D3.5: Data Collection, Security, Storage, Governance & Management Services Bundles – Release 1.00

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Big Energy Data Value Creation within SYNergetic enERGY-as-a-service Applications through trusted multi party data sharing over an AI big data analytics marketplace

Deliverable nº:	**D3.5**
Deliverable name:	**Data Collection, Security, Storage, Governance & Management Services Bundles – Release 1.00**
Version:	**1.0**
Release date:	**30/09/2021**
Dissemination level:	**Public**
Status:	**Final**
Author:	**UBI,** Suite5, MAG, HEDNO, ICCS, FVH, CUE, COBRA, CIRCE, KRK, KBZ, KONCAR, VERD, ENES, VTT, CAV, IPTO, EPA, ETRA, UCY

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Document history:

Version	Date of issue	Content and changes	Edited by
0.1	18/01/2021	Template version for DEM deliverables	Suite5
0.2	06/08/2021	Initial D3.5 Table of Contents	UBITECH
0.3	15/09/2021	First contributions	UBITECH
0.4	17/09/2021	Initial partners contributions	Suite5, MAG
0.5	23/09/2021	Final partners contributions	UBITECH, Suite5, MAG
0.6	24/09/2021	Full draft available for review	UBITECH
0.7	29/09/2021	Reviewed version	EPA, CIRCE
1.0	30/09/2021	Final version	UBITECH

Peer reviewed by:

Partner	Reviewer
CIRCE	Gregorio Fernandez Aznar
EPA	Panagiotis Kontogiorgos

Deliverable beneficiaries:

WP / Task	WP / Task	WP / Task
WP2 / T2.1, T2.4	WP5 / T5.1-T5.4	WP8 / T8.1, T8.2
WP3 / T3.5	WP6 / T6.1-T6.4	WP9 / T9.1
WP4 / T4.1-T4.5	WP7 / T7.1-T7.4	

Table of contents

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Abbreviations and Acronyms

Acronym	Description
AI	Artificial Intelligence
API	Application Programming Interface
CIM	Common Information Model
CSV	Comma-separated values
JSON	JavaScript Object Notation
OPE	On-Premise Environment
TSV	Tab-separated values
UI	User Interface
XML	Extensible Mark-up Language

Executive summary

The current deliverable D3.5 entitled "Data Collection, Security, Storage, Governance & Management Services Bundles – Release 1.00" constitutes a report of the efforts and the produced results till M21 of Task 3.2 "Platform Backbone Infrastructure, On-Premise and Secure Experimentation Playground Data Containers Development", Tasks 3.3 "Core Big Data Ingestion, Curation and Management Services" and Task 3.4 "Data Assets Security, Encryption and Privacy Mechanisms" of WP3 "End-to-end Interoperable Big Data Management Platform". The purpose of this deliverable is to deliver the first official release of the Services Bundles of WP3, namely the **Data Collection**, **Data Security**, **Data Storage**, **Data Governance** and **Platform Management Services Bundles**.

As the deliverable type is demonstrator per the SYNERGY Description of Action, the main objective of this deliverable is to deliver the updated technical documentation of the delivered Services Bundles, that supplements the information documented in deliverable D3.2, focusing on the updates from the previous iteration and the advancements in the implementation of the Services Bundles.

To this end, the provided technical documentation presents the complete implementation details of each Services Bundle on M21 documenting in detail:

- **The updated implementation status of the assigned functionalities of the components of each Services Bundle**. In accordance with the SYNERGY platform's architecture and the components' design specifications as documented in the SYNERGY deliverables D2.6 "SYNERGY Framework Architecture including functional, technical and communication specifications v1" and D3.2 "Data Collection, Security, Storage, Governance & Management Services Bundles – Beta Release", the deliverable reports the advancements from M16 till M21 in the implementation of each specific functionality offered by the components involved on each Services Bundle.
- **The updated technical specifications of the internal design and architecture of each Services Bundle.** The deliverable reports the updated technical documentation of the internal designs and architectures of each Services Bundle, as formulated during the implementation phase from M16 till M21, highlighting the updates from the previous iteration where needed, as well as the role and scope of each component within each Service Bundle along with the interactions between the various involved components.

- **The exploited technology stack for the implementation of each Services Bundle.** The deliverable presents the updated list of technologies, tools and frameworks which were leveraged for the implementation of each Services Bundle, listing also their exact versions and license details.
- **The updated technical documentation of the technical interfaces offered by each Services Bundle.** The deliverable provided the updated technical details of the provided APIs utilising the dominant Swagger framework.
- **The updated installation instructions for each Services Bundle.** The deliverable presents the updated details related to the deployment of each Services Bundle within the SYNERGY platform.
- **The assumptions made and possible restrictions identified.** The deliverable documents the updated documentation of the various assumptions that were made during the implementation phase of each Services Bundle till M21, as well as the imposed restrictions.
- **The updated licensing and access details of the delivered Services Bundles.** The deliverable documents the updated licensing information, as well as the latest access information, for each produced software artefact.
- **The advancements from the previous release of each Services Bundle.** The deliverable highlights the updates and enhancements that have been introduced in Release 1.00 in comparison to the Beta release of each Services Bundle.
- **The detailed implementation plan of the upcoming functionalities of each of the components of the Services Bundles.** The deliverable presents the updates in the list of planned functionalities per component involved on each Services Bundle for the upcoming release.

The deliverable constitutes the second iteration of the report of the work performed within the context of tasks T3.2, T3.3 and T3.4 and presents the Release 1.00 of the Services Bundles of WP3 on M21. The delivered Services Bundles, along with the ones delivered within the context of WP4 “Big Data Analytics and Data Sharing Mechanisms”, will be the main pillars of the first official release of the SYNERGY platform that will be documented with deliverable D3.6 “SYNERGY Integrated Platform & Open APIs – Release 1.00” on M24. As the implementation of the Service Bundles is living process that will last until M33 as per the SYNERGY Description of Action, Tasks 3.2, 3.3 and 3.4 will remain active in order to continuously provide updates and enhancements, and produce the final release of Services Bundles, namely Release 2.00, which will be documented with deliverable D3.7 “Data Collection, Security, Storage, Governance & Management Services Bundles – Release 2.00”. The final iteration of this deliverable will contain the final complete documentation of the Services Bundles and will contain all optimisations and enhancements that will be introduced during the last implementation phase in order to address any new requirements that may arise, as well as to better address the ones collected and documented in deliverable D2.2 “End-user and Business requirements analysis for big data-driven innovative energy services and ecosystems v2”, and the demo partners’ and energy applications’ feedback from their experience with the SYNERGY Platform.

1 Introduction

1.1 Purpose of the document

This deliverable presents the efforts undertaken within the context of the T3.2 “Platform Backbone Infrastructure, On-Premise and Secure Experimentation Playground Data Containers Development”, T3.3 “Core Big Data Ingestion, Curation and Management Services” and T3.4 “Data Assets Security, Encryption and Privacy Mechanisms” of WP3 ”End-to-end Interoperable Big Data Management Platform”. Hence, its main purpose is to deliver the first official release (Release 1.00) of the Data Services Bundles developed within the context of WP3, namely the Data Collection, Data Security, Data Storage, Data Governance and Platform Management Services Bundles. Towards this end, the document at hand builds directly on top of its previous iteration in order to provide the updated technical documentation and implementation details for each of the aforementioned Data Services Bundles.

1.2 Scope of the document

The deliverable D3.5 “Data Collection, Security, Storage, Governance & Management Services Bundles – Release 1.00” documents the work performed within the context of T3.2, T3.3 and T3.4 of WP3 till M21. It constitutes the second iteration of the report that documents the outcomes of the specific tasks during the period from M16 till M21 and delivers the first official release of the Data Collection, Data Security, Data Storage, Data Governance and Platform Management Services Bundles on M21 in accordance with the SYNERGY Description of Action.

Towards this end, the scope of the deliverable is to deliver the updated detailed technical documentation of the first official versions that constitute the Release 1.00 of each Data Services Bundle produced in WP3. In detail, the deliverable builds on top of the information included in the previous iteration of the deliverable, namely deliverable D3.5, in order to report the updated implementation details of the services involved in the data check-in process that compose the Data Collection Services Bundle along with the services offering the data safeguarding, data privacy preservation and data access control functionalities that compose the Data Security Services Bundle. In the same manner, the documents reports the updated implementation details of the Data Storage Services Bundle that consists of several storage and indexing tools which are leveraged as the wide range of storage solutions of the SYNERGY platform, as well as of the services composing the Data Governance Services Bundle that offers the required set of data management-related functionalities within the SYNERGY Cloud Infrastructure. Finally, the deliverable reports the updated implementation details of the Platform Management Services Bundle which is composed of complementary components that provide the services related to the effective and secure management of the underlying infrastructures.

Following the same approach as in the previous iteration, the deliverable provides for each service bundle the updated status of implementation of the related functionalities, as well as the technical details of the internal design and architecture of each Services bundle. In addition to this, the deliverable presents the updated list of frameworks and tools which are leveraged in the implementation of each bundle, as well as the updated technical documentation of the offered APIs by each Services Bundle. Moreover, the latest installation instructions are documented together with the updated information with regards to the assumptions made or restrictions imposed. Additionally, the deliverable reports the updated licensing and access information of each Services Bundle. The deliverable presents also the updates from the previous release, namely the Beta Release, highlighting the advancements made during this period. Finally, the deliverable providing an overview of the planned functionalities of each Services Bundle for their upcoming release.

The deliverable D3.5 aims at building upon the outcomes of the work performed within the context of WP2. In detail, during the implementation of the first official release of the Services Bundles of WP3 the technical requirements and the use cases of the SYNERGY project, as documented in deliverable D2.1 “End-user and Business requirements analysis for big data-driven innovative energy services and ecosystems” were taken as input along with the detailed design specifications of the components of the integrated SYNERGY platform, as documented in deliverable D2.6 “SYNERGY Framework Architecture including functional, technical and communication specifications v1”.

As the deliverable type of D3.5 is demonstrator per the SYNERGY Description of Action, the current deliverable is the second iteration of the accompanying report that documents the first official release of the Services Bundles which are implemented and delivered within the context of WP3, documenting the advancements from the first iteration for the beta release. The Services Bundles of WP3 are delivered hand-in-hand with the respective Data Services Bundles of WP4 “Big Data Analytics and Data Sharing Mechanisms” and are providing the pillars of the first release of the SYNERGY platform that will be documented with deliverable D3.6 “SYNERGY Integrated Platform & Open APIs – Release 1.00” on M24 as per the SYNERGY Description of Action.

1.3 Structure of the document

Sections 2, 3, 4, 5 and 6 are following the same structure in order to present the detailed technical documentation of the delivered Services Bundles in a unified manner. Hence, at first the implementation status of the functionalities of the specific bundle are documented. Then, the internal design and architecture of the specific bundle is documented, followed by the exploited technology stack, the offered API's documentation and the installation instructions. Finally, the assumptions made and restrictions imposed are documented, the updates from the previous release are highlighted and the planned features of the specific bundle are presented. To this end, the structure of the document is organized as follows:

- Section 2 documents the detailed technical documentation of the Data Collection Services Bundle.
- Section 3 documents the detailed technical documentation of the Data Security Services Bundle.
- Section 4 documents the detailed technical documentation of Data Storage Services Bundle.
- Section 5 documents the detailed technical documentation of Data Governance Services Bundle.
- Section 6 documents the detailed technical documentation of Platform Management Services Bundle.
- Section 7 concludes this deliverable D3.5 "Data Collection, Security, Storage, Governance & Management Services Bundles – Release 1.00" and summarises the next steps.

2 Data Collection Services Bundle Release 1.00

2.1 Overview

The SYNERGY Data Collection Services Bundle offers a range of functionalities that are involved in the data check-in process and delivered by five distinct components, namely: (a) the Data Handling Manager, (b) the Matching Prediction Engine, (c) the Data Ingestion Service, (d) the Mapping & Transformation Service, and (e) the Cleaning Service.

The Data Handling Manager enables the data check-in job configuration, according to which data providers from the electricity data value chain are able to define the data ingestion method and schedule, whereupon the data (that they own) shall arrive at the SYNERGY Platform. Furthermore, the Data Handling Manager drives the data check-in configuration process for the necessary pre-processing steps before the data are stored, i.e. data mapping and transformation, data cleaning, data anonymization and data encryption. The Data Handling Manager is practically instrumental to collect the configuration and coordinate all data ingestion and pre-processing steps at design time and make it available, through the Master Controller, to the different services (i.e. Data Ingestion Service, Mapping & Transformation Service, Cleaning Service, Anonymization Service, Encryption Engine) at execution time.

The Matching Prediction Engine uses information from the Common Information Model (CIM) Manager in order to determine the way that the fields of a dataset will be mapped to the concepts of the CIM. Such a matching between the CIM and the dataset results to a common domain-specific understanding over the data that facilitates their use for data-related services in the SYNERGY Platform.

The Data Ingestion Service collects the data into the SYNERGY Platform according to the configurations made by the data providers through the Data Handling Manager. The collection of the data can be performed based on five different ingestion modalities: (a) through direct file upload, (b) through application programming interfaces (APIs) exposed on the data provider’s side, (c) through application programming interfaces (APIs) provided by the SYNERGY Platform, (d) through streaming mechanisms offered by the SYNERGY Platform, and (e) through streaming mechanisms offered on the data provider’s side.

The Mapping & Transformation Service performs the mapping and transformation rules on the datasets, that is to rename dataset's field names, convert measurement units, and reformat data (if needed) as they were defined in the configuration file generated by the Matching Prediction Engine, in order to comply with the CIM.

The Cleaning Service provides the functionalities to clean the data ingested into the SYNERGY Platform by performing a set of cleaning rules configured by the data provider, e.g. simple value substitutions, reformatting, and outlier detection and substitution, in order to address quality issues within the data.

2.2 Implemented Functionalities

The first official release of the Data Collection Services implements a set of functionalities that complement the features already available since their beta release as documented in the SYNERGY Deliverable D3.2. The following table explains the exact status of implementation in release 1.00.

Table 1: Implemented Functionalities in Release 1.00 of the SYNERGY Data Collection Services

Feature		Status	Notes
DHM_1	Step-by-step definition of the data ingestion configuration in an intuitive manner	Implemented	Step-by-step definition of the data ingestion configuration is possible both in the back-end and in the front-end through the five (5) ingestion methods described above: Batch Files, External APIs, Platform's APIs, External PubSub mechanism (Kafka Broker), Platform's PubSub mechanism (Kafka Broker).
DHM_2	Step-by-step data mapping configuration	Implemented	The mapping and transformation configuration is available in both the back-end and the front-end, allowing the user (data provider) to check the mapping predictions (as provided by the Matching Prediction Engine), insert the necessary mapping and transformation details or update the mapped concepts to the correct concepts of the SYNERGY Common Information Model (CIM).
DHM_3	Step-by-step data cleaning configuration	Implemented	The cleaning configuration is fully supported at design time, allowing the user (data provider) to define a range of cleaning rules depending on the field's data type in order to declare how incomplete, incorrect, inaccurate or irrelevant parts of the data should be handled.
DHM_4	Step-by-step data anonymisation configuration	Implemented	The anonymisation configuration is fully supported, allowing for the definition of anonymisation rules for "identifying", "quasi-identifying" and "sensitive" columns/fields in the data by the data providers. "Privacy-risky" columns are identified in the SYNERGY Common Information Model (CIM) and appropriate checks are in place to prevent the user from finalizing the anonymisation configuration

Feature		Status	Notes
			without defining how they will be handled by the SYNERGY Platform.
DHM_5	Step-by-step data encryption configuration	Implemented	The encryption configuration has been defined and is available for batch file upload through the On-Premise Environment only. The data asset provider can also set their encryption parameters, selecting which columns/fields will be indexed.
DHM_6	Data Storage configuration in a flexible and informed manner	Implemented	The storage configuration has been defined and allows data providers to define the permanent data storage where their data will be persisted (either on the SYNERGY Cloud infrastructure or in their own On-Premise Environment) if they have opted for local processing of their data. The intermediate, non-permanent data storage for processing their data is dependent on whether the data check-in job is executed in the cloud or on-premise.
DHM_7	Secure handling of advanced authentication aspects, while protecting sensitive data	Implemented	Secure handling of sensitive parameters is ensured through the presence of a vault that is responsible for tightly controlling access to such resources.
DHM_8	Lifecycle management of data check-in configurations	Implemented	The functionalities for updating and / or deleting a data check-in job configuration have been implemented. They are accompanied by appropriate consistency checks and constraints (e.g. which parts of the configuration can be updated and when). For files upload, a data append functionality is also available to the users as part of the update mechanism for the resulting data assets (since in the rest of the data ingestion methods it is enabled by default).
MPE_1	Generation of proposed mapping from the dataset fields to the CIM concepts	Partially implemented	Linguistic and sample-based techniques to match a data sample to the corresponding CIM concepts have been implemented, creating appropriate mapping predictions and a confidence score. Such techniques have been optimized for the final CIM.
MPE_2	Updates and refinements to the generated mapping	Implemented	Appropriate front-end and back-end functionalities to support the update of the mapping configuration with complementary information have been delivered.
MPE_3	Easy exploration of the CIM concepts	Implemented	Appropriate endpoints to retrieve the latest version of the SYNERGY CIM and selected concepts have been created.
MPE_4	Mapping validation	Implemented	The appropriate mapping validation checks are in place in order to properly guide the user to correct any invalid mappings or provide any necessary complementary information.
MPE_5	Mapping template export, update and re-use	Not implemented	(Functionality to be eventually provided in release 2.00 as it requires significant changes across different components)

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Feature		Status	Notes
MPE_6	Mapping template revision once a new CIM version is available	Implemented	All necessary checks for traceability of the changes introduced between different CIM versions and whether they affect an incomplete mapping, are implemented. Such checks result into appropriate actions taking into consideration whether there are non-backwards compatible changes (e.g. removal of a concept or a field included in a draft mapping) or only backwards-compatible changes (e.g. updates in the definition of a concept or a field used in a draft mapping). All finalized mapping configurations are not affected by the changes in the CIM versions.
DIS_1	Flexible configuration of data ingestion process	Implemented	The data ingestion configuration (that is provided through the Data Handling Manager) is complete with all necessary information and appropriately utilized by the Data Ingestion Service to perform its execution.
DIS_2	Data collection from files	Implemented	Data collection for different types of files (e.g. csv/tsv, json, xml) that are to be processed is complete. Other file types (e.g. images) are stored as-is without further processing. The configuration provided and the service implemented allows for updating the data collected through files, by appending new data to the existing data asset.
DIS_3	Data collection from APIs	Implemented	Data collection from APIs that are exposed from 3rd parties (i.e. electricity data value chain stakeholders, open data portals), but also from the SYNERGY Platform APIs is possible through different customized configurations (for retrieving vs pushing data). The configuration provided and the service implemented are flexible and, especially in the case of 3rd party APIs, have tried to anticipate support for as many APIs as possible by providing different options for authentication, pagination, query parameters configuration and selection of data.
DIS_4	Data collection from PubSub mechanisms	Implemented	Data collection through the SYNERGY Platform's PubSub mechanism and the data provider's PubSub mechanism is possible through customized configurations. Kafka is the PubSub mechanism that is supported in both cases.
DIS_5	Selection, pre-processing and storage of data payload subset	Implemented	The current back-end and front-end implementation clearly separates the data that are ingested from the data that need to be processed in order to allow for appropriate user selection (from external APIs and external PubSub mechanisms without requiring any changes in their implementation). The emerging data payload from data ingestion is temporarily stored in an object store and available to be consumed by the subsequent services in the data check-in pipeline. The necessary information among the services is passed through an appropriate messaging mechanism that has been implemented in the backend.
DIS_6	Secure and reliable data transfer	Implemented	All necessary mechanisms to efficiently transfer the datasets an organisation owns or has legitimately acquired (according to an active smart contract in the Contract Lifecycle Manager) to the Secure Experimentation Playground (under

Feature		Status	Notes
			the coordination of the Master Controller) have been implemented.
MTS_1	Mapping of ingested data to the SYNERGY CIM	Implemented	The data mapping and transformation service is complete and its implementation is based on the mapping configuration (that is provided through the Data Handling Manager).
MTS_2	Transformation of ingested data to comply with the SYNERGY CIM	Implemented	The mechanism to support measurement unit transformations, datetime format transformations and time-zone transformations has been implemented. Support for different measurement units, datetime formats and all time-zones is already provided.
MTS_3	Insights into the results of the transformation rules performed over the ingested data	Implemented	During the execution of the transformation rules, specific metrics are collected per field in order for the data providers to be aware of the interventions that were performed on their data, but also eventually for data consumers to be potentially informed about the changes introduced in the original data.
MTS_4	Clear failure indications of transformation rules	Implemented	The mapping and transformation service may fail with appropriate failure codes. The errors per field are also collected and are introduced to the data providers in the user interface. If the whole mapping step failed (once it was configured, without any past successful execution), then the data providers need to unlock the specific step, perform changes in the mapping configuration and re-submit it for execution.
CS_1	Flexible configuration of cleaning rules	Implemented	The data cleaning service has been implemented and is executed on demand based on the cleaning step configuration (that is provided through the Data Handling Manager).
CS_2	Data cleaning rules execution	Implemented	The cleaning rules that are applied in each field depend on its data type and allow the data providers to enforce different options: allowed value ranges, uniqueness constraints, mandatory constraints (for handling missing data values), regular expression patterns, and outliers identification. They are also accompanied by corrective measures: dropping entries and replacing values. It needs to be noted that the options that are actually allowed in the cleaning step depend on the data ingestion method.
CS_3	Feedback from the executed cleaning rules	Implemented	During the execution of the cleaning rules, different metrics are collected per field in order for the data providers to be aware of the interventions that were performed on their data, but also eventually for data consumers to be potentially informed about the changes introduced in the original data.
CS_4	Easy testing of defined cleaning rules on sample data	Implemented	Different templates have been created to translate each cleaning rule to a user-friendly language in order for a data provider to be able to understand the expected interventions on their data. It is also possible to run the

Feature		Status	Notes
			cleaning rules on the sample data in order to see such interventions in practice before finalizing the cleaning step.

2.3 Internal Design and Architecture

The Data Collection Services extend over the Data Presentation layer (that is currently ongoing towards the integrated SYNERGY Platform Release 1.00 on M24, but its beta release has been already presented in D3.4 on M18), the Business Logic layer (separated between back-end and services) and the Data Access layer (that essentially represents the Data Storage Services). As shown in the Figure 1, the internal architecture that the Data Collection Services follow indicates how the different components, namely the Data Handling Manager, the Matching Prediction Engine, the Data Ingestion Service, the Mapping & Transformation Service, and the Cleaning Service, are positioned across the different layers.

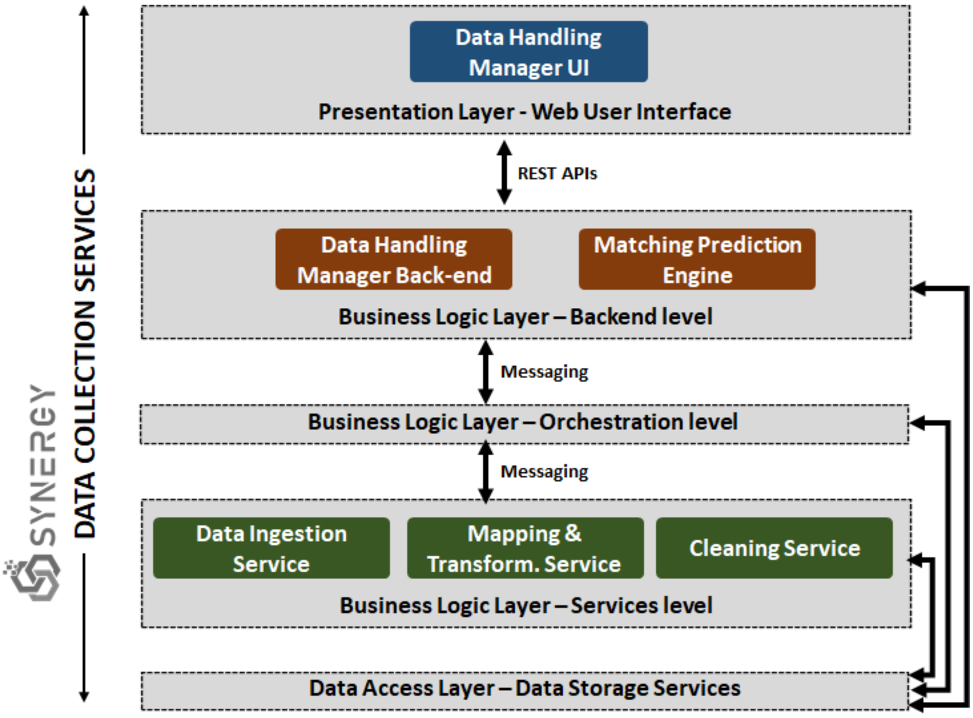

SYNERGY
DATA COLLECTION SERVICES
Data Handling Manager UI
Presentation Layer - Web User Interface
REST APIs
Data Handling Manager Back-end
Matching Prediction Engine
Business Logic Layer – Backend level
Messaging
Business Logic Layer – Orchestration level
Messaging
Data Ingestion Service
Mapping & Transform. Service
Cleaning Service
Business Logic Layer – Services level
Data Access Layer – Data Storage Services

Figure 1: SYNERGY Data Collection Services in Layers

2.4 Technology Stack and Implementation Tools

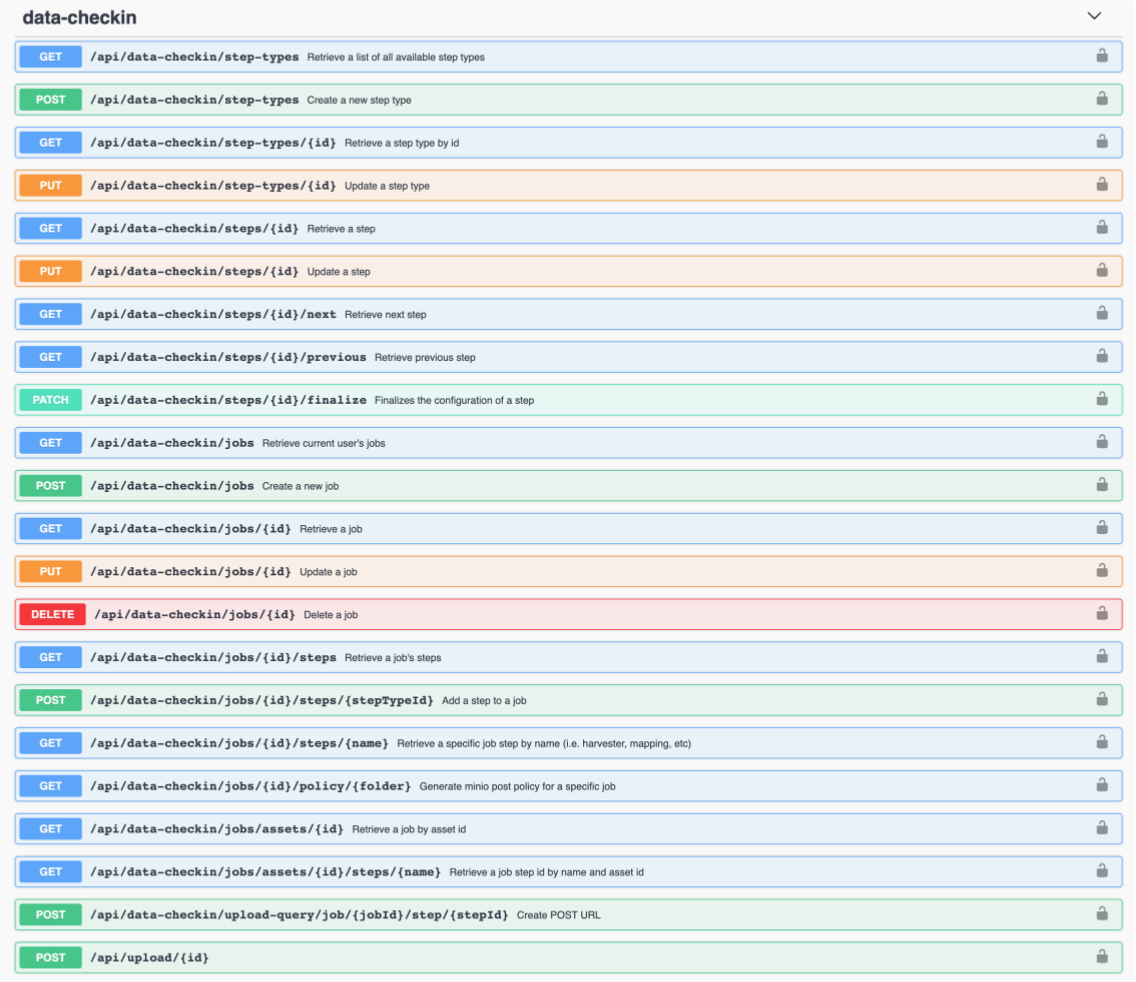
The Data Collection Services are written in TypeScript and Python 3.8.2 and, in their beta release, leverage a number of open source technologies as depicted in the following table.

Table 2: Technology Stack for the SYNERGY Data Collection Services

Library	Version	License
Nest NodeJS Web Framework	12	MIT
TypeORM	-	MIT
Flask	1.1.1	BSD 3-Clause
Flask RESTful extension	0.3.8	BSD 3-Clause
Flask CORS support	3.0.8	MIT
Vue.js	2.6.11	MIT
TailwindCSS	-	MIT
Pandas	1.0.3	BSD 3-Clause
Pika	1.1.0	BSD 3-Clause
NumPy	1.18.1	BSD
Kafka	2.7.0	Apache License 2.0
Zookeeper	3.6.2	Apache License 2.0

2.5 API Documentation

The Data Collection Services communicate with the other components and services in the SYNERGY Platform through the Master Controller's messaging functionality, as well as through selected APIs that have been created and are documented in Swagger as depicted in the following figures.



data-checkin

GET /api/data-checkin/step-types Retrieve a list of all available step types

POST /api/data-checkin/step-types Create a new step type

GET /api/data-checkin/step-types/{id} Retrieve a step type by id

PUT /api/data-checkin/step-types/{id} Update a step type

GET /api/data-checkin/steps/{id} Retrieve a step

PUT /api/data-checkin/steps/{id} Update a step

GET /api/data-checkin/steps/{id}/next Retrieve next step

GET /api/data-checkin/steps/{id}/previous Retrieve previous step

PATCH /api/data-checkin/steps/{id}/finalize Finalizes the configuration of a step

GET /api/data-checkin/jobs Retrieve current user's jobs

POST /api/data-checkin/jobs Create a new job

GET /api/data-checkin/jobs/{id} Retrieve a job

PUT /api/data-checkin/jobs/{id} Update a job

DELETE /api/data-checkin/jobs/{id} Delete a job

GET /api/data-checkin/jobs/{id}/steps Retrieve a job's steps

POST /api/data-checkin/jobs/{id}/steps/{stepTypeId} Add a step to a job

GET /api/data-checkin/jobs/{id}/steps/{name} Retrieve a specific job step by name (i.e. harvester, mapping, etc)

GET /api/data-checkin/jobs/{id}/policy/{folder} Generate minio post policy for a specific job

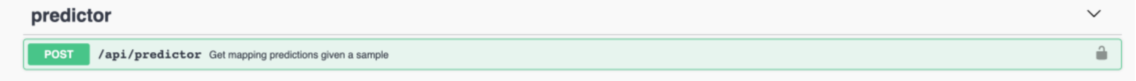
GET /api/data-checkin/jobs/assets/{id} Retrieve a job by asset id

GET /api/data-checkin/jobs/assets/{id}/steps/{name} Retrieve a job step id by name and asset id

POST /api/data-checkin/upload-query/job/{jobId}/step/{stepId} Create POST URL

POST /api/upload/{id}

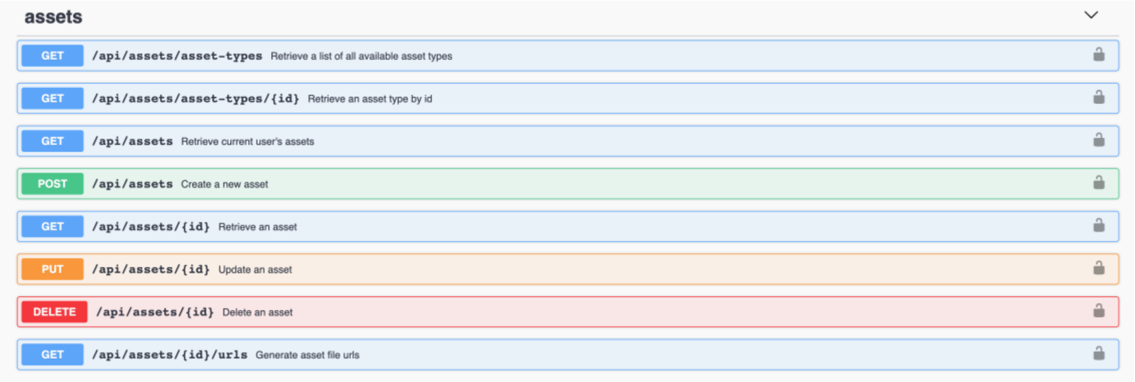
Figure 2: Swagger API Documentation for Data Check-in Jobs (Data Handling Manager)



predictor

POST /api/predictor Get mapping predictions given a sample

Figure 3: Swagger API Documentation for Mapping Prediction Engine



assets

GET `/api/assets/asset-types` Retrieve a list of all available asset types

GET `/api/assets/asset-types/{id}` Retrieve an asset type by id

GET `/api/assets` Retrieve current user's assets

POST `/api/assets` Create a new asset

GET `/api/assets/{id}` Retrieve an asset

PUT `/api/assets/{id}` Update an asset

DELETE `/api/assets/{id}` Delete an asset

GET `/api/assets/{id}/urls` Generate asset file urls

Figure 4: Swagger API Documentation for Assets (Data Handling Manager)

2.6 Installation Instructions

The Data Collection Services are served as a web application in the SYNERGY Cloud Platform, but their use can be also accompanied by the installation of the On-Premise Environment.

Detailed instructions for the Data Collection Services deployment are provided in the related private code repository even though all subcomponents are already packaged as Docker containers.

2.7 Assumptions and Restrictions

In the current release of the Data Collection Services (where the development activities for the different components are still ongoing), certain assumptions (that inevitably represent restrictions in certain cases) have been made:

- All data ingestion modalities are offered in the SYNERGY Cloud Platform. When it comes to the Server and Edge On-Premise Environments (OPE), only selected data ingestion methods are supported. For example, only files upload is currently supported in the Server OPE since core features that are provided by the Server OPE, such as end-to-end encryption and on-premise storage, are mostly critical for batch data.
- The application of the pre-processing steps affects the real-time availability of data ingested through streaming mechanisms or through APIs in the SYNERGY Platform.
- The data ingestion method naturally limits the options offered in the cleaning step to avoid inconsistency of the data stored, e.g. replacing outliers with min/max/mean values in data ingested through streaming mechanisms or through APIs cannot be calculated on-the-go while

the service is executed, and is thus not offered as an option during the cleaning configuration in such cases.

2.8 Licensing and Access

Each component of the Data Collection Services bundle is closed source. The deployed version of the Data Collection Services bundle is made available through the integrated SYNERGY platform (as documented in D3.4 for the beta release and in the upcoming D3.6 for Release 1.00).

2.9 Updates from Beta Release

The main changes that have been introduced in release 1.00 of the Data Collection Services in comparison to the beta release (that was documented in D3.2) include:

- Front-end implementation of all services: (a) the Data Handling Manager, (b) the Matching Prediction Engine, (c) the Data Ingestion Service, (d) the Mapping & Transformation Service, and (e) the Cleaning Service.
- Updates in the data check-in job configuration that were considered as necessary during the integration of the different services and components.
- Configuration of the Mapping Prediction Engine based on the final structure and concepts of the SYNERGY Common Information Model (CIM).
- End-to-end tests and adjustments on how sensitive parameters are handled in the vault to facilitate all services that need to utilize them.
- Refined mapping and cleaning validation checks as well as updated mapping and cleaning reports.
- Restarting a failed data check-in step that has not been successfully executed in order to allow its data provider to revise the specific step's configuration and re-execute the specific step and its forthcoming steps.
- Secure and reliable data transfer (as Parquet files) across the different tiers of the SYNERGY architecture, especially to the organization's Secure Experimentation Playground.

2.10 Planned Features for Next Release

The first official release of the Data Collection Services is currently being complemented by front-end and back-end updates that are considered necessary during the integration of the different services and components in the SYNERGY Integrated Platform Release 1.00, as mentioned in the following table.

Table 3: Planned Functionalities in the next release of the SYNERGY Data Collection Services

Feature	Notes
DHM_1-8, MPE_2-4, MPE_5, DIS_1-5	Updates and enhancements as needed based on integration tests and stress testing of all provided functionalities towards the official release of the SYNERGY Integrated Platform on M24 as well as through feedback initially provided by the SYNERGY demo partners and the energy application developers (who are the only beta platform users until access is provided to external stakeholders).
MPE_1	The supported schema matching techniques are expected to be updated and further fine-tuned as actual data start being checked in in the SYNERGY Platform.
MPE_5	The mapping template support in order to allow reusing and updating an existing mapping configuration to a new data asset will be added in release 2.00.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

3 Data Security Services Bundle Release 1.00

3.1 Overview

The SYNERGY Data Security Services Bundle offers data security-related functionalities in order to safeguard and protect the privacy of data assets, to eliminate the risk of unauthorized data access or data leakage, and to allow full control over their access on the data asset providers. These core functionalities are included in three components that constitute the Data Security Services Bundle, namely: (a) the Anonymisation Service, (b) the Encryption Engine, and (c) Access Policy Engine.

The Anonymisation Service is responsible for allowing data providers to handle any potentially sensitive information that might be available within the datasets, by providing the functionalities needed to protect such information by anonymizing a part (i.e., fields) of the dataset. Note that, although the Anonymisation Service is included in the Data Security Services Bundle, the configuration of data anonymisation is performed through the Data Handling Manager as part of the data check-in job configuration.

The Encryption Engine is responsible for offering full protection of the data that were loaded to the SYNERGY Platform by the data providers through their On-Premise Environment. These mechanisms aim at eliminating the risk of unauthorized data access or data leakage, by encrypting the input datasets (or part of them), based on data encryption rules that were defined by the data providers. In addition, the Encryption Engine provides decryption functionalities on the data once they are transferred to the appropriate location according to the terms of an active data contract (e.g. in the Secure Experimentation Playground of the data asset consumer) so that they can be accessed by authorized users.

The Access Policy Engine allows data asset providers to define the access rules that permit or deny the access requests on their data assets within the SYNERGY Platform. During the search of a data asset by a data consumer within the SYNERGY Data & AI Marketplace as well as during the retrieval of data through the API Gateway, the Access Policy Engine evaluates the applicable access policies in order to filter (hide or make visible) the corresponding results.

3.2 Implemented Functionalities

The first official release of the Data Security Services implements a set of functionalities that complement the features already available since their beta release as documented in the SYNERGY Deliverable D3.2. The following table explains the exact status of implementation in release 1.00.

Table 4: Implemented Functionalities in Release 1.00 of the SYNERGY Data Security Services

Feature		Status	Notes
AS_1	Identification of fields that need anonymisation	Implemented	The characterisation of each field that appears in a dataset as identifier, quasi-identifier and sensitive is supported in the data check-in job configuration and the anonymisation service.
AS_2	Flexible configuration of anonymisation rules	Implemented	The configuration of the anonymisation rules has been completed and integrated with the Data Handler Manager.
AS_3	Data anonymisation rules execution	Implemented	The anonymisation step execution has been developed in accordance with the anonymisation configuration and the execution of all data check-in services. The anonymisation method that is currently supported in the Anonymisation Service is k-anonymity that allows different grouping, masking, generalisation techniques to be applied at different levels over selected fields depending on their data type.
AS_4	Insights extracted from the executed anonymisation rules	Implemented	During the execution of the anonymisation rules, different metrics are collected (e.g. the level that anonymisation had to reach per field) and the achieved information loss is computed. In case the achieved information loss is above the expected information loss by the data provider or the desired k-anonymity is not achieved, appropriate triggers are in place to cause the failure of the anonymisation step.
ES_1	Flexible configuration of encryption rules	Implemented	The configuration of the encryption rules is implemented in collaboration with the Data Handling Manager and followed by the Encryption Engine.
ES_2	Data encryption rules execution	Implemented	Symmetric encryption over the data has been implemented based on state-of-the art approaches. The encrypted data are stored as a binary file (that is transferred from an On-Premise Environment to the Core Cloud Platform).
ES_3	Computations on data prior to encryption for search purposes	Implemented	Different computations are performed on each field depending on its data type prior to the execution of the encryption rules.
ES_4	Data decryption	Implemented	Decryption of the encrypted binary file files is possible in the cloud. Asymmetric encryption is applied on the key exchange (for encrypting/decrypting the data payload) between the On-Premise Environment and the Core Cloud Platform.

Feature		Status	Notes
ES_5	Key revocation	Not implemented	(Functionality to be eventually provided in release 2.00 as it requires significant changes across different components in the Data Collection, Security, Sharing and Analytics Services Bundles in addition to the Platform Management Services)
APE_1	Flexible definition, configuration, and update of data asset access policies	Implemented	Access policies can be fully defined and stored, following an overarching allow-all or deny-all access strategy. They practically take the form of exceptions that are expressed over properties of the requestor and can be combined into complex rules using Boolean logic.
APE_2	Enforcement of asset access policies	Implemented	The mechanism for enforcing access policies has been developed and appropriate endpoints have been created while it has been fully integrated with the other components in the SYNERGY architecture (e.g. Data & AI Marketplace, Contracts Lifecycle Manager, API Gateway).

3.3 Internal Design and Architecture

The Data Security Services extend over the Data Presentation layer (that is currently under development within the SYNERGY Integrated Platform), the Business Logic layer (separated between back-end and services) and the Data Access layer (that essentially represents the Data Storage Services). As shown in the Figure 5, the internal architecture that the Data Security Services follow indicates how the different components, namely the Anonymisation Service, the Encryption Engine, and the Access Policy Engine, are positioned across the different layers. It needs to be noted that the Data Handling Manager that conceptually belongs to the Data Collection Services also appears in the architecture as it offers front-end and back-end functionalities that are leveraged by all services related to data check-in.

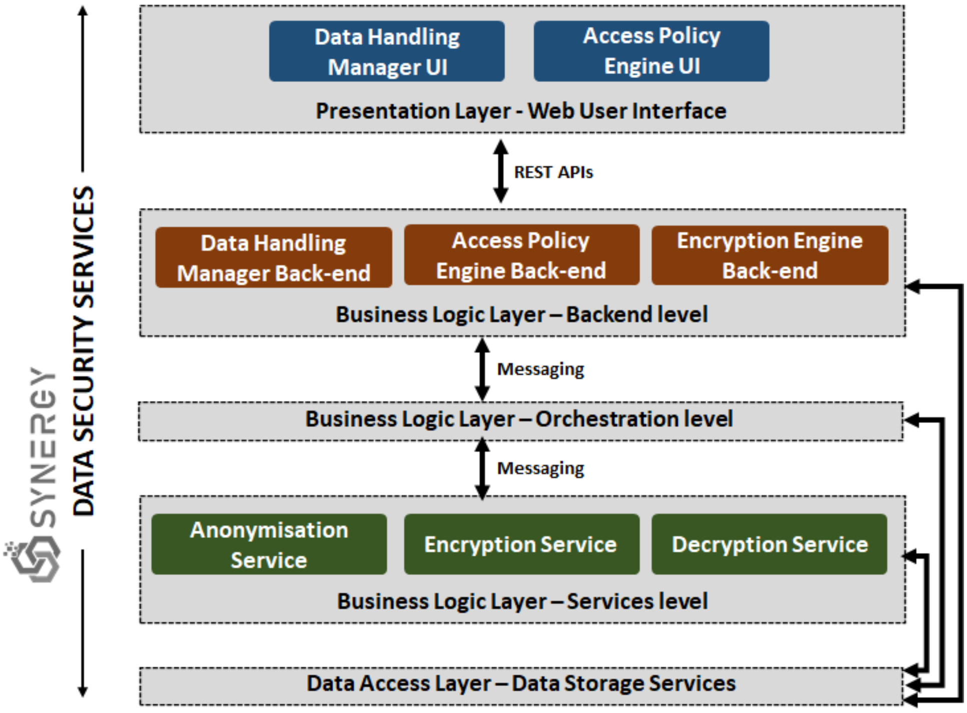

SYNERGY
DATA SECURITY SERVICES
Data Handling Manager UI
Access Policy Engine UI
Presentation Layer - Web User Interface
REST APIs
Data Handling Manager Back-end
Access Policy Engine Back-end
Encryption Engine Back-end
Business Logic Layer – Backend level
Messaging
Business Logic Layer – Orchestration level
Messaging
Anonymisation Service
Encryption Service
Decryption Service
Business Logic Layer – Services level
Data Access Layer – Data Storage Services

Figure 5: SYNERGY Data Security Services in Layers

3.4 Technology Stack and Implementation Tools

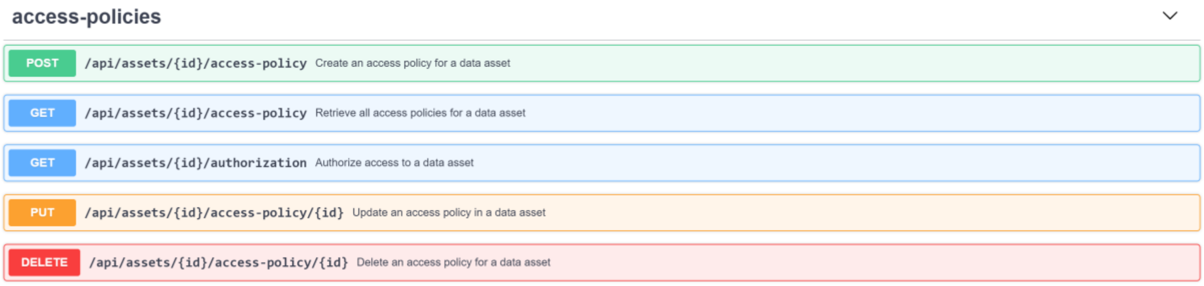
The Data Security Services are written in TypeScript and Python 3.8.2 and, in their release 1.00, leverage a number of open source technologies as depicted in the following table.

Table 5: Technology Stack for the SYNERGY Data Security Services

Library	Version	License
Nest NodeJS Web Framework	12	MIT
TypeORM	-	MIT
Flask	1.1.1	BSD 3-Clause
Flask RESTful extension	0.3.8	BSD 3-Clause
Flask CORS support	3.0.8	MIT
Pandas	1.0.3	BSD 3-Clause
Pika	1.1.0	BSD 3-Clause
NumPy	1.18.1	BSD
casbin	5.2.3	Apache 2.0 License

3.5 API Documentation

The Data Security Services communicate with the other components and services in the SYNERGY Platform through the Master Controller's messaging functionality, as well as through selected APIs that have been created and are documented in Swagger as depicted in the following figure.



access-policies

POST /api/assets/{id}/access-policy Create an access policy for a data asset

GET /api/assets/{id}/access-policy Retrieve all access policies for a data asset

GET /api/assets/{id}/authorization Authorize access to a data asset

PUT /api/assets/{id}/access-policy/{id} Update an access policy in a data asset

DELETE /api/assets/{id}/access-policy/{id} Delete an access policy for a data asset

Figure 6: Swagger API Documentation for Access Policies (Access Policy Engine)

It needs to be noted that the configuration for the Anonymisation and Encryption steps utilize the APIs that have been defined in Figure 2 (Section 2.5) for Data Check-in Jobs.

3.6 Installation Instructions

The Data Security Services are served as a web application in the SYNERGY Cloud Platform, but their use can be also accompanied by the installation of the On-Premise Environment.

Detailed instructions for the deployment of the Data Security Services are provided in the related private code repository even though all subcomponents are already packaged as Docker containers.

3.7 Assumptions and Restrictions

In the current release of the Data Security Services, where the development activities for the different components are still ongoing, certain assumptions (that inevitably represent restrictions in certain cases) have been made:

- The data ingestion method naturally limits whether the anonymisation step can be applied or not to avoid inconsistency of the data stored. Anonymisation is practically allowed only in the case of batch data upload through files since all data become available at once. The existence of the anonymisation step restricts the update of the data ingested since any attempt to append new data may also result in inconsistent data.

- Data encryption is permitted only through the On-Premise Environment (OPE) and concerns only files upload in order to avoid any further complexity on the encryption key storage (that should be kept "locally").
- Data that may be encrypted for transfer between the On-Premise Environment and the Cloud Environment are decrypted in the Secure Experimentation Playground of the organization to whom they belong in order to be readily available for running analytics.

3.8 Licensing and Access

Each component of the Data Security Services bundle is closed source. The deployed version of the Data Security Services bundle is made available through the integrated SYNERGY platform (as documented in D3.4 for the beta release and in the upcoming D3.6 for Release 1.00).

3.9 Updates from Beta Release

The main changes that have been introduced in release 1.00 of the Data Security Services in comparison to the beta release (that was documented in D3.2) include:

- Front-end implementation of the Access Policy Engine, as well as of the reporting functionalities for the Anonymisation Service and the Encryption Engine.
- Development of the key exchange mechanism, and end-to-end tests and adjustments on how symmetric encryption (for the data) and asymmetric encryption (for the symmetric key exchange) are applied between the different layers of the SYNERGY architecture (On-Premise Environment, Secure Experimentation Playgrounds).
- Refined anonymisation validation checks as well as updated anonymisation reports.
- The encrypted data are stored in a binary file to address the significant volume overhead that was introduced in the beta release.
- Restarting a failed data anonymisation and encryption step that have not been successfully executed in order to allow its data provider to revise the specific steps' configuration and re-execute them.

3.10 Planned Features for Next Release

The first official release of the Data Security Services is currently being complemented by front-end and back-end updates that are considered necessary during the integration of the different services and components in the SYNERGY Integrated Platform Release 1.00, as mentioned in the following table.

Table 6: Planned Functionalities in the next release of the SYNERGY Data Security Services

Feature	Notes
AS_1-4, ES_1-4, APE_1-2	Updates and enhancements as needed based on integration tests and stress testing of all provided functionalities towards the official release of the SYNERGY Integrated Platform on M24 as well as through feedback initially provided by the SYNERGY demo partners and the energy application developers (who are the only beta platform users until access is provided to external stakeholders).
ES_5	Support for key revocation (e.g. when a contract expires) will be added in release 2.00.

4 Data Storage Services Bundle Release 1.00

4.1 Overview

The Data Storage Services Bundle consists of several storage and indexing tools that are responsible to maintain, among others, anything related to configurations, data assets, contracts, sensitive data, and logs. In particular, this service bundle is responsible for a wide range of storage solutions used in the SYNERGY Platform, that are described in brief below.

a) Processed dataset (along with the processed sample data) permanent storage in an inherently scalable database.

b) Metadata storage, maintaining information about dataset metadata, and creating appropriate connections (links) between the stored datasets and their relevant metadata information.

c) Analytics models storage, maintaining a dedicated storage space for pre-trained algorithms, model parameters etc.

d) Data check-in and data analysis job configurations storage, maintaining a space for storing configurations to be used when needed for logging, cloning, updating, or re-execution.

e) Temporary object storage of intermediate files, maintaining a space for storing temporary objects that are needed for potential continuation of an interrupted job, but also for storing other files (e.g. in case image files are to be uploaded).

f) Sensitive data and credentials storage, maintaining a dedicated secure database that stores and manages sensitive data such as tokens, usernames, passwords, and API keys.

g) CIM and energy vocabulary storage, maintaining a space for storing the SYNERGY CIM concepts and energy vocabularies that are used within the SYNERGY Platform.

h) Data asset contracts storage, maintaining smart data asset sharing contracts in a distributed ledger, while ensuring non-repudiation, full traceability, and privacy of the contracts' details and the involved stakeholders' information.

i) Platform's operational data storage, maintaining a space for all platform data regarding users, organizations, operations' metadata, and other operational data of the platform.

4.2 Implemented Functionalities

The first official release of the Data Storage Services includes a set of functionalities that complement the features already available since their beta release as documented in the SYNERGY Deliverable D3.2. The following table explains the exact status of implementation in release 1.00.

Table 7: Implemented Functionalities in the current release of the SYNERGY Data Collection Services

Feature		Status	Notes
DSS_1	Storage of Datasets in Trusted Data Containers	Implemented	The storage of the processed data occurs in separate MongoDB databases per organization (within the Secure Experimentation Playgrounds) in the SYNERGY Cloud Platform.
DSS_2	Metadata Storage for SYNERGY Data Assets	Implemented	The metadata that have been defined in the SYNERGY metadata model (as briefly introduced in the SYNERGY Deliverable D4.1 and D4.3) are appropriately stored and indexed in Elasticsearch.
DSS_3	Storage of AI models	Implemented	Trained AI models that have been created internally and externally to the SYNERGY Platform and become available in different formats depending on the execution library, are stored in the SYNERGY object storage in MinIO.
DSS_4	Storage of all data check-in and data analysis jobs configurations	Implemented	The configuration of data check-in jobs and data analytics jobs / pipelines is appropriately broken down in steps and stored in the SYNERGY relational storage layer in PostgreSQL.
DSS_5	Temporary object storage of intermediate files	Implemented	Any intermediate data file that is created during the data check-in or the data analytics jobs execution, is stored in the SYNERGY object storage that is powered by MinIO.
DSS_6	Storage of sensitive data and credentials	Implemented	Sensitive credentials are handled with appropriate care, stored in an encrypted form and used with the help of a Vault implementation in SYNERGY.
DSS_7	Storage of the SYNERGY CIM and energy vocabularies	Partially implemented	The different minor and major versions of the SYNERGY Common Information Model are properly stored and indexed in Elasticsearch. Although energy vocabularies can potentially be also introduced and stored, they are not yet part of the CIM thus the implementation of the storage of energy vocabularies remains open in this release.
DSS_8	Storage in a distributed Contracts Ledger	Implemented	Full traceability of the data assets contracts is ensured through a contracts distributed ledger built on Ethereum (as presented in the SYNERGY Deliverable D4.1). Initially, a “private” Ethereum network is used in SYNERGY with the option to transition to the “public” Ethereum network by the end of the SYNERGY project.
DSS_9	Centralized storage for all platform’s operational data	Implemented	The operational data of the SYNERGY Platform (e.g. organisations, users) are stored in the SYNERGY relational storage layer in PostgreSQL.

4.3 Internal Design and Architecture

The Data Storage Services concern the Data Access layer (that essentially represents the Data Storage Services). As shown in Figure 7, the internal architecture that the Data Storage Services follow indicates the different databases that have been leveraged.

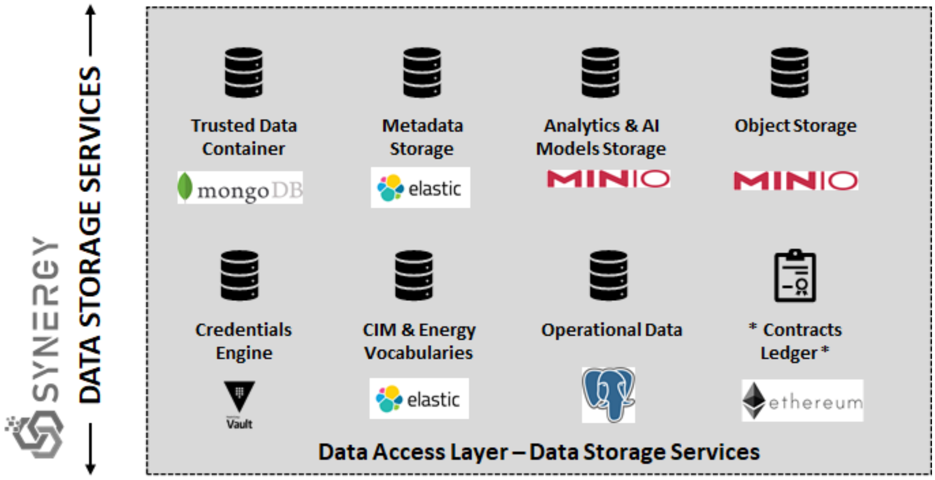


Figure 7: SYNERGY Data Storage Services in Layers

4.4 Technology Stack and Implementation Tools

The Data Storage Services, in their current release, leverage a number of open source databases as depicted in the following table.

Table 8: Technology Stack for the SYNERGY Data Storage Services

Library	Version	License
PostgreSQL	12.2	PostgreSQL License (similar to BSD/MIT)
MinIO	-	Apache License 2.0
MongoDB	4.4	Apache License 2.0
PyMongo	3.10.1	Apache License 2.0
Elasticsearch	7.10	Elastic License
Vault	-	Mozilla Public License 2.0
Ethereum	1.9.24	LGPL-3.0 License

4.5 API Documentation

There are no external APIs that are exposed by the Data Storage Services to accompany its initial version except for the APIs that have been presented in sections 2, 3, 5, 6.

4.6 Installation Instructions

Detailed instructions for the deployment of the Data Storage Services are provided in the related private code repository even though all loaders to the different databases are already packaged as Docker containers.

4.7 Assumptions and Restrictions

In SYNERGY, the Data Storage Services are designed and implemented as a multiglot data persistence layer bringing the best of breed of different data storage and indexing solutions. Even though such a decision increases the complexity of the data access layer, it was considered as essential in order to increase reliability, trust and effectiveness of the diverse end-to-end data functionalities.

In addition, Python code for AI models cannot be directly uploaded by any user in the SYNERGY platform due to security reasons.

4.8 Licensing and Access

Each component of the Data Storage Services bundle is closed source. The deployed version of the Data Storage Services bundle is made available through the integrated SYNERGY platform (as documented in D3.4 for the beta release and in the upcoming D3.6 for Release 1.00).

4.9 Updates from Beta Release

The main changes that have been introduced in release 1.00 of the Data Storage Services in comparison to the beta release (that was documented in D3.2) include:

- A cluster of MongoDB databases is set up. A dedicated MongoDB database is automatically created at the moment when the registration of an organization is approved, and is part of the Secure Experimentation Playground for the specific organization.

- Replicas of the different database instances (e.g. for MongoDB, Elasticsearch, MinIO) is set up for scalability and increased security.
- Increased functionalities for the pre-trained models' storage (to embrace models created in various libraries as described in the SYNERGY Deliverable D4.3).
- Removal of support from Gitlab as it was considered prudent from a security perspective, not to allow any code functionalities on the Analytics Workbench.

4.10 Planned Features for Next Release

The first official release of the Data Storage Services is currently being complemented by any updates that are considered necessary during the integration of the different services and components in the SYNERGY Integrated Platform Release 1.00, as mentioned in the following table.

Table 9: Planned Functionalities in the next release of the SYNERGY Data Storage Services

Feature	Notes
DSS_1 – DSS_9	Stress testing, performance and scalability tested will be extensively assessed in release 1.00 of the SYNERGY Integrated Platform.
DSS_7	Storage of energy vocabularies will be fully supported in an end-to-end manner, across different components (CIM Manager, Data Handling Manager, Analytics Workbench, Data & AI Marketplace).

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

5 Data Governance Services Bundle Release 1.00

5.1 Overview

The Data Governance Services Bundle provides a set of data management-related functionalities through the SYNERGY Cloud Infrastructure, and it consists of three main components: (a) the Master Controller, (b) the Data Lineage Service, and (c) the CIM Manager.

The Master Controller is instrumental in orchestrating the services that are executed in the SYNERGY Core Cloud Platform, the Secure Experimentation Playgrounds, and the On-Premise Environments, based on the users' (data asset provider or data asset consumer) configuration. In addition, the Master Controller is responsible for scheduling the execution of the data check-in and analytics services in an automatic manner, based on the scheduling configurations defined by the users.

The Data Lineage Service handles and tracks all the changes or actions performed on a data asset regarding its content, structure or metadata. As a result, the evolution of a data asset and its connection with other data assets are recorded and can be used by authorized users to identify any potential issues with the data asset.

The Common Information Model (CIM) Manager is responsible for maintaining and managing the lifecycle of the Common Information Model that contains domain-specific information related to the data exchanged among the electricity data value chain stakeholders, and involves its creation, validation, update and deprecation. Hence the model can be continuously enriched and updated by the model administrator in such a way that its use is consistent and sustainable across the different functionalities offered by the SYNERGY Platform.

5.2 Implemented Functionalities

The first official release of the Data Governance Services implements a set of functionalities that complement the features already available since their beta release as documented in the SYNERGY Deliverable D3.2. The following table explains the exact status of implementation in release 1.00.

Table 10: Implemented Functionalities in the current release of the SYNERGY Data Governance Services

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Feature		Status	Notes
MC_1	Cloud orchestration of services adapted to the needs of each data check-in job	Implemented	The orchestration of the execution of data check-in services in a cluster in the SYNERGY cloud is implemented. The different services are scheduled and executed, according to their configuration, while their communication is facilitated through a messaging mechanism.
MC_2	On-premise orchestration of services adapted to the needs of each data check-in job	Implemented	The execution of data check-in services in Server and Edge On-Premise Environments (that support Docker) is implemented and orchestrated by the Master Controller.
MC_3	Cloud orchestration of services adapted to the needs of each data analysis job	Implemented	The orchestration of the execution of data analytics pipelines in a cluster in the SYNERGY cloud is implemented. The pipeline is scheduled and executed, according to its configuration, while the communication of errors and status is facilitated through a messaging mechanism.
MC_4	On-premise orchestration of services adapted to the needs of each data analysis job	Implemented	The execution of data analytics pipelines in Server and Edge On-Premise Environments (that support Docker) is implemented and orchestrated by the Master Controller that resides in the cloud.
MC_5	Managing the secure transfer of data from/to the Secure Experimentation Playgrounds	Implemented	The data transfer from the Cloud Storage or the On-Premise Environment to the Secure Experimentation Playground of an organisation is implemented. For performance purposes, the Parquet format is used. If the data are uploaded in an encrypted form from the Server On-Premise Environment, they are decrypted once the secure key exchange mechanism is applied.
MC_6	Monitoring and managing the services execution status	Implemented	Appropriate messaging on the execution status is enabled between the different services through the Master Controller for the cloud and (server / edge) on-premise execution. Info, feedback and error messages are appropriately collected and stored.
DLS_1	Retrieval and presentation of dataset changes	Not implemented	(Functionality to be eventually provided in release 2.00 as full traceability of the dataset changes requires significant back-end changes that could not be introduced in this release)
DLS_2	Retrieval and presentation of connections among data assets	Partially implemented	The connections between different data assets in terms of derivative data assets that may emerge from executing a data analytics pipeline in the Analytics Workbench, are properly stored and managed in the back-end (to support the creation of multi-party data asset contracts in the Contract Lifecycle Manager).
CIMM_1	Definition, update and deprecation of model concepts, fields and hierarchies	Implemented	The lifecycle management of the SYNERGY Common Information Model, including its concepts, fields and hierarchies, has been completed from a back-end and front-end perspective. Different minor and major versions of the CIM are created based on the changes that are introduced

Feature		Status	Notes
			as different evolution rules are applied as described in the SYNERGY Deliverable D3.1.
CIMM_2	User-driven suggestions of new model concepts or updates in the existing CIM	Implemented	The necessary mechanism for proposing new concepts/fields (during the data check-in job configuration), as well as for retrieving and handling suggestions provided by users (i.e. data providers) to the CIM, has been implemented.
CIMM_3	Informed decision over CIM suggestions	Implemented	The approval or rejection of the CIM suggestions by the CIM manager have been already implemented. Relevant notifications are issued to the appropriate organizations that provided each CIM suggestion.
CIMM_4	Navigation to the CIM concepts, properties, hierarchies and relationships	Implemented	It is a functionality that is currently delivered through the mapping step configuration (in the Data Handling Manager UI) where data providers are able to navigate to the relevant CIM concepts (depending on the category they have selected) and check their fields, associated description and data type, as well as their related concepts.

5.3 Internal Design and Architecture

The Data Governance Services extend over the Data Presentation layer (that is under development as part of the beta SYNERGY integrated platform to be documented in D3.4), the Business Logic layer (separated between back-end, orchestration and services) and the Data Access layer (that essentially represents the Data Storage Services). As shown in Figure 8, the internal architecture that the Data Governance Services follow indicates how the different components, namely the Master Controller, the Data Lineage Service and the CIM Manager, are positioned across the different layers.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

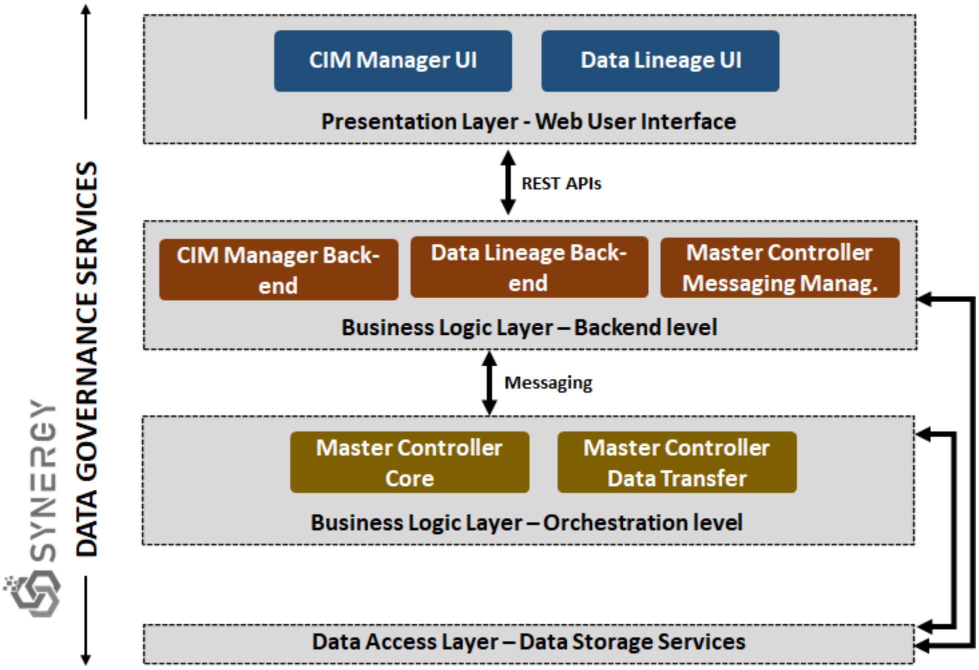

SYNERGY
DATA GOVERNANCE SERVICES
CIM Manager UI
Data Lineage UI
Presentation Layer - Web User Interface
REST APIs
CIM Manager Back-end
Data Lineage Back-end
Master Controller Messaging Manag.
Business Logic Layer – Backend level
Messaging
Master Controller Core
Master Controller Data Transfer
Business Logic Layer – Orchestration level
Data Access Layer – Data Storage Services

Figure 8: SYNERGY Data Governance Services in Layers

5.4 Technology Stack and Implementation Tools

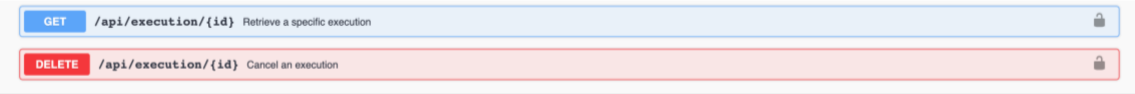
The Data Governance Services are written in TypeScript and, in their current release, leverage a number of open source technologies as depicted in the following table.

Table 11: Technology Stack for the SYNERGY Data Governance Services

Library	Version	License
Nest NodeJS Web Framework	12	MIT
TypeORM	-	MIT
Kubernetes	1.18	Apache License 2.0
RabbitMQ	3.8.2	Mozilla Public License

5.5 API Documentation

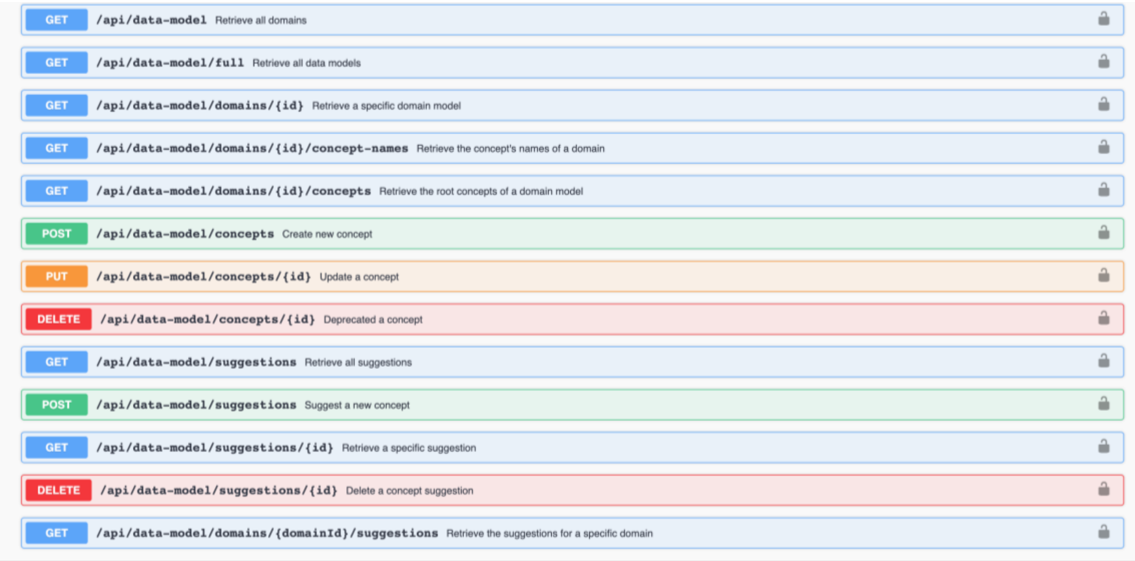
The Data Governance Services communicate with the other components and services in the SYNERGY Platform through selected APIs that have been created and are documented in Swagger as depicted in the following figures.



GET /api/execution/{id} Retrieve a specific execution

DELETE /api/execution/{id} Cancel an execution

Figure 9: Swagger API Documentation for Execution of a Data Check-in Job or a Data Analytics Job (Master Controller)



GET /api/data-model Retrieve all domains

GET /api/data-model/full Retrieve all data models

GET /api/data-model/domains/{id} Retrieve a specific domain model

GET /api/data-model/domains/{id}/concept-names Retrieve the concept's names of a domain

GET /api/data-model/domains/{id}/concepts Retrieve the root concepts of a domain model

POST /api/data-model/concepts Create new concept

PUT /api/data-model/concepts/{id} Update a concept

DELETE /api/data-model/concepts/{id} Deprecated a concept

GET /api/data-model/suggestions Retrieve all suggestions

POST /api/data-model/suggestions Suggest a new concept

GET /api/data-model/suggestions/{id} Retrieve a specific suggestion

DELETE /api/data-model/suggestions/{id} Delete a concept suggestion

GET /api/data-model/domains/{domainId}/suggestions Retrieve the suggestions for a specific domain

Figure 10: Swagger API Documentation for CIM Manager

5.6 Installation Instructions

The Data Governance Services are served as a web application in the SYNERGY Cloud Platform so no installation steps need to be taken by an end user (data asset provider or data asset consumer). Detailed instructions for their deployment are provided in the related private code repository even though all subcomponents are already packaged as Docker containers.

5.7 Assumptions and Restrictions

In the current release of the Data Governance Services (where the development activities for the different components are still ongoing), certain assumptions (that inevitably represent restrictions in certain cases) have been made:

- The execution of data check-in services in an Edge On-Premise Environment has the limitation that the edge devices (gateways) need to be able to support Docker while the data are checked in and directly stored in the cloud (without communication with a Server On-Premise Environment).

- The Master Controller operation, when it comes to the On-Premise Environments (in both the server and edge editions), is affected by their constant connectivity.

5.8 Licensing and Access

Each component of the Data Governance Services bundle is closed source. The deployed version of the Data Governance Services bundle is made available through the integrated SYNERGY platform (as documented in D3.4 for the beta release and in the upcoming D3.6 for Release 1.00).

5.9 Updates from Beta Release

The main changes that have been introduced in release 1.00 of the Data Governance Services in comparison to the beta release (that was documented in D3.2) include:

- Front-end implementation of the CIM Manager.
- End-to-end tests and adjustments on how changes in the CIM are handled across the different components and layers of the SYNERGY architecture (On-Premise Environment, Secure Experimentation Playgrounds).
- Updated schedule management in the Master Controller.
- Automated restart of services in the Master Controller (e.g. in case maintenance activities occur by the cloud provider in which the SYNERGY Platform is hosted). Option for manually restarting long running data check-in jobs (through Platform's PubSub mechanism and for the API polling option of the 3rd party APIs).
- Facilitation of the key exchange process through the Master Controller.
- Development of the headless Edge On-Premise Environment that can be installed in gateways that support Docker.
- Support for execution of data check-in jobs and data analytics pipelines in the Edge On-Premise Environments.

5.10 Planned Features for Next Release

The first official release of the Data Governance Services is currently being complemented by front-end and back-end updates that are considered necessary during the integration of the different services and components in the SYNERGY Integrated Platform Release 1.00, as mentioned in the following table.

Table 12: Planned Functionalities in the next release of the SYNERGY Data Governance Services

Feature	Notes
MC_1-6, CIMM_1-4	Updates and enhancements as needed based on integration tests towards release 1.00 of the SYNERGY Integrated Platform, as well as through stress testing of the provided functionalities.
DLS_1	This feature will become available in release 2.00 of the Data Governance Services.
DLS_2	This feature will become fully available in release 2.00 of the Data Governance Services, but its early implementation in the current release will support various platform functionalities under the hood (e.g. for multi-party contracts, for controlling access to derivative data assets).

6 Platform Management Services Bundle Release 1.00

6.1 Overview

The Platform Management Services Bundle offers a set of complementary components that allow for effective and secure management of the underlying infrastructures. It consists of five main components namely: (a) the Resources Orchestrator, (b) the Notifications Engine, (c) the Security, Authentication & Authorisation Engine, (d) the Platform Analytics Engine, and (e) the API Gateway.

The Resources Orchestrator is responsible for allocating the appropriate resources (i.e., memory, compute, and storage capacity) to the data check-in and data analysis jobs, so that they are executed successfully in an isolated manner. Although the Resources Orchestrator has full control over the data check-in and data analysis jobs that are to be executed on the SYNERGY Cloud Infrastructure, it cannot interfere with jobs that are to be executed on the On-Premise Environments.

The Notifications Engine is responsible for detecting events related to the status and progress of data check-in, acquisition and analytics workflows, in order to notify the corresponding users of the SYNERGY Platform, with appropriate content.

The Security, Authentication & Authorisation Engine component is involved with security-related functionalities within the SYNERGY Platform, while enabling the registration, authorization, and authentication of organisations and users. In addition, this component handles the verification of API keys (for external applications), and generates the tokens that are used for secure data exchange among the components of the platform.

The Platform Analytics Engine acts as the monitoring component for different services of the SYNERGY Platform in order to provide significant insights to the data asset providers as well as to the platform administrators.

The API Gateway is responsible to handle all API requests that are coming from any external application, fetching and aggregating the appropriate data from the SYNERGY Data Storage Services and finally responds to the requests accordingly.

6.2 Implemented Functionalities

The first official release of the Platform Management Services implements a set of functionalities that complement the features already available since their beta release as documented in the SYNERGY Deliverable D3.2. The following table explains the exact status of implementation in release 1.00.

Table 13: Implemented Functionalities in the current release of the SYNERGY Platform Management Services

Feature		Status	Notes
RO_1	Management of resources in the SYNERGY Cloud infrastructure	Implemented	The necessary mechanisms to automatically deploy the relevant services according to the schedule of execution of the respective job and manage the appropriate resources in the cloud are in place.
RO_2	Dynamic calculation and allocation of resources to the data check-in jobs in the SYNERGY Core Cloud Platform	Implemented	Depending on the size of the data that are to be uploaded (esp. for batch files upload), certain thresholds on the resources allocated have been created and are accordingly provisioned.
RO_3	Spawning the SYNERGY Secure Experimentation Playgrounds	Implemented	A Secure Experimentation Playground with permanent, isolated storage is created per organisation upon their registration. The transfer of data to/from such playgrounds is available and allows, for example, data to be transferred to SEP whenever a data check-in job is executed, as well as results to be stored in SEP whenever a data analytics pipeline is executed.
RO_4	Dynamic scaling of resources for running the data analysis jobs in the SYNERGY Secure Experimentation Playgrounds	Implemented	Depending on the "nominal" resources that the data preparation and analytics blocks constituting an analytics pipeline require, appropriate computation infrastructures are allocated in the cloud. Currently, if the optimal resources are not available, the analytics job shall be executed with less resources, wait to be executed till minimal resources become available or fail.
NE_1	Evaluate different incoming events in the SYNERGY Platform to find the interested users/organisation	Implemented	The different events that are collected during the execution of the services by the Master Controller are appropriately evaluated to be channelled to the concerned audience.
NE_2	Issue notifications for different events in the SYNERGY Platform	Implemented	The mechanism to handle notifications for different events, indicatively regarding the progress of the data check-in job execution, the data analytics jobs execution, the data asset contract lifecycle management and the proposed CIM concepts management, has been implemented.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 872734.

Feature		Status	Notes
NE_3	Deliver messages to users with appropriate content per event in the SYNERGY Platform	Implemented	Notifications are triggered in the SYNERGY platform and via email. Appropriate templates have been created for each type of notification in order for it to be clear and understandable for a user.
NE_4	Management of notifications by the users in the SYNERGY Platform	Implemented	The management of notifications has been implemented allowing users to mark as read and delete notifications.
SAAE_1	Manage identity information for the SYNERGY platform's users and organisations	Implemented	The management of organisations has been fine-tuned in release 1.00 in order to underpin the whole platform operation. Organisation-based access is granted upon checking the initial registration and the organisation manager is able to invite users to his/her organisation.
SAAE_2	Provide fine-grained auth. and author. services	Implemented	The communication between different services and components in the SYNERGY Platform is secured using one-time tokens.
SAAE_3	Establish integrity and trustworthiness across the different layers of the SYNERGY platform	Partially implemented	In the current release (as in the beta release), the integrity and origin of data between the Cloud Platform and the Server On-Premise Environments is ensured by registering the On-Premise Environment in the user's account in the Cloud Platform and creating an one-time access token that is valid for a few minutes to authorize the registration between the 2 layers. The secure registration of Edge On-Premise Environments is currently in progress as part of the integration activities towards the SYNERGY Integrated Platform, Release 1.00.
SAAE_4	Facilitate the secure exchange of keys across the different layers of the SYNERGY platform	Implemented	The secure exchange of keys among the different architecture layers in the SYNERGY Platform has been implemented, ensuring that the symmetric key (for data encryption/decryption) is properly encrypted for its transfer (through asymmetric key encryption).
PAE_1	Monitoring the SYNERGY Services	Partially implemented	It is possible to track the status of different services in the SYNERGY Platform in order to become aware for services that may be unavailable at any moment.
PAE_2	Understand the data asset's use per organisation	Not implemented	(Functionality to be eventually provided in release 2.00 as it requires a number of actual datasets to be already populated and available in the SYNERGY Platform)
APIG_1	User-driven configuration of the raw and derivative energy data retrieval	Implemented	Configuration of the retrieval of data and results through the SYNERGY APIs is possible for data that are stored in the cloud, in the Secure Experimentation Playground of an organization.

Feature		Status	Notes
APIG_2	Direct access to raw and derivative energy data for the authorised applications	Implemented	Data retrieval is possible with the help of API endpoints that are created dynamically based on the configuration of a retrieval query by a data consumer.
APIG_3	Management of failures	Partially implemented	Error codes have been created to anticipate certain failures in the retrieval.
APIG_4	Management of API keys	Implemented	API keys are created and effectively managed to get authorisation to retrieve data corresponding to a specific retrieval query.
APIG_5	Handling of concurrent API calls	Implemented	Pagination, CORS support and traffic management aspects have been investigated and addressed in the current release.

6.3 Internal Design and Architecture

The Platform Management Services extend over the Data Presentation layer, the Business Logic layer (separated between back-end, orchestration and services) and the Data Access layer (that essentially represents the Data Storage Services). As shown in Figure 11, the internal architecture that the Platform Management Services follow indicates how the different components, namely the Resources Orchestrator, the Notifications Engine, the Security, Authentication & Authorisation Engine, the Platform Analytics Engine, and the API Gateway, are positioned across the different layers.

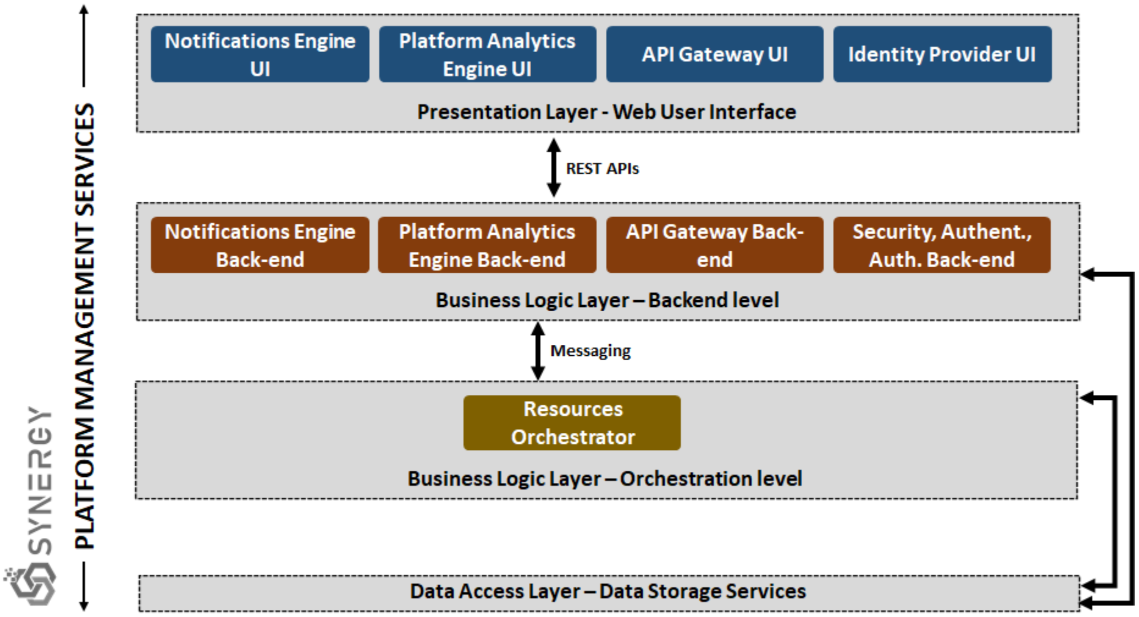

Notifications Engine UI
Platform Analytics Engine UI
API Gateway UI
Identity Provider UI
Presentation Layer - Web User Interface
REST APIs
Notifications Engine Back-end
Platform Analytics Engine Back-end
API Gateway Back-end
Security, Authent., Auth. Back-end
Business Logic Layer – Backend level
Messaging
Resources Orchestrator
Business Logic Layer – Orchestration level
Data Access Layer – Data Storage Services
SYNERGY
PLATFORM MANAGEMENT SERVICES

Figure 11: SYNERGY Platform Management Services in Layers

6.4 Technology Stack and Implementation Tools

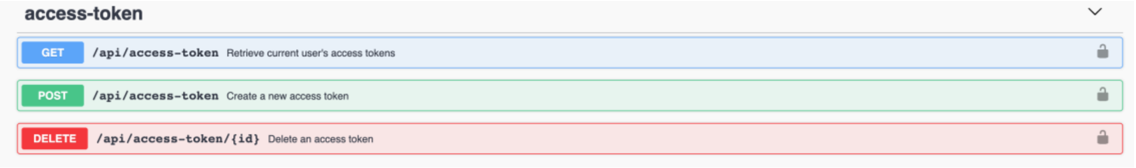
The Platform Management Services are written in TypeScript and, in their current release, leverage a number of open source technologies as depicted in the following table.

Table 14: Technology Stack for the SYNERGY Platform Management Services

Library	Version	License
Nest NodeJS Web Framework	12	MIT
TypeORM	-	MIT
Kubernetes	1.18	Apache License 2.0
SSE	12.18.4	MIT (Included in the libs of NodeJS)

6.5 API Documentation

The Platform Management Services communicate with the other components and services in the SYNERGY Platform through the Master Controller's messaging functionality, as well as through selected APIs that have been created and are documented in Swagger as depicted in the following figures.



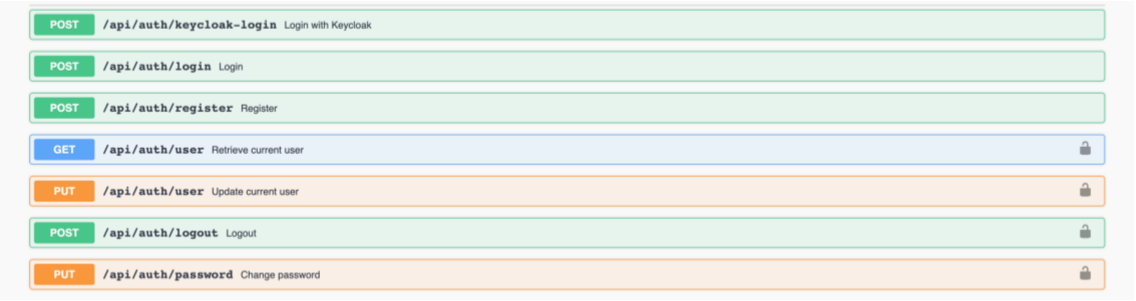
access-token

GET /api/access-token Retrieve current user's access tokens

POST /api/access-token Create a new access token

DELETE /api/access-token/{id} Delete an access token

Figure 12: Swagger API Documentation for Access Tokens (SAAE)



POST /api/auth/keycloak-login Login with Keycloak

POST /api/auth/login Login

POST /api/auth/register Register

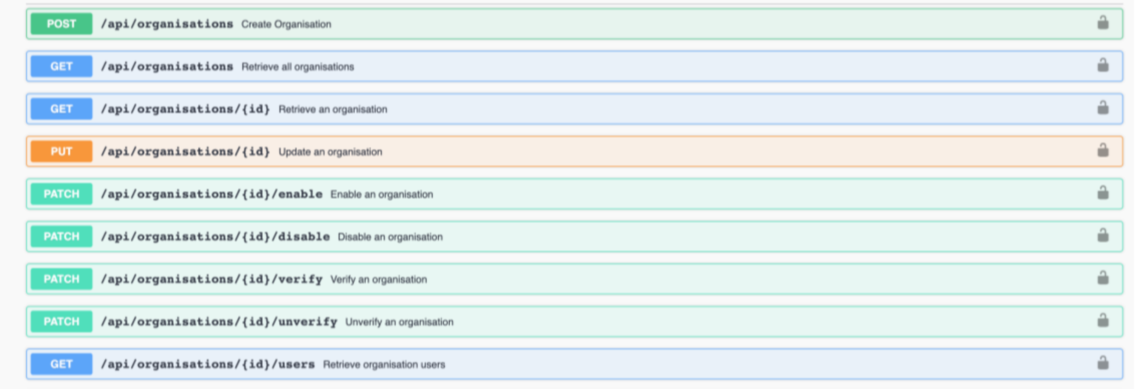
GET /api/auth/user Retrieve current user

PUT /api/auth/user Update current user

POST /api/auth/logout Logout

PUT /api/auth/password Change password

Figure 13: Swagger API Documentation for Authentication (SAAE)



POST /api/organisations Create Organisation

GET /api/organisations Retrieve all organisations

GET /api/organisations/{id} Retrieve an organisation

PUT /api/organisations/{id} Update an organisation

PATCH /api/organisations/{id}/enable Enable an organisation

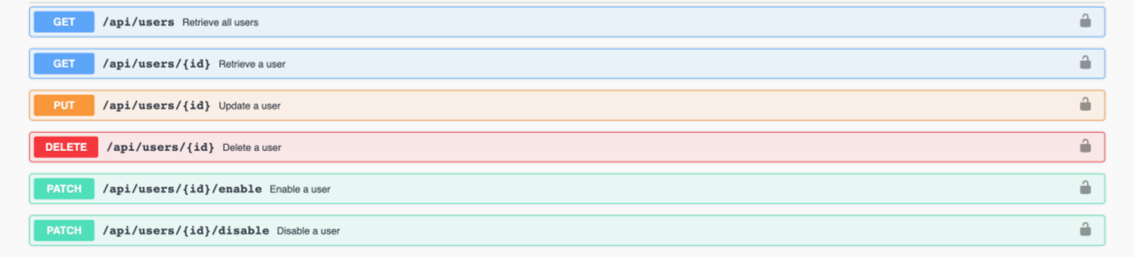
PATCH /api/organisations/{id}/disable Disable an organisation

PATCH /api/organisations/{id}/verify Verify an organisation

PATCH /api/organisations/{id}/unverify Unverify an organisation

GET /api/organisations/{id}/users Retrieve organisation users

Figure 14: Swagger API Documentation for Organisations (SAAE)



GET /api/users Retrieve all users

GET /api/users/{id} Retrieve a user

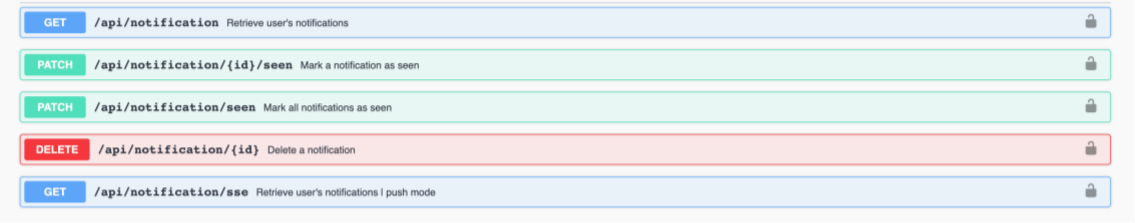
PUT /api/users/{id} Update a user

DELETE /api/users/{id} Delete a user

PATCH /api/users/{id}/enable Enable a user

PATCH /api/users/{id}/disable Disable a user

Figure 15: Swagger API Documentation for Users (SAAE)



GET /api/notification Retrieve user's notifications

PATCH /api/notification/{id}/seen Mark a notification as seen

PATCH /api/notification/seen Mark all notifications as seen

DELETE /api/notification/{id} Delete a notification

GET /api/notification/sse Retrieve user's notifications | push mode

Figure 16: Swagger API Documentation for Notifications (Notifications Engine)

It needs to be noted that the rest of the components do not expose any APIs in this release.

6.6 Installation Instructions

The Platform Management Services are served as a web application in the SYNERGY Cloud Platform, so no further installation of software is required on the end user side (data asset provider or data asset consumer).

Detailed instructions for the deployment of the Platform Management Services are provided in the related private code repository even though all subcomponents are already packaged as Docker containers.

6.7 Assumptions and Restrictions

In the current release of the Platform Management Governance Services, the development activities for the different components are still ongoing, so certain assumptions (that inevitably represent restrictions in certain cases) have been made:

- The API Gateway currently creates tokens with retrieval and upload scope that do not expire. Although this is a common practice in many cloud-based services, such a decision may be reconsidered in the final release.
- The retrieval of data and results can be configured (through the API Gateway) only for data that stored in the Cloud Platform, not for data and results residing in the On-Premise Environments.

6.8 Licensing and Access

Each component of the Platform Management Services bundle is closed source. The deployed version of the Platform Management Services bundle is made available through the integrated SYNERGY platform (as documented in D3.4 for the beta release and in the upcoming D3.6 for Release 1.00).

6.9 Updates from Beta Release

The main changes that have been introduced in release 1.00 of the Platform Management Services in comparison to the beta release (that was documented in D3.2) include:

- Front-end implementation of the Notifications Engine; the Security, Authentication & Authorisation Engine; and the API Gateway.
- Updated resources management in the Resources Orchestrator.
- Full organization-based support across all platform's components enabled through the Security, Authentication & Authorisation Engine.
- Facilitation of the key exchange process.
- Support for more notification types and issuing notifications via email (in addition to the platform notifications).

6.10 Planned Features for Next Release

The first official release of the Platform Management Services is currently being complemented by front-end and back-end updates that are considered necessary during the integration of the different services and components in the SYNERGY Integrated Platform Release 1.00, as mentioned in the following table.

Table 15: Planned Functionalities in the next release of the SYNERGY Platform Management Services

Feature	Notes
RO_1-4	Scalability aspects need to be further investigated in different situations, e.g. very large amounts of data check in jobs with small data running very frequently, many data check-in jobs with large volumes of data concurrently executed.
NE_1-4	Updates and enhancements as needed based on the results of integration and stress testing of the provided functionalities. Support for issuing additional notification types, if required.
SAAE_1-2, 4	Updates and enhancements as needed based on the results of integration and stress testing of the provided functionalities.
SAAE_3	Any further mechanisms to increase trust and integrity between the Secure Experimentation Playgrounds, the On-Premise Environments and the Cloud Platform are under investigation at the moment.
PAE_1-2	To be partially supported in the forthcoming platform release (1.00) and fully supported in the final platform release (2.00).
APIG_1-5	Updates and enhancements as needed based on the results of integration and stress testing of the provided functionalities in release 1.00 of the integrated platform.

7 Conclusions

The purpose of the deliverable D3.5 entitled “Data Collection, Security, Storage, Governance & Management Services Bundles – Release 1.00” was to deliver the first official release of the Data Services Bundles which were implemented within the context of WP3. To this end, the deliverable included the updated documentation that supplemented the information documented in deliverable D3.2 with the detailed accompanying report which presents the updated technical documentation and implementation details of these Services Bundles. The implementation of these Services Bundles was driven by the outcomes of the work performed within the context of WP2.

Toward this end, the deliverable at hand presented the updated technical details of the first official release of the five (5) Services Bundles which were implemented in WP3, namely the Data Collection, Data Security, Data Storage, Data Governance and Platform Management Services Bundles. The deliverable built directly on top of the information presented in the previous iteration in order to documented in detail for each specific Services Bundle:

- The updated implementation status of each functionality undertaken by the components composing each Services Bundle for their Release 1.00, in accordance with the design specifications defined in deliverable D2.6.
- The updated technical details of the internal design and architecture of each Service Bundle, highlighting the updates from the previous version where needed with regards to the positioning of each component and their interactions within the Service Bundle.
- The updated technology stack and the implementation tools which were leveraged during the implementation phase of the Release 1.00 of each specific Service Bundle, providing the details of the version of each utilised tool and the respective license.
- The updated technical documentation of the provided APIs from each Service Bundle for their Release 1.00 using the well-established Swagger framework.
- The updated installation instructions for each Services Bundle indicating their deployment details.
- The updated information with regards to the different assumptions made during the development of the Release 1.00 of the Services Bundle, as well as the possible restrictions imposed.

- The updated licensing information for the produced software artefacts and the access details for each Service Bundle.
- The main updates that were introduced as part of Release 1.00 of each Service Bundle on top of their Beta Release.
- The list of planned functionalities for the next final release of each Service Bundle, namely Release 2.00, again in accordance with the design specifications defined in deliverable D2.6.

The first official release of the Services Bundles of WP3 constitutes the second iteration of the produced Services Bundles that are delivered on M21 per the SYNERGY Description of Action. The delivered release was produced building on top of the previous release, namely the Beta Release, with the introduction of several optimisations and enhancements. The delivered Services Bundles along with the Services Bundles implemented within the context of WP4 are providing the foundations for the first official release of the SYNERGY platform which will be delivered on M24 and will be documented with deliverable D3.6 "SYNERGY Integrated Platform & Open APIs – Release 1.00". Nevertheless, the delivery of the Services Bundles in WP3 is a living process that will last till M33. Hence, one additional iteration is expected on M33 per the SYNERGY Description of Action which will be documented with deliverable D3.7. This final iteration will contain the necessary refinements and updates that will be based on the project's advancements and the new requirements that may arise based on the experience and feedback collected by the demo partners and the energy applications from the WP8 and WP5-WP7 activities respectively, as well as the results of the testing and pre-validation activities in WP8.

8 References

SYNERGY Consortium. (2020). SYNERGY D2.1 "End-user and Business requirements analysis for big data-driven innovative energy services and ecosystems v1"

SYNERGY Consortium. (2021). SYNERGY D2.2 "End-user and Business requirements analysis for big data-driven innovative energy services and ecosystems v2"

SYNERGY Consortium. (2020). SYNERGY D2.6 "SYNERGY Framework Architecture including functional, technical and communication specifications v1"

SYNERGY Consortium. (2021). SYNERGY D3.2 "SYNERGY Data Collection, Security, Storage, Governance & Management Services Bundles – Beta Release"

SYNERGY Consortium. (2021). SYNERGY D3.3 "SYNERGY Integrated Platform – Alpha, Mock-ups Release"

SYNERGY Consortium. (2021). SYNERGY D3.4 “SYNERGY Integrated Platform & Open APIs – Beta Release”

SYNERGY Consortium. (2021). SYNERGY D4.1 " SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles – Beta Release"