

D4.3 SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles - Release 1.00







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: D4.3

Deliverable name: **SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles**

- Release 1.00

Version: 1.0

Release date: 13/10/2021

Dissemination level: **Public** Status: **Final**

Author: \$5, KBZ, MAG, UBI, ETRA, ICCS, VTT, COBRA, URB, CIRCE, CAV,

KONCAR, ENES, UCY



Document history:

Version	Date of issue	Content and changes	Edited by
0.1	18/01/2021	Template version for DEM deliverables	Suite5
0.2	10/08/2021	Initial Table of Contents	Suite5
0.3	03/09/2021	Initial partners contributions	Suite5, KBZ, MAG
0.4	24/09/2021	Final partners contributions	Suite5, KBZ, MAG
0.5	28/09/2021	Full draft available for review	Suite5
0.6	04/10/2021	Reviewed version	VERD, URB
0.7	05/10/2021	Updated version	Suite5
1.0	13/10/2021	Final version ready for submission Suite5	

Peer reviewed by:

Partner	Reviewer	
URB	Alvaro Lacuey Cruz	
VERD	Marily Efstratiadi	

Deliverable beneficiaries:

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



Table of contents

Ex	ecutive	summary	7
1	Introdu	ıction	9
	1.1	Purpose of the document	9
	1.2	Scope of the document	9
	1.3	Structure of the document	10
2	Data M	atchmaking Services Bundle Release 1.00	11
	2.1	Overview	11
	2.2	Implemented Functionalities	11
	2.3	Internal Design and Architecture	12
	2.4	Technology Stack and Implementation Tools	14
	2.5	API Documentation	15
	2.6	Installation Instructions	17
	2.7	Assumptions and Restrictions	17
	2.8	Licensing and Access	19
	2.9	Updates from Beta Release	19
	2.10	Planned Features for Next Release	19
3	Data SI	haring Services Bundle Release 1.00	22
	3.1	Overview	22
	3.2	Implemented Functionalities	22
	3.3	Internal Design and Architecture	24
	3.4	Technology Stack and Implementation Tools	25
	3.5	API Documentation	26
	3.6	Installation Instructions	28



	3.7	Assumptions and Restrictions	28		
	3.8	Licensing and Access	28		
	3.9	Updates from Beta Release	29		
	3.10	Planned Features for Next Release	29		
4	Data Aı	nalytics Services Bundle Release 1.00	31		
	4.1	Overview	31		
	4.2	Implemented Functionalities	31		
	4.3	Internal Design and Architecture	34		
	4.4	Technology Stack and Implementation Tools	35		
	4.5	API Documentation	36		
	4.6	Installation Instructions	37		
	4.7	Assumptions and Restrictions	37		
	4.8	Licensing and Access	38		
	4.9	Updates from Beta Release	38		
	4.10	Planned Features for Next Release	39		
5	5 Conclusions41				
_	D-6		40		



Abbreviations and Acronyms

Acronym	Description
Al	Artificial Intelligence
API	Application Programming Interface
CIM	Common Information Model
CLM	Contract Lifecycle Manager
CRUD	Create Read Update Delete
DL	Deep Learning
DoA	Description of Action (annex I of the Grant Agreement)
ML	Machine Learning
OPE	On-Premise Environment (provided by the SYNERGY Platform)
SEP	Secure Experimentation Playground (as part of the SYNERGY Cloud Platform)
TF-IDF	Term Frequency – Inverse Document Frequency
WP	Work Package



Executive summary

The current deliverable D4.3 entitled "SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles – Release 1.00" constitutes a report on the produced results of three tasks, namely Task 4.1 "Big Data Analytics Workbench and Jobs Execution Engines", Task 4.4 "Blockchain-enabled, Trusted Multi-Party Data Sharing Services" and Task 4.5 "Big Data Exploration and Matchmaking Services", executed under the WP4 (Big Data Analytics and Data Sharing Mechanisms) activities. The purpose of this deliverable is to document the first official release (Release 1.00) of the different WP4 Services Bundles: (a) Data Analytics Services Bundle, (b) Data Sharing Services Bundle and (c) Data Matchmaking Services Bundle.

As the deliverable type is demonstrator per the SYNERGY Description of Action, the scope of the document at hand is to provide the detailed technical documentation of the delivered Services Bundles in Release 1.00 in the form of a supplementary report, as it was the case with D4.1 which reported in the Beta Release of the WP4 Services Bundles. For each of the three aforementioned services bundles, the current report provides information along the following dimensions: (a) The current implementation status of the assigned functionalities of the components that are related to the bundle, taking into consideration the functionalities documented in the SYNERGY deliverable D2.6 "SYNERGY Framework Architecture including functional, technical and communication specifications v1" and any additional (new / revised) functionalities that will be reported in D2.7 (due on M24); (b) The technical specifications of the internal design and architecture of the specific Services Bundle; (c) The exploited technology stack for the implementation of the Services Bundle; (d) The APIs provided by the services bundle, which have been documented using Swagger; (e) The installation instructions and deployment guidelines for the Services Bundle; (f) Potential assumptions made and restrictions identified during the development process; (g) The licensing information and access details of the delivered Services Bundle; (h) The major updates between the current release (Release 1.00) and the previous Beta Release of the specific bundle; (i) The detailed implementation plan of the upcoming functionalities of each of the components of the Services Bundles, which will become available in the next final release of the bundle.

The current deliverable presents the Release 1.00 of the WP4 Services Bundles on M21. The delivered Services Bundles will be integrated in Release 1.00 of the SYNERGY platform that will be documented in deliverable D3.6 "SYNERGY Integrated Platform & Open APIs – Release 1.00"



on M24. The implementation activities of the WP4 Service Bundles are and will remain continuous to provide additional functionalities, as well as 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 (across WP5-WP8). The final iteration of this deliverable is expected on M33 and will contain the final complete documentation of the Services Bundles in D4.4 "SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles – Release 2.00".



1 Introduction

1.1 Purpose of the document

The current deliverable, D4.3 SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles – Release 1.00, documents the work performed in the context of three tasks of WP4 "Big Data Analytics and Data Sharing Mechanisms": Task 4.1 "Big Data Analytics Workbench and Jobs Execution Engines", Task 4.4 "Blockchain-enabled, Trusted Multi-Party Data Sharing Services", and Task 4.5 "Big Data Exploration and Matchmaking Services". The purpose of this report is to provide the accompanying technical documentation of the Release 1.00 of the three WP4 Data Services Bundles:

- Data Matchmaking Services that provide search and recommendation functionalities to discover data assets in the SYNERGY platform.
- Data Sharing Services that provide functionalities required for sharing and trading data assets in a secure and trustful manner among stakeholders of the electricity data value chain.
- Data Analytics Services that offer data processing and analysis functionalities with the aim of extracting insights from own and acquired data.

1.2 Scope of the document

This deliverable provides the supportive report for the first, official release of the Data Analytics Services Bundle, the Data Matchmaking Services Bundle, and the Data Sharing Services Bundle. For the components of the presented data services bundles, the implemented functionalities in the current release are explained in this document, referring to the full expected functionalities that were introduced in D2.6 "SYNERGY Framework Architecture including functional, technical and communication specifications" and shall be updated with D2.7 on M24. Apart from the current implementation status and the updates of this release compared to the M15 beta release, the plan for the delivery of additional functionalities is also provided.

Detailed specifications per each services bundle, including the technology stack, the architecture, and interfaces to the SYNERGY platform for this release are described, also



providing licensing and access information, as well as any applicable assumptions and/or restrictions identified during the design and implementation activities.

It needs to be noted that the development of the WP4 Data Services Bundles proceed in close collaboration with the WP3 Data Services Bundles to ensure alignment towards their integration in the SYNERGY platform.

This deliverable builds upon the work presented in D4.1 "SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles – Beta Release" and will be followed by an additional release to be documented in "D4.4-SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles – Release 2.00" in M33 that will introduce new features, modifications and enhancements according to the implementation plan and the feedback received.

1.3 Structure of the document

This document is structured as follows:

- Sections 2, 3, and 4 follow the same structure and provide the implementation details including technologies and tools, internal design and architecture, licensing information, access details and installation instructions for three Data Services Bundles covered in this document, namely Data Matchmaking, Data Sharing, and Data Analytics Services Bundles respectively. Furthermore, they present the current implementation status, the planned functionalities and the major updates between the Beta Release and Release 1.00 of the respective Services Bundles.
- Section 5 concludes the document and summarises the next steps.



2 Data Matchmaking Services Bundle Release 1.00

2.1 Overview

The Data Matchmaking Services Bundle provides to the data asset consumers the means to discover data assets in the SYNERGY platform through concrete search functionalities, and recommendations to explore more assets available in the energy domain and also provides to the data asset providers suggestions for creating new datasets (based on dataset profiles needed).

The Data Matchmaking Services Bundle in the SYNERGY Cloud Infrastructure consists of: (a) the Query Builder, and (b) the Matchmaking Engine. The Query Builder provides asset discoverability and exploration functionalities and enables the users to search for data assets of interest, browse the results and explore the selected data assets in more detail in order to identify useful candidates for acquisition. The Matchmaking Engine is responsible for recommending data assets including existing energy-related datasets and potential combinations of the datasets to the users.

2.2 Implemented Functionalities

Table 1 indicates a summary of the functionalities listed in D2.6 that are implemented or partially implemented in the current release. It needs to be noted that the functionalities that have not been implemented yet are not mentioned in the table, but are presented in detail in section 2.9.

Table 1: Implemented functionalities of Data Matchmaking Services Bundle

Feature		Status	Notes
QB_1	Intuitive asset search supporting both keyword-based queries and faceted search	Implemented	A dual search functionality has been implemented: (a) keyword-based search on the assets' titles and description and (b) a faceted search on (selected properties from) the assets' metadata and the concepts included in the assets based on the SYNERGY Common Information Model (CIM).
QB_2	Asset search extending beyond data	Partially implemented	The mechanism to search for trained ML/DL models that have been registered as assets, has been developed. Searching



Feature		Status	Notes
			also over data analytics pipelines will be supported in the next release.
QB_3	Execution of search queries and provision of results	Implemented	A service has been implemented to execute search queries against the query engine and retrieve and parse the results.
QB_4	Search history and search query update and re- execution	Implemented	Create-Read-Update-Delete (CRUD) functionalities have been developed for search queries.
MME_2	Providing data asset suggestions to the users	Partially implemented	The Matchmaking Engine recommends relevant datasets to the users according to the similarity of the profile of datasets as well as the knowledge that is provided by data asset providers in the data check-in including categories and concepts from the SYNERGY CIM.
MM_3	Providing data provider suggestions to the users	Implemented	If no results are found in the existing datasets from the search endpoint provided by Query Builder, the Matchmaking Engine recommends a list of data asset providers according to the similarity of the search with the profile of datasets.
MME_4	View the list of recommendations	Partially implemented	The users will get the results of the recommendation process as a list of suggested datasets.

2.3 **Internal Design and Architecture**

The overall internal architecture of the Data Matchmaking Services bundle contains three layers as shown in Figure 1.

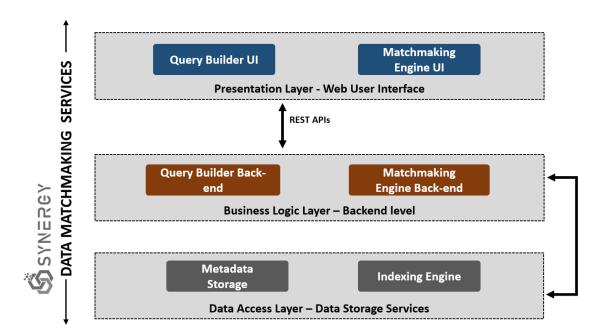


Figure 1: Data Matchmaking Services Bundle Internal Architecture

The operations of the fundamental elements are explained below:

Data Access Layer: The indexing engine that provides the infrastructure for the
matchmaking bundle services provides the indexing of data elements, index storage,
and index search mechanism which runs the queries created by the Query Builder, and
Matchmaking Engine. The provided search and recommendation services exploit the
metadata of data assets. Both the metadata storage and the indexing engine are shared
across the platform's components and services as described in the Data Storage Services
Bundle (in D3.2 and D3.5).

• Business Logic Layer:

- Query Builder: The Query Builder is responsible for: (a) transforming the user-defined configuration of a search for assets, including both search keywords and selected facets, to a query in the language expected by the query engine, (b) executing the search query and (c) parsing and returning the search results, i.e., the assets that matched the user's search.
- Matchmaking Engine: This recommendation logic makes use of the data stored in the Indexes created in the previous steps (by the Query Builder) and stored by the indexing engine to retrieve the similar data assets to the query as



potential recommendations for the user. The recommendation algorithms will increase the chances of satisfaction of queries through matchmaking, with suggestions for data asset providers based on the profile of the data assets accessed or requested by the user who is currently conducting a query. In this way, mostly distance measures have been used to find and rank similar items based on the item features. The ranking mechanism selects a set of k result items with the highest score according to the scoring function defined based on the recommendation logic.

It needs to be noted that the necessary APIs that are exposed by each engine have been developed and are available for use in the Presentation Layer, but also by other components and services in the SYNERGY Platform as mentioned in section 2.5.

Presentation Layer: The presentation layer is responsible for making the functionalities
of the Data Matchmaking Services Bundle available to the data asset consumers and to
other services/applications in the SYNERGY platform.

2.4 Technology Stack and Implementation Tools

Table 2 presents the technology stack used in the development of the different layers of the beta release of the Data Matchmaking Services Bundle.

Table 2: The technology Stack of Data Matchmaking Services Bundle

Library	Version	License
Nest NodeJS Web Framework	12	MIT
Elasticsearch (As the indexing Engine, part of the storage layer, but used by this Services Bundle)	7.10	Apache 2.0
Pandas	1.1.3	3-Clause BSD
Scikit-learn	0.23.2	New BSD
Flask	1.1.2	3-Clause BSD
TensorFlow	2.6.0	Apache 2.0



2.5 API Documentation

The API provided by the Query Builder, shown in Figure 2, remains the same as reported in D4.1, although it can now support through the same endpoints searching over additional assets, e.g. models, apart from datasets.



Figure 2: Query Builder API - Swagger screenshot

As previously reported, the API offers six (6) endpoints, as follows:

- Four (4) endpoints to support CRUD functionalities for search queries.
- One (1) endpoint to retrieve all past saved queries of a particular user.
- One (1) endpoint to perform search, which supports free-text and facetted.

The Matchmaking Engine (related to MME_2, MME_3 and MME_4 features) is responsible for the Data Matchmaking Services Bundle and also exposes an API, shown in Figure 3.



Figure 3: Matchmaking Engine API in Swagger

This release contains four (4) main endpoints that can be used:

- One (1) endpoint to create the indexes used in the recommendation mechanism.
- One (1) endpoint to invoke a data asset recommender that finds similar datasets to a selected dataset (by its ID) with the highest TF-IDF as a basic matching approach (Figure 4).





- One (1) endpoint to invoke a data asset recommender that finds k nearest neighbour datasets of a selected dataset (by its ID) as the main service of the current release of the Matchmaking Engine (Figure 5). The default value for the k is 3 but can be sent as a parameter to the service.
- One (1) endpoint to invoke a data provider recommender that finds data providers that
 provide similar datasets with the highest TF-IDF as a basic matching approach to the
 search query of the user in the case there in case no results are found for existing
 datasets (Figure 6).

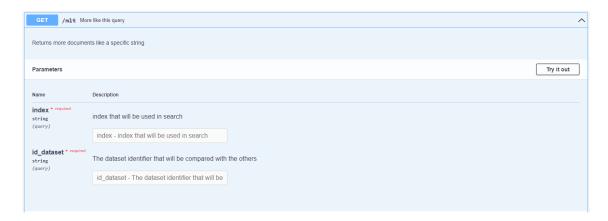


Figure 4: Matchmaking Engine (More Like This) API endpoint

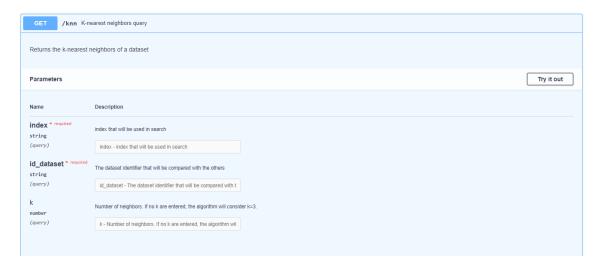


Figure 5: Matchmaking Engine (K-Nearest Neighbour) API endpoint



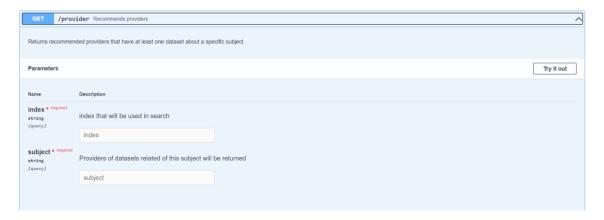


Figure 6: Matchmaking Engine (Provider recommender) API endpoint

2.6 Installation Instructions

Detailed instructions for the installation and deployment of the library that implements the Data Matchmaking Services Bundle functionalities are provided in the related private code repository. In order to use the current release, an instance of each component (Matchmaking Engine, and Query Builder) including an instance of the indexing engine needs to be created. To facilitate the installation process components are provided as Docker containers.

2.7 Assumptions and Restrictions

The Matchmaking Services Bundle provides the foundations that will be used to offer dataset search and recommendation services to the users in the Marketplace. Such datasets must have already passed the data check-in process and become compatible with the SYNERGY CIM. Consequently, the implementation has been made according to the available metadata and data structure for the datasets.

Although the implemented algorithms are flexible on the structure of the underlying datasets, future modification might be applied to adopt previously unforeseen structure needs. For this release, Table 3 (also reported in D4.1 and included here for completeness) presents the expected attributes of the dataset, according to the metadata schema that has been adopted in



SYNERGY (considering state-of-the art approaches such as the Dublin Core Metadata Initiative and DCAT-AP).

Table 3: Data Assets meta data assumed by Data Matchmaking Services Bundle

Category	Attributes			
	id (number)			
	name			
	description			
General Information	assetTypeId (number)			
	date created			
	date updated			
	tags			
	type (e.g., text)			
Distribution	format (e.g., JSON)			
	language			
	temporal coverage			
Futont	temporal resolution			
Extent	spatial coverage			
	spatial resolution			
	domain (e.g., energy)			
	streaming status			
Structure	primary concept			
	other concepts			
	concepts			
	id			
Created by	first name			
Created by	last name			
	organisation id			
	list of concepts			
Schema	list of UID of concepts			
	hierarchy of concepts			



2.8 Licensing and Access

As explained in the previous sections, the SYNERGY Data Matchmaking Services Bundle delivers the backend functionalities for asset search and recommendation for the SYNERGY Marketplace. The deployed version of the 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).

The library that offers these backend functionalities for data matchmaking, i.e. the implementation of the Data Matchmaking Services Bundle, is a closed source component.

2.9 Updates from Beta Release

The main changes that have been introduced in release 1.00 of the Data Matchmaking Services in comparison to the beta release (that was documented in D4.1) include:

- Front-end implementation of the Query Builder.
- Updates in the available search services that were considered as necessary during the integration of the different services and components in order to be compatible with the SYNERGY CIM concepts.
- Support for searching over non-data assets, particularly to trained ML and DL models.
- Extension of the recommendation algorithms to Knowledge-based recommendations (in addition to content-based filtering that was supported in the beta release). The knowledge-based recommendation uses knowledge models including categories and concepts from the SYNERGY CIM and is leveraged in the current release to recommend a performed search to data asset providers, that may potentially provide relevant data assets in case no results are found.

2.10 Planned Features for Next Release

The first official release of the Data Matchmaking 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 4: Planned Functionalities in the next release of the SYNERGY Data Matchmaking Services

Feature (identifier from D2.6 or newly assigned)	Description & Notes	
QB_1, QB_3-4, MME_3	Updates and enhancements as needed based on integration tests 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).	
QB_2	Asset search extending beyond data Searching over data analytics pipelines will be supported in the next version.	
MME_1	Settings of the user preferences for recommendations This feature allows users to choose between personalized or generic recommendations. The personalised recommendations provide dataset/data provider suggestions based on the user's preferred data concepts from the Common Information Model.	
MME_2	Providing data asset suggestions to the users This is the main functionality of the Matchmaking Engine and as described above will be iteratively completed based on the availability of required data. The final version will also provide collaborative or item-based filtering based on the availability of data	
MME_4	View the list of recommendations In the current release (1.00), this feature gets the results of the dataset recommendation process as a list of suggested data assets. In case there is no suggested dataset in the list, it gets the results of the suggested data providers. Selected recommendations and favorite lists will be reused (depend on the availability) as feedback for the future release.	

A clarification should be added in respect to MME_2, which is planned to leverage collaborative filtering in the final release of the Matchmaking Services bundle (as the main feature of the Matchmaking engine), however this will be further evaluated depending on the availability of data. Collaborative techniques use past interactions of users (data aaset consumers) with data assets to recommend Data Assets to the users with the most similar interactions profile and suggest the most popular assets among similar user groups. Since collecting the history of user activities will not be feasible until the final steps of the project, and according to low expected user interactions with the items that is normal in a dataset marketplace compared to the high



level of user interactions required for the collaborative approaches, the collaborative filtering has to be scaled down to items_based recommendations in the meantime.



3 Data Sharing Services Bundle Release 1.00

3.1 Overview

As described in D4.1, the SYNERGY Data Sharing Services Bundle is responsible for all functionalities pertaining to the sharing and trading in a secure and trustful manner of data assets among stakeholders of the electricity data value chain, where the term data assets refers to any of the following: Datasets (in the same form they were made available within SYNERGY or as outputs of data processing/analysis performed through the SYNERGY tools and services), ML/DL models and analytics pipelines.

The three components involved in the Data Sharing Services Bundle are: (a) the Data & Al Marketplace, (b) the Contracts Lifecycle Manager, and (c) the Remuneration Engine. The way these three components interact and operate within the SYNERGY Platform to provide data asset acquisition functionalities has been described in D2.6 and will be updated in D2.7. The two main workflows for data asset acquition, also reported in D4.1, remain valid and are briefly presented here again for completeness:

- 1. Bilateral data asset acquisition, involving a single data asset provider and a data asset consumer.
- 2. Multi-party data asset acquisition, involving multiple data asset providers and a data asset consumer.

3.2 Implemented Functionalities

Table 5 presents the implementation status for the functionalities foreseen to be provided through the three components that constitute the backbone of the data sharing services bundle in its first, official release (Release 1.00). It needs to be noted that the feature identifiers refer to the features of the respective components, as these were initially documented in D2.6 and D4.1.

It needs to be noted that any functionalities that have not been implemented yet are not mentioned in the following table, but are presented in detail in section 3.9.

Table 5: Implemented functionalities of Data Sharing Services Bundle





Feature		Status	Notes
DAIM_1	Clear definition of data asset metadata	Implemented	CRUD functionalities have been implemented to handle metadata for the SYNERGY assets, namely datasets, pre-trained ML/DL models and data analytics pipeline results.
DAIM_2	Handling of licenses, IPR handling and pricing schemes for a data asset	Implemented	Licenses and IPR can be thought of as metadata, but the functionalities implemented here include more targeted functions, such as the availability of license templates and the drafting of custom licenses. Support for licenses and IPR handling of pre-trained ML/DL models and data analytics pipeline results has been added.
DAIM_3	Data asset search and recommendations	Implemented	Searching over data assets and provision of recommendations are listed as features relevant to data sharing, since they are core part of the functionalities to be offered by the Data and AI Marketplace and its user interface. Indeed, searching in the Marketplace is available (free-text, faceted, advanced) and the mechanism to include recommendations is implemented. The exact status of the relevant Data Matchmaking Services Bundle is detailed in Section 2.
DAIM_4	Data asset browsing	Implemented	The backend functionalities enabling retrieval of data assets' information were part of the services' beta release. They have now been adapted to address the needs of the corresponding UI which was made available in the Beta release of the integrated platform and also extended to support the full flow of asset browsing according to the desired user experience (again pertaining to the needs of the UI).
DAIM_5	Management of data asset sharing contracts	Implemented	The basic mechanisms for drafting, signing, updating, storing, negotiating on and retrieving bilateral and multi-party data asset sharing contracts have been implemented.
CLM_1	Smart contract drafting and update	Implemented	The basic mechanisms for drafting, signing, updating, storing, negotiating on and retrieving bilateral and multi-party data asset sharing contracts have been implemented.
CLM_2	Smart contracts validity (status) check	Implemented	A service has been implemented to enable communication with the SYNERGY distributed ledger (Ethereum platform) to check the status (validity) of a specific bilateral data asset contract.
CLM_3	Enforcement of smart contract terms	Partially implemented	Some basic rules and corresponding validation mechanisms have been implemented to check for and prevent actions that violate a data asset contract, however enforcing terms that require validation across asset lineage still needs investigation and additional services to be implemented.



Feature		Status	Notes
CLM_4	Downloading smart contracts	Implemented	The functionality to download a data asset sharing contract as a pdf file has been implemented.
RE_1	Management of frictionless payments (in a cryptocurrency) for multi-party data asset contracts based on their pricing terms.	Implemented	A functionality to enable payments in the SYNERGY cryptocurrency in the wallets of the respective platform users has been developed.
RE_2	Calculation of remuneration payments in multi-party data asset contracts	Implemented	A mechanism to estimate and settle the amount to be paid to each asset provider involved in a multi- party data contract based on the smart contract's terms has been developed.
RE_3	Confirmation of offline payments in bilateral smart contracts	Implemented	An endpoint has been developed to allow users that act as data providers in a bilateral smart contract to confirm that an offline payment has been made for the specific contract.
WM_1	Secure creation and storage of a wallet	Partially implemented	In order to sign the contracts in the Contract Lifecycle Manager, the functionality of creating and storing an organisation's wallet has been developed in the cloud. The Wallet is stored in an encrypted form in the Cloud Platform, yet its storage in the On-Premise Environment also needs to investigated for increased security.

3.3 Internal Design and Architecture

The internal design of the Data Sharing Services (depicted in Figure 7) remains unchanged compared to the beta release and is composed by:

- The Presentation Layer, which is made available through the SYNERGY integrated platform. This layer includes the user interfaces of the Data & Al Marketplace and the Contracts Lifecycle Manager.
- The Business Logic Layer (backend level), which encapsulates the core functionalities
 of the services, as it provides the backend mechanisms (i.e. the code implementing the
 foreseen processes) and exposes the APIs that make the functionalities available for use



by the presentation layer of the involved components, but also by any other component of the SYNERGY architecture.

The Data Access Layer, which provides the data storage services. As reported in D4.1, apart from the contracts' ledger which conceptually belongs to this services bundle (as the CLM is the only component directly accessing it), the Metadata Storage and the Indexing Engine are shared across the platform's components and services through the Data Storage Services Bundle (described in D3.2 and D3.5).

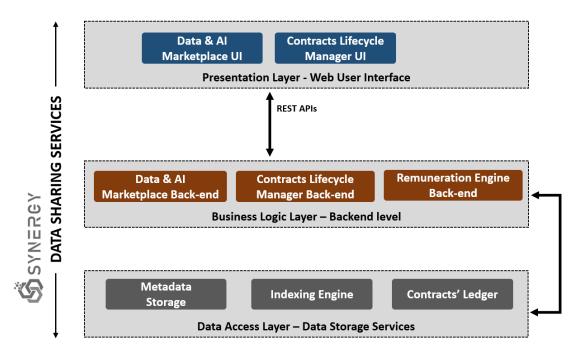


Figure 7: Internal design of the data sharing services bundle (Release 1.00)

3.4 Technology Stack and Implementation Tools

Table 6 presents the technology stack used in the development of the Presentation, Business Logic, and Data Access layers of the current release of the Data Sharing Services Bundle.

Table 6: Technology stack of Data Sharing Services Bundle

Library	Version	License
Nest (NodeJS) web framework ¹	12	MIT

¹ https://nestjs.com





Library	Version	License
Go Ethereum ²	1.9.25	LGPLv3
EthersJS	5	MIT
Truffle Framework	5.2.0	MIT
Elasticsearch (part of the Data Storage services bundle, but included here for completeness)	7.10	Apache 2.0
PostgreSQL (part of the Data Storage services bundle, but included here for completeness)	12	PostgreSQL License (a liberal Open-Source license, similar to the BSD or MIT licenses)

3.5 API Documentation

Swagger is used to facilitate designing, building, documenting, and consuming the REST APIs of the Data Sharing Services Bundle.

Figure 8 presents a screenshot of the Swagger specification of the Data Sharing Services Bundle APIs. It should be noted that the previously used data-contract endpoints that were reported in D4.1 have been adapted to support sharing of assets beyond datasets. Multi-party contracts are handled through the same endpoints. Additional endpoints have been created to allow more fine-grained update actions over contracts (that were previously managed through the playload of the request in the beta release).

² https://ethereum.org/en



Page 26

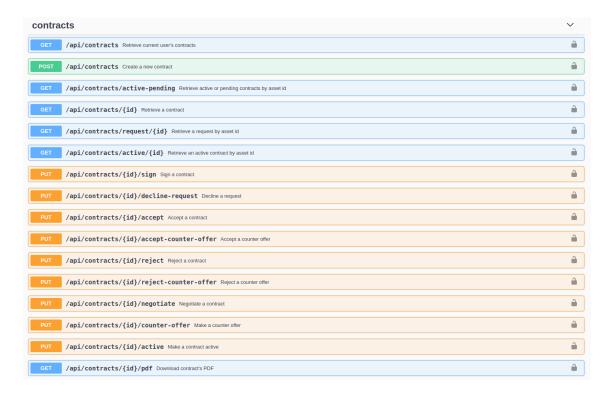


Figure 8: Data Sharing API - Swagger screenshot

The services bundle offers 16 endpoints as follows:

- Two (2) endpoints that correspond to the creation and retrieval functionalities of a specific data asset contract. No explicit delete functionality is offered in alignment with the underlying distributed ledger principles. Deleting a contract can be implicitly performed by changing its status, i.e. through the update (PUT) endpoint.
- Nine (9) endpoints correspond to updating functionalities for the status of the contract, e.g. allowing to reject it, activate it, perform a counter-offer, etc.
- One (1) endpoint retrieves all data contracts for a specific user (or organisation).
- One (1) endpoint retrieves the pdf that corresponds to a specific contract
- Three (3) endpoints to address specific contract and asset request information needs,
 e.g. to retrieve a contract only if active. These endpoints directly address needs of the respective user interface.



3.6 Installation Instructions

Detailed instructions for the installation and deployment of the library that implements the Data Sharing Services Bundle functionalities are provided in the related private code repository. All sub-components (Data & Al Marketplace, Contract Lifecycle Manager, Remuneration Engine) are packaged as Docker containers to speed up the installation process.

3.7 Assumptions and Restrictions

In the current release of the Data Sharing 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 implementation is based on the SYNERGY metadata model and the license and pricing attributes it provides.
- In case a data asset consumer has opted for their datasets to always remain on-premise, i.e. stored to the Server and Edge On-Premise Environments (OPE), they are not available in the SYNERGY Marketplace at the moment as they are considered as not available for sharing. Once support for multi-party computations is provided (in the final release), such assets will be also visible through the Marketplace.
- Multi-party contracts involve multiple data asset providers that need to be reimbursed for the acquisition of a data asset (esp. in the case of data analytics results). In order for the SYNERGY Platform to be able to properly handle such multi-party contracts, it is compulsory that payments are performed through the supported crypto-currency. The option for "offline" payment (through bank transfer, credit card) that is available for settling bilateral contracts is not available for multi-party contracts.

3.8 Licensing and Access

The respective information remains unchanged compared to what was reported in D4.1:



- The deployed version of the services bundle is made available through the integrated SYNERGY platform, which is documented in D3.4 for the beta release and in the upcoming D3.6 for Release 1.00.
- The library that offers these backend functionalities for data sharing, i.e., the implementation of the Data Sharing Services Bundle is a closed source component.

3.9 Updates from Beta Release

The main updates of the Data Sharing Services Bundle in its current Release 1.00 compared to its Beta Release, evolve around the following axes:

- Front-end implementation of the Data & Al Marketplace, and the Contracts Lifecycle Manager.
- The provision of functionalities for multi-party contracts and specifically the extension
 of the available underlying mechanisms to support handling (drafting, negotiating upon,
 accepting, rejecting, updating and signing) data asset contracts among multiple
 providers and one consumer.
- The provision of functionalities for sharing assets that extend beyond datasets and specifically allow for trained models to be also shared between members of the SYNERGY platform.
- The provision of crypto-currency based payment functionalities as a remuneration mechanism for the asset contracts.

3.10 Planned Features for Next Release

Table 7 presents the features of the data sharing services bundles that have been planned for release 2.00.

Table 7: Planned functionalities for the next release of Data Sharing Services Bundle





Feature (identifier from D2.6 or newly assigned)	Description & Notes
DAIM_1-5, CLM_1-2, CLM_4, RE_1-3	Updates and enhancements as needed based on integration tests 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).
CLM_3	Enforcement of smart contract terms As described in D2.6, planned functionalities include assessing attempted asset manipulation/handling/processing actions against the contract's terms to ensure that no terms are violated, for those terms that are enforceable through the overall SYNERGY platform.
RE_4	Credit payment to the data asset consumers in case the multi-party smart contract terms are violated As described in D2.6, planned functionalities include services to ensure that part of the funds is returned from the wallet of the defaulting data asset provider(s) to the wallet of the data asset consumer, in cases where involved parties fail to uphold the contract terms.

It needs to be noted that a light version of the Wallet Manager has been developed in the cloud as mentioned in section 3.2. Its storage and availability through the Server On-Premise Environment has been postponed for the final release due to certain implications it enforces on signing contracts that require further investigation.



4 Data Analytics Services Bundle Release 1.00

4.1 Overview

As described in D4.1, the Data Analytics Services Bundle in the SYNERGY Cloud Infrastructure offers all functionalities around data processing and analysis with the aim of insights' extraction from data. The components that are mainly involved in the provision of this services bundle are:

(a) the Analytics Workbench, (b) the Visualization & Reporting Engine, (c) the Data Manipulation Service, (d) the Analytics Execution Service, and (e) the Secure Results Export Service.

The aforementioned five components interact to enable:

- 1. The design of a data manipulation and analysis pipeline comprising potentially numerous and diverse data processing steps.
- 2. The execution of the designed pipeline, either within Secure Experimentation Playgrounds of the SYNERGY cloud infrastructure or in the premises of the data consumer.
- 3. The retrieval and visualization of the pipeline execution results.

The three workflows have been documented in D2.6, along with details for the interacting components and their foreseen features.

4.2 Implemented Functionalities

Table 8 indicates a summary of the functionalities listed in D2.6 that are implemented or partially implemented in the current release.

Table 8: Implemented functionalities of Data Analytics Services Bundle

Feature		Status	Notes
AW_1	Definition of reusable and customisable data manipulation blocks	Implemented	A service has been implemented that enables the configuration of data manipulation blocks according to a strictly defined template (which defines how parameters are configured, input and output types, etc.)
AW_2	Definition of re- usable and	Implemented	As an extension to the AW_1 functionality, a specific type of block that applies trained energy ML models





Feature		Status	Notes
	customisable energy analytics models blocks based on pre- trained algorithms		has been implemented and can be configured to support different models (from the supported ML libraries).
AW_3	Definition of re- usable and customisable energy analytics models blocks based on basic algorithms	Implemented	Wrappers around Spark MLlib, Keras (Tensorflow), scikit-learn, Statsmodels and XGBoost have been implemented to allow configuring and training ML/DL models from the specific libraries.
AW_4	Re-usable and customisable data input blocks	Implemented	Similar to the functionality implemented for AW_1 but targeting data ingestion processes for data assets that are required as inputs to a data manipulation and analysis pipeline.
AW_5	Re-usable and customisable data output blocks	Implemented	Similar to the functionality implemented for AW_1 but targeting data export processes to make the results of a data manipulation and analysis pipeline available for consumption.
AW_6	Configuration and validation of data analysis pipelines	Implemented	Data manipulation and analysis pipelines can be configured, i.e. blocks can be added and their parameters defined. Configured pipelines are validated through a dedicated service that: (a) checks if all required parameters in the block configurations have been defined and (b) that the blocks form a consistent directed acyclic graph.
AW_7	Update of data analysis pipelines	Implemented	Changes to selected properties in the configuration of a saved data manipulation and analysis pipeline are enabled.
AW_8	Insights into the execution of data analysis pipelines	Implemented	Three message queues have been implemented (in collaboration with the Master Controller in D3.2) to enable the communication between the pipeline execution process and the pipeline configuration/handling process. The communication mainly involves appropriate status updates and execution logs generated by the execution engine being sent to the queues.
AW_9 (new)	Registration of externally trained compatible ML/DL models	Implemented	An additional functionality has been developed allowing ML/DL models trained externally to the platform to be registered in (uploaded and stored to) the platform and enriched with metadata that allow them to be later on applied by the Analytics Execution Service in the same way that models trained within the platform are applied.



Feature		Status	Notes	
DMS_1	Definition of modular data manipulation functions	Implemented	A library to execute various data manipulation functions (filters, aggregations, replacements, transformations) has been implemented leveraging the pandas library and pyspark framework.	
DMS_3	Extraction of summary statistics for each data manipulation function applied in a dataset	Implemented	Extraction of summary statistics (e.g. count of entries, count of nulls, min/max/avg for numeric values, indicative values for categorical features) for the result of each data manipulation function applied as part of a data manipulation and analysis pipeline is possible upon request.	
AES_1	Implementation, configuration and application of predefined, pre-trained ML/DL models for energy problems	Implemented	A library to load and execute trained ML and DL models has been implemented.	
AES_2	Application of generic algorithms from popular ML/DL libraries	Implemented	A library to configure and train ML and DL algorithms and then store the trained models has been implemented.	
AES_3	Unified multi- purpose analytics API	Implemented	The data manipulation and analysis libraries offer an API that allows the configuration of the pipelines to be executed as sequences of steps that correspond to the configuration of the Analytics Workbench blocks.	
AES_4	Feedback from the executed analytics job	Implemented	Refer to AW_8 notes above	
SRES_1	Securely export data analytics results as a file	Implemented	Leveraging the unified API of the DMS and AES and the data export functionality of the analytics workbench (AW_5), a service to export the results of a pipeline in the form of a file has been implemented.	
SRES_2	Securely export data analytics results through an API	Implemented	Leveraging the unified API of the DMS and AES and the data export functionality of the analytics workbench (AW_5), a service to export the results of a pipeline through an API has been implemented.	
VRE_1	Support for built-in charts	Implemented	Various built-in charts (pie charts, line charts, radar charts, etc.) are supported through dedicated data retrieval and aggeragtion functionalities.	
VRE_3	Saving and exporting visualisations and reports	Implemented	A functionality to export and save a visualisation in an image format has been developed.	



4.3 Internal Design and Architecture

The Data Analytics Services, as depicted in Figure 9 and also reported in D4.1, extend over:

- The Presentation Layer, which is made available through the SYNERGY integrated platform releases. As described, the presentation layer of the analytics services includes the user interfaces of the Analytics Workbench and of the Visualisation & Reporting Engine.
- The Business Logic Layer, which comprises: (a) the backend level that offers the backend functionalities of the Analytics Workbench and exposes REST APIs needed both by the workbench and the visualisation & reporting engine, (b) the orchestration level which is responsible for coordinating the execution of the defined analytics pipelines in the appropriate execution framework and environment (i.e. Secure Experimentation Playgrounds in the Cloud, On-Premise Environments installed locally in the stakeholders' premises), and (c) the services level that encapsulates the functionalities of the Data Manipulation Service, the Analytics Execution Service and the Secure Results Export service.
- The Data Access Layer, which is responsible for the data storage services and includes
 the Trusted Data Container, the Metadata Storage, the Analytics & Al Models Storage
 and the Indexing Engine.

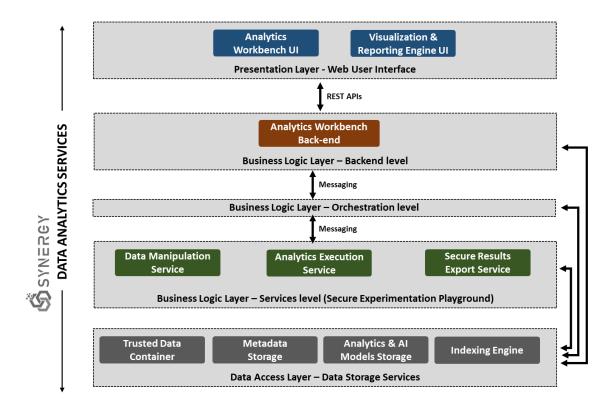


Figure 9: Internal design of Data Analytics Services Bundles

It should be noted that the Secure Experimentation Playground has been now fully implemented, hence the relevant indication of partial implementation included in the respective D4.1 image has been removed.

4.4 Technology Stack and Implementation Tools

Table 9 presents the technology stack used in the development of the current, beta release of the Data Analytics Services Bundle.

Table 9: Technology Stack of Data Analytics Services Bundle

Library	Version	License
Spark	3.0.1	Apache License 2.0
Pandas	1.2.1	BSD-3-Clause License
Scikit-Learn	0.24.1	BSD-3-Clause License
MinIO	8.0.0	Apache V2 license and Affero General Public License Version 3 (AGPLv3)
RabbitMQ	3.8	Mozilla Public license
NestJS	12	MIT





Library	Version	License
PostgreSQL (part of the Data Storage services bundle, but included here for completeness)	12	PostgreSQL License (a liberal Open Source license, similar to the BSD or MIT licenses)
Keras	2.4.3	Apache License 2.0
Tensorflow	2.4.1	Apache License 2.0
Statsmodels	0.12.1	BSD 3-Clause
XGBoost	1.4.2	Apache License 2.0

4.5 API Documentation

The services currently implemented in the Data Analytics Services bundle evolve around the following two entities, as shown in Figure 10 and previously documented in D4.1:

- Blocks, which are the smallest data handling units and correspond to a particular data processing/analysis step. There are three endpoints to retrieve, update and delete blocks by id. Creating blocks is done through the pipeline endpoints (as every block belongs to a specific pipeline).
- Pipelines, which consist of multiple blocks and define the way in which the processing steps defined by these blocks should be performed, e.g., their order and the framework that will be used to execute the defined sequence of blocks (steps). There are endpoints to create, update, retrieve and delete pipelines, and to add blocks to/ retrieve blocks of a pipeline (by id). It should be noted that the endpoint used to request the execution of a particular pipeline in the background invokes the data manipulation and analytics execution engines through the Master Controller component. (In general the Data Analytics Services communicate with the other components and services in the SYNERGY Platform through the Master Controller's messaging functionality, presented in D3.2 and D3.5).



Figure 10: Swagger specification of the Data Analytics Services Bundle

4.6 Installation Instructions

The Data Analytics Services bundles are offered as a web application (through the SYNERGY Integrated Platform), hence no installation will be required by the users. Detailed instructions for the installation and deployment of the libraries that implement the aforementioned functionalities (services) are provided in the related private code repositories. All subcomponents are packaged as Docker containers to speed up the installation process.

4.7 Assumptions and Restrictions

In the current release of the Data Analytics 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 asset consumers can configure the available Data Manipulation, ML and DL blocks in the pipelines they create or register pre-trained ML/DL models they have created outside the SYNERGY Platform, but cannot provide their own code in dedicated blocks due to security constraints of the overall SYNERGY Platform.
- Data analytics pipelines can be executed on-premise in the Server/Edge On-Premise environments only for the case of Python execution and for data that are already stored



locally. In this case, it is not possible to visualize the analytics pipeline results or retrieve them via API.

4.8 Licensing and Access

The deployed version of the Data Analytics Services bundle is made available through the integrated SYNERGY platform, which is documented in D3.4 for the beta release and in the upcoming D3.6 for Release 1.00.

The libraries that offer the backend functionalities for data analytics, which have been documented in the current section, are a closed source component.

4.9 Updates from Beta Release

The main updates in the current release are summarised as follows:

- Front-end implementation of the Analytics Workbench, and the Visualization & Reporting Engine.
- Support for training and applying deep learning models has been added, using the python library keras, which operates over Tensorflow.
- Support for applying models from two additional ML libraries, specifically from statsmodels and xgboost, has been provided.
- Numerous new data manipulation functions have been added (e.g. advanced math
 operations and extracting additional time-related information from datetime columns)
 and various existing functions have been extended and/or improved to offer an
 enhanced user experience (e.g. allowing multiple lag features to be created through the
 same block, which was previously accomplished through multiple blocks).
- Support for visualisation has been added through data retrieval functionalities which enable aggregations to be computed over the selected data subsets.
- The execution workflow has been updated to enhance the user experience during training & evaluation, i.e. to enable more intuitive interactions. Furthermore, executing



over a subset of the data in order to quickly return some insights to the user, is made possible in this release.

• A registration functionality for externally trained ML/DL models has been developed.

4.10 Planned Features for Next Release

As previously explained in D4.1, due to the nature of the Data Analytics Services Bundle, it is expected that the provided functionalities will be continuously improved and extended. These extensions may correspond to making additional libraries available in the execution services, to improving the configurations of the available blocks to offer a more intuitive user experience, to enriching the available data manipulation and analytics functions with new ones etc. Therefore, features AW_1-6, AW_9, DMS_1, DMS_3, AES_1-3, VRE_1 and VRE_3, although fully implemented in the current release, are expected to receive updates in the next one as well, to address emerging data analysis needs and the constantly evolving landscape. These updates are additional to the ones documented in the following table, i.e. the ones that emerge based on testing and integration activities.

Table 10: Planned functionalities for the next release of Data Analytics Services Bundle

Feature (identifier from D2.6 or newly assigned)	Description & Notes
AW_1-9, DMS_1, DMS_3, AES_1-4, SRES_1-2, VRE_1, VRE_3	Updates and enhancements as needed based on integration tests and stress testing 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).
DMS_2	Creation of configurable data manipulation function templates Allowing users to define and use their own data manipulation functions is an advanced feature that will be offered in subsequent releases.
VRE_2	Visualization through custom reports over the data and their derivatives Creating multiple visualisations per analytics pipeline has been implemented, yet support for report creation requires further investigation.



Finally, it should be noted that the above features regarding the data manipulation and analysis functionalities are (and will be) adapted, whenever needed, during the integration activities in order to be executed within the on-premise (server and edge) environments.



5 Conclusions

The current deliverable D4.3 entitled "SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles – Release 1.00" constitutes a report on the produced results of three tasks: Task 4.1 "Big Data Analytics Workbench and Jobs Execution Engines", Task 4.4 "Blockchain-enabled, Trusted Multi-Party Data Sharing Services" and Task 4.5 "Big Data Exploration and Matchmaking Services" of WP4 "Big Data Analytics and Data Sharing Mechanisms". It essentially accompanies the first official release (Release 1.00) of the three (3) Services Bundles that are implemented in WP4, namely Data Analytics Services Bundle, Data Sharing Services Bundle and Data Matchmaking Services Bundle, providing their up-to-date technical documentation and implementation details.

The present, first official release of the WP4 Services Bundles constitutes the second iteration of the produced Services Bundles that are delivered on M21 per the SYNERGY Description of Action. This deliverable has built directly on top of the documentation presented for the beta release in order to provide detailed information for each specific Services Bundle along the following dimensions:

- 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 with additional (new) functionalities being added when needed (that shall be also reported in D2.7).
- 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.
- 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 (extensions and improvements) that have been introduced in Release
 1.00 of the Services Bundles compared to the functionalities included in the Beta
 Release that was delivered on M15
- The list of planned functionalities for the next final release of each Service Bundle, namely Release 2.00.

As far as next steps are concerned, the delivered Services Bundles in WP4 along with the Services Bundles implemented within the context of WP3 will be integrated in Release 1.00 of the SYNERGY platform that will be documented in deliverable D3.6 "SYNERGY Integrated Platform & Open APIs – Release 1.00" on M24. The implementation activities of the WP4 Service Bundles will continue to provide additional functionalities, as well as enhancements and modifications of the existing ones, 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. An additional final release of the WP4 Services Bundles is expected on M33 and will be documented in D4.4 "SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles – Release 2.00".



6 References

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

SYNERGY Consortium. (2021). SYNERGY D3.2 "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 D3.5 "Data Collection, Security, Storage, Governance & Management Services Bundles – Release 1.00"

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