---|---|---|
| Description | Retrieves the data mapper settings | | |
| Reference Code | DCS07 | | |
| Service | Data Collection | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |
| | GET | id: The identifier of the data mapper retrieval | id (string) step-id (string) title (string) configuration (json) created (datetime) status (string) |
| Endpoint URI | /twinergy-api/data-collection/data-mapper-config/{id} | | |
| Interaction with Services | Data Collection Service | | |

*Table 10 Curator Configuration interface specification*

| Name | Curator Configuration | | |
|---|---|---|---|
| Description | Creates the data curator settings | | |
| Reference Code | DCS08 | | |
| Service | Data Collection | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

| | POST | id: The identifier of the data curator | 200 OK<br>201 Created<br>401 Unauthorized<br>403 Forbidden<br>404 Not Found |
|---|---|---|---|
| **Endpoint URI** | /twinergy-api/data-collection/data-curator-config /{id} | | |
| **Interaction with Services** | Data Collection Service | | |

*Table 11 Curator Retrieval interface specification*

| **Name** | Curator Retrieval | | |
|---|---|---|---|
| **Description** | Retrieves the data curator settings | | |
| **Reference Code** | DCS09 | | |
| **Service** | Data Collection | | |
| **Type of interface** | REST | | |
| **Inputs** | Methods of the interface | Parameters of the method | Return Values of the method |
| | GET | id: The identifier of the data curator retrieval | id (string)<br>step-id (string)<br>title (string)<br>configuration (json)<br>created (datetime)<br>status (string) |
| **Endpoint URI** | /twinergy-api/data-collection/data-curator-config/{id} | | |
| **Interaction with Services** | Data Collection Service | | |

*Table 12 Metadata Editor Configuration interface specification*

| **Name** | Metadata Editor Configuration | | |
|---|---|---|---|
| **Description** | Creates the metadata editor settings | | |
| **Reference Code** | DCS10 | | |
| **Service** | Data Collection | | |
| **Type of interface** | REST | | |
| **Inputs** | Methods of the interface | Parameters of the method | Return Values of the method |
| | POST | id: The identifier of the metadata editor configuration | 200 OK<br>201 Created<br>401 Unauthorized<br>403 Forbidden<br>404 Not Found |
| **Endpoint URI** | /twinergy-api/data-collection/metadataeditor-config/{id} | | |
| **Interaction with Services** | Data Collection Service | | |

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

*Table 13 Metadata Editor Retrieval interface specification*

| Name | Metadata Editor Retrieval | | |
|---|---|---|---|
| **Description** | Retrieves the metadata editor settings | | |
| **Reference Code** | DCS11 | | |
| **Service** | Data Collection | | |
| **Type of interface** | REST | | |
| **Inputs** | Methods of the interface | Parameters of the method | Return Values of the method |
| | GET | id: The identifier of the metadata editor retrieval | id (string) step-id (string) title (string) configuration (json) created (datetime) status (string) |
| **Endpoint URI** | /twinergy-api/data-collection/metadataeditor-config/{id} | | |
| **Interaction with Services** | Data Collection Service | | |

## 2.2.3 Data Security Service

The TwinERGY CDMP users' data security and privacy concerns about data that are ingested and handled in the TwinERGY Core Data Management Platform are addressed by the Data Security Service. As a result, it includes a number of functions and functionalities that are easily configured in the TwinERGY CDMP, such as the Access Policies Controller and the Data Anonymization Handler.

The Access Policies Controller functionality enhances user confidence in the entire TwinERGY Core Data Management Platform by allowing users to apply access policies that flexibly manage access requests to data within the TwinERGY CDMP. The access policy controller effectively handles policy enforcement at run-time so that only authorized platform users can view and access the data. This functionality allows platform users to define and configure access policy routines, allowing these routines to be based on data fields and requestor properties. Routines are saved and users uploading data can easily modify them. After the access policy is defined and configured, the policy is applied when access to the data contained in the CDMP is requested.

As for the Anonymization handler functionality, it ensures the protection of the data from inadvertent disclosure of personal and company information. Before the data becomes available on the TwinERGY platform, anonymization actions are performed and reflected in the defined part of the data (i.e., fields). The users of the CDMP are provided with the appropriate interventions on the data, in case they need to anonymize fields in their data that contain personally identifiable information.

TWIN ERGY

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

### 2.2.3.1 State of Development

The Data Security Service release, as specified in this deliverable, introduces a range of functionalities that are described below.

The functionalities that were part of release 1.00 of the Data Security Services are depicted as follows:

- Backend access policy management and handling: Policies are used to determine when access is permitted.
- Backend enforcement of data access policies: The access policies that have been set are applied when a request is made to access data that has been ingested into the TwinERGY CDMP.

The functionalities that are part of the 2.00 release, extending the functionalities of the release 1.00 of the Data Security Service are depicted as follows:

- Data Anonymization: Prior to the data being made available on the CDMP, users of the platform perform anonymization routines on the data or its relevant elements (i.e., dataset fields). The TwinERGY Platform, via the Data Anonymization Handler functionality, enables the CDMP users to anonymize any segments of their datasets that they assume may include personally identifying information.

### 2.2.3.2 Technological Infrastructure

The back-end of the Data Security Service is developed with **Python 3.10**, using also a range of open-source technologies, namely **Django 4.0**, and **Pandas 1.3.0**, as documented in 2.2.2.2 Technological Infrastructure.

The access policy management library of the Data Security Service is **CASL v4**[6].

### 2.2.3.3 Interface specification

*Table 14 Access Policy Creation interface specification*

| Name | Access Policy Creation | | |
|---|---|---|---|
| Description | Creates an access policy for a dataset | | |
| Reference Code | DSS01 | | |
| Service | Data Security | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |

---

[6] https://casl.js.org/v4/en/guide/intro

TWIN ERGY

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

| | POST | id: The identifier of the dataset | 200 OK<br>201 Created<br>401 Unauthorized<br>403 Forbidden<br>404 Not Found |
|---|---|---|---|
| **Endpoint URI** | /twinergy-api/dataset/{id}/policy | | |
| **Interaction with Services** | Data Security Service | | |

*Table 15 Access Policy Retrieval interface specification*

| **Name** | Access Policy Retrieval | | |
|---|---|---|---|
| **Description** | Retrieves all access policies for a dataset | | |
| **Reference Code** | DSS02 | | |
| **Service** | Data Security | | |
| **Type of interface** | REST | | |
| **Inputs** | Methods of the interface | Parameters of the method | Return Values of the method |
| | GET | id: The identifier of the dataset | id (string)<br>step-id (string)<br>title (string)<br>policies (json)<br>created (datetime)<br>status (string) |
| **Endpoint URI** | /twinergy-api/dataset/{id}/policy | | |
| **Interaction with Services** | Data Security Service | | |

*Table 16 Access Policy Update interface specification*

| **Name** | Access Policy Update | | |
|---|---|---|---|
| **Description** | Updates an access policy for a dataset | | |
| **Reference Code** | DSS03 | | |
| **Service** | Data Security | | |
| **Type of interface** | REST | | |
| **Inputs** | Methods of the interface | Parameters of the method | Return Values of the method |
| | POST | id: The identifier of the dataset<br><br>policy_id: The identifier of the access policy | 200 OK<br>201 Created<br>401 Unauthorized<br>403 Forbidden<br>404 Not Found |
| **Endpoint URI** | /twinergy-api/dataset/{id}/policy/{id} | | |

| Interaction with Services | Data Security Service |
|---|---|

*Table 17 Access Policy Deletion interface specification*

| Name | Access Policy Deletion | | |
|---|---|---|---|
| Description | Deletes an access policy for a dataset | | |
| Reference Code | DSS04 | | |
| Service | Data Security | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |
| | DELETE | id: The identifier of the dataset<br><br>policy_id: The identifier of the access policy | 200 OK<br>401 Unauthorized<br>403 Forbidden<br>404 Not Found |
| Endpoint URI | /twinergy-api/dataset/{id}/policy/{id} | | |
| Interaction with Services | Data Security Service | | |

*Table 18 Anonymization Handler Configuration interface specification*

| Name | Anonymization Handler Configuration | | |
|---|---|---|---|
| Description | Creates the data anonymization handler settings | | |
| Reference Code | DSS05 | | |
| Service | Data Security | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |
| | POST | id: The identifier of the data anonymization handler configuration | 200 OK<br>201 Created<br>401 Unauthorized<br>403 Forbidden<br>404 Not Found |
| Endpoint URI | /twinergy-api/data-collection/data-anonymization-handler-config/{id} | | |
| Interaction with Services | Data Security Service | | |

*Table 19 Anonymization Handler Retrieval interface specification*

| Name | Anonymization Handler Retrieval |
|---|---|
| Description | Retrieves the data anonymization handler settings |
| Reference Code | DSS06 |

| Service | Data Security | | |
|---|---|---|---|
| **Type of interface** | REST | | |
| **Inputs** | Methods of the interface | Parameters of the method | Return Values of the method |
| | GET | id: The identifier of the data anonymization handler retrieval | id (string) step-id (string) title (string) configuration (json) created (datetime) status (string) |
| **Endpoint URI** | /twinergy-api/data-collection/data-anonymization-handler-config/{id} | | |
| **Interaction with Services** | Data Security Service | | |

## 2.2.4 Data Storage Service

This service meets the demand for dependable data storage and indexing by providing resilient measures and different indexing methodologies. Depending on the type of information stored in the TwinERGY Platform and the means for retrieval, different storage solutions are utilised to meet various needs. The Data Storage Service is in charge of securely storing a wide range of data, as well as the related metadata, in a reliable and secure manner.

### 2.2.4.1 State of Development

The Data Storage Service release, as specified in this deliverable, introduces a range of functionalities that are described below.

The functionalities that were part of release 1.00 of the Data Storage Services are depicted as follows:

- Data Store: This functionality entails storing various types of data (e.g., datasets) together with their associated metadata, making it accessible to all TwinERGY Core Data Management Platform services.
- CIM Store: Storage of the TwinERGY Common Information Model, as well as all necessary concepts and fields.
- CIM Extension: Extension of the TwinERGY CIM with the creation of additional fields, based on the data samples from the TwinERGY pilot sites.

TWIN ERGY

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

### 2.2.4.2 Technological Infrastructure

The back-end of the Data Storage Service is developed with the utilization of **PostgreSQL 14**[7] and **Elasticsearch 7.14**[8].

### 2.2.4.3 Interface Specification

There are no external APIs that are exposed by the Data Storage Service to accompany its initial version except for the APIs that have been presented in sections 2.2.2.3, 2.2.3.3 and 2.2.5.3.

## 2.2.5 Platform Management Service

This service establishes the tools and processes for users to register on the platform in a safe and reliable manner. Through appropriate authentication and security measures, users are given access to the data that they are authorized to use. When data ingestion takes place in the platform, users are alerted via platform notifications based on their preferences. The service has three main functionalities: the (i) Identity Manager, the (ii) Search & Retrieval Manager, and the (iii) Notifications Handler.

Summarized, with the development of the Platform Management Service:

- Users register, authorize, and authenticate themselves using the TwinERGY CDMP's identity management functionality, which is interconnected with security-related processes.
- Users are able to search for and find data that may be proven valuable, as well as determine and define which of these available data are relevant to their needs, using the Search & Retrieval functionality.
- The Notifications handling functionality is in charge of detecting events related to the status and progress of data ingestion in order to provide the TwinERGY Core Data Management Platform's users with appropriate material.

### 2.2.5.1 State of Development

The Platform Management Service release, as specified in this deliverable, introduces a range of functionalities that are described below.

The functionalities that were part of release 1.00 of the Platform Management Services are depicted as follows:

- Identity Management: This functionality enables the creation and management of Core Data Management Platform users allowing for the establishment and

---

[7] https://www.postgresql.org/about/news/postgresql-14-released-2318/

[8] https://www.elastic.co/blog/whats-new-elasticsearch-7-14-0

controlling of identifying information for those who are permitted to use the platform.

## 2.2.5.2 Technological Infrastructure

The back-end of the Platform Governance Module is developed with **Python 3.10**, using also a range of open-source technologies, namely **Django 4.0**, as documented in 2.2.2.2 Technological Infrastructure

**Websockets 10.1**[9] and **Kubernetes 1.22.2**[10] are utilized for server-side notifications implementation and for the system deployment respectively.

## 2.2.5.3 Interface Specification

*Table 20 User Retrieval interface specification*

| Name | User Retrieval | | |
|---|---|---|---|
| Description | Retrieves the specifics for a user | | |
| Reference Code | PMS01 | | |
| Service | Platform Management | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |
| | GET | id: The identifier of the user | username (string) name (string) e-mail (string) |
| Endpoint URI | /twinergy-api/user/{id} | | |
| Interaction with Services | Platform Management Service | | |

*Table 21 User Update interface specification*

| Name | User Update | | |
|---|---|---|---|
| Description | Updates the specifics for a user | | |
| Reference Code | PMS02 | | |
| Service | Platform Management | | |
| Type of interface | REST | | |
| Inputs | Methods of the interface | Parameters of the method | Return Values of the method |

---

[9] https://websockets.readthedocs.io/en/stable/

[10] https://kubernetes.io/blog/2021/08/04/kubernetes-1-22-release-announcement/

| | POST | id: The identifier of the user | 200 OK<br>201 Created<br>401 Unauthorized<br>403 Forbidden<br>404 Not Found |
|---|---|---|---|
| **Endpoint URI** | /twinergy-api/user/{id} | | |
| **Interaction with Services** | Platform Management Service | | |

*Table 22 User Deletion interface specification*

| **Name** | User Deletion | | |
|---|---|---|---|
| **Description** | Deletes a user | | |
| **Reference Code** | PMS03 | | |
| **Service** | Platform Management | | |
| **Type of interface** | REST | | |
| **Inputs** | Methods of the interface | Parameters of the method | Return Values of the method |
| | DELETE | id: The identifier of the user | 200 OK<br>401 Unauthorized<br>403 Forbidden<br>404 Not Found |
| **Endpoint URI** | /twinergy-api/user/{id} | | |
| **Interaction with Services** | Platform Management Service | | |

*Table 23 User Password Update interface specification*

| **Name** | User Password Update | | |
|---|---|---|---|
| **Description** | Updates the password of a user | | |
| **Reference Code** | PMS04 | | |
| **Service** | Platform Management | | |
| **Type of interface** | REST | | |
| **Inputs** | Methods of the interface | Parameters of the method | Return Values of the method |
| | PUT | id: The identifier of the user | 200 OK<br>401 Unauthorized<br>403 Forbidden<br>404 Not Found |
| **Endpoint URI** | /twinergy-api/user/{id}/password | | |
| **Interaction with Services** | Platform Management Service | | |

## 2.3 Services Installation and Access

### 2.3.1 Services Installation

A private repository contains the Data Collection, Data Security, Data Storage, and Platform Management Services. As all elements are already packed as Docker containers, the relevant private code repository includes explicit instructions for the service's deployment.

### 2.3.2 Licensing and access

Only closed-source code has been used in the development of the Data Collection, Data Security, Data Storage, and Platform Management services. The versions of the different services that are to be deployed will be accessed through the TwinERGY project's Core Data Management Platform.

# 3. Conclusions

Deliverable D5.7, "Data Collection, Security, Storage, and Management Services Bundles - Release 2.00," served as the updated TwinERGY Core Data Management Platform Services Bundles and detailed the additions and improvements made in comparison to Deliverable D5.5, "Data Collection, Security, Storage, and Management Services Bundles - Release 1.00." It also included a brief report summarizing the technological infrastructure, established interface specifications, and the implementation status of each service's functionalities. The deliverable defined the revised versions of the four (4) developed TwinERGY CDMP services, namely the Data collection, Data security, Data storage, and Platform Management Services, and offered additional information for each of the services, including:

- The current status of various functionalities provided from each service, expanding the functionalities introduced in D5.5 "Data Collection, Security, Storage & Management Services Bundles – Release 1.00".
- The technological infrastructure of each service, including the version used.
- The interface specifications that were created.

The current deliverable is the TwinERGY DoA-compliant 2.00 version of the Services, made available on M28, which gives back-end implementation details and prepares the ground for the upcoming 2.00 release of the TwinERGY Integrated Data Management Platform on M32 (D5.8 – June 2023). It is crucial to highlight that the TwinERGY Platform's services are continuously being improved and will do so till M32. Any updates and improvements made as the project moves forward will be taken into consideration in the subsequent TwinERGY Core Data Management Platform 2.00 release.

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

# References

9.   DoA Part A, TwinERGY Consortium Agreement No 957736

10.  TwinERGY Consortium. (2020). TwinERGY D5.5 "Data Collection, Security, Storage & Management Services Bundles – Release 1.00"

11.  TwinERGY Consortium. (2020). TwinERGY D5.8 "TwinERGY Integrated Data Management Platform – Release 2.00"

12.  DoA Part B, TwinERGY Consortium Agreement No 957736

13.  Data Catalog Vocabulary (DCAT) - Version 2 (w3.org)

14.  https://kafka.apache.org/intro

Twinergy has received funding from the European
Union's Horizon 2020 research and innovation
programme under grant agreement No. 957736

# Annex

## TwinERGY Functional Requirements

*Table 24 Functional Requirements of the TwinERGY CDMP*

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

| Req_012 | The platform shall provide the definition of metadata of available datasets. | Completed |
| --- | --- | --- |
| Req_013 | The platform should allow the user to define data anonymization routines. | Completed |
| Req_014 | The platform should allow the user to define data curation routines on the data that the user owns. | Completed |
| Req_015 | The platform shall formulate and enforce a data access control decision based on the associated data access policies. | Completed |
| Req_016 | The platform should enable the periodic data retrieval from Open Data APIs. | Completed |
| Req_017 | The platform should ensure an intuitive, easy-to-use user interface. | Ongoing |
| Req_018 | The platform should ensure a user-interface with minimum latency. | Ongoing |
| Req_019 | The platform should be able to handle big volumes of data. | Ongoing |
| Req_020 | The platform should allow users to append new data to already created datasets. | Completed |