

D3.4 SYNERGY Integrated Platform & Open APIs – Beta Release







Digitising and transforming European industry and services: digital innovation hubs and platforms

Deliverable nº:	D3.4	
Deliverable name:	SYNERGY Integrated Platform & Open APIs – Beta Release	
Version:	2.0	
Release date:	17/11/2021	
Dissemination level:	Public	
Status:	Draft	
Author:	MAG, Suite5, UBITECH, KBZ, ETRA, ICCS, FVH, CUE, COBRA, CIRCE,	
CAV, IPTO, KRK, KONCAR, ENES		





Document history:

Version	Date of issue	Content and changes	Edited by
0.1	18/01/2021	Template version for DEM deliverables	Suite5
0.2	14/05/2021	Initial D3.3 Table of Contents	MAG
0.3	11/06/2021	Initial contributions to Sections 2, 3, 4	MAG, Suite5, UBITECH
0.4	18/06/2021	Updated contributions to Sections 2, 3, 4, 5	MAG, Suite5, UBITECH, KBIZ
0.5	22/06/2021	Draft section 6	MAG, Suite5
0.6	23/06/2021	Draft circulated for internal review	MAG
0.7	29/06/2021	Updates based on the internal review feedback	MAG, Suite5, UBITECH
1.0	02/07/2021	Final version to be submitted to the EC	MAG
1.1	10/11/2021	Updated version adding the background technologies (Table 3) in Section 2	MAG, Suite5, UBITECH
1.2	15/11/2021	Draft circulated for internal review	MAG
2.0	17/11/2021	Final version to be submitted to the EC	MAG

Peer reviewed by:

Partner	Reviewer
Suite5	Fenareti Lampathaki
UBITECH	Dimitris Miltiadou
URBENER	Alvaro Lacuey

Deliverable beneficiaries:

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





Table of contents

Exe	ecutive su	mmary		8
1	Introduo	ction		9
	1.1	Purpos	e of the document	9
	1.2	Scope	of the document	9
	1.3	Structu	ire of the document	10
2	SYNERG	Y Beta P	Platform at a Glance	12
3	Data Ch	eck-in U	ser Journey	19
	3.1	View a	ll Data Check-in Jobs	19
	3.2	Create	a New Data Check-in Job	20
		3.2.1	Data Ingestion Configuration	22
	3.2.	1.1	File Upload Method	22
	3.2.	1.2	Application's own (external) API method	24
	3.2.	1.3	Platform's (internal) API method	. 27
	3.2.	1.4	Streaming data to platform's (internal) mechanism	. 29
	3.2.	1.5	Streaming data from application's own (external) mechanism	. 30
		3.2.2	Pre-processing Rules Definition	31
	3.2.	2.1	Mapping Configuration	32
	3.2.	2.2	Cleaning Configuration	. 37
	3.2.	2.3	Anonymisation Configuration	. 40
	3.2.	2.4	Encryption Configuration	. 43
		3.2.3	Define New Data Asset Profile	44
	3.3	Execut	e a Data Check-in Job	50
		3.3.1	Pre-processing Rules Execution	50
	3.3.	1.1	Mapping Execution	51
	3.3.	1.2	Cleaning Execution	. 52



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



	3.3.	1.3	Anonymisation Execution	53
	3.3.	1.4	Encryption Execution	54
	3.4	Upload	d data through the On-Premise Runner	55
	3.5	Manag	ge Data Check-in Jobs	56
		3.5.1	Edit a Data Check-in Job	56
		3.5.2	Delete a Data Check-in Job	57
	3.6	Manag	ge Data Asset Profiles	57
		3.6.1	View all Data Asset Profiles	57
		3.6.2	Delete a Data Asset Profile	58
4	Data Sea	arch and	Acquisition User Journey	59
	4.1	Naviga	te to the SYNERGY Marketplace	60
	4.2	Acquii	re a Data Asset	62
		4.2.1	Review a Data Asset Request (Data Asset Provider Perspective)	64
		4.2.2	Prepare a draft contract (Data Asset Provider Perspective)	65
		4.2.3	Review a draft contract (Data Asset Consumer Perspective)	66
		4.2.4	Negotiate a draft contract (Data Asset Consumer Perspective)	68
		4.2.5	Review a revised contract (Data Asset Provider Perspective)	68
		4.2.6	Settle a finalized contract (Data Asset Provider Perspective)	69
	4.3	Retriev	ve a Data Asset	70
		4.3.1	Create a Retrieval Query	70
		4.3.2	View all Saved Queries	74
5	Data An	alvtics l	Jser Journey	75
-		-		-
	5.1		Il Analytics Pipelines	76
	5.2	Create	an Analytics Pipeline	76
	5.3	Config	ure an Analytics Pipeline	77



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



		5.3.1 Graph View	78
		5.3.2 Table View	81
		5.3.3 Results View	82
		5.3.4 Schedule Execution	83
		5.3.5 Execution History	84
	5.4	Visualization Results of an Analytics Pipeline	84
6	Additio	nal Platform Functionalities	86
	6.1	Common Information Model (CIM) Manager	86
	6.2	Edit Organisation Profile	88
	6.3	Edit User Profile	90
		6.3.1 Generate Access Tokens	91
		6.3.2 Register an On-Premise Execution Environment	91
	6.4	Wallet Manager	93
		6.4.1 Import an Existing Wallet	93
		6.4.2 Create a New Wallet	94
7	Platfor	m Integration and Support Activities	96
	7.1	Integration Approach	96
	7.2	Platform Availability	98
	7.3	Platform Support Mechanisms	98
8	Conclus	sions & Next Steps	101
9	Referei	nces	103





Abbreviations and Acronyms

Acronym	Description	
AI	Artificial Intelligence	
API	Application Programming Interface	
CIM	Common Information Model	
CSV	Comma-Separated Values	
DL	Deep Learning	
DLT	Distributed Ledger Technologies	
DoA	Description of Action (annex I of the Grant Agreement)	
DCJ	Data Check-in Job	
JSON	JavaScript Object Notation	
ML	Machine Learning	
OPE	On-Premise Environment	
SASL	Simple Authentication and Security Layer	
TSV	Tab-Separated Values	
XML	Extensible Markup Language	
WP	Work Package	





Executive summary

This document D3.4 "SYNERGY Integrated Platform & Open APIs – Beta Release" presents the beta release of the SYNERGY Integrated Platform that integrates the beta release of the services bundles developed under WP3 "End-to-end Interoperable Big Data Management Platform", and WP4 "Big Data Analytics and Data Sharing Mechanisms", in order to deliver a seamless user experience. These services bundles include the Data Collection Services Bundle, the Data Security Services Bundle, the Data Sharing Services Bundle, the Data Matchmaking Services Bundle, the Data Analytics Services Bundle, the Data Governance Services Bundle, the Data Storage Services Bundle, and the Platform Management Services Bundle.

The purpose of this deliverable is to document the functionalities of the beta version of the SYNERGY Integrated Platform that brings added value to the different electricity data value chain stakeholders for efficient data management, trusted data sharing and insightful data analytics. To this end, the set of core functionalities that are offered by the SYNERGY Integrated Platform are described from a user-oriented perspective across three core user journeys: (a) the data check-in user journey that describes the process that data asset providers need to follow to make their assets available in the SYNERGY platform; (b) the data search and acquisition user journey which describes the process of searching, exploring, and acquiring data assets (that belong to other organizations) through the SYNERGY Data & AI Marketplace; and (c) the data analytics user journey that describes the workflow of designing data analytics pipelines and their execution, in order for the data asset providers / consumers to gain valuable insights regarding their own and/or acquired data assets. Moreover, the beta version of the SYNERGY Platform enables organization-based access, prepares the ground for single sign-on functionalities with the SYNERGY energy apps and allows the lifecycle management of the SYNERGY Common Information Model (CIM).

The current deliverable presents the Beta version of the SYNERGY Integrated Platform that provides the foundations for the implementation of the upcoming versions. Towards this end, additional refinements and enhancements will be analysed and introduced as needed in release 1.00 (expected on M24), according to the SYNERGY integration plan and requirements, as well as to the feedback that will be collected by the SYNERGY energy apps (in WP5-WP7), the demonstration activities (in WP8), and the living lab activities (engaging external stakeholders) in WP9.



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



1 Introduction

1.1 Purpose of the document

The SYNERGY Deliverable D3.4 "SYNERGY Integrated Platform & Open APIs – Beta Release" accompanies the beta release of the SYNERGY Integrated Platform and intends to document its functionalities from a user-oriented perspective. The SYNERGY Integrated Platform is mainly delivered under task T3.5 "Platform and Services Bundles Continuous Integration and SYNERGY Open APIs Delivery", in collaboration with tasks T3.2 "Platform Backbone Infrastructure, On-Premise and Secure Experimentation Playground Data Containers Development", T3.3 "Core Big Data Ingestion, Curation and Management Services", T3.4 "Data Assets Security, Encryption and Privacy Mechanisms", T4.1 "Big Data Analytics Workbench and Jobs Execution Engines", T4.4 "Blockchainenabled, Trusted Multi-Party Data Sharing Services", and T4.5 "Big Data Exploration and Matchmaking Services". The documentation presented in D3.4 focuses on the end-to-end, integrated functionalities delivered through the different data services bundles, namely the Data Collection Services Bundle, the Data Security Services Bundle, the Data Sharing Services Bundle, the Data Matchmaking Services Bundle, the Data Analytics Services Bundle, the Data Governance Services Bundle, the Data Storage Services Bundle, and the Platform Management Services Bundle. Specifically, this deliverable focuses on providing a thorough description of the workflows and their corresponding functionalities supported by the beta release of the SYNERGY Platform, and any other additional platform functionality that was developed under WP3 "End-to-end Interoperable Big Data Management Platform" and WP4 "Big Data Analytics and Data Sharing Mechanisms" as reported in the beta release of all services bundles that was documented in the SYNERGY Deliverables D3.2 and D4.1. Essentially, this documentation provides a detailed step-by-step explanation of the various developed platform functionalities and associated workflows supported in the beta release of the SYNERGY Integrated Platform & Open APIs.

1.2 Scope of the document

The scope of this deliverable is to document the beta release of the SYNERGY Integrated Platform, by building on the initial, planned functionality of the alpha version (documented in the SYNERGY Deliverable D3.3), and by introducing the on-demand APIs that provide either full access or access to selected data assets that can be legitimately uploaded or retrieved by external legacy systems of the



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

SYNERGY

energy stakeholders or other third parties, according to their needs. More precisely, the scope of this document is:

- To present the beta version of the SYNERGY Integrated Platform including a thorough description of the platform's core functionalities. These functionalities are described by three different core user journeys namely: (a) the data check-in user journey that denotes the process of uploading data assets to the platform; (b) the data search and acquisition user journey that involves any process related to data search, exploration and acquisition; and (c) the data analytics user journey that describes the process of creating and executing data analytics pipelines and visualizations to gain valuable insights from certain data assets.
- To present the additional functionalities that are provided along with the core platform's functionalities, to support different tasks related to the common information model lifecycle management, user and organization profile management, and wallet creation.
- To provide description regarding the open APIs that are customized per stakeholder and data asset and delivered along with the SYNERGY Integrated Platform beta release. The process of generating customized, open APIs will be available to SYNERGY stakeholders to support their energy-related applications that are to be developed.

The beta version of the SYNERGY Integrated Platform provides a well-documented description of the platform functionalities designed and developed under WP3 and WP4, and described in D3.2 "Data Collection, Security, Storage, Governance & Management Services Bundles – Beta Release", D3.3 "SYNERGY Integrated Platform – Alpha, Mock-ups Release", and D4.1 "SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles – Beta Release". Moreover, the development of the SYNERGY Integrated Platform is properly aligned to the technical requirements and use cases of the SYNERGY project as documented in D2.1 "End-user and Business requirements analysis for big data-driven innovative energy services and ecosystems", as well as the SYNERGY Architecture defined in D2.6 "SYNERGY Framework Architecture including functional, technical and communication specifications v1".

1.3 Structure of the document

The structure of the remainder of this document is organized as follows:

• Section 2 provides a high-level view of the SYNERGY Integrated Platform, describing the main components and services bundles that are already integrated within the platform.



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

SYNERGY

- Section 3 presents the data check-in user journey that guides the energy data value chain stakeholders, acting as data asset providers, through the data collection and management functionalities including both the design and the execution phases.
- Section 3 presents the data search and acquisition user journey that guides the energy data value chain stakeholders, acting as data consumers, through the search and acquisition functionalities that are available in the SYNERGY Data & AI Marketplace.
- Section 4 presents the data analytics user journey that the SYNERGY Integrated Platform
 offers to the energy data value chain stakeholders, acting as data providers or consumers,
 for designing and executing data analytics pipelines in order to gain valuable insights and
 visualizations for a particular data asset.
- Section 5 presents all the additional functionalities that are provided along with the core functionalities of the SYNERGY Integrated Platform. These functionalities support various management mechanisms for platform's users, organization profiles, and the lifecycle of the common information model.
- Section 6 summarises the overall integration and support approach followed in the SYNERGY Platform.
- Section 7 concludes this deliverable and provides future steps that are to be followed in the project.
- Section 8 includes the list of relevant references.





2 SYNERGY Beta Platform at a Glance

In accordance with the integration plan that was presented in D3.3, the SYNERGY Integrated Platform on its beta release has delivered the planned functionalities for the Cloud Platform (including the Secure Experimentation Playgrounds that are spawn per organisation) and the Server On-Premise Environment. The core workflows concerning data check-in, data search and sharing, and data analytics, are supported as described in the SYNERGY Integrated Platform Alpha Release (in D3.3) and the beta release of the various Data Services Bundles (as described in D3.2 and D4.1). An overview of the core functionalities associated with each services bundle is depicted in Figure 1. The features delivered in the beta release appear as underlined in the figure.

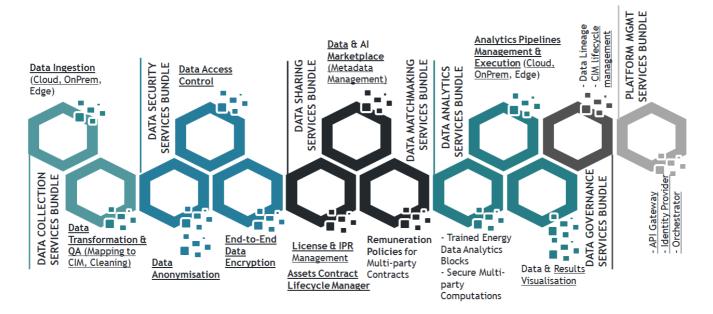


Figure 1: SYNERGY Services Bundles

In more detail, the main functionalities delivered through the SYNERGY Beta Platform release towards the different electricity data value chain stakeholders who at any moment may assume the role of data asset providers and / or data asset consumers, are briefly described in Table 1 below.

Data Collection Services Bundle				
	Configuration of the data check-in process by the data provider, and its proper execution in the SYNERGY Cloud Infrastructure and / or the Server On-Premise Environment.			
	Configuration and execution of different data ingestion, mapping and transformation and cleaning services that are invoked as-needed to appropriately handle batch, near real-time (through push and pull mechanisms) and streaming data collection.			





The status of integration among the different SYNERGY platform components is depicted in the following table. Such interactions essentially remain the same for the Cloud Platform (including the Secure Experimentation Playgrounds that are spawn per organisation) and the On-Premise Environments (in their server and edge edition). Obviously, the integration of the Data Storage Services Bundle has been ensured and it effectively communicates with all components and services in the SYNERGY Platform. It needs to be noted that the beta release integrated the core





functionalities of the SYNERGY Platform, so added-value functionalities that were not at the core (e.g. recommendations for additional data assets of interest or for electricity data value chain stakeholders who could potentially have/create the requested data asset) were postponed for integration in the next releases (even in cases when the beta release of the specific components had been already available as documented in D3.2 and D4.1).

Component	Related Components with which there is interaction/integration	Status in SYNERGY Platform Beta Release		
Data Collection Services Bundle				
Data Handling	Matching Prediction Engine	Integration Completed		
Manager	Access Policy Engine	Integration Completed		
	Master Controller	Integration Completed		
Matching Prediction	Mapping & Transformation Service	Integration Completed		
Engine	CIM Manager	Integration Completed		
Data Ingestion Service	Master Controller	Integration Completed		
Mapping & Transformation Service	Master Controller	Integration Completed		
Cleaning Service	Master Controller	Integration Completed		
Data Security Services	Bundle			
Anonymisation Service	Master Controller	Integration Completed		
Encryption Engine	Master Controller	Integration Completed		
	Wallet Manager	Partial Integration		
Access Policy Engine	Data Handling Manager	Integration Completed		
	Data & Al Marketplace	Integration Completed		
	Analytics Workbench	Integration Completed		
	API Gateway	Integration Completed		
Data Sharing Services	Bundle			
Data & Al	Query Builder	Integration Completed		
Marketplace	Matchmaking Engine	Integration Completed		
	Contract Lifecycle Manager	Integration Completed		
Contract Lifecycle Manager	Remuneration Engine	Not Planned for the Beta Release		
	Notifications Engine	Integration Completed		
	Master Controller	Integration Completed		
Remuneration Engine	Wallet Manager	Not Planned for the Beta Release		
Wallet Manager	Encryption Engine	Partial Integration		
	Remuneration Engine	Not Planned for the Beta Release		
Data Matchmaking Ser	rvices Bundle			
Query Builder	Data & AI Marketplace	Integration Completed		
	Matchmaking Engine	Integration Completed		
	API Gateway	Integration Completed		
Matchmaking Engine	Query Builder	Integration Postponed for the upcoming releases		

Table 2: SYNERGY Components Integration Status





	Data & Al Marketplace	Integration Postponed for the upcoming releases
Data Analytics Services	s Bundle	
Analytics Workbench	Visualization & Reporting Engine	Integration Completed
	Data & Al Marketplace	Integration Completed
	API Gateway	Integration Completed
Visualization & Reporting Engine	Service Secure Results Export Service	Integration Completed
Data Manipulation Service	Master Controller	Integration Completed
Analytics Execution	Master Controller	Integration Completed
Service Secure Results Export Service	Master Controller	Integration Completed
Data Governance Serv	ices Bundle	
Master Controller	Access Policy Engine, Contract Lifecycle Manager, Data Ingestion Service, Mapping & Transformation Service, Cleaning Service, Anonymisation Service, Encryption Engine, Data Manipulation Service, Analytics Execution, Service Secure Results Export Service, Data Handling Manager, Analytics Workbench	Integration Completed for all components
Data Lineage Service	Master Controller	Not Planned for the Beta Release
CIM Manager	Matching Prediction Engine	Integration Completed
	Query Builder	Integration Completed
Platform Management	Services Bundle	
Resources Orchestrator	Master Controller	Integration Completed
Notifications Engine	Master Controller	Integration Completed
	Contract Lifecycle Manager	Integration Completed
Security, Authentication & Authorisation Engine	All Components & Services	Integration Completed
Platform Analytics	Data Handling Manager, Data & Al	Not Planned for the Beta
Engine	Marketplace, Contract Lifecycle Manager	Release
API Gateway	Access Policy Engine	Integration Completed
	Contract Lifecycle Manager	Integration Completed
	Query Builder	Integration Completed
	Service Secure Results Export Service	Integration Completed

As an Innovation Action project, SYNERGY builds on different background technologies that have been brought by the core technical partners (namely Suite5, UBITECH and MAG). The following table presents the background technologies that have been leveraged in each component and the updates that have been already introduced in the beta release of the SYNERGY Platform. It needs to be noted that the SYNERGY Data Storage Services as well as the SYNERGY Server On-Premise Environment, have been developed from scratch based on open-source technologies (as described in D3.2).





Component	Background Technologies	Updates Introduced in Beta Release
Component Data Collection Service		Opuates introduced in Beta Release
		Europhianalities described in detail in
Data Handling Manager	Developed from scratch (leveraging the experiences from past big data projects, especially ICARUS, and open-source technologies)	Functionalities described in detail in D3.2
Matching Prediction Engine	Developed from scratch (leveraging the experiences from past big data projects, especially ICARUS, and open-source technologies)	Functionalities described in detail in D3.2
Data Ingestion Service	Developed from scratch (leveraging the experiences from past big data projects, especially ICARUS, and open-source technologies)	Functionalities described in detail in D3.2
Mapping & Transformation Service	ICARUS Mapper	 Refactoring for performance improvements (for big batch data and real-time data) More sophisticated and domain- specific transformations for data harmonization (e.g. supported datetime formats, timezone conversions, measurement unit transformations) Collection of mapping service execution metrics and generation of mapping reports
Cleaning Service	ICARUS Cleanser	 Refactoring for performance improvements (for batch data and real-time data) Inclusion of consistency checks between the data harvesting method and the permitted cleaning rules (per data type) Collection of cleaning service execution metrics and generation of cleaning reports
Data Security Services	Bundle	
Anonymisation Service	ICARUS Anonymiser	 Refactoring for performance improvements (for batch data) Inclusion of consistency checks between the data harvesting method and the permitted anonymization rules (per data type) Collection of anonymization service execution metrics and generation of anonymization reports
Encryption Engine	ICARUS Encryption Manager, ICARUS Decryption Manager, ICARUS Key Pair Administrator	 Refactoring for performance improvements in the required encryption/decryption time Redesigned key exchange process

Table 3: Background technologies for the SYNERGY Components





Component	Background Technologies	Updates Introduced in Beta Release
Access Policy Engine	Developed from scratch (leveraging	Functionalities described in detail in
	the knowledge gained from the ICARUS Policy Manager)	D3.2
Data Sharing Services	Bundle	
Data & Al	CYBELE Data Policy and Assets	• Redesigned user experience in the
Marketplace	Brokerage Engine	SYNERGY Data Marketplace
		• Extended metadata schema for
		data assets
Contract Lifecycle	CYBELE Data Policy and Assets	• Updated smart contracts structure
Manager	Brokerage Engine	and mechanism in blockchain (e.g.
		inclusion of license metadata and
		payable addresses, effective
		handling of expired contracts)
		• Support for transferring acquired
		datasets between the Secure
		Experimentation Playgrounds of different organizations
Remuneration Engine	None - To be developed from scratch	
Remuneration Engine	in future releases	-
Wallet Manager	CYBELE Data Policy and Assets	Adopted with minor adaptations,
Wallet Mallager	Brokerage Engine	required during the integration
		activities
Data Matchmaking Se	rvices Bundle	
Query Builder	None - Developed from scratch	Functionalities described in detail in
	(based on open-source technologies)	D4.1
Matchmaking Engine	None - Developed from scratch	Functionalities described in detail in
	(based on open-source technologies)	D4.1
Data Analytics Service		
Analytics Workbench	Suite5 Enterprise Analytics Suite	• Support for pipelines to be configured and executed in Spark
		MLlib
		• Support for pipelines to be
		executed in Server On-Premise
		Environments
		• Partially implemented support for
		pre-trained energy data analytics
		models blocks – to be fully
		supported in release 1.00 (and
Visualization &	None - Developed from scratch	documented in D4.2, D4.3) Functionalities described in detail in
Reporting Engine	(based on open-source technologies)	D4.1
Data Manipulation	Suite5 Enterprise Analytics Suite	• Support for additional data
Service		manipulation functions (e.g. for
		more advanced datetime
		transformations and aggregations)
Analytics Execution	Suite5 Enterprise Analytics Suite	Generation of detailed log for each
		pipeline's execution
		• Storage of results as data assets (in
		 Storage of results as data assets (in order to be eventually available in





Component	Background Technologies	Updates Introduced in Beta Release
Service Secure	None - Developed from scratch	Functionalities described in detail in
Results Export	(based on open-source technologies)	D4.1
Service	(
Data Governance Serv	ices Bundle	
Master Controller	None - Developed from scratch	Functionalities described in detail in
	(leveraging the experiences from past	D3.2
	big data projects, especially ICARUS,	
	and open-source technologies)	
Data Lineage Service	None – To be developed from scratch	-
	in future releases	
CIM Manager	None - Developed from scratch	Functionalities described in detail in
	(based on open-source technologies)	D3.2
Platform Management	Services Bundle	
Resources	None - Developed from scratch	Functionalities described in detail in
Orchestrator	(leveraging the experiences from past	D3.2
	cloud projects, especially UNICORN	
	and MATILDA, and open-source	
	technologies)	
Notifications Engine	None - Developed from scratch	Functionalities described in detail in
	(based on open-source technologies)	D3.2
Security,	None - Developed from scratch	Functionalities described in detail in
Authentication &	(based on open-source technologies)	D3.2
Authorisation Engine		
Platform Analytics	None - To be developed from scratch	-
Engine	in future releases	
API Gateway	None - Developed from scratch	Functionalities described in detail in
		D3.2





3 Data Check-in User Journey

This section describes the data check-in user journey including the various steps a data asset provider needs to follow for loading data on the SYNERGY Platform. In the data check-in workflow (depicted in Figure 2), the data asset provider starts by creating a new data check-in job, configures the data ingestion process that is responsible to load the data on the SYNERGY Platform, defines the preprocessing rules that are to be applied on the data asset, defines the data asset's metadata and licensing details, and finally executes the data check-in job that triggers the data loading to the SYNERGY Platform.

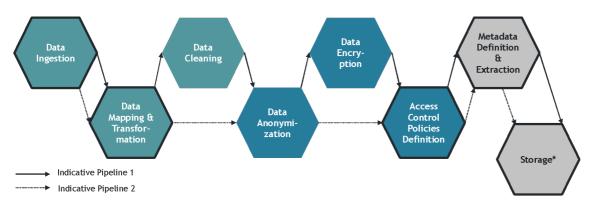


Figure 2: Data Check-in Pipelines

It needs to be noted that there is a distinct separation between the creation and configuration (design) phase of the data check-in job as described in Section 3.1, and the actual execution of the configured data check-in job by the SYNERGY Platform as described in Section 3.3 and 3.4.

3.1 View all Data Check-in Jobs

A summary list including all the created/configured data check-in jobs appears in the Data Check-in Jobs view as shown in Figure 3. The data asset provider may view all the details regarding the data check-in jobs in a list view, including the job status and available actions depending on the job status. The date of creation and the user (within the organization) that has created the data check-in job, as well as the status of execution for each data check-in job, are also presented. In particular, when the data check-in job is already executed, a tick mark, cross mark, or exclamation mark will appear next to the corresponding data check-in step (i.e. Harvester, Mapping, Cleaning, Anonymiser, Encryption, Loader) if the execution for this step was successful, failed, or not processed in case of wrong configuration, respectively, for the steps that are relevant for the specific job. In the case where the





data check-in job is not executed yet, but instead it is in the queue for execution, then a clock icon will appear denoting an execution pending status.

					(9	💓 John Doe 🗸
Data Check-in Jobs						+ Create
🛓 Created by John Do	e 🗎 Last modified on June 28, 2021				Configuration: Mapping	Â
✓ Harvester	III Mapping	441 Cleaning	III Anonymiser	III Loader		
Test Created by John Do asd	e 🗎 Last modified on June 24, 2021				Configuration: Harvester	
III Harvester	III Mapping	III Loader				
					Queued: Loader	
✓ Harvester	Mapping	 Anonymiser 	0 Loader			
🛓 Created by John Do	e 🗎 Last modified on June 23, 2021				(Idle: Mapping)	
✓ Harvester	- Mapping	- Cleaning	- Anonymiser	- Loader		
	Solar Power Gene 4. Created by John De Solar power generation What wester 1. Created by John De acr 11. Harvester Solar Power Gene 4. Created by John De spo What wester Solar Power Gene 4. Created by John De spo Solar Power Gene 4. Created by John De spo Solar Power Gene	Solar Power Generation Data (Platform's API) Created by John Doe Last modified on June 28, 2021 Solar Power Generation Marvester Mapping Test Created by John Doe Last modified on June 28, 2021 ad If Harvester Mapping Solar Power Generation (File Upload) Created by John Doe Last modified on June 23, 2021 age A Created by John Doe Last modified on June 23, 2021 age Marvester Mapping Solar Power Generation Data (Platform's API) Created by John Doe Last modified on June 23, 2021 age Created by John Doe Last modified on June 23, 2021 	Solar Power Generation Data (Platform's API) Created by John Doa Last modified on June 28, 2021 Solar power generation Mapping Created by John Doa Last modified on June 24, 2021 adr Harvester Mapping Last modified on June 24, 2021 adr Solar Power Generation (File Upload) Created by John Doa Last modified on June 23, 2021 Created by John Doa Last modified on June 23, 2021 Created by John Doa Last modified on June 23, 2021 Created by John Doa Last modified on June 23, 2021 Created by John Doa Last modified on June 23, 2021 Solar Power Generation Data (Platform's API) Created by John Doa Last modified on June 23, 2021 	Solar Power Generation Data (Platform's API)	Solar Power Generation Data (Platform's API)	Solar Power Generation Data (Platform's API) Centegrantion: Marging:

Figure 3: View All Data Check-in Jobs

3.2 Create a New Data Check-in Job

The first step of the data check-in workflow is the creation and configuration of a new data check-in job. The data asset provider may create a new data check-in job by selecting the Create button at the Data Check-in Jobs page which lists all the created data check-in jobs, as depicted in Figure 3. First, the data asset provider needs to provide basic information (i.e., title, and description) about the data check-in job in the Data Check-in Job Details section, as shown in Figure 4. Additionally, the data asset provider needs to select the pre-processing rules, that are to be applied on the loaded data, ranging from Mapping, Cleaning, Anonymiser. It needs to be noted that the selection of the Mapping step activates the selection of the other pre-processing steps (i.e., Cleaning, Anonymisation, Encryption). The final step for the creation of the data check-in job, is to select the environment where the data check-in job is to be executed. In particular, the data asset provider may select either Cloud Execution to run the data check-in job on the cloud, or On-premise Execution to run the data check-in job and the create button at the top right part of the page. Finally, the data check-in job is created and shown in the list of Data Check-in Jobs page (as depicted in Figure 3).

The data asset provider can proceed to configure the different pre-processing rules that were selected before. First, it is mandatory to configure the Harvester step that was pre-selected by



SYNERGY

default. The configuration can be done by selecting the appropriate data check-in job from the list, and then by selecting the Harvester option. The Harvester configuration determines the data ingestion method that is to be followed, which sets the way of loading the data on the SYNERGY Platform. As depicted in Figure 5, the data provider may select the data loading method as follows:

- <u>File Upload Method</u>, allowing data asset providers to upload files in different formats (i.e., CSV, JSON, XML, and other formats) containing batch historical and operational data.
- <u>Application's Own (External) API Method</u>, which supports the use of an external API, owned by the data provider or any other third party, for importing the API request results.
- <u>Platform's (Internal) API Method</u>, which allows data uploading to API endpoints provided by the SYNERGY Platform and customized to the data that are to be pushed.
- <u>Platform's (Internal) Kafka PubSub Method</u>, which can be configured to upload streaming data to the Kafka Publish-Subscribe (PubSub) mechanism provided by the SYNERGY Platform.
- <u>Application's Own (External) Kafka PubSub Method</u>, allowing data providers to use their own Kafka PubSub mechanism and grant access to the SYNERGY Platform to collect the streaming data they publish in a specific topic.

By selecting the "Finalize Configuration" button, the corresponding configuration page depending on the selection will appear.

5			
	Data Check-in Job		× Cancel ✓
	Job Details	NAME	
Basic information about the job Solar Power Generation Basic information about the job Basic information about the job Pre-processing Basic information about the job Data solar power jahn in india. MARVESTER Calent taba through files, extremal APIs or other services Data solar power jahn in india. Mary Solar to the common data model Basic india diagraphi jasses, such as compt or inaccurate records Ampring your data to make it unreadable to a person or entity without permission Basic india diagraphi jasses, such as compt or inaccurate records Ampring your data to make it unreadable to a person or entity without permission Basic india diagraphi jasses, such as compt or inaccurate records Biocyty your data to make it unreadable to a person or entity without permission Browy is person or entity without permission Biocyty your data to make it unreadable to a person or entity without permission Browy is person or entity without permission Biocyty is person or entity without permission			
		DESCRIPTION	
ets		Data collected from a solar power plant in India.	
	Pre-processing	HARVESTER	
	Before importing your data to the platform, you can use additional		
	tools to better prepare them		In from a solar power plant in India. ER ER Trough files, extnemal APIs or other services. data to the common data model Safar so that quality issues, such as corrupt or inaccurate records. SER SIGN SIGN IN
		ANONYMISER	
		Encrypt your data to make it unreadable to a person or entity without permission	
		LOADER Load the processed data to the data storage	
	Execution	CLOUD EXECUTION	
	You can choose whether you want the data-checkin job to be run on	The data-checkin job will run on the cloud.	
	the cloud or on-premise through a registered runner	 ON-PREMISE EXECUTION The data-checkin job will run on-premise through a registered runner. 	

Figure 4: Create a new Data Check-in Job (DCJ)





3.2.1 Data Ingestion Configuration

3.2.1.1 File Upload Method

Once the data asset provider has selected the data upload option (shown in Figure 5), the Data Ingestion configuration page appears as depicted in Figure 6.

Ŝ							👃 🛛 ě John Doe 🗸
G) Home	Configure Har	vester: Solar Power Generation			< Back to Jobs	Save	✓ Finalize Configuration
A Data Check-In	STEP 1 Setup Harvest Servi	ce		STEP 2 Test and Review Configuration			
88 My Assets		Data Loading How do you plan to load your data to the platform?	FILE UPLOAD Direct file upload (CSV, JSON	,XML)			
A Marketplace			STREAMING DATA TO INTEI Upload streaming data to the STREAMING DATA FROM 0	internal PubSub mechanism provided by the platform			
حی Analytics			Collect streaming data from the DATA PROVIDER'S AVAILAB	he PubSub mechanism that is available on your end	open APIs		
Data Models			PLATFORM'S API Upload data to the Platform's	APIs			

Figure 5: Data loading method

In the configuration page, the file format of the data that are to be uploaded needs to be selected. Following that, the data asset provider should upload a sample file including an indicative number of entries from the whole data asset (applies to CSV, TSV, JSON, and XML files only), and the actual file that is to be uploaded. In the case that the data provider selects not to upload the data directly in





the cloud, the location of the file where it will be stored locally and retrieved from the On-Premise Environment, needs to be provided accordingly.

ගී						👃 (🛞 John Doe 🗸
G	Configure Harv	vester: Solar Power Generation			X Cancel Save	✓ Finalize Configuration
Data Check-In	STEP 1 Setup Harvest Servio	20		STEP 2 Test and Review Configuration		
88 My Assets		Data Loading How do you plan to load your data to the platform?	 FILE UPLOAD Direct file upload (CSV, JSON 	, XML)		
C Marketplace		Format				
Analytics		Select the format of the file you will use	O CSV ○ JSON ○ XML (Other		
Data Models		Sample Upload Upload a sample of your data to be used in next steps	BROWSE Plant_1_Weath	er_Sensor_Data_sample.csv 22.7 KB		
		Upload File(s) Upload your file(s) to be processed (if in csv, json, xml format)	BROWSE Plant_1_Weath	er_Sensor_Data.csv 230.1KB		
					Ne	xt >

Figure 6: Data Check-in – File Upload Method - Setup Harvest Service (Step 1)

By selecting the Next button, the data asset provider may view the details of the data sample that was uploaded in a tabular view, or tree view for non-flat formats, as depicted in Figure 7. Finally, the data provider may save the configuration provided so far or finalize the configuration button that essentially completes the Harvester configuration.

STEP 1 Setup Harvest	Service			STEP 2 Test and Review Configuration	n	
	ADDED FILES Plant_1_Weather_Sensor_	"Data.csv				230.1KB ×
	DATE_TIME	PLANT_ID	SOURCE_KEY	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION
	15/05/2020 0:00	4135001	HmiyD2TTLFNqkNe	25.18431613	22.8575074	0
	15/05/2020 0:15	4135001	HmiyD2TTLFNqkNe	25.08458867	22.76166787	0
	15/05/2020 0:30	4135001	HmiyD2TTLFNqkNe	24.9357526	22.59230553	0
	15/05/2020 0:45	4135001	HmiyD2TTLFNqkNe	24.8461304	22.36085213	0
	15/05/2020 1:00	4135001	HmiyD2TTLFNqkNe	24.62152536	22.16542264	0

Figure 7: Data Check-in - File Upload Method - Review (Step 2)





3.2.1.2 Application's own (external) API method

Once the data asset provider has chosen to ingest data from an own (external) API that is exposed by his/her organisation's systems, the corresponding configuration page will appear as shown in Figure 8. Initially, the data asset provider is asked to select the format of the API response (e.g. JSON or XML), and the authentication details regarding the API access policies (e.g. None, if no authentication is needed; Bearer if authentication that involves security tokens is needed; and Custom, if the external API support the use of a custom URL for authentication). Following that, more details need to be provided, depending on the authentication type selection (e.g. for Custom Authentication, the Authentication URL and the Authentication Query Body, in order to retrieve the Access Token, allowing the data asset provider to test the authentication policies inserted, by selecting the Test Login button). Then, the full API path including the base URL, along with the appropriate method (i.e., GET, POST, PUT), need to be inserted. If a POST method is selected, the query body of the request needs to be also provided.

Ŝ							👃 🛛 ě John Doe 🗸
۵ Home	Configure Har	rvester: Solar Power Generatin			Save	✓ Finalize Configuration	
Data Check-In	STEP 1 Setup Harvest Serv	vice		STEP 2 Test and Review Configuration			
88 My Assets		Data Loading How do you plan to load your data to the platform?	O DATA PROVIDER'S AVAILABLE Collect data from the APIs provid	API fed by applications and systems of the data provider or from open AP	lis		
Analytics		Response Format Select the format of the API response	O JSON 🔿 XML				
Data Models		Authentication Details Details about the authentication policies of the API	None Bearer Custom				
		Method, URL & Body ③ The method, URL and query body of the request.	GET http://api.openwestherma	p.org/data/2.5/weather?q=London&appid=2x93/28cd067/5ec1100as97963/fs94	18/imit=1008page=		C
		Pagination Select your API pagnation options	None Offset O Page PARAMETER Lant How many items to include in each page Page The starting page	NAME limit page		IGNC	

Figure 8: Data Check-in - API (external) – Setup Harvest Service (Step 1a)

Once the full API path is provided, the corresponding request parameters will appear automatically in the Request Parameters section where one can edit them, remove them, or even add new query parameters, as depicted in Figure 9. Different options for pagination of the API responses, such as offset or page, are also provided to define how paginated API responses should be handled. By inserting the API method URL and query body, the corresponding request parameters will appear in the Request Parameters sections, where the data asset provider may also insert additional request parameters. In addition to the request parameters, the data asset provider may also add extra



SYNERGY

headers to the API calls by selecting the Add Header button. It needs to be noted that any request parameter and extra header can be treated as sensitive (that essentially means that it is stored in an encrypted form), depending on the data asset provider's preference.

e Harvester: Solar Power Generatin				× Cancel	Save 🗸 F	inalize Configurat
st Service		STEP 2 Test and Review Configuration				
	The starting page	1				
Request Parameters Any url, query or body parameters that will be used on the API calls	PARAMETER	VALUE	TYPE	SENSITIVE		
	D 5	100	Query		ď	
	Ø 1	1	Query		e	
	🦁 q	London	Query		e o	
	Ø appid	2a93f28cd087f5ec1100aa9796	Query		2 8	
				+ ADD C	QUERY PARAMETER	
Extra Headers	HEADER	VALUE		SENSITIVE		
Any extra header that needs to be part of the API calls	demo	demo ×			1) + ADD HEADER	
	Request Parameters Any urit, query or body parameters that will be used on the API calls	The starting page The starting page Request Parameters Any uri, query or body parameters that will be used on the API calls D S Q 1 Q appid Extra Headers HEAGE	t Service Test and Review Configuration The starting page The starting page The starting page Request Parameters Any uri, query or body parameters that will be used on the API calls Any uri, query or body parameters that will be used on the API calls PARAMETER VALUE Q Q Q London Q Q Q London Q Q Q ULUE Extra Headers HEADER VALUE VALUE	t Service Test and Review Configuration The starting page The starting page Request Parameters Any uri, query or body parameters that will be used on the API calls Any uri, query or body parameters that will be used on the API calls PARAMETER VALUE V	th Service Test and Review Configuration The starting page The starting page Request Parameters Any sift, query or body parameters that will be used on the API calls Any sift, query or body parameters that will be used on the API calls PARAMETER VALUE VALU	A Service Test and Review Configuration The standing page Request Parameters Any uit, query or body parameters that will be used on the API calls Request Parameters Any uit, query or body parameters that will be used on the API calls Request Parameters Any uit, query or body parameters that will be used on the API calls Request Parameters Any uit, query or body parameters that will be used on the API calls Request Parameters Request Par

Figure 9: Data Check-in - API (external) – Setup Harvest Service (Step 1b)

Another important step of the external API data ingestion method is to define the retrieval settings regarding the schedule and periodicity of the data ingestion that the SYNERGY Platform should initiate to retrieve data from the specific API, as shown in Figure 10. In particular, the data asset provider may select the start and end date of the retrieval, and its periodicity (i.e., retrieve once, periodic retrieval according to schedule, and polling every 60 seconds). By selecting the periodic retrieval option, the data asset provider can configure the retrieval according to a schedule (or potentially multiple schedules), and the retrieval periodicity (i.e., hourly, daily, weekly, or monthly). The polling retrieval method is configured by default for ingesting data every minute. The next step of this workflow is to set the processing periodicity of the data retrieved, by selecting one of the options: (a) immediately, (b) on an hourly basis, (c) on a daily basis, or (d) on a weekly basis. Finally, the last option of this configuration is to determine the way that errors are handled. Currently there are two options available: (a) No action, and (b) Retry a specified number of retries every 30 seconds, in case that an error has occurred.



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



Configure Har	vester: Solar Power Generatin						× Cancel	Save 🗸	inalize Confi
STEP 1 Setup Harvest Serv	ice		STEP : Test ar	d Review Config	uration				
	Retrieval Settings Select how ofter you want to retrieval through the API? Until when?	Retrieve from Period Schedule 1 Schedule 2	Periodic Retrieval (ac 28 June 2021 Hourty Daily V Wednesday O Ars at 04:00 (UTC) on Wedn Monday 5 ns at 15:00 (UTC) on Mond	to month eekty Month esdays	End Date	۵		+ ADD SCHEDULE	
	Processing How often should we process your data?	• Immediately	On an hourly basis	On a daily basis	s 🔿 On a weekly	y basis			
	Error Handling Strategy How should we handle any errors harvesting data?	O No action	Retry 5 times	(every 30 secon	ds)				

Figure 10: Data Check-in - API (external) – Setup Harvest Service (Step 1c)

By selecting the Next button, the SYNERGY platform will make a call to the API endpoint according to the configuration that has been set. If the API call is successful, the data asset provider comes across the Review step as depicted in Figure 11. During this step, the data asset provider may select the response handling method that determines whether each API response will be handled as a single object and stored as one individual record containing the concepts selected by the data asset provider, or whether each API response should be stored as multiple records that can be separated based on the selected path of the response. Additionally, the data asset provider may insert additional response data, associated with a value and its value type (e.g. static or dynamic). A static parameter added in each record/row of the response data, is a fixed value that does not change each time the API is called, while a dynamic parameter added in each record/row of the response data, contains datetime values that are updated each time the API is called.





Ŝ					Q.	l 🛞 John Doe 🗸
습 Home	Configure Ha	rvester: Solar Power Generatin			X Cancel Save V Fit	inalize Configuration
Crime Check-In	STEP 1 Setup Harvest Serv	vice		STEP 2 Test and Review Configuration		
88 My Assets		Response Handling How should the response be handled?	AS A SET OF RECORDS UNDER	a single object and stored as one record R THE PATH liple records that appear under the selected pr	ath	
ංදි Analytics 🗔 Data Models		Store Additional Parameter within the Data Details about any state or dynamic parameters that should be added in each record/row.	Add additional data PARAMETER storedDatetime DYNAMIC VALUE current_timestamp	VALUE TYPE Dynamic	~	
		API Response Selection The complete API response reviewed when testing the API connection. The user neets to select the concepts that should be further processed and stored. The concepts that are not selected will be discarded.	<pre>{ "cord": { "lor": "number", "lar": "number" }, "uesther": [[(fit: "number", "mestin": "string", "mestin": "string",</pre>			

Figure 11: Data Check-in - API (external) – Review (Step 2a)

Finally, the data asset provider needs to review and select the concepts that should be further processed and stored to the SYNERGY platform, while a summary of the API's response that will be permanently stored, will be shown accordingly at the bottom of the page as shown in Figure 12.

Configure Harvester: Solar Power Generatin		X Cancel Save V Finalize C
STEP 1 Setup Harvest Service	STEP 2 Test and Review Configuration	
	<pre>vi mater,</pre>	C RECORCUTE AN CALL
Selected API Response Summary A summary of the part of the API response that will be permanently stored.	{ "timerone": 3600, "id": 2643743, "name": "London", "cod": 280, "cod": 28	

Figure 12: Data Check-in - API (external) – Review (Step 2b)

3.2.1.3 Platform's (internal) API method

Once the data asset provider has selected the Platform's API data harvesting method during the creation of the Check-in Job, its configuration page is shown as Figure 13 depicts. The data asset provider should select the type of the data that are to be uploaded (e.g. text, text and binary). In particular, the data asset provider may choose whether to upload text data (e.g. JSON, or XML) or





text data along with binary data (including any file format, e.g. JPG, PDF, IFC, etc) through the generated API. In addition to this, the data asset provider needs to upload a sample file containing an indicative number of entries (rows) from the "text" data that will be sent to the SYNERGY Platform's API. Additionally, the data asset provider needs to select the processing periodicity of the data that are to be fetched among the options: (a) immediately, (b) on an hourly basis, (c) on a daily basis, or (d) on a weekly basis. This setting defines the frequency of the processing of the data, that are to be pushed to the SYNERGY Platform.

ගී							۵ (John Doe 🗸
G Home	Configure Harv	ester: Solar Power Generation			× Cancel	Save	✓ Finalize (Configuration
Cr Data Check-In	STEP 1 Setup Harvest Servic	0		STEP 2 Test and Review Configuration				
88 My Assets		Data Loading How do you plan to load your data to the platform?	PLATFORM'S API Upload data to the Platform's	APIs				
Aarketplace		Data Type What type of data do you intend to load?	• Text only	y .				
Analytics		what type of oata do you intend to load?						
Data Models		Processing How often should we process your data?	Immediately On an hourt	y basis 🚫 On a daily basis 🚫 On a weekly basis				
		Sample Upload Upload a sample of your data to be used in next steps	BROWSE Openweather_	sample.json 758.0 B				
							Next >	

Figure 13: Data Check-in - API (internal) – Setup Harvest Service (Step 1)

During the second step (Test and Review phase) of the Platform's API data harvesting method that is shown in Figure 14, the data asset provider is able to see an auto-generated API endpoint in the SYNERGY platform. However, to use this generated API, the application should be authenticated using an already generated access token that needs to be added into an "X-API-TOKEN" header in the API request. If the application is not authenticated, the data asset provider should generate a new access token (with "upload" scope) by selecting the "generate a new one button" link as shown in Figure 14. Instructions on how to use the POST endpoint are provided accordingly. To upload binary files through the generated API, the data asset provider should select "Multipart/Form data" as the body type of the request, and to include an additional key-value pair (with key name "_uploaded_file" and value the binary file that is to be uploaded) in the request body accordingly. It needs to be noted that one API request can only send one binary file. Finally, the data asset provider may review the details of the configuration and the data sample that was uploaded in the previous step, and select Finalize to proceed to the configuration of next steps of the Data Check-in workflow (as described in Section 3.2.2).



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



Š	Configure Ha	rvester: Solar Power Generation		Cancel Sve V						
Home A ta Check-In	STEP 1 Setup Harvest Ser	vice		STEP 2 Test and Review Configuration						
Assets		Method & URL The API method and the full URL	POST https://synergy-bigda	ta.eu/api/upioad/0a879bb4-bb1b-444d-8010-a94225947ce7		œ				
ta Models		Instructions How to use the POST endpoint	 Use an already generated ac requests. Add the created access toke 	you should be authenticated. To do so, you need to: cess token with upland scope or generate a new one. This token w into an <u>scape-tokke</u> header in your request. locad to the API, into the body of your request, in JSON format.	III be used to authenticate ;	your				
		Data Sample The details of the data sample that was uploaded	[{ "coord": { "lor": -122.08, "lar": 127.08 } , weather": [{ ("d": 000,							

Figure 14: Data Check-in - API (internal) – Review (Step 2)

3.2.1.4 Streaming data to platform's (internal) mechanism

Once the data asset provider has selected to harvest data through the Platform's Kafka PubSub Mechanism, the corresponding configuration page is shown as in Figure 15.

Ŝ					👃 🍈 John Doe 🗸
ل Home	Configure Harvester: Test kafka			X Cancel Save	✓ Finalize Configuration
A Data Check-in	STEP 1 Setup Harvest Service	STEP 2 Test and R	Review Configuration		
OD OO My Assets	Data Loading How do you plan to load your data to the platform?	STREAMING DATA TO INTERNAL PUBSUB MER Upload streaming data to the internal PubSub me			
Marketplace	Format Select the format of the data you will upload to the kafka topic	O JSON () XML			
Data Models	Connection Details The connection details that you should use to upload streaming data	Connection URL Topic topic353 SASL mechanism screm-sha-512 Username job99 Password Please make sure to save this password to	because it will not be shown again.		×
	Retrieval Settings Until when you want to retieve data from Kafka topic?	Retrieve until End Date			

Figure 15: Data Check-in - Streaming data (internal) – Setup Harvest Service (Step 1)

During this configuration, the data asset provider should select the format of the data to be published (i.e., JSON, or XML). Then the data asset provider may view the connection details (i.e., the connection URL that refers to the platform's Kafka PubSub mechanism, the topic name in which the data should be published, the Simple Authentication and Security Layer (SASL) mechanism that is used, and the credentials that the data asset provider should use to publish data to the specific topic).





The retrieval end date that defines the date that the streaming data retrieval has to be completed, the periodicity that defines how often the data are processed, and the error handling strategy have to be set accordingly. Finally, the data asset provider needs to upload a sample of the streaming data according to the file format selected (i.e., JSON, or XML), in order to proceed to view the sample in the next step.

During the review step, the data asset provider may finalize the harvesting configuration, after reviewing the streaming data structure (in a tree-view), and by selecting the Finalize button as depicted in Figure 16.

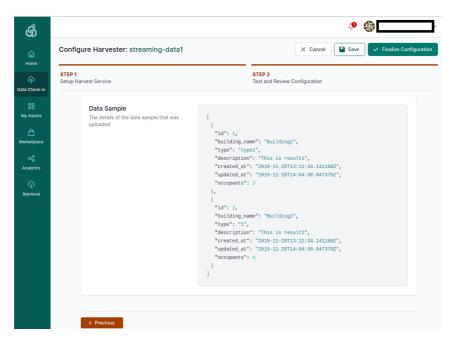


Figure 16: Data Check-in - Streaming data (internal) – Review (Step 2)

3.2.1.5 Streaming data from application's own (external) mechanism

Once the data asset provider has selected to use a Kafka PubSub mechanism provided by an external application (owned by the data provider), the corresponding configuration page for subscribing the SYNERGY Platform to the already published streaming data, as shown in Figure 17 is revealed. Initially, the data asset provider needs to select the format of the streaming data that is published (i.e., JSON, or XML) in the specific topic, and to upload a sample of the streaming data including a few entries from the streaming data. Moreover, the data asset provider shall insert the connection URL of the external Kafka that shall be used by the SYNERGY platform, the topic name to which the data will be published, the SASL mechanism that is used, the credentials that the SYNERGY Platform should use to access the Kafka mechanism and the group id, if applicable.





conngaren	Harvester: Weather Data (Streaming from Application)			X Cancel Save V Finalize Cont
STEP 1 Setup Harvest	Service		STEP 2 Test and Review Configuration	
	Data Loading How do you plan to load your data to the platform?		ATA FROM OWN PUBSUB MECHANISM g data from the PubSub mechanism that is available on your end	
	Format Select the format of the data you will stream to the kafka topic	O JSON 🔿 XML		
	Connection Details	Connection URL	https://application.com/streaming/weatherdata	
	The connection details which will be used to collect your streaming data	Торіс	current	
		Group Id	john.doe@example.com	
		SASL mechanism	PLAIN	
		Username	john.doe	
		Password	•••••	

Figure 17: Data Check-in - Streaming data (external) – Setup Harvest Service (Step 1a)

The retrieval end date and the processing periodicity need to be defined as well, as shown in Figure 18. It needs to be noted that the processing periodicity will be by-passed if needed (e.g. if the volume of streaming data is very high or low).

Ĝ						¢	🌔 John Doe 🗸	
G Home	Configure Ha	vester: Weather Data (Streaming from Application)			X Cancel Save) V Fin	alize Configuration	
Criter Check-In	STEP 1 Setup Harvest Serv	ice		STEP 2 Test and Review Configuration				
88			SASL mechanism	PLAIN				1
My Assets			Username	john.doe				
			Password					
Analytics								
Data Models		Retrieval Settings Until when you want to retieve data from Kafka topic?	Retrieve until	02 July 2021				i
		Processing How often should we process your data?	• Every Hour	Every Day				l
		Error Handling Strategy How should we handle any errors harvesting data?	No action	Retry 5 times (every 30 seconds)				
						Next >		*

Figure 18: Data Check-in - Streaming data (external) – Setup Harvest Service (Step 1b)

3.2.2 Pre-processing Rules Definition

As mentioned in Section 3.1 which describes the creation of a new data check-in job, the data asset provider is asked to select some pre-processing rules to be applied on the data upon their injection





to the SYNERGY Platform. The pre-processing rules that are available for selection, as shown in Figure 4, are Mapping, Cleaning, Anonymisation, and Encryption. Although mapping is not a mandatory step, it is strongly encouraged to be selected in order for the data to be mapped to the SYNERGY CIM. In addition, by enabling Mapping, the Cleaning, Anonymiser and Encryption steps also become available for selection, in order to solve data quality issues (e.g. corrupt or inaccurate records), and to anonymise the data to prevent an individual from being identified, respectively. It needs to be noted that the Encryption step is available only if the data asset provider has opted for on-premise execution. This section describes the workflow that a data asset provider needs to follow in order to configure these pre-processing rules, prior to their actual execution which is described in the subsequent Section 3.3.

3.2.2.1 Mapping Configuration

The mapping configuration, provided by the Data Collection Services Bundle as described in Section 2 of D3.2 "Data Collection, Security, Governance & Management Services Bundles – Beta Release", is divided into three main steps namely the Mapping Info, the Mapping Configuration, and the Mapping Review and Confirmation. In the first step, the data asset provider is asked to select the standard to which the ingested data comply, if applicable, and the main category in which the data refer to, as depicted in Figure 19. Once the data asset provider selects a category, he/she needs to define the most appropriate concept that represents the data that will be uploaded.

Mapping fo	or: Solar Power Generation			
STEP 1 Info		STEP 2 Configuration	STEP 3 Review and Confirmation	
	Domain Main domain to which the data refer to	1623938113056-ENERGY A data model representing the energy conte: ELECTRICITY DATA MODEL FOR SYNERG' The data model (Common Information Mode		
	Standards Basic information about the standards to which the data comply	NONE OPENADR (2.0) SAREF4ENER (1.12) IF0 (4.1) USEF UFTP (1.01) SSN (090 18-070)	IEC CIM (16) OBXXML (10.0) SAREF (v3.13) SAREFABLOC (11.2) UN/CEFACT CCTS (3)	
	Category Main concept to which the data refer to	Building Device		>
		Event Flexibility		>

Figure 19: Mapping - Provide Mapping Info (Step 1)





In the next step, the data asset provider comes across the predictions that the SYNERGY Platform has made about how the source data (from the sample provided in the Data Ingestion step) map to the SYNERGY Common Information Model (CIM) (that has been specified in the SYNERGY Deliverable D3.1) in the Mapping Playground, as depicted in Figure 20.

Ŝ									👃 🛛 🂓 John Doe
ය Home	Mapping for: Solar Power Gene	ration					× Cance	Valida	ate Save V Finalize Mapping
ch ata Check-In	STEP 1 Info	STEP 2 Configuration				STEP 3 Review and	Confirmatio	in	
	DATA MODEL	PLAYGROUND	All	Predicted Co	orrected	Unidentified	Invalid	Selected	MAPPING DETAILS
ly Assets	Q, Search	SOURCE DATA	CONFIDENCE LEVEL	COMMON DATA	A MODEL				Click on one or more fields in playground to start
nrketplace	✓ SensingMeasurement Me temperatureChangeRate	DATE_TIME String		observedTe SensingMeasur			erature	×	prayground to start
alytics	we averageWindSpeed we minTemperature we maxTemperature	PLANT_ID Number		SensingMeasur	urement				
D Models	meanTemperature meanTemperature meanTemperature meanTemperature	SOURCE_KEY String		SensingMeasur	urement				
	recastMinTemperature forecastMaxTemperature forecastTemperatureRange	AMBIENT_TEMPERATURE Number		minTemper SensingMeasur				×	
	Do you think that something is missing? Let us know! MinTemperature Double The minimum temperature within a	MODULE_TEMPERATURE Number		minTemper SensingMeasur				×	
	certain period of time.			SensingMeasur	urement				

Figure 20: Mapping - Configuration (Step 2a)

These predictions are accompanied by certain confidence levels denoting how much uncertainty there is for a particular field/concept to be correctly predicted and mapped to the corresponding field/concept of the common data model. In case there is no predicted mapping, or the predicted mapping is wrong, the data asset provider can search for a particular field/concept under the Data Model section located at the left side, as shown in *Figure 20*, while a short description of the selected field/concept appears at the bottom left side. Additionally, the data asset provider can propose a new field/concept in case the CIM has not already foreseen it, by selecting the "Let us know" link which pop-ups a new window for proposing a new field/concept to be added in the common data model, as shown in Figure 21.





Ĝ						🗘 🛛 🍥 John Doe 🗸
	Mapping for: Solar Power Gene				× Cancel	Validate Save ✓ Finalize Mapping
	STEP 1 Info		Suggest New Model Concept	t	EP 3 view and Confirmation	
	DATA MODEL	PLAYGROUN	NAME	PARENT CONCEPT	entified Invalid Sele-	cted MAPPING DETAILS
		SOURCE DATA	relativeHumidity	WeatherMeasurement		
	 SensingMeasurement temperatureChangeRate 	DATE_TIME (DATA TYPE boolean		ouble edTemperature	
	we averageWindSpeed we minTemperature we maxTemperature	PLANT_ID	DESCRIPTION The humidity that is measured with resp	pect to a certain reference point.		
	w observedTemperature w meanTemperature w forecastTemperature	SOURCE_KEN	RELATED TERMS			
	 forecastMinTemperature forecastMaxTemperature forecastTemperatureRange 	AMBIENT_TE	STANDARD (IF ANY) SSN - OGC 16-079		operature	
	Do you think that something is missing? Let us know! MinTemperature Double	MODULE_TEN		Cancel	operature	
	The minimum temperature within a certain period of time.	IRRADIATION	Number			
	< Previous					Next >

Figure 21: Mapping - Suggest New Model Concept

A particular column of the source (original) data may be mapped to a specific field or concept that is listed in the Data Model section. A column in the source data can be mapped to a field of the CIM by placing the appropriate field from CIM to the corresponding source data (row in the Mapping Playground) using the drag n' drop functionality. A column in the source data can be mapped to the fields of a related concept by selecting the corresponding row in the Mapping Playground and by selecting the related concept (along with a related prefix) from the dropdown menu that appears in the Mapping Details section located at the right side of the page, as shown in Figure 22. In cases the related concept can be customised (according to the CIM provisions), the data asset provider can provide his/her own prefix along with its definition.

	Mapping for: Solar Power Gene	eration					[× Cancel	💠 Valida	te Save	🗸 Finalize Mapp
ck-in	STEP 1 Info		STEP 2 Configuration				STEP 3 Review and	Confirmation	1		
	DATA MODEL	PLAYGROUND		Clear Selection All	Predicted	Corrected	Unidentified	Invalid	Selected	MAPPING	DETAILS
ets	Q Search	SOURCE DATA		CONFIDENCE LEVEL	COMMON	DATA MODEL				SET RELATED	CONCEPT
lace	100 realFeelTemperature	DATE_TIME String				redDateTime	Datetime measuredDate	īme	×	PREFIX	ion
ics	 reportedDateTime seaLevelPressureMin temperature 	PLANT_ID Number		Drag and drop a field here	WeatherN	Measurement				The locati station.	on of the metereologic nce location for the w
dels	temperatureMax temperatureMin ve temperatureMin ve uvIndex	SOURCE_KEY String				ature Doubl Aeasurement >			×	C Enter yo	ents/forecasts. Jr own prefix ONCEPT SET & PRED
	NO visibilityMax	SAMPLE VALUES					-			5210	SET & PREI
	Do you think that something is missing?	PLANT_ID "4135001" "4135001" "4135001" "4									
	RelatedLocation Location An object that includes the properties of a certain location that is related to weather conditions.	4133001 4133001 4133001 4	10001 410001								

Figure 22: Mapping – Configuration (Step 2b)



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



By selecting the Set Concept, the corresponding list of all the fields included in the related concept is revealed in the left Data Model section, allowing the data asset provider to map the input data (row in the Mapping Playground) to a particular field that is included in the related concept selected. This allows the connection of fields in the source data with other fields that are included in the selected related concept of the base concept that was selected in Step 1. The same procedure should be followed when the data asset provider selects the Set Concept & Predict, although in this case, the prediction service is executed once more to predict mappings between the input data and the selected related concept.

In the right section, the Transformation Details are shown providing more information including the transformations rules that are to be applied on the different fields depending on their data type and the CIM provisions. For numeric fields, the applicable measurement unit (whenever applicable according to the CIM) needs to be provided as shown in Figure 23. Depending on the selection that the data asset provider makes, he/she is informed whether the data will need to be transformed to the baseline measurement unit (followed in the SYNERGY CIM).

Mapping for: Solar Power Gen	eration			X Ci	ncel 🗳 Valid	ate Save	🗸 Finali
STEP 1 Info	STEP 2 Configuration			STEP 3 Review and Confirm	nation		
DATA MODEL	PLAYGROUND	Clear Selection All	Predicted Corrected	Unidentified Inva	lid Selected	MAPPING DI	ETAILS
Q Search	SOURCE DATA	CONFIDENCE LEVEL	COMMON DATA MODEL			TITLE	
~ WeatherMeasurement	AMBIENT_TEMPERATURE Number		temperature Double		×	temperature	
NO seaLevelPressureMean			WeatherMeasurement >			Celcius	JUREMEN'
NO absoluteHumidity NO atmosphericPressure NO cloudiness	MODULE_TEMPERATURE Number		WeatherMeasurement			No transformation	sieguieu
Ser conditionDetails Ser conditionIntensity Ser conditionStatus	IRRADIATION Number		WeatherMeasurement				
See description	SAMPLE VALUES						
Do you think that something is missing? Let us know!	AMBIENT_TEMPERATURE "25,18431613" "25,08458867" "24,9357526" "24,84613	104" "24 62152536" "" " " " " " " " "					
RelatedLocation Location An object that includes the properties of a certain location that is related to weather conditions.							

Figure 23: Mapping Configuration - Numeric Field Transformation

For datetime fields, the desired datetime format and the applicable time-zone (if enabled) should be defined as shown in Figure 24. It needs to be noted that during the mapping step, data type casting according to the CIM provisions is also performed (e.g. from integer to double, from string to datetime, etc.).





Mapping for: Solar Power Gene	eration		X Cancel	Validate Save ✓ Finalize Map
STEP 1 Info	STEP 2 Configura	tion	STEP 3 Review and Confirmation	
DATA MODEL	PLAYGROUND	Clear Selection All	Predicted Corrected Unidentified Invalid Selecte	MAPPING DETAILS
Q, Search	SOURCE DATA	CONFIDENCE LEVEL	COMMON DATA MODEL	TITLE
✓ WeatherMeasurement seaLevelPressureMean	DATE_TIME String		measuredDateTime Datetime WeatherMeasurement > measuredDateTime	X DATE FORMAT
seaLevelPressureMax				YYYY-MM-DD hh:mm:ss
ne absoluteHumidity	PLANT_ID Number	* ~	id String	× REFERENCE TIMEZONE
N2 atmosphericPressure		-	WeatherMeasurement > stationLocation > id	UTC
No cloudiness	SOURCE_KEY String		temperature Double	Will be ignored if already available in you
SeconditionDetails			WeatherMeasurement > temperature	
Sr conditionStatus				
Ser description				
Se id	SAMPLE VALUES			
Do you think that something is missing?	DATE_TIME "15/05/2020 0:00" "15/05/2020 0:15" "15/05/202	0.0.20" "15/05/2020 0.45" "15/05/2020 1:0	An	
	13/03/2020 0.00 13/03/2020 0.13 13/03/202	0 0.30 13/03/2020 0.43 13/03/2020 1.0	•	
RelatedLocation Location An object that includes the properties of a certain location that is related to weather conditions.				

Figure 24: Mapping Configuration - Datetime Field Transformation

In the Mapping Playground section, the data asset provider may filter the view of the concepts by different categories (i.e., Predicted Mappings, Unidentified Mappings, Corrected Mappings, Invalid Mappings, and Selected Mappings). Once the data asset provider is satisfied with the mapping configuration, by clicking the Next button, the Mapping Review and Confirmation page is loaded as depicted in Figure 25. During this step, the data asset provider may view a summary of the concepts that are mapped to the SYNERGY CIM (described in D3.1 "SYNERGY Common Information Model") and will be transformed according to the configuration provided during the previous step. Additionally, the data asset provider may view more details regarding the mapping and transformation for the source data (per field) by selecting a particular row as shown in Figure 25. In addition, users are able to view the Unidentified Concepts at the bottom of the page, where the title and data type of the concepts that are not mapped to the CIM, and thus they will be excluded from the mapped data that will be uploaded in the SYNERGY Platform and proceed to further processing.





ගී						4	🕽 🔅 John Doe 🗸
۲. Home	Mapping for: Solar Power Generation				× Cancel	Save	✓ Finalize Mapping
م Data Check-in	STEP 1 Info	STEP 2 Configuration		STEP 3 Review and Confirmation			
00 My Assets	MAPPED CONCEPTS SOURCE DATA		COMMON DATA MODEL				
A Marketplace	DATE_TIME String		measuredDateTime Datetime WeatherMeasurement > measuredDate				
حم Analytics	 Field values will be transformed to datetime data type, from YYYY-MM-DD hhamn Field values will be transformed to UTC timezone, if there is no timezone information 						
Data Models	PLANT_ID Number		id String WeatherMeasurement > stationLocal	ion > id			
	Field values will be transformed to string data type.						
	SOURCE_KEY String		temperature Double WeatherMeasurement > temperature				
	UNIDENTIFIED CONCEPTS MODULE_TEMPERATURE number IRRADIATION number						
	IRRADIATION number						

Figure 25: Finalize Mapping (Step 3)

3.2.2.2 Cleaning Configuration

If the cleaning step has been selected during the creation of the data check-in job, as shown in Figure 4, the corresponding workflow for manipulating and cleaning the data ingested into the SYNERGY Platform needs to be configured. This functionality aims at providing to the SYNERGY Platform, accurate, complete, and consistent data that could be used by the specific organization that owns the data asset, but also the overall electricity value chain that can potentially acquire it. The workflow on this functionality involves several features offered by the Cleaning Service described in Section 2 of D3.2 "Data Collection, Security, Governance & Management Services Bundles – Beta Release". In particular, the data cleaning workflow involves the definition of data cleaning rules depending on the data type of each field in the dataset, in order to eventually store a high-quality dataset. The cleaning rules are divided into validation options and corrective actions. The former involves the definition of allowed value ranges, uniqueness constraints, mandatory constraints, regular expression patterns, and outliers identification, while the latter involves dropping unnecessary columns/entries and replacing values when needed. The first step of the data cleaning workflow allows the data asset





provider to select the fields/columns to which he/she will define the cleaning rules and constraints, as shown in Figure 26.

ගී		👃 🏾 🍥 John Doe 🗸
) Home	Cleaning for: Solar Power Generation	X Cancel Save V Finalize Cleaning
ہم Data Check-in	STEP 1 Configuration	STEP 2 Review and Report
88	FIELDS Clear Selection All Datetime String Double	RULES & CONSTRAINTS
My Assets	measuredDateTime Datetime	No cleaning rules and constraints have been defined for the selected field
A Marketplace	WeatherMeasurement > measuredDateTime	Add a constraint
~	id String	
Analytics	WeatherMeasurement > stationLocation > id	
8	temperature Double	
Data Models	WeatherMeasurement > temperature	
	SAMPLE VALUES WeatherMeasurement > measuredDateTime	
	"15/05/2020 0:00" "15/05/2020 0:15" "15/05/2020 0:30" "15/05/2020 0:45" "15/05/2020 1:00" "*** *** *** *** *** *** *** ****	
	Click a field to select • Hold down Ctri and click to select	multiple fields • Press Esc to clear selection Next >

Figure 26: Cleaning Configuration – Select Fields for Cleaning (Step 1a)

It needs to be noted that the data asset provider is able to select multiple columns based on the same data type, in order to define rules that are applicable to more than one fields. As shown in Figure 27, the data asset provider may define different constraints (e.g. mandatory constraint where field values must not be null, regular expression pattern constraint where field values must have an exact match with a particular regular expression, and unique constraint where field values must be unique).





Upon this selection, he/she may define an outlier rule (e.g. drop, or replace with a particular default value) in case a value is considered as outlier.

இ		🗘 🔅 John Doe ~
습 Home	Cleaning for: Solar Power Generation	X Cancel Save V Finalize Cleaning
Data Check-in	STEP 1 Configuration	STEP 2 Review and Report
88	FIELDS Clear Selection All Datetime String Double	RULES & CONSTRAINTS
My Assets	measuredDateTime Datetime WeatherMeasurement > measuredDateTime	CONSTRAINT Mandatory Constraint Field values must not be null
مح Analytics	Id Strong WeatherMessurement > stationLocation > id	OUTLERS RULE Drop
Data Models	temperature Docete WeatherMeasurement > temperature	Cancel Create constraint
	SAMPLE VALUES	
	WeatherMeasurement > measuredDataTime "15/05/2020 0.00" "15/05/2020 0.15" "15/05/2020 0.30" "15/05/2020 0.45" "15/05/2020 1.00"	
	Click a field to select + Hold down Cttt and click to select	multiple fields - Press Ere to clear selection Next >

Figure 27: Cleaning Configuration - Set Rules & Constraints (Step 1b)

Each time a cleaning rule and constraint is added, the data asset provider can view what exactly will be applied in the data in simple language. In addition, the data asset provider may add a new rule and constraint, edit or remove an existing rule or constraint, or even change the order of the rules or constraints that are to be applied as shown in the Rules & Constraints section at the right part of Figure 28.

ஞ்		🗘 🛛 🏈 John Doe 🗸
ن Home	Cleaning for: Solar Power Generation	X Cancel Save V Finalize Cleaning
Data Check-in	STEP 1 Configuration	STEP 2 Review and Report
88	FIELDS Clear Selection All Datetime String Double	RULES & CONSTRAINTS
My Assets	measuredDateTime 2 Clearing Rules Defined WeatherMeasurement > measuredDateTime 2	Field values must not be null, otherwise the whole row will be dropped. Field values must be unique, otherwise the whole row will be dropped.
Analytics	Id [Strms] (Cleaning Rule Defined WeatherMeasurement > stationLocation > id	+ Add another constraint
Data Models	temperature Double (1 Cleaning Rule Defined) VestherMeasurement > temperature	
	SAMPLE VALUES	
	WeatherMeasurement > measuredDateTime	
	*15/05/2020 0:00" *15/05/2020 0:15" *15/05/2020 0:30" *15/05/2020 0:45" *15/05/2020 1:00" *** *** *** *** ***	
	Click a field to select • Hold down Ethil and click to sele	ct multiple fields • Press Esc to clear selection Next >

Figure 28: Cleaning Configuration - View Rules (Step 1c)





Once the data asset provider is satisfied with the cleaning configuration, by selecting the Next button, the final step of reviewing the cleaning rules and constraints will appear as shown in Figure 29.

Ŝ			👃 🛛 🂓 John Doe 🗸
6 Home	Cleaning for: Solar Power Generation	X Cancel	✓ Finalize Cleaning
P Data Check-in	STEP 1 Configuration	STEP 2 Review and Report	
88	FIELDS WITH CONSTRAINTS		
My Assets	measuredDateTime Datetime WeatherMeasurement > measuredDateTime		2 Cleaning Rules Defined
Marketplace	 Field values must not be unique, otherwise the whole row will be dropped. Field values must be unique, otherwise the whole row will be dropped. 		
Data Models	id Grane WeatherMeasurement > stationLocation > id		1 Cleaning Rule Defined
	⊘ Field values must be unique , otherwise the whole row will be dropped .		
	temperature Deuble WeatherMeasurement > temperature		1 Cleaning Rule Defined
	Field values must be between -50 and 70, otherwise the whole row will be dropped.		
	< Previous		

Figure 29: Finalize Cleaning (Step 2)

3.2.2.3 Anonymisation Configuration

Another optional pre-processing step of the data check-in job configuration is the Anonymisation step, offered by the Data Collection Services Bundle as described in Section 2 of D3.2 "Data Collection, Security, Governance & Management Services Bundles – Beta Release". In particular, if the anonymisation step is enabled in the data check-in job configuration, the execution of the Anonymisation Service, which is described in Section 4.2.1 of D2.6 "SYNERGY Framework Architecture including functional, technical and communication specifications v1", will be triggered, and the data asset provider shall define the data anonymisation rules to be executed as shown in Figure 30. During this step, a data asset provider defines which of the fields that have been mapped to the SYNERGY CIM during the Mapping step are sensitive, quasi-identifiers and identifiers since by default all fields are considered as insensitive. Sensitive data (that reveal private information such as genetic data, health data, ethnic origin, etc) should be protected by an anonymisation algorithm that ensures that an individual cannot be identified. If the data asset provider selects identifier (i.e. field that can be directly used to identify an individual) as the anonymisation type, then the particular column will be dropped from the dataset. Otherwise, if the data asset provider selects quasi-identifier (i.e. field that is not an identifier itself, but combined with other quasi-identifiers may uniquely identify an individual), the generalization method needs to be defined depending on the field type.



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



ගී		Q 🛞 John Doe 🗸
습 Home	Anonymisation for: Solar Power Generation	X Cancel Save 🗸 Finalize Anonymisation
CP Data Check-In	STEP 1 Configuration	STEP 2 Review and Report
88	FIELDS Clear Selection All Insensitive Identifier Quasi-Identifier Sensitive	ANONYMISATION RULES
My Assets	measuredDateTime DateIme WeatherMessurement > messuredDateTime	INSENSITIVE This field will remain unchanged and it will not have any effect on the anonymisation algorithm. Edit
Analytics	Id String Insenditive WeatherMeasurement > stationLocation > Id	88.
Data Models	temperature Touste Insensitive WestherMeasurement > temperature	
	SAMPLE WeatherMeasurement > measuredDateTime	
	"15/05/2020 0:00" "15/05/2020 0:15" 15/05/2020 0:30" "15/05/2020 0:45" "15/05/2020 1:00" " " " " " " " " " " "	
		Next >

Figure 30: Anonymisation Configuration – Select Fields for Anonymisation (Step 1a)

To select one of the anonymisation rules, the data asset provider needs to select the field from the left Fields section, and then select the appropriate rule by selecting the Edit button from the Anonymisation Rules section at the right section as shown in Figure 31. Depending on the data type and the anonymisation type that the data asset provider selected for a field, different anonymisation methods become available providing different configuration options. Since the cleaning step is not compulsory and the null values may not be handled till this step, the data asset providers may need to select the way that the anonymisation step will handle the null values (i.e., keep, or replace with a value).

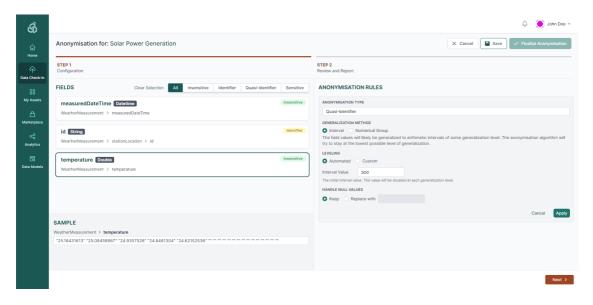


Figure 31: Anonymisation Configuration - Set Anonymisation Rules (Step 1b)





As soon as the data asset provider is satisfied with the definition that he/she has provided for a field, he/she may select the Apply button in order for the Anonymisation Rules Preview section to appear as shown in Figure 32. In this section, the different anonymisation levels and examples are provided in order to make clear to the data asset provider what exactly will happen to the data.

Ŝ		🔔 🛛 🌞 John Doe 🗸
6 Home	Anonymisation for: Solar Power Generation	X Cancel Save V Finalize Anonymisation
P Data Check-in	STEP 1 Configuration	STEP 2 Review and Report
88	FIELDS Clear Selection All Insensitive Identifier Quasi-Identifier Sensitive	ANONYMISATION RULES
My Assets	measuredDateTime Datetime Insensitive	QUASI-IDENTIFIER
Marketplace	WeatherMeasurement > measuredDateTime	INTERVAL The field values will likely be generalized to arithmetic intervals of some generalization level. The anonymisation algorithm will
~	id String Identifier	try to stay at the lowest possible level of generalization.
Analytics	WeatherMeasurement > stationLocation > id	Level 0: The initial values of temperature before anonymisation. O Level 1: Values will be generalized to arithmetic intervals of size 500.
8	temperature Double Quasi-Identifier	 Level 2: Values will be generalized to arithmetic intervals of size 1000. Level 3: Values will be generalized to arithmetic intervals of size 2000.
Data Models	WeatherMeasurement > temperature	Example
		Level 0 Level 1 Level 2 Level 3
		25.18431613 0-500 0-1000 0-2000
		25.08458867 0-500 0-1000 0-2000
	SAMPLE	24.9357526 0-500 0-1000 0-2000
	WeatherMeasurement > temperature	24.8461304 0-500 0-1000 0-2000
	"25.18431613" "25.08458867" "24.9357526" "24.8461304" "24.62152536" "" " " " " " " " " " " " " " " " " "	24.62152536 0-500 0-1000 0-2000
		Edit
		eur
		Next >

Figure 32: Anonymisation Configuration - View Anonymisation Rules (Step 1c)

Once the data asset provider has defined anonymisation rules for the fields of interest, he/she may proceed to the next step where the full set of anonymisation rules can be reviewed per anonymisation type, as displayed in Figure 33. The data asset provider may also define the acceptable information loss threshold and specific parameters of the selected anonymisation algorithm (e.g. for k-anonymity, the k-value needs to be filled in). It needs to be noted that if the desired anonymisation is not achieved or the resulting data loss is above the acceptable information loss threshold, the SYNERGY Platform will impose failure of the anonymisation step on purpose at execution time.





Ś		👃 🛛 🍥 John Doe 🗸
ل Home	Anonymisation for: Solar Power Generation	X Cancel Save V Finalize Anonymisation
Data Check-In	STEP 1 Configuration	STEP 2 Review and Report
88 My Assets	Acceptable Information Loss Threshold: 70% 2	Anonymisation Algorithm: k-anonymity (k=2) 🥒
Â	QUASI-IDENTIFIERS	
Marketplace	temperature Double. WeatherMeasurement > temperature	Interval
Analytics	 Field values are likely to be generalized in numerical intervals of size 500. Additional generalization levels may be applied. 	
Data Models	IDENTIFIERS	
	id String WeatherMeasurement > stationLocation > id	Drop Column
	⊙ This column will be dropped from the dataset.	
	INSENSITIVE measuredDateTime datetime	
	< Previous	

Figure 33: Finalize Anonymisation (Step 2)

3.2.2.4 Encryption Configuration

The data encryption functionality is available to data asset providers that require end-to-end security in the SYNERGY Platform and the On-Premise Environment (that has been installed locally) to eliminate the risk of unauthorized data access or leakage during the data transfer to the cloud. As Section 4.2.2 of D2.6 describes, data asset providers are able to (optionally) set encryption parameters to be applied on the whole dataset. It needs to be noted that, during the creation of the data check-in job, the data asset provider is asked whether to include the data encryption in the preprocessing steps (as in Figure 4).





As depicted in Figure 34, the data asset provider is able to select the concepts of the dataset whose values are to be indexed to facilitate search (as presented in section 4.1).

		🧶 🌞 Ко	onstantinos Charalam
Encrypting for: Energy	× Cancel	Save	✓ Finalize Encry
Indexing Information The fields that are toggied to be indexed, will have their metadata stored unencrysted, to enable search capabilities in the dataset. This may include minimum, maximum or unique values of the indexed Index all fields Index none of the fields Index none of the fields	l fields.		
FIELDS			INDEX
id (String) Event > id			
name Skring Event > name			
stateCount Integer Event > relatedStatus() > stateCount			
name String: Event > relatedStatus[] > name			
CreatedDateTime Datatime Event > createdDateTime			
description (States) Event > description			

Figure 34: Encryption Configuration

3.2.3 Define New Data Asset Profile

The last step of the data check-in job configuration is to store the final processed data asset in the SYNERGY Platform. Hence the data asset provider is asked to provide a title and a short description for the processed data asset that is to be stored in the SYNERGY Platform, as shown in Figure 35.

ஞ்					🧶 🔅 John Doe 🗸
습 Home	Data Loading f	or: Solar Power Generation		X Cancel	✓ Finalize Dataset
CP Data Check-in		Destination How do you want your data to be handled?	NEW DATASET Create a new dataset and load the processed data		
SS My Assets		Dataset Information Enter a title and a short description for your asset. You will be able to	наме		
Aarketplace		change these once the asset is created	Weather Data Descentrion Collected weather data.		
Analytics					
Data Models					

Figure 35: Create a new Data Asset Profile





However, a more detailed data asset profile needs to be defined by the data provider, according to the SYNERGY Metadata Schema that is presented in D4.1. Thus, the data asset provider needs to navigate to Assets on the main navigation bar. As depicted in Figure 36, the data asset provider may change the title and the description of the stored data asset, and insert additional information such as tags, distribution details, extent details, licensing information, and pricing details, according to the SYNERGY metadata schema. Depending on the access level of the data asset, the data asset provider is requested to define the applicable access policies.

Ś				Q	9 🛞 John Doe 🗸
ل Home	Edit Asset Det	tails: Weather Demo Data Uploading		X Cancel	Delete Save
لې Data Check-in		General Information General information about the profile of the specific data asset	TITLE The name of the data asset by which it can be easily identified.		
88 My Assets			Weather Demo Data		
Â			DESCRIPTION A brief overview that acts as an account of the data asset's contents.		
Marketplace			Demo data collected from openweathermaps.	h	
Data Models			TAGS A list of keywords and/or arbitrary textual tags associated with the data asset by its data provider.		
Data Models			weather x	*	
			REFERENCE The external data assets (only Other Files) to which a data asset is linked.		
			Select reference	Ψ	
		Distribution Details Information regarding the availability and access to the specific data	TYPE The nature or genre of the data asset using a controlled vocabulary.		
		asset	Text Image Model Video Audio Text and Binary		
			FORMAT The format to which the data will be available through the platform. It is not necessarily the same format with which the data that are	uploaded comply.	
			JSON ×	-	*

Figure 36: Define a new Data Asset Profile

In particular, the data asset provider needs to add tags (i.e., list of keywords, or arbitrary textual tags) that are associated with the data asset, as well as any potential reference to external data asset (already stored in the SYNERGY Platform) to which a data asset is linked. The type which describes the nature of dataset, the format of the data asset to which the data will be available, and the language of the data asset, need to be defined appropriately by selecting an item from the corresponding drop-down lists. Details regarding the temporal coverage, spatial coverage, temporal resolution, and spatial resolution units need to be defined as well by selecting the appropriate values for these units from the corresponding drop-down lists. It needs to be noted that for cases that the temporal or spatial coverage value cannot be specified in advance, the platform allows data asset providers to select temporal and spatial coverage based on the actual data that are uploaded, by selecting the "Calculated based on data" option, as shown in Figure 37.



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



	Edit Asset De	tails: Weather Demo Data Uploading			X Cancel	🖶 Delete	•
			LANGUAGE The language of the data asset.				
-in			English			-	
8							
ce		Extent Details Information regarding the coverage and granularity of the data asset from a temporal and spatial perspective	TEMPORAL COVERAGE UNIT The time period during which the data were collected or the time period the data are referring to.	TEMPORAL COVERAGE The period that the data asset covers is extracted fr concept/field in the data.	om a specific		
			Calculated based on data	WeatherMeasurement > measuredDateTime		*	
cs dels			SPATIAL COVERAGE UNIT The location/area the data refer to or were collected from: either defined directly (6.g. geographical area) or indirectly (6.e. place of interest or an activity that is the subject of the collection).	SPATIAL COVERAGE The spatial coverage of the data asset, in terms of c data asset refers.	ity code(s) to whic	h the	
			Specific City Code ~	LON			
			TEMPORAL RESOLUTION UNIT The frequency of acquiring new data from the same data source (e.g. as part of a dynamic process from a system/location/sensor).	TEMPORAL RESOLUTION The typical temporal granularity/frequency of the d asset.	ata included in a d	lata	
			Per Day 👻	1			
			SPATIAL RESOLUTION UNIT The granularity applied within the data allowing to distinguish different spaces using the data; either in terms of actual space/ground area (e.g. room, zone, building, country etc.) or as defined by a sensor/sensor network.				
			Not applicable ~				

Figure 37: Define a New Data Asset Profile - Extent Details

Then, the data asset provider needs to set the access and licensing information which corresponds to the desired visibility of a data asset and the applicable access policies. Thus, the data asset provider needs to select the access level from three different options: (a) Public, allowing the access of the data asset to any organisation without requiring any access policies to be satisfied or the existence of a data contract; (b) Private, allowing other organisations to access the data asset if the access policies are satisfied and there is an active contract; and (c) Confidential, denying access of the data asset to external organisations as it is only intended for use within the organisation that owns it. In case that the access level is set to Confidential, there are no available options regarding licensing and its associated terms under which the data asset is made available, since it will be visible only to the data asset owner.

In contrast, if the access level is set to Public or Private, the data asset provider needs to fill the corresponding licensing information, as shown in Figure 38 and Figure 39, respectively. In particular, the name of the copyright owner and the data license stating the legal terms and giving the official permission to the data asset should be selected (or provided in case of custom licenses). If the data asset provider selects an already existing well-defined data license, then the licensing details are automatically filled, otherwise the data asset provider should select Custom to fill the licensing details according to his/her needs, as shown in Figure 39. Public access level allows the data asset profile to be available to all and thus the pricing details and access policy sections are disabled since the asset is accessible by everyone.



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



Edit Asse	et Details: Weather Demo Data Uploading			X Cancel	Delete S
-in	Licensing Information Information reparding the license and its associated terms under which a data asset is made available	Access Level The desired visibility of a data asset, i.e. Confidential (not to be shared). Private (to be shared with appropriate licensing). Public (available to all).	COPYRIGHT OWNER The organization owning or managing rights over the officially acting as the data asset provider.	e data asset and	
5		Public Private Confidential	Openweather		
		O License options are based on the selected Access Level.			
ace		LICENSE The legal statement/terms giving official permission to the data asset in a	custom manner or according to well-defined data license	es.	
		Creative Commons Attribution (CC BY)		Ψ	
s		LINK A link to the exact legal terms of the specific license.			
		https://creativecommons.org/licenses/by/4.0/legalcode			
		DERIVATION An indication whether the creation and distribution of any update, adapta data asset that constitutes a derivative data asset is allowed, with permiss			
		🗹 Modify 💟 Excerpt 💟 Annotate 💟 Aggregate			
		ATTRIBUTION An indication whether it is required to give credit to copyright owner and/or asset provider. Required Not required	REPRODUCTION An indication whether from a given data asset, tempt reproductions can be created by any means, in any fo part. Allowed Prohibited		
		DISTRIBUTION An indication whether restricted or unrestricted publication and distribution of the data asset is allowed.	SHARE ALIKE An indication whether any transformation or building requires distribution under the same license.	g upon the dataset	

Figure 38: Define a New Data Asset Profile - Access Level - Public

ର୍ଭ				🧢 🔅 John Doe 🗸	
۵ Home	Edit Asset Details: Weather Demo Data Uploading			X Cancel	
Data Check-In	Licensing Information Information regarding the license and its associated terms under which a data asset is made available	Access Level The desired visibility of a data asset. i.e. Confidential (not to be shared). Private (to be shared with appropriate licensing). Public (available to all).	COPYRIGHT OWNER The organization owning or managing rights over the e officially acting as the data asset provider.	data asset and	
88 My Assets		Public O Private Confidential	Openweather		
<u> </u>		O License options are based on the selected Access Level.			
Aarketplace		LICENSE The legal statement/terms giving official permission to the data asset in a	custom manner or according to well-defined data licenses.		
~~ Analytics		Custom			
Data Models		LINK A link to the exact legal terms of the specific license.			
		Enter link			
		DERIVATION An indication whether the creation and distribution of any update, adaptat data asset that constitutes a derivative data asset is allowed with permissis Image: Modify			l
		ATTRIBUTION An indication whether it is required to give credit to copyright owner and/or asset provider. Required Not required 	REPRODUCTION An indication whether from a given data asset, tempor reproductions can be created by any means, in any for part. Allowed Prohibited		
		DISTRIBUTION An indication whether restricted or unrestricted publication and distribution of the data asset is allowed.	SHARE ALIKE An indication whether any transformation or building u requires distribution under the same license.	upon the dataset	÷

Figure 39: Define a New Data Asset Profile - Access Level - Private

By selecting Private access level, the data asset provider needs to add all the corresponding licensing information as mentioned before, along with pricing details regarding the payment and acquisition of the data asset, and the access policy defining who has access to the data asset or not, as shown in Figure 40. In particular the data asset provider needs to select the payment method (e.g. Credit / Debit Card, Bank Transfer, Online Payment Method, or Other), along with the calculation scheme (e.g. Fixed per Data Asset, Fixed per Row, or Request Depended) and the cost of the data asset with its currency.





	Edit Asset De	etails: Weather Demo Data Uploading			X Cancel	🖶 Delete	💾 Save
me			An indication whether restricted or unrestricted use of the data asset in a different context is allowed.	An indication whether storage beyond the platform downloading) is permitted for the data asset.	n (i.e. local		
ck-in			Allowed Prohibited	Allowed Prohibited			
esets			TARGET PURPOSE The intended use that the data provider allows, i.e. for business purposes, for profit purposes.	or academic purposes, for scientific purposes, for pers	onal purposes, for r	ion-	
ft atplace			Select target purpose			*	
o							
Models		Pricing Details Information regarding the payment and acquisition of the specific data asset	PAYMENT METHOD The applicable payment method that the data provider has defined in order credit/debit card, bank transfer, online payment services. Credit / Debit Card ×	r for the payment to be conducted "offline" (outside th	ne platform), e.g.	-	
			CALCULATION SCHEME The applicable cost calculation scheme for the data asset that may range from fixed per row and fixed per asset to request dependent.	COST The price for the acquisition of the data asset inclu fixed per asset.	ding its currency, if	it is	
			Fixed per Data Asset	5000	EUR	~	
		Access Policy Define who has and doesn't have access to this data	 Allow everyone to view this data 	O Deny everyone to view this data			

Figure 40: Define a New Data Asset Profile - Pricing Details

Finally, the access policies should be defined dictating whether certain users (or everyone) should be allowed or denied to view the data asset in the Marketplace (note: viewing a data asset in the Marketplace only means that the data asset consumer is eligible to acquire it – a contract needs to be put in place to get actual access to it as described in Section 4.2), by adding exceptions on certain organization or user parameters using logical conditions, e.g. on the user email or the organization type, as depicted in Figure 41.

ගී								@ (💓 John Doe 🗸
ل Home	Edit Asset De	tails: Weather Demo Data Uploading					X Cancel	聞 Delete	Save
Data Check-in			Credit / Debit Card x					•	Â
88			CALCULATION SCHEME The applicable cost calculation scheme for the data from fixed per row and fixed per asset to request d		COST The price for the acquisition of fixed per asset.	the data asset includin	ig its currency, if it	is	
My Assets			Fixed per Data Asset	v	5000		EUR	•	
Marketplace		Access Policy							
Analytics		Define who has and doesn't have access to this data	 Allow everyone to view this data 		 Deny everyone to vie 	w this data			
Data Models			+ Add an exception						
			User email	is equal to		e.doe@example.com		8	
					AND				
			Organisation type	is equal to	Selec	t	- (Ð	
				+ AND (add	another condition)				
			 Anyone with email "johnie.doe@example 	e.com"					
					Can	Create po	olicy exception		

Figure 41: Define a New Data Asset – Private Access – Deny Everyone with Exceptions





By selecting Save, the data asset profile is finalized, and the data asset provider may navigate to the overview, the license details and the data structure of the data asset profile that has been created and stored. The data asset profile overview, the license details and the data structure pages are depicted in Figure 42, Figure 43, Figure 44, and Figure 45, respectively.

Ĝ	ot 🛞 🌒	nn Doe 🗸
G Home		
Data Check-In		
SS My Assets	Weather Demo Data	
A Marketplace	Overview Access Policies License Details Data Structure	
Analytics	General General information regarding the data asset.	
Data Models	Description Demodulate of from openweathermaps.	
	Togs weather optimization	
	Distribution Details Information regarding the availability and access to the specific data asset.	
	Type © Format © Text JSON	
	Velocity © Accessibility © Batch Through an API	

Figure 42: Define a new Data Asset Profile – Overview (visible to all)

Ĝ	🧶 🍥 John Doe 🛩
G Home	
Data Check-In	
US My Assets	Weather Demo Data
A Marketplace	Overview Access Policies License Details Data Structure
Analytics	Active Policies Information about the access policies that are enforced over this data asset.
Data Models	Deny everyone to view this data
	Exceptions (Evaluated sequentialy) O Anyone with email "johnie.dee@example.com"

Figure 43: Define a new Data Asset Profile – Access Policies (visible only to the respective data asset provider)





ගී	5.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1	🧔 🛞 John Doe 🗸
G Home		
Data Check-In		
BB My Assets	Weather Demo Data	🛞 Edit 🗘 Retrieve
Ameriketplace	Overview Access Policies License Details Data Structure	
حچ Analytics	Licensing Information Information regarding the license and its associated terms under which the specific data asset is made available.	
Data Models	Access Level © License © Private Custom	
	Copyright Owner () Link () Openweather N(A	
	Derivation ① Attribution ① Modify Required	
	Reproduction © Distribution © Allowed	
	Share Alike © Re-context © Required Allowed	
	Offline Retention © Target Purpose © Allowed Business	

Figure 44: Define a new Data Asset Profile - License Details

Ŝ	🥵 🍥 John Doe 🗸
G Home	
Cryphone Data Check-In	
88 My Assets	Weather Demo Data
A Marketplace	Overview Access Policies License Details Data Structure
ବ୍ଦ୍ୱି Analytics ରିପ Data Models	Weather/Measurement Owner An object containing the statubiccondition of the atmosphere at a particular location or area over a certain period of time. Weather is usually described at any given time by six main sub-components: temperature, atmospheric pressure, wind, hunding yu, registration, and conditions. These sub-components: can represent either current weather measurements, or the weather forecast for a particular location during a certain time period. Measure@DateTime Exetume A datatime sequence identifying when a weather condition has been measured. Id String An objective instance of the structure of the structur

Figure 45: Define a new Data Asset Profile – Data Structure

Obviously, the data asset provider is able to edit the metadata (with the exception of the Data Structure) at any point through the Assets view (described in Section 3.5).

3.3 Execute a Data Check-in Job

3.3.1 Pre-processing Rules Execution

As mentioned already, the pre-processing steps configuration (at design time) has a distinct separation from the actual execution of each pre-processing step. In this section, a short description





regarding the execution of each pre-processing step is provided accordingly. Note that, all the preprocessing steps (i.e., Mapping, Cleaning, Anonymisation, and Encryption) can be only executed if each pre-processing configuration has been finalised.

3.3.1.1 Mapping Execution

If the Mapping step has been successfully executed, the data asset provider may view what transformations happened on the data per field in the source data, as shown in Figure 46.

Ĝ				Ø	🂓 John Doe 🗸				
ل Home	Mapping for: Solar Power Gener	Mapping for: Solar Power Generation							
	MAPPING & TRANSFORMATION E	EXECUTION REPORT	Domain: Electricity Data Model for SYNERGY-Early Standard: None	e Category: Wea	therMeasurement				
Crophone Check-In	SOURCE DATA	COMMON DATA MODEL							
88 My Assets	DATE_TIME String	measuredDateTime Datetime WeatherMeasurement > measuredDateTime	Transformation Successfully Executed	Transformed 3.2k	Set to Null O				
۵	Field values were transformed to dateting	ne data type, from YYYY-MM-DD hh:mm:ss to ISO 8601							
Marketplace	Field values wwere transformed to UTC	timezone, if there was no timezone information in the data.							
ح Analytics	PLANT_ID Number	id String WeatherMeasurement > stationLocation > id	Transformation Successfully Executed	Transformed 3.2k	Set to Null				
Data Models	Field values were transformed to string.								
	AMBIENT_TEMPERATURE Numb	temperature Double WeatherMeasurement > temperature	Transformation Successfully Executed	Transformed 3.2k	Set to Null O				
	Field values were transformed to double	e data type.							

Figure 46: Mapping Execution - Successful

Otherwise, if the Mapping step failed due to wrong transformation rule definitions, the data asset provider may view the failed transformations that have been defined (as shown in Figure 47) and make the appropriate corrections in a new Data Check-in Job. It needs to be noted that in the case of recurring jobs, the Mapping step may have run successfully for n times and failed in the n+1 time, which is why exact number of rows/values are provided per execution.





		ation Data		< Go bac	k 🔛 Update
	Aapping step failed! The mapping step for the data check-in job	plailed. You may see a summary of what went wrong in order to revise the transformations that have been defined, accordingly.			
	IAPPINGS DURCE DATA	COMMON DATA MODEL			
D	ATE_TIME String	measuredDateTime Datetime WeatherMeasurement > measuredDateTime	Transformation Failed	Transformed 0 (+0)	Set to Null 0 (+0)
۲	Field values were transformed to datetin	ne data type, from Unix Timestamp (nanoseconds) to ISO 8601	Field values cann	ot be parsed in the spe	cified date format
۲	Field values wwere transformed to UTC	timezone, if there was no timezone information in the data.	Transformation not attempted		
PL	LANT_ID Number	VeatherMessurement > id	Transformation Successfully Executed	Transformed O (+3.2k)	Set to Null 0 (+0)
ø	Field values were transformed to string	data type.			
A		temperature Double Weather/Measurement > temperature	Transformation Successfully Executed	Transformed 0 (+3.2k)	Set to Null 0 (+0)
	Field values were transformed to double	data type.			

Figure 47: Mapping Execution - Failed

3.3.1.2 Cleaning Execution

Once the data provider completes the Cleaning configuration and it is executed, a pop-up appears notifying the user that the Cleaning step has been executed. An indicative example of a successful execution of the Cleaning rule is shown in Figure 48. In this step, the data asset provider may view what cleaning constraints were met and resulted into transformations per field in the source data.

Cleaning	g for: Solar Power Generation					< Go I	Nack 📔 Upda
CLEANIN	IG STEP STATISTICS FOR ALL SUCCES	SSFUL EX	ECUTIONS				
3	3182 Rows Initially in the Dataset		99.97% of the Dataset Dropped	0.00% of the Dataset Transformed	1 Row	s Remaining in t	he Dataset
FIELDS	VITH CONSTRAINTS						
	eadDateTime Datetime			All Cleaning Rules Su	ccessfully Executed	Drops O	Replacement 0
	values which were null, were dropped.					0	_
id Strin Weatherly	g leasurement > stationLocation > id			All Cleaning Rules Su	ccessfully Executed	Drops 3181	Replacement 0
Field	values which were not unique were dropped.					3181	-
	ature Double leasurement > temperature			All Cleaning Rules Su	ccessfully Executed	Drops O	Replacemen O
@ Field	values between -50 and 70, were dropped.					0	_

Figure 48: Cleaning Execution - Successful

However, in the case where the field values of a column do not follow the defined pattern, as configured in the Cleaning configuration rules, a failed Cleaning execution report page will appear,





as shown in Figure 49, notifying the data asset provider to view the errors that have occurred, in order to revise the cleaning rules and constraints that have been defined.

ගී		🧶 🔅 Johr	in Doe 🗸
ሬ	Cleaning for: Solar Power Generation	< Go back 🗎 U	pdate
Home	Cleaning step failed! Generic error during transformation		
88 My Assets	FIELDS WITH CONSTRAINTS		
A Marketplace	measuredDateTime DateLine WeatherMeasurement > measuredDateTime	1 Cleaning Rule De	efined
~	O Field values must have an exact match with the asd regular expression, otherwise the whole row will be dropped.		
Analytics	temperature Double WeatherMeasurement > temperature	1 Cleaning Rule De	efined
Data Models	⊙ Field values must be between 20 and 50, otherwise the whole row will be dropped.		

Figure 49: Cleaning Execution - Failed

3.3.1.3 Anonymisation Execution

If the Anonymisation step has been executed and the achieved information loss is less than the acceptable information loss threshold that has been defined during the configuration, the Anonymisation step is considered as successful as Figure 50 depicts.

ගී				🧔 🐞 John Doe 🗸
G Home	Anonymisation for: Solar Power Generation (File Upload)			< Go back Update
Pata Check-In	Achieved Information Loss: 33.33% (lower than the Acceptable Information Loss Threshold: 70%)		Anonymisation Algo	orithm: k-anonymity (k=2)
88	QUASI-IDENTIFIERS			
My Assets	name (Sting) WeatherMeasurement > stationLocation > name	Aasking	Generalization Level	Information Loss
Marketplace	⊘ Field values remained unchanged.			
Analytics				
Data Models				
	INSENSITIVE			
	ld string temperature double			

Figure 50: Anonymisation Execution - Successful



SYNERGY

On the contrary, if the achieved information loss is higher than the acceptable information loss threshold that had been defined during the configuration or if the desired level of anonymisation was not achieved, the Anonymisation step is considered as failed as Figure 51 suggests.

Ŝ		🦉 🐞 John Doe 🗸
ل Home	Anonymisation for: Solar Power Generation	< Go back
Data Check-in	Anonymisation step failed!	
88 My Assets	Acceptable Information Loss Threshold: 70%	Anonymisation Algorithm: k-anonymity (k=2)
A	QUASI-IDENTIFIERS	
Marketplace	temperature Double. Weather/Measurement > temperature	Interval
Analytics	⊘ Field values are likely to be generalized in numerical intervals of size 500. ⊗ Additional generalization levels may be applied.	
Data Models	IDENTIFIERS	
	id (String) WeatherMeasurement > stationLocation > id	Drop Column
	⊘ This column will be dropped from the dataset.	
	INSENSITIVE	
	measuredDateTime datetime	

Figure 51: Anonymisation Execution - Failed

3.3.1.4 Encryption Execution

Similarly to the rest of the pre-processing steps, the data asset provider is able to view the report of the encryption execution, displaying whether the encryption step was executed and applied successfully or not (according to the configuration provided), as depicted in Figure 52.

		Ø Konstantinos Charalam
E	ncrypting for: Energy	X Cancel
In	Indexing Information The fields that are toggled to be indexed, will have their metadata stored unencrypted, to enable search capabilities in the dataset. This may include minimum, maximum or unique values of the indexed fields.	
F	IELDS	INDEX
	ici String	
	Event > Id	
1	name String	
	Event > name	
	stateCount Integer Event > relatedStatus[] > stateCount	
	name String	
	Event > relatedStatus[] > name	
	createdDateTime Datetime	
	Event > createdDateTime	_
ſ	description String	
	Event > description	

Figure 52: Encryption Execution





3.4 Upload data through the On-Premise Runner

In the case that the data asset provider has already installed an On-Premise Runner and he/she needs to upload data through it, the On-Premise Execution option should be selected as shown in Figure 4. After selecting the pre-processing steps that are to be applied on the uploaded data, by selecting Save, the Setup Harvester page for the On-Premise Execution appears, as shown in Figure 53. It needs to be noted that the only data loading option that the data asset provider has using the On-Premise Execution method, is to upload files.

Ĝ						۵ 🏶	Konstantinos Charalampous 🗸
fair Home	Configure Harv	vester: Energy			× Cancel	Save	✓ Finalize Configuration
Pata Check-In	STEP 1 Setup Harvest Service	ce		STEP 2 Test and Review Configuration			
88 My Assets		Data Loading How do you plan to load your data to the platform?	 FILE UPLOAD Direct file upload (CSV, JSON, 	XML)			
A Marketplace		Format					
Analytics		Select the format of the file you will use	○ CSV O JSON ○ XML ●	Other			
Data Models Q Retrieval		Sample Upload Upload a sample of your data to be used in next steps	BROWSE Denergy_sample.	ison 135.3 KB (Sample has been cropped)			
		File(s) path Provide your file(s) paths to be processed (if in csv, json, xml format)	FILE PATH /home/konstantinos/Downloads/	datasets/energy_data.json			
							Next >

Figure 53: File Upload Method - On-Premise Execution - Setup Harvest Service (Step 1a)

Although the same procedure as in the File Upload Method in the Cloud Execution, for this case the data asset provider should include the full path to the file including the data, and in the next step the sample is uploaded in Step 1 and viewed in Step 2, as shown in Figure 54.





Configure Harvester: Energy	X Cancel Save V Finalize Configu
STEP 1 Setup Harvest Service	STEP 2 Test and Review Configuration
Data Sample The details of the data sample that was uploaded	r
	("meta": ("view": ("id": "8yq3-m0wp",
	"name": "Energy Usage 3030", "attribution: "City of Chicago", "attributioni.Mr: "http://www.cityofchicago.org", "averageRating": 100,
	"Category": "Environment & Sustainable Development", "createdatt": 30600588, "description": "Displays several units of energy consumption for households, businesses, and industries in the City of Chicago during 2020. Electric The data was aggregated from ComEd and Peoples Natural Gas
	by Accenture. Electrical and gas usage data comprises 88 percent of Chicago's buildings in 2010. The electricity data comprises 68 percent of overall electrical usage in the city while gas data comprises 81 percent of all gas consumption in Chicago for 2010. Census blocks with less than 4 accounts is displayed at the domunity Area without further geographic identifiers. This dataset also contains
	<pre>selected variables describing selected characteristics of the Census block population, physical housing, and occupancy.", "displaytype": "table",</pre>
	"domloadount: 7889, "hideFrontLato: 'false", "hideFrontLatJson': "false",
	"indexUpdatedAt": 1453833034, "newBatckend": "true", "numberOfCommerts": 0,

Figure 54: File Upload Method - On-Premise Execution - Setup Harvest Service (Step 2)

The rest of the steps are configured as presented in Section 3.2.

3.5 Manage Data Check-in Jobs

3.5.1 Edit a Data Check-in Job

Data asset providers are able to edit an existing data check-in job by navigating to the Data Check-in Jobs view, as shown in Figure 3. In particular, a data check-in job can be edited by selecting Edit from the options menu that is located at the right side of each data check-in job. Directly the Update Data Check-in Job view will be appear as depicted in Figure 55. Although the data asset provider is able to update the name and description of the data check-in job, the pre-processing rules and the data check-in execution location cannot be updated. By selecting the Update button at the top right part of this view, the data check-in job is updated and the data asset provider returns back to the Data Check-in Jobs view.





ගී				🧔 🐞 John Doe 🗸
G Home	Update Data C	heck-in Job		X Cancel V Update
Data Check-In		Job Details Basic information about the job	NAME Solar Power Generation	
88 My Assets			DESCRIPTION Data to be checked-in using the platform's API.	
Marketplace		Pre-processing Before importing your data to the platform, you can use additional tools to better prepare them	HARVESTER Collect data through files, extnernal APIs or other services MAPPINO dista to the common data model CLEANING CLEANING Detect and address data quality issues, such as compt or inaccurate records AnonYMISER AnonYMISER CLEANING CLEANING	
		Execution You can choose whether you want the data-checkin job to be run on the cloud or on-premise through a registered runner	CLOUD EXECUTION The data-checkin job will run on the cloud. ON-PREMERE EXECUTION The data-checkin job will run on-premise through a registered runner.	

Figure 55: Update a Data Check-in Job

3.5.2 Delete a Data Check-in Job

In the Data Check-in Jobs view, depicted in Figure 3 by selecting Delete from the options menu at the right side of each data check-in job, the data asset provider can delete a data check-in job. This will delete completely and permanently the configuration which means that if a data asset provider wants to execute again a data check-in job, he/she will not be able to do so. It needs to be noted that deleting a data check-in job does not affect the data that have been already stored in the SYNERGY Platform which can be deleted as described in Section 3.6.2.

3.6 Manage Data Asset Profiles

3.6.1 View all Data Asset Profiles

The data asset provider may view the list of all the data asset profiles that belong to his/her organisation or have been acquired by his/her organisation in the Assets View as shown in Figure 56, regardless its status (e.g. deprecated, incomplete, available). By selecting the options icon located at the right side of each row, the data asset provider may delete, or edit a particular data asset profile. The editing of a data asset profile, involves the same procedure followed during the definition of a new data asset profile as described in Section 3.2.3.





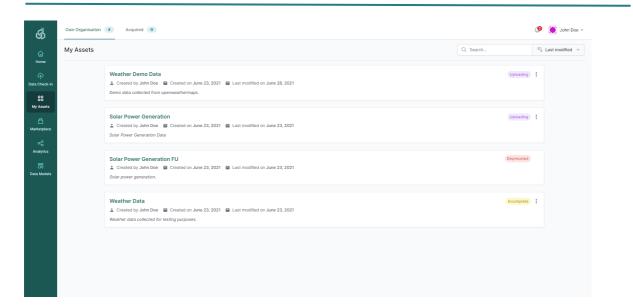


Figure 56: View all Data Asset Profiles

3.6.2 Delete a Data Asset Profile

A data asset provider may delete a Data Asset Profile by selecting Delete from the options menu located at the right side of each Data Asset Profile, as depicted in Figure 56. It needs to be noted that the actual data associated with the Data Asset Profile are deleted in case there is no active data asset contract. Any data retrieval operations associated with the specific dataset (as described in Section 4.3) are not available after the deletion of a particular Data Asset Profile, while the corresponding entry (in terms of metadata) will remain in the SYNERGY Platform as a deprecated Data Asset.





4 Data Search and Acquisition User Journey

A fundamental goal of the SYNERGY Platform is to provide all the functionalities needed for sharing and trading data assets in a secure and trustful manner, to stakeholders of the electricity data value chain. It needs to be noted that data assets can range from raw data, to datasets, analytics results, and others. Hence, it is of high importance that data asset consumers that need to acquire data assets from data asset providers, reach an agreement stating the data sharing terms that is prepared based on the contract preparation process followed in the SYNERGY Platform. The contract preparation process is described in the subsequent subsections, while the different steps are depicted in Figure 57. In particular, a data asset consumer may request for quotation for a particular (or possibly multiple) data assets; the data asset provider receives the request for this data asset, checks its content, and prepares a draft data asset contract which is then sent back to the asset consumer to accept, negotiate, or reject the prepared draft contract accordingly. This process is repeated until the data asset consumer and the data asset provider reach consensus, and hence they can settle the contract which is then activated by the SYNERGY Platform.

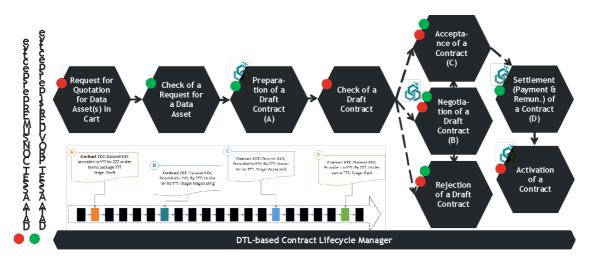


Figure 57: Contract Preparation Process

Once the data asset consumer has an active contract with the provider, the acquired data asset is available for use in the SYNERGY Platform (in the consumer's organization Secure Experimentation Playground and / or locally in consumer's On-Premise Environment, if the contract's terms allow it). It needs to be noted that both stakeholders of the electricity data value chain and the SYNERGY energy apps developers are expected to utilize the data search and acquisition user journey to get access to data from different stakeholders in the SYNERGY Platform. In this section, the core three components involved in the Data Sharing Services Bundle, namely (a) the Data & AI Marketplace, (b)



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



the Contracts Lifecycle Manager, and (c) the Remuneration Engine are described through the Data Search and Acquisition User Journey.

4.1 Navigate to the SYNERGY Marketplace

Data asset consumers may access the SYNERGY Data & AI Marketplace where they can navigate through the data assets that they are eligible to view (upon satisfying the access policies set by their respective providers).

As depicted in Figure 58, data asset consumers view highlight info regarding each data asset such as the type of data asset (i.e., dataset in this beta release), the title of the data asset, a short description regarding the data asset, the organisation that owns or manages the data asset, and a cover image related to the data asset. It needs to be noted that data assets that are stored locally are considered as not to be shared and are not shown in the SYNERGY Marketplace.

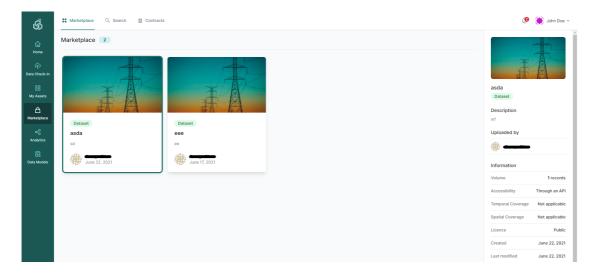


Figure 58: SYNERGY Data & AI Marketplace

Data asset consumers navigating through the SYNERGY Marketplace, can potentially search for a particular data asset by typing in the search bar, and/or by browsing data assets by different parameters (e.g. domain, category, accessibility, type, format, and language). The corresponding results from the search functionality are filtered and displayed directly in the Results View, as depicted in Figure 59, where one can sort the results based on relevance, title, date available, date modified, and volume.





ගී	SS Marketplace Q Search 🗄 Co	ontracts		<i>.</i> 9	🂓 John Doe 🗸
G) Home	Results 1		Save Relevance V		- Â
Data Check-in	< FILTERS CLEAR	R 📕		I	
Data Check-In	All ~	Q		asda	A.
My Assets		~	A A A A A A A A A A A A A A A A A A A	Dataset	
۵	 Electricity Data Model for SYNERGY-Early 1623938113056-energy 			Description	
Marketplace	Show concepts	>	Dataset	sd	
~	CATEGORIES		asda	Uploaded by	
Analytics	AirConditioningSystem	1	sd		
8	ACCESSIBILITY		Alte		
Data Models	Through an API		June 22, 2021	Information	
				Volume	1 records
	🔲 Text			Accessibility	Through an API
				Temporal Coverage	Not applicable
	JSON			Spatial Coverage	Not applicable
		>		Licence	Public
				Created	June 22, 2021
				Last modified	June 22. 2021

Figure 59: Search Page

If the data asset consumer selects a particular data asset in the Marketplace, he/she will be redirected in a new browser tab where the overview and more detailed information of the data asset regarding license information, and data structure (as described in Section 3.1.3) will appear.

Additionally, a data asset consumer may perform advanced queries by selecting a domain of his/her interest which reveals the Show concepts button that enables the Concepts slideover as shown in Figure 60. Hence the data asset consumer may select multiple fields/concepts to build a data query, which enables the Data Query slideover for adding, editing, or removing conditions to the query.

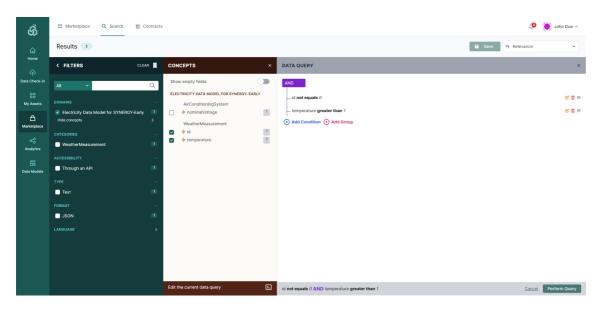


Figure 60: Search Page - Advanced Data Query





Selecting to "Perform a Query", a search within the selected concepts and the added conditions is triggered, and the corresponding results are shown as in Figure 61.

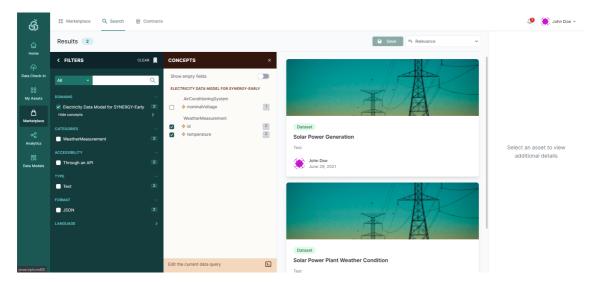


Figure 61: Search Page - Advanced Data Query - Results

4.2 Acquire a Data Asset

As organisations in the SYNERGY Platform may undertake the role of both data asset providers and consumers, their members may view the contracts of the data assets they are involved in (as data asset providers or consumers), search for a particular contract, filter the contracts based on their status, and sort them, as shown in Figure 62. These contracts are visible in a list view, while for each contract, basic information such as the data asset title, the contract details (i.e., date created, date last updated, data asset provider, data consumer), the contract status, and any available action.





Contracts 0			
٩		=rh Title ≎	
ASSET TITLE	CONTRACT DETAILS	STATUS ACTIONS	
Data Asset 1 Description of contract 1	Created on Jun 11, 2021 - Last updated on Jun 11, 2021 Provider: Organization X 2 Consumer: Organization Y	Active ©	
Data Asset 2 Description of contract 2	Created on Jun 11, 2021 - Last updated on Jun 11, 2021 Provider: Organization X 2 Consumer: Organization Y	Draft	
Data Asset 3 Description of contract 3	Created on Jun 11, 2021 - Last updated on Jun 11, 2021 Created on Jun 11, 2021 Consumer: Organization Y	Negotiate	
Data Asset 4 Description of contract 4	Created on Jun 11, 2021 - Last updated on Jun 11, 2021 Provider: Organization X 2 Consumer: Organization Y	Rejected	
Data Asset 5 Description of contract 5	Created on Jun 11, 2021 - Last updated on Jun 11, 2021 Created on Jun 11, 2021 Consumer: Organization Y	Request ©	
Showing 1 to 5 of 7 results		< 1 2 >	

Figure 62: SYNERGY Data & AI Marketplace – Contracts List

Data asset consumers may initiate the process for formally acquiring a data asset by creating a request to the data asset provider as shown in Figure 63. In particular, the data asset consumer needs to insert the details regarding the data asset request, and specifically to select the duration of use of the data asset (including any updates available during the respective period), to select whether the data asset will be used exclusively in the SYNERGY platform or if the data consumer would like to download the data asset "locally" through the SYNERGY Platform APIs. In addition to these, the data asset consumer should optionally insert a message to be provided to the data asset provider, describing how the data asset will be used. Then, the data asset consumer may view the data asset structure and the processing rules that have been applied on the data, and select the exact fields to be acquired, if not the whole data asset. During this step, the data asset consumer may also apply filters to the (unencrypted) fields to obtain a data asset "slice", whenever applicable. Once the data





asset consumer is satisfied with the details inserted regarding the request of the data asset, he/she is able to submit the request to the organisation's manager at the data asset provider side.

ஞ்				۹
Ga Home ⊕	Create a new Request for a Data Asset	Request Create Draft Contract	Finalize Contract Activate Contract	X Cancel + Submit
Data Check-in				
OO My Assets	Wind Production Data Asset Provider: Test Organisation Ltd Data related to wind production			
≙				
Marketplace	Request Details	Duration of Use ①	○ For 1 days ○ For 1 months ○ For 1 years ◎ F	orever
%	Details about your request for the specific data asset(s).	Offline Retention ③	Use exclusively in the SYNERGY Platform Use locally	
Analytics		Message ()	Test message	
Φ				
Retrieval	Data Asset Info	Transformation Rules Cleaning Ru	les Anonymisation Rules Encryption Applied	
	Details about the data asset structure and processing rules that have been applied on the data. You need to select what fields you would like to acquire.	WindTurbine Object A device that converts the wind's kit relatedEnergyGenerationMee The association of the wind turbin observedDateTime DateIrim The observation time for the mee netGenerationWind Double	vetic energy into electrical energy: surements with its measurements aurement	

Figure 63: Request for Data Asset Acquisition by the Organisation's Manager (Consumer Perspective)

4.2.1 Review a Data Asset Request (Data Asset Provider Perspective)

The manager of the organisation (acting as the data asset provider) is able to review the details of each request made by a potential data asset consumer. He/she proceeds to reject or accept the request for the data asset based on the information provided, as depicted in Figure 64.





Ŝ			Ø	*
습 Home	Review a Request for a Data Asset (Request)		Back To Contracts X Decline Request	Offer a Contract
ta Check-in	Review Request	Create Draft Contract	Finalize Contract Activate Contract	
00 Iy Assets	Wind Production Data Asset Consumer: Suite5			
arketplace	Data related to wind production			
	Request Details Details about your request for the specific data asset(s).	Duration of Use ()	Forever Use exclusively in the SYNERGY Platform	
Analytics		Message [©]	Test message	
ata Models				
Ω Retrieval	Data Asset Info The fields of the dataset included in the contract and their applicable filters	 WindTurbine Ob A device that convert relatedEnergyGe 	Cleaning Rules Anonymisation Rules Encryption Applie stewind's kinetic energy into electrical energy. nerationMeasurements Object he wind turbine with its measurements	1
			Time Datetime	All values 🖓

Figure 64: Review Request for Data Asset Acquisition by the Organisation's Manager (Provider Perspective)

4.2.2 Prepare a draft contract (Data Asset Provider Perspective)

In case the organisation's manager acting as data asset provider has accepted the request, a draft contract for the corresponding data asset is prepared as shown in Figure 65. The manager may review once more the details regarding the data asset and its licensing, define the payment details, and prepare the contract terms that concern the data asset acquisition between the data asset provider and consumer. It needs to be noted that the terms that are enforced by the SYNERGY Platform are also listed and cannot be edited by any party.





			۹ 🎆
	repare a Draft Contract for a Data Asset	O	Back To Contracts Signate Contract
-in	Wind Production 2 Data Asset Consumer: Suite5 Data related to wind production		
e	Contract Details Details about your contract for the specific data asset(). 🗎 Pending 🕓 Forever	8217 records
IS	Data Asset Info The fields of the dataset included in the contract and their applicable filters	Transformation Rules Cleaning Rules Anonymisation Rules WindTurbine Object A device that converts the wind's kinetic energy into electrical relatedEnergyGenerationMeasurements ObservedDateTime The observation time for the measurement netGenerationWind Double output	
Ρ	repare a Draft Contract for a Data Asset	•	Back To Contracts
e	Data Asset License The dataset's license metadata		notate O Aggregate
S	Payment Details The payment details of dataset	Price € 100 EUR ∨ Payment Method Through a bank transfer	VAT 20 %
	Contract Terms The contract terms that concern the data asset acquisition between the data asset provider and consumer	Heading 2 ◆ B I U ↔ = = = = >> ↔	IΞ ≔ <u>A</u> ﷺ <i>I</i> _x

Figure 65: Create Draft Contract for Data Asset Acquisition by the Organisation's Manager (Provider Perspective)

The organisation's manager is then able to sign the draft contract upon inserting the password of the organisation's wallet (which he/she manages), in order to unlock the wallet and sign the draft contract in the blockchain.

4.2.3 Review a draft contract (Data Asset Consumer Perspective)

Once a draft contract is signed by the data asset provider organisation, the manager of the data consumer organisation is notified and may view the contract details, as shown in Figure 66. Upon





reviewing the details of the draft contract prepared by the data asset provider, the manager of the data asset consumer may accept, negotiate or reject the offer accordingly. During this step, the data asset consumer is also able to download the draft contract as a pdf file.

Ŝ		🧶 🚸 👘]
යි Home	Review the Draft Contract for a Data Asset	Draft Contracts Reject Pregotiate Accept	
Data Check-in	Create Request	Create Draft Contract Finalize Contract Activate Contract	
88 My Assets Marketplace & Analytics Data Models Retrieval	Contract Terms The contract terms that concern the data asset acquisition between the data asset provider and consumer	Term 1 Term 1.1 Iorem ipsum Term 1.2 Iorem ipsum Term 2 Iorem ipsum	
	SYNERGY Platform Terms The contract terms that are enforced by the SYNERGY Plarform	In its current release, the SYNERGY Platform does not impose any additional terms in comparison to the <u>platform's terms of use</u> .	

Figure 66: Review Draft Contract for Data Asset Acquisition (Consumer Perspective)

If the manager representing the data asset consumer has opted to accept the draft contract, he/she is prompted to insert the organisation's wallet password and write the contract to the SYNERGY blockchain (as depicted in Figure 67). The process continues as described in Section 4.2.7.

Ŝ				🧢 🌐 Mario	s Phinikettos 🗸
	Prepare a Draft Contract for a Data Asset	equest		K Back To Contracts	✓ Sign
	Review Request				
	Data Asset License The dataset's license metadata Wait	ng for transaction to be mined and confirmed.	Annotate	8 Aggregate	
	_	This might take a while Re-context Offline Retention	Oistribution	Share Alike	
	Payment Details The payment details of dataset	Price € 1000	EUR 🗸 VAT	20 %	
		Payment Method Through a bank transfer			
	Contract Terms The contract terms that concern the data asset	Heading 2 \Rightarrow B I U \Rightarrow \equiv \equiv \equiv \equiv	"** 🗄 🖂 🧍	A Ix	

Figure 67: Sign a Contract for Data Asset Acquisition (Consumer/Provider Perspective)





On the contrary, if the manager representing the data asset consumer has rejected the draft contract's terms, then the process concludes, and the data asset consumer does not get access to the requested data asset. The data asset provider also views the contract with "rejected" status.

4.2.4 Negotiate a draft contract (Data Asset Consumer Perspective)

In case that the data asset consumer chooses to negotiate any part of the contract, e.g. regarding payment and/or terms (as displayed in Figure 68), he/she is able to change the payment details such as the cost, as well as update the contract terms by adding new terms, editing or removing the existing ones. Following that, the data consumer manager will be able to sign the revised contract, by inserting the organisation's wallet password and writing the contract to the SYNERGY blockchain.

Ś			۹ 🔅
G Home	Negotiate the Draft Contract for a Data Asset	Draft	K Back To Contracts Sign
ጥ	Create Request	Create Draft Contract Finalize Contract Activate Contract	
Data Check-in	Payment Details The payment details of dataset	Price € 500 EUR ∨ VAT Payment Method Through a bank transfer VAT	20 %
Analytics	Contract Terms The contract terms that concern the data asset acuisition between the data asset provider and consumer	Heading 1 \Rightarrow B I U \Rightarrow	. (A) Ix

Figure 68: Negotiate Draft Contract for Data Asset Acquisition (Consumer Perspective)

4.2.5 Review a revised contract (Data Asset Provider Perspective)

Once a contract is revised, the data asset provider manager needs to review the changes that the data consumer asked for, as shown in Figure 69. As it happened in the case of the draft contract on behalf of the data asset consumer, though, in Section 3.3.4, a data asset provider is able to accept, reject or further negotiate the revised contract.





Ĝ		
A Home	Review Contract Negotiation for a Data Asset Negotiate	e C Back To Contracts X Reject P Negotiate Accept
م ata Check-in	Review Request Create D	raft Contract Finalize Contract Activate Contract
My Assets	Payment Details Price The payment details of dataset VAT Payment I Payment I	€500 Price Decreased - Previous Price: €1000 20% Method Through a bank transfer
arketplace & Analytics ata Models Retrieval	acquisition between the data asset provider and Cerrm for the data asset provider asset provider as as as asse	m 1 - new n 1.1 psum loremipsum n 1. 2 3 p sum Lorem
		m 2 - updates

Figure 69: Review Updated Draft Contract for Data Asset Acquisition (Provider Perspective)

It needs to be noted that this procedure described in Sections 4.2.4, 4.2.5 and 4.2.6 may iterate until both the data asset provider and data asset consumer are satisfied with the contract for the data asset acquisition.

4.2.6 Settle a finalized contract (Data Asset Provider Perspective)

Once the data asset provider and the data asset consumer have reached consensus and signed the same version of the contract, the respective payment, if any, needs to be settled according to the payment method that is mentioned in the signed contract. In the initial version of the SYNERGY platform, when the data consumer has paid the price of the data asset as dictated in the contract's terms, the data asset provider needs to confirm the payment and thus the contract will be activated as shown in Figure 70.





Ŝ	🧶 🎆
G Home	Activate the Contract for a Data Asset Signed Create Draft Contract Finalize Contract Activate Contract
Data Check-in	
88 My Assets	Wind Production Download Contract as PDF Download Contract as PDF Data Asset Consumer: Suite5 Data related to wind production Data related to wi
Marketplace	Contract Details Pending Forever
Analytics	Details about your contract for the specific data asset(s).
Data Models	Data Asset Info Transformation Rules • Cleaning Rules • Anonymisation Rules • Encryption Applied The fields of the dataset included in the contract and the fields of the dataset included in the contract and
C Retrieval	their applicable filters VindTurbine Object A device that converts the wind's kinetic energy into electrical energy. relatedEnergyGenerationMeasurements Object The association of the wind turbine with its measurements
	Image: Second Secon

Figure 70: Activate Draft Contract for Data Asset Acquisition (Provider Perspective)

As soon as the contract has been activated by the data asset provider, the data consumer is eligible to acquire the data asset according to the contract's terms, e.g. the SYNERGY platform transfers the data asset to the data asset consumer's Secure Experimentation Playground (for use only in the SYNERGY platform).

4.3 Retrieve a Data Asset

4.3.1 Create a Retrieval Query

Data asset consumers may create a new Retrieval Query defining its title, description and acquisition method as described in the following figure.





Create	new Retrieval Query		× Cancel
In	Retrieval Query Details Basic information about the Retrieval Query	ти	
		Enter Retrieval Query title DESCRIPTION	
		Enter a short description for your retrieval query	
	Retrieval Accessibility The retrieval accessibility mechanism	THROUGH AN API Expose an API to retrieval data from assets - applicable to all assets that have been mapped to a data model	
s		THROUGH A DATA STREAMING MECHANISM	
		Retrieve data by subscribing to a stream - applicable only to assets with active streaming jobs THROUGH A DOWNLOADABLE FILE	
		Download assetS as fileS - applicable only to assets that have not been mapped to a data model	

As depicted in Figure 71, he/she needs to select the concepts from the selected data assets that will be extracted and returned as search query results. In the next step, the retrieval configuration needs to be provided by selecting the concepts which the members of an organisation want to extract from each selected dataset. In addition, the concepts that will be used as query parameters to filter the query results need to be defined (with different options becoming available depending on the concept's data type). In case where two or more data assets were selected, the users need to define the join parameters (concepts) from each dataset that will be used to merge the different datasets in the data to be exposed through the SYNERGY APIs. The query results that will be returned can be filtered by different concepts/fields that can be used as query parameters.

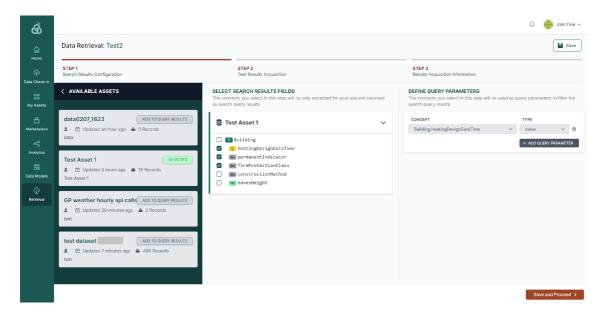


Figure 71: Search Query Definition (Step 1)



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



Following that, as depicted in Figure 72, at the left side of this page, the users shall be able to edit the body with indicative values for the concepts/fields, to be used as the query parameters in order to appropriately filter the dataset, while at the right side of the page, a sample preview of the result is shown after selecting Run Query to acquire the sample of the query results.

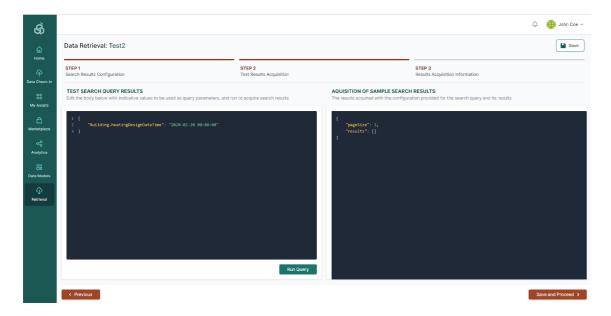


Figure 72: Search Results Configuration (Step 2)

The last step of the retrieval configuration, after selecting Save and Proceed, instructs the users how to acquire the retrieval results through the SYNERGY APIs as depicted in Figure 73. Instructions regarding authentication, API usage, pagination and sorting of results are provided in addition to the endpoints (i.e., for GET and POST methods) including the full API paths containing the BASE URL and the different path segments to be used for the query, and the query body (for the POST method). It needs to be noted that the platform requires the user requesting retrieval from API, and receiving the associated query results, to be authenticated by adding an already generated retrieve access token into an X-API-TOKEN header of the request. In case that there the data asset consumer does not have an already generated retrieve access token, he/she needs to generate one by selecting the "generate a new one" link, and then to follow the same procedure to add it into the request header. Finally, the API query results can be saved by selecting the Save button, while additional options for sorting and paginating the results are available. For sorting the results in a specific order, the data asset consumer needs to provide sorting query parameters (e.g. orderBy that specifies the full path of the field to be ordered by, and orderDirection that specifies the ordering direction - ASC for ascending ordering - and DESC for descending ordering) in the request. For paginating the results, the data asset consumer needs to provide a value for the pageSize parameter which is by default 50.



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



Once the data asset consumer has executed the request, he/she will receive a lastRecordId variable that is a reference to the last retrieved record which needs to be included in the subsequent request as a parameter. The same process needs to be followed for paginating the next results. It needs to be noted that by default, a maximum of 100 results per dataset can be returned if the data asset consumer has selected multiple datasets in the first step of the Query Retrieval Creation, and hence he/she is responsible to adjust the query parameters as mentioned already to get more results.

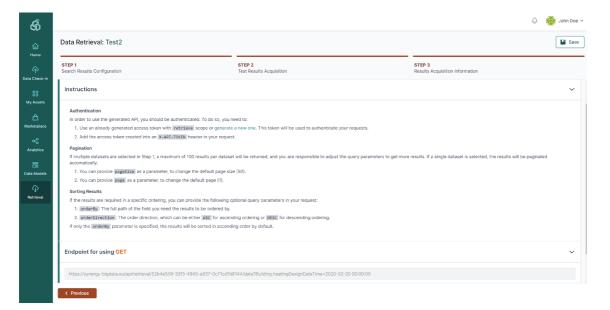


Figure 73: Test Results Acquisition (Step 3)

It needs to be noted that if any of the data assets appearing in a retrieval query has been deprecated or their respective provider(s) have revoked access to them (through changes in the access policies or the involved contract has expired), a notification is displayed to the data asset consumer.





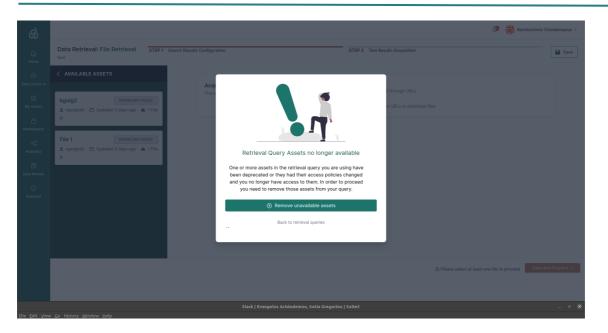


Figure 74: Effect of changes in data asset availability in retrieval queries

4.3.2 View all Saved Queries

A user may view all the queries that have been saved in the SYNERGY Platform, following the creation of a new retrieval query as described in the previous section. The list of saved queries can be accessed through the bookmark button located in the search bar as shown in Figure 75. The Saved Queries list allows the user to rename or delete a saved query by clicking on the options available for each query. Moreover, the user may access the saved query by clicking on its name from the Saved Queries list, and directly the data assets that were added to the query results will be automatically selected, allowing the user to edit the current saved query by adding or removing query results from the saved query.





ගී			@ @
G Home	Data Retrieval Queries		+ Creat
ta Check-In	Retrieval 4 Created by Description 4	🕳 🖨 Last modified on June 23, 2021	Configuration :
My Assets	S My Asset 1	My Asset 2	
larketplace	Retrieval 5 Created by Description 5	Last modified on June 23, 2021	Configuration :
Analytics	No Assets Specified		
ta Models	Retrieval 3 Created by Description 3	Last modified on June 23, 2021	(Asset(s) Available) :
	My Asset 1		
	Retrieval 1 Created by Description 1	🖶 📾 Last modified on June 23, 2021	(Assetts) Unavailable 👔
	S My Asset 2		
	Retreival 2	🗯 Last modified on June 23, 2021	Asset(s) Deprecated :

Figure 75: View Saved Queries

5 Data Analytics User Journey

The SYNERGY Platform provides the Data Analytics Services Bundle, described in D4.1 "SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles - Beta Release", that offers all the functionalities around data processing and analysis to gain valuable insights over data assets that have been checked-in in the SYNERGY Platform as discussed in Section 3, or have been acquired from other electricity data value chain stakeholders (according to the Data Search & Acquisition User Journey described in Section 4). As mentioned in D4.1 "SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles – Beta Release", there are five main components included in the Data Analytics Services Bundle, namely: (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. These components are described in this section divided into four different workflows (i.e. View all Analytics Pipelines, Create an Analytics Pipeline, Configure an Analytics Pipeline, and Visualization Results of an Analytics Pipeline). Similarly, as in the case of Data Check-in workflow, there is a distinct separation between the design of a data analytics pipeline and its execution either in the organisation's Secure Experimentation Playground in the cloud, or on-premise. Although both data asset providers and data asset consumers may proceed to the Data Analytics User Journey, they will be referred to as data asset consumers, for brevity. It needs to be noted that both stakeholders





of the electricity data value chain and the SYNERGY energy apps developers are expected to utilize the data analytics user journey to run analytics on the SYNERGY Platform.

5.1 View all Analytics Pipelines

The data asset consumers may see all the analytics pipelines they have defined as depicted in Figure 76, where the details, the status and the available actions for each analytics pipeline is displayed. A data asset provider may sort the list of analytics pipelines by the title, the date it was created, the date it was executed, the latest execution(s) status (since a pipeline may be executed multiple times according to a schedule) and the data asset provider (within the organisation) who created it. Clear indications appear when the execution workflow definition of the pipeline is still incomplete or not valid, as well as in the cases of pending pipelines that are scheduled to be executed in the future, while successful and failed executions are highlighted with appropriate icons. Finally, a data asset provider may select to configure or remove a pipeline, to view the results of the pipeline in the pipeline results view (as described in Section 4.4), and view the pipeline execution log, from the Actions column.

Ĝ	٥	👂 💓 John Doe 🗸
G) Home	Analytics Workflows	+ Create
Data Check-In	Configuration :	
CC My Assets	nt. Visualisation 1	
Marketplace	 Solar Power Genaration Forecasting Created by John Doe Last modified on June 28, 2021 View all execution logs > Test 	
8	at Visualisation 1	
Data Models		
	all Visualisation 1	
	Configuration : Configuration : A Created by John Do Last modified on June 28, 2021 An analytics workflow to predict the power generation from solar panels.	
	at Visualisation 1	
	Contiguration] .

Figure 76: View all Analytics Pipelines

5.2 Create an Analytics Pipeline

A data asset provider may create a new analytics pipeline through the Data Analytics view by selecting the Create button, as shown in Figure 76. Initially, the view for creating a new Workflow





appears as depicted in Figure 77, where the data asset provider needs to insert some basic details such as a title, and a short description that essentially provides an overview of this data analytics pipeline job. Then, the data asset provider needs to select the execution framework of their analytics pipeline, that is the framework (i.e., Apache Spark engine, or Python environment) that will be used for the data processing tasks included in the pipeline. In addition, the location where the analytics pipeline job will be executed (i.e., Cloud Execution, or On-Premise Execution), needs to be defined as well. It needs to be noted that, running the analytics pipeline on the Cloud Execution is the only available option if the data asset consumer selects the Apache Spark framework for running the analytics pipeline. In contrast, the data asset consumer is able to run the analytics pipeline on all the available execution locations (i.e., Cloud Execution, On-Premise Execution) using a Python environment. It needs to be noted that local execution can run analytics only over data that are stored locally at the moment (since no transfer of data from the cloud to the On-Premise Environment is allowed in this release). By clicking on the Save button, the Analytics Workbench page loads accordingly as described in the subsequent section.

ගී			🦉 🐞 John Doe 🗸
لک Home	Create new Workflow		X Cancel V Save
Control Check-In Big My Assets Marketplace	Workflow Details Basic information about the workflow	NAME Solar Power Generation Forecast Edecembrican Edecembrican An analytics workflow to predict the power generation from solar panels.	
Analytics	Execution Framework The execution framework that will be used for the data processing tasks in the workflow	Spark Execution in the Apache Spark Engine (v3.0) Python Execution in a python environment	
	Execution Location You can choose whether you want the workflow to be run on the cloud or on-premise through a registered runner	CLOUDEXECUTION The data-checkin job will run on the cloud. ON-PREMISE EXECUTION The data-checkin job will run on-premise through a registered runner.	

Figure 77: Create a new Analytics Pipeline

5.3 Configure an Analytics Pipeline

Through the Analytics Workbench, data asset providers are able to appropriately configure a complete analytics pipeline, by defining the necessary input data (input blocks), the data manipulation functions (data preparation), the machine learning (ML) algorithms that are to be applied on the data, and the output data (output block) for storing the results of an analytics pipeline.





This configuration is done by connecting these different pipes/function blocks such that the whole analytics pipeline is designed in a visual manner, in the form of easily configurable blocks that are appropriately connected to form the final data analysis pipeline. Towards the design of a complete data analytics pipeline, three main views are available to support visually the configuration, the validation of changes on the data sample after each function block, and the configuration for visualising the results.

5.3.1 Graph View

The Graph view of the Analytics Workbench allows users to design (in a visual way) an analytics pipeline as shown in Figure 78.

Ŝ			🧶 🔅 John Doe 🗸
6 Home	Workflow Design for: Solar Powe	r Generation Forecast	< Back
Home A	Graph View Table View	Results View	👯 Configure 🗸 💾 Schedule 🗸 🛕 5 Errors 🗸 🕝 Finalise
Data Check-In	LIBRARY	< GRAPH VIEW	14 Z
CC My Assets	- P 🗈 🔅 🖸		
	Search	Q ⊕ ReadCSV ▲	
Marketplace	ExportCSV Store dataframe in a csv file	+ Append A	
Analytics	ExportJSON Store dataframe in a json file	÷	
Data Models	ExportParquet Store dataframe in a parquet file	Apply Trained Keras Model	
	ExportToMongo Write dataset to MongoDB.	+ ExportCSV ▲	
	StoreCloudResult Store a dataframe in the cloud and create a Result	*	
	StoreLocalResult Store a dataframe locally (in Parquet format and create a Result		

Figure 78: Configure an Analytics Pipeline – Graph View

By searching, browsing and filtering in the Library located at the left side of the view, the data asset providers may find the relevant blocks for their pipeline, i.e. input blocks, data preparation blocks, machine learning (ML) blocks, and output blocks. Upon locating a relevant block needed for the analytics pipeline in the Library, the data asset provider can add them in the main Graph View by selecting the add button (cross icon). Adding a function results in a block that appears in the main Graph View without any connections (directed arrows between blocks), since the connection between the added blocks, is added when the data asset provider configures the block connections in the right slideover. Indicatively the data asset provider should define the method for reading the input data, by adding a particular input block (e.g., Read Dataset which reads a stored data asset, and Read Result which reads an existing result from another data analysis pipeline that has been already





executed), and inserting its associated parameters accordingly in the right slideover revealed after the block selection, as shown in Figure 79.

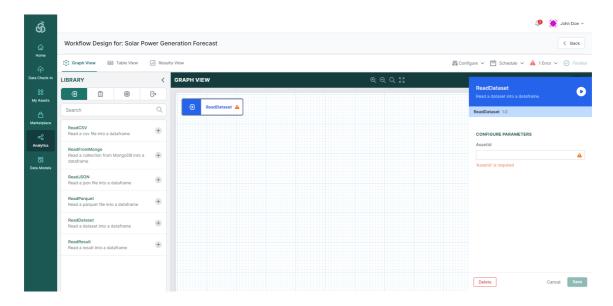


Figure 79: Analytics Pipeline - Input Blocks

Following, the data asset provider may define data preparation rules by adding different processing functions (e.g., sort, drop null values, etc.), that are to be executed before the actual analytics functions, as shown in Figure 80.

Ŝ				I (19) (10) (10) (10) (10) (10) (10) (10) (10	John Doe 🗸
ம் Home	Workflow Design for: Solar Power G		< Back		
	🛟 Graph View 🖃 Table View 💷 Results View			👯 Configure 🗸 💾 Schedule 🗸 🔺 3 Errors 🗸	Finalise
Data Check-In	LIBRARY	GRAPH VIEW	ଭ୍ର୍ଠ୍ 👯		
88 My Assets	ج (1) (1) (1) (1) (1) (1) (1) (1) (1) (1)	ReadDataset A		Compute Aggregations Compute aggregations of specified types given columns. The result is a new datase with the computed values grouped by the defined columns.	t 🕑
Marketplace	Aggregations				
Analytics	Compute Aggregations Compute aggregations of specified types on given columns. The result is a new dataset with the computed values grouped by the defined columns.	Compute Aggregations		aggregations ComputeAggregations CONFIGURE TASK INPUTS	
Data Models	ComputeTransformAggregations Compute aggregated values of given columns and store the results in new columns in the initial dataframe, according to the group by columns				
	Math				
	Aggregations over Lists Compute selected aggregations on the lists + in the given column(s)				
				Delete Canc	el Save

Figure 80: Analytics Pipeline – Data Preparation Blocks

Once the data asset provider has added the input and data preparation blocks, the appropriate machine learning block should be selected depending on the actual analysis that is intended to be done, as shown in Figure 81. The various machine learning blocks are classified into three categories



SYNERGY

based on their use (e.g. evaluate, train, or apply), while the available machine learning algorithms (e.g., Binary Classification, Regression, Clustering, etc.) are allocated accordingly to these categories.

ගී			🧔 🐞 John Doe 🗸
۵	Workflow Design for: Solar Power	eneration Forecast	< Back
Home	🛟 Graph View 📰 Table View 🖃	suits View	👯 Configure 🗸 💾 Schedule 🗸 🔺 4 Errors 🗸 📀 Finalise
Data Check-In	LIBRARY	GRAPH VIEW ଭ୍ର୍ଠ୍ତଃ	
BB My Assets	-7 🕯 😐 🕞	ReadDataset	Clustering (SKlearn - labeled data) Apply sklearn model for clustering on labeled data.
≙	Search	ReadDataset	apply Clustering (SKlearn - labeled data) 1.0
Marketplace	Apply		CONFIGURE TASK INPUTS
Analytics	Apply Trained Keras Model Apply Keras model on labeled or unlabeled data	ComputeAggregations 🔺	df 🔺 🏠
Data Models	Classification (SKlearn - labeled data) Apply sklearn model for classification on labeled data	😥 Clustering (SKleam - labeled data) 🔺	
	Classification (SKlearn - unlabeled data) Apply sklearn model for classification on unlabeled data		
	Clustering (SKlearn - labeled data) Apply sklearn model for clustering on labeled data		
	Clustering (unlabeled data) Apply sklearn model for clustering on unlabeled data		
	Regression (SKlearn - labeled data) Appty sklearn model for regression on labeled data		Delete Cancel Save

Figure 81: Analytics Pipeline - Machine Learning Blocks

Finally, the data asset provider should define the output block that denotes the way the data will be stored, or obtained from other mechanisms to create a result (e.g., through visualization, through the SYNERGY Platform's APIs, etc.), as shown in Figure 82.

Ĝ			🧔 🔅 John Doe 🗸
۵	Workflow Design for: Solar Power G	eneration Forecast	< Back
Home	Graph View 🗊 Table View 💷 Res	ults View	👯 Configure 🗸 📋 Schedule 🗸 🔺 5 Errors 🗸 🕑 Finalise
Data Check-in	LIBRARY	GRAPH VIEW	
CC My Assets	ə 🗈 🕸 🕞		Store CloudResult Store a dataframe in the cloud and create a
£	Search Q	- 관 ReadDataset 🔺	Result.
Marketplace	ExportCSV		StoreCloudResult 1.0
~~	ExportCSV Store dataframe in a csv file	ComputeAggregations	CONFIGURE TASK INPUTS
Analytics	ExportJSON Store dataframe in a json file		df 🔺 🔺
Data Models	ExportParquet Store dataframe in a parquet file	Clustering (SKlearn - labeled data) A	
	ExportToMongo Write dataset to MongoDB.	C StoreCloudResult	
	StoreCloudResult Store a dataframe in the cloud and create a Result		
	StoreLocalResult Store a dataframe locally (in Parquet format) + and create a Result		
			Delete Cancel Save

Figure 82: Analytics Pipeline - Output Blocks

Practically, in the Library View, by selecting a block, the data asset provider may view and change the description of that block, see the type and version of the block, define the connections of the block (i.e., upstream tasks, downstream tasks), and define the parameters depending on the block type. By





selecting the Save button, a quick run over the sample will be executed for the configured block in order to propagate the data structure to next block and a relevant status icon will appear in the right side of the block in the Graph View. Once the blocks are configured correctly, a Valid icon will appear at the top bar of the Analytics Pipeline View. Additionally, the data asset provider may change the name or the description of the Analytics Pipeline by selecting the Configure button.

The available actions in the blocks include: (a) test run in order to execute the pipeline up to the specific block for a small data sample, (b) settings in order to edit or delete a block from the pipeline. In case there are any validation errors in the pipeline, they will be displayed in the appropriate blocks, and a summary of the validation results will be displayed in the Validation Results as shown in Figure 83. The data asset provider may proceed to view the results of the test run by selecting the Table View (Section 5.3.2) from the Workflow Designer top bar, and follow the workflow as described in the subsequent section.

.TS n on the tasks and workflow we have found the follo	lowing issues		
et' is missing required parameters ggregations' is missing required parameters			
ExportCSV Store dataframe in a csv file	•		CONFIGURE TASK INPUTS
ExportJSON Store dataframe in a json file	•		df
ExportParquet Store dataframe in a parquet file	+	Clustering (SKlearn - labeled data) 🛕	
ExportToMongo Write dataset to MongoDB.	•	C+ StoreCloudResult	
StoreCloudResult Store a dataframe in the cloud and create a Result	•		
StoreLocalResult Store a dataframe locally (in Parquet format) and create a Result	÷		
	s 4 different chains. You can have a single et is missing required parameters groupstores is missing required parameters (Stakam - Lubeled data)'s missing required fifesult's missing required parameters scortCSV scortCSV Store dataframe in a joon file scortCSV Store dataframe in a parquet file cates dataframe in a parquet file score dataframe in the cloud and create a tesuit Store a dataframe in the cloud and create a tesuit Store a dataframe in the cloud and create a tesuit Store a dataframe (bally (in Parquet format	s 4 different chains. You can have a single chain only et's in missing required parameters geogradions' is missing required parameters (Stekam - Labeled data) is missing required parameters Result is missing required parameters SteportCSV Store dataTame in a csv file ExportCSV Store dataTame in a pron file ExportToMongo Write dataset to MongoOB. Store CloudResult Store dataTame in the cloud and create a the Tesult Store dataTame localy (in Parquet format) the	s 4 different chains. You can have a single chain only et a missing required parameters geographorse in missing required parameters Stream - tabeled data) is missing required parameters

Figure 83: Analytics Pipeline - Validation Results

5.3.2 Table View

In the Table view, depicted in Figure 84, data asset providers obtain an overview of a data sample, and how data change each time they are processed according to the function blocks defined in the previous step. In particular, the data asset provider may view the blocks that constitute the pipeline in the left slideover, while by selecting a particular block the corresponding workflow is highlighted, and the results of the test run appear in the main Table View. Additionally, by selecting a particular block, the data asset provider may change the settings that were defined previously in the Graph





View. In the main Table View, the data asset consumer may view the different columns of the outcomes and may sort them, search for a specific value or resize them.

ගී					🧔 🔅 John Doe 🗸
<u>ن</u>	Workflow Design for: Solar Power Genaration Forecasting		< Back		
Home	(2) Gra	ph View 📰 Table View	al Results	/iew	👯 Configure 🗸 📛 Schedule 🗸 🖒 Valid 🧭 Finalise
Data Check-In	OVERV	IEW	<	TABLE VIEW	
88 My Assets	linput	Data Preparation 🔶 Machine Lea	rming • OutputCk		Last updated a few seconds ago 😋
<u>A</u>	Ð	ReadDataset 1 column	۲	WeatherMeasurement (string) \diamond	
Marketplace			0	("id": "4135001.0", "temperature": 25.18431613)	
~%		ExportJSON		("id": "4135001.0", "temperature": 25.08458867)	
Analytics		Exportation		("id": "4135001.0", "temperature": 24.9357526)	
Data Models				("id": "4135001.0", "temperature": 24.8461304)	
Data Models				("id": "4135001.0", "temperature": 24.62152536)	
				("id": null, "temperature": null }	
				("id": null, "temperature": null)	
				("id": null, "temperature": null)	
				("id": null, "temperature": null)	
				("id": null, "temperature": null)	
				No tasks currently running	

Figure 84: Configure an Analytics Pipeline – Table View

5.3.3 Results View

Once the data asset provider is satisfied with the defined and configured data analytics pipeline, he/she may configure the desired output / visualization of the results from the Results View as shown in Figure 85. In particular, the data asset provider may select the output block that he/she is interested to visualize, and then the type of diagram that needs to be visualized from the left slideover menu that includes all the available visualization types. This menu provides several types of diagrams for visualization such as line graph, scatter plot, and bar chart. Following that, the data asset provider needs to select the appropriate output blocks in order configure the parameters that are to be visualized, from the Configuration right slideover menu. The parameter selection needs to be done for each axis according to the type of the diagram, and the needs of the data asset provider with regards to what is to be visualized. In addition, the data asset provider may add a title for the diagram, legends for the data, axes labels for the diagram, as well as to select among various other options such as to add grids, and tooltip to the diagram.





ີ້ Workflow Design for: Solar P	ower Generation Forecast	Ø so John Doe ~
arme	an Rosults View	tt Configure ∨ 💾 Schedule ∨ 🛕 2 Errors ∨ ⊘ Finalise
Check-in OPTIONS	< RESULTS VIEW	ଷ୍ର୍ଦ୍ 🕹 CONFIGURATION
Assets	how more	CENERAL Y AXIS X AXIS
	LINE Line graphs can be used to show how data change over time	Main Title
VISUALIZATION TYPE	5	LINE
Nytics Search	Q	Desription Line graphs can be used to show how data char
Models Line Line graphs can be used to s how data change over time	now	Legend C Legend C Labels C Grid C
Area Area graphs can be used to how data change over time		
Bar Bar graph	3	
Radar radar	No data to display	Delete Cancel Save

Figure 85: Configure an Analytics Pipeline – Results View

5.3.4 Schedule Execution

Apart from the configuration of the different input, processing, analytics and output blocks that need to be added in a complete analytics pipeline, the data asset provider may also configure the scheduling of the pipeline execution, by selecting the Schedule option from the Analytics Workbench bar, as shown in Figure 86. A top slideover to define a schedule (or multiple schedules) is displayed allowing the data asset provider to select the execution period (i.e., start date, end date), the frequency of the execution (i.e., hourly, daily, weekly, monthly), and the exact time of execution depending on the frequency.

ஞ்					ø	🂓 John Doe 🗸
ര	Workflow Design for: Solar Power Generatio	on Forecast				< Back
SCHEDULES			SCHEDULE EDITOR			
Runs at 05:*	(UTC) every day, from 29 June 2021 to 2 July 2021	~ °	Start Date 15 July 2021	End Date 31 July 2021	Frequency Hourly Daily We	ekly Monthly
	۲		Schedule Details	31 July 2021	Houry Daily W	montany
						Cancel Save
8	Search				Line graphs can be used to sh	ow how data char
Data Models	Line				Legend Labels	
	Line graphs can be used to show how data change over time				Grid	
	Area Area graphs can be used to show how data change over time	4				
	Bar Bar graph	3				
	Radar radar	Axis title	No data to display		Delete	Cancel Save

Figure 86: Configure an Analytics Pipeline – Schedule Execution





5.3.5 Execution History

By selecting the Execution History option for a particular Analytics Pipeline, from the Analytics Workbench bar, a new screen including the execution logs will appear as shown in Figure 87. Within this view, the data asset provider can see the execution summary information such as the number of total executions, the number of successful executions, the number of failed executions, and the average execution time. Below these, the data asset provider may view more details regarding each execution including its timestamp, status, and available actions. When an execution is queued, there is only one available option that is to cancel the next execution as the schedule shall continue (skipping the next execution only). When an execution has successfully been completed or failed, the available options are: a) to view the details which opens a table that includes the task details, b) to delete the execution including the accompanied results and log files, and c) to visualize the results in the Results View. It needs to be noted that the execution logs are organized in a listed view that can be sorted by the date of execution or status accordingly.

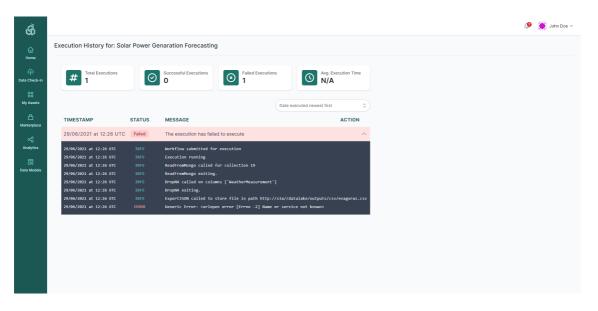


Figure 87: Analytics Pipeline - Execution History

5.4 Visualization Results of an Analytics Pipeline

Once an analytics pipeline has been finalised and has been successfully executed, the data asset consumer may view the results as shown in Figure 88. The results are adapted based on the configuration of the visualization, as defined in the Results View page described in Section 4.3.3. Moreover, the user is able to download the results locally, or to send the results to an API.





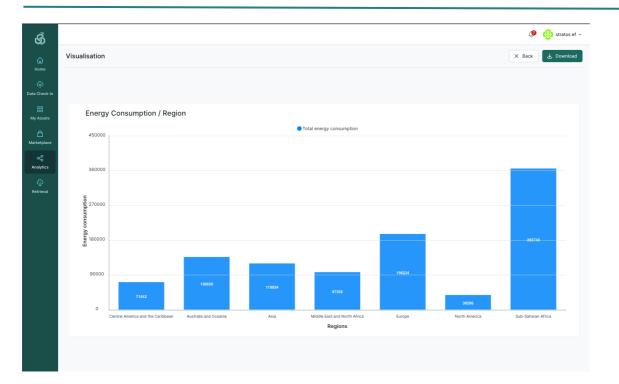


Figure 88: Analytics Pipeline – View Results of an Analytics Pipeline





6 Additional Platform Functionalities

In addition to the functionalities that were described in the former sections, the SYNERGY Platform provides some additional functionalities to different types of users (i.e., platform's administrators, data asset consumers, data asset providers, and others) enabling platform support functions. Such support functionalities include: (a) Common Information Model Lifecycle Management for the CIM administrators, (b) Edit Organization Profile, (c) Edit User Profile, and (d) Wallet Management.

6.1 Common Information Model (CIM) Manager

The Platform's administrators are allowed to create, edit, update and deprecate data models representing data that are exchanged between the stakeholders of the energy data value chain, in accordance with the CIM lifecycle management approach that was defined in the SYNERGY Deliverable D3.1. This functionality is offered by the CIM Manager component, and can be accessed through the Models view.

Within the CIM Manager, depicted in Figure 89, the CIM manager is able to view: (a) the list of concepts that are included in the data model at the left side of the page, (b) the fields/concepts that are included (or linked) in the selected concept at the middle of the page, and (c) more details about the selected field such as its description, its mapping to standards, and its related terms, at the right side of the page.

ගී			
	Electricity Data Model For SYNERGY-	Early 1.0.0 - stable	< Back
	CONCEPTS Add <	CONCEPT EDITOR	CONCEPT DETAILS
Pata Check-in	Q Search	ACLine	TITLE
	Active Proposed Deprecated All	DESCRIPTION A wire or combination of wires, with consistent electrical characteristics building a single electrical system.	ACLine DESCRIPTION
My Assets	ACLine Address	RELATED TERMS ac line power line Category:Network	A wire or combination of wires, with consistent electrical characteristics building a single electrical system.
	Aggregator AggregatorPortfolio	Fields Add new field 🖓 Active 🗸	STANDARDS MAPPING Name Standard & Version Type
Analytics	AirConditioningSystem	baseVoltage Double RELATED TERMS voltage level base voltage	ACLIneSegment IEC CIM 16 Object @ + Add a standards mapping RELATED TERMS
Data Models	AutomatedOperationProfile BalancingResponsibleParty	id String RELATED TERMS identifier	ac line x power line x Category:Network x + Add a related term
	Battery BatteryControlAction Boiler	impedance Double RELATED TERMS length impedance	
	BoilerControlAction Building	length Double RELATED TERMS line length AC line length	
	BuildingSpace BuildingStorey	name String RELATED TERMS identification name	
	BuildingZone		Deprecate

Figure 89: Common Information Model (CIM) Manager – Concepts/Fields View





In this page the CIM manager may add, edit, or delete existing concepts or fields according to its needs. In order to maintain the CIM up-to-date according to the latest needs of the data asset providers in the SYNERGY Platform, the CIM manager is able to edit the concepts and fields of the selected data model, which generates a new major or minor version of the CIM (depending on the evolution rules that are triggered under the hood) and makes the new data model and its associated concepts/fields available in the Mapping configuration stage. It is worth mentioning that any proposed concepts by the data asset providers are also visible in the CIM Manager, allowing the CIM manager to take appropriate actions.

In particular, the CIM manager may add a concept to the model by selecting the Add button in the Concepts section to enable the right Concept Creator section where he/she needs to insert the title, the description, the mapping to a particular standard (if any), as well as to add related terms for the concept. Additionally, the CIM manager can select to import an existing concept from another data model by selecting the Import from another Data Model option (e.g. from a deprecated CIM version) as shown in Figure 90. Finally, the concept is created by selecting the Create button which enables the Concept Editor section in order for the CIM manager to provide the fields to be included in the concept.

ගී			- ● <	\supset
۵	Electricity Data Model For SYNERGY-	Early 1.0.0 - stable		< Back
Home	CONCEPTS Add <	CONCEPT EDITOR CONCEPT CREATOR		
🔶 Data Check-in	Q Search	ACLine TITLE*		
88	Active Proposed Deprecated All	DESCRIPTION A wire or combination of wires, with consistent electrical		
My Assets	ACLine	characteristics building a single electrical system. DESCRIPTION* RELATED TERMS ac line power line Category:Network Enter a description for the concept		
۵	Address	RELATED TERMS actine power line Categoly steework		
Marketplace	Aggregator	Fields Add new field V Active Add new field		
Analytics	AggregatorPortfolio	object	~	
8	AirConditioningSystem	baseVoltage Double RELATED TERMS Voltage level base voltage New Concept		
Data Models	AirConditioningSystemControlAction	Import from another Data Model		
	AutomatedOperationProfile	id String Standards Mapping		
	BalancingResponsibleParty	RELATED TERMS Identifier + Add a standards mapping RELATED TERMS		
	BatteryControlAction	impedance Double + Add a related term		
	Boiler	RELATED TERMS length impedance		
	BoilerControlAction	length Double		
	Building	RELATED TERMS line length AC line length		
	BuildingSpace	name String		
	BuildingStorey	RELATED TERMS identification name		
	BuildingZone		Create Concept	

Figure 90: Common Information Model (CIM) Manager – Concept Creator

Once the concept is created, the CIM manager can add the fields or sub-concepts, by selecting the "Add new Fields" option in the Concept Editor section, which will enable the Field Creator section at the right side of the page as Figure 91 depicts. In this section, the CIM manager can provide the title and description for the field, select its data type (e.g. string, double, integer, object, etc), add a





standard mapping (referring to the fields of certain standards), add related terms, and add metadata to the field. Appropriate metadata can be added to the field depending on the data type and including information regarding different aspects of the data that could be potentially uploaded to the particular field. Indicative metadata options are: (a) Allow multiple occurrences of this field; (b) Order of multiple occurrences is important; (c) Allow indexing of this field; (d) Denoting if the data can be customized (for object data type); (e) Denoting the temporal coverage of the data (for date, time, and datetime data type); (f) Measurement Type and Unit (for double and integer data types).

Finally, the CIM manager can create the field by selecting the Create Field button at the bottom of the page.

ώ	Electricity Data Model For SYNERGY-	arly 1.0.0 - stable	
Home	CONCEPTS Add <	CONCEPT EDITOR	FIELD CREATOR
۲۷ Data Check-in	Q Search	Address	TITLE*
	Active Proposed Deprecated All	DESCRIPTION The location at which a particular stakeholder	Enter field title
80 My Assets		or person may be found or reached.	DESCRIPTION*
	ACLine	RELATED TERMS location Category:General	Enter a description for the field
A Marketplace	Address		
	Aggregator	Fields Add new field Add new field	DATA TYPE*
Analytics	AggregatorPortfolio		Select data type
	AirConditioningSystem	addressLine String	STANDARDS MAPPING
Data Models	AirConditioningSystemControlAction	RELATED TERMS Full Address	+ Add a standards mapping
Data Models	AutomatedOperationProfile	apartmentID String	RELATED TERMS
	BalancingResponsibleParty	RoomID ApartmentID	+ Add a related term
	Battery	RELATED TERMS ApartmentNumber	
	BatteryControlAction	buildingName String	
	Boiler	RELATED TERMS BuildingIdentifier	
	BoilerControlAction		
	Building	buildingNumber String House Number Building No	
	BuildingSpace	RELATED TERMS House No	
	BuildingStorey	1000110	
	BuildingZone	city String	Create Field

Figure 91: Common Information Model (CIM) Manager – Field Creator

6.2 Edit Organisation Profile

The organisation's manager (who is also the organisation's legal representative, eligible to sign any data asset contracts in the SYNERGY Platform) may edit the organisation's profile, by navigating to the Edit Organisation Profile page as shown in Figure 92. The organisation's manager may edit the organisation's business name, by which it is typically identified. In addition, the manager can add a cover photo, and a description that provides a brief overview regarding the organisation's activities. The organisation type that distinguishes the organisation within the electricity data value chain, needs to be selected as well (even though this functionality may be deactivated upon the initial





organization registration to avoid potential misuse). The organisation's manager may add, edit, or remove a department (including its name, address, city, and country) to the Departments List as well.

Ŝ				Ϋ́,
6 Home	Edit Organization Details 🥥			K Back Save
Data Check-in	🖆 Overview	General Information Basic information about the	LEGAL NAME	BUSINESS NAME
	Kinders	organization	S5 Devs	S5 Devs
My Assets			ORGANIZATION TYPE	WEBSITE www.example.com
Aarketplace			DESCRIPTION	
æ			Testing organisation!	
Analytics				
Data Models				
¢		Location Information Basic information about the location of	ADDRESS	POSTAL CODE
Retrieval		the organization	Somewhere 124	1234
			COUNTRY	СІТҮ
			🖾 Cyprus	Limassol ~

Figure 92: Edit Organisation Profile - Overview

In addition, users can be added in the Users List by defining the department they work, their role in the organisation, and their status (i.e., Active, Invited, Blocked), while a user can be activated (if Blocked) or deactivated (if Invited or Active) by the manager by selecting the corresponding buttons, as depicted in Figure 93.





Ŝ				¢	
G Home	Edit Organization Details	5 📀		<	Back 🗸 Save
ፍ	Cverview	NAME	EMAIL	STATUS	ACTION
Check-in	Members			Active	Deactivate
Assets				Active	Deactivate
۵				Active	Deactivate
etplace				Active	Deactivate
alytics				Active	Deactivate
Models				Active	
Ð				Active	Deactivate
trieval				Active	Deactivate
				Active	Deactivate

Figure 93: Edit Organisation Profile - Users List

6.3 Edit User Profile

Users may edit their individual user profiles, by navigating to the Personal Profile page that is loaded by selecting the name of the user at the top right side of the SYNERGY Platform. As depicted in Figure 94, the user may update basic information such as the first and last name, as well as to change their current password to a new one, by filling the corresponding text boxes.

ගී						🥔 🐞 John Doe 🗸
ය Home	Edit User Details Not verified					K Back V Save
Data Check-in	USER PROFILE		ACCESS TOKENS		ON-PREMISE RUNNER	
BB Children BB My Assets	General Information Basic information about the user	FIRST NAME		IST NAME		
A Marketplace		John		Doe		
می Analytics			•			
Data Models						
	Change password	Enter current password	Enter new password	d Repeat new password		
				Change Password		

Figure 94: Edit User Profile – Overview





6.3.1 Generate Access Tokens

By selecting the Access Tokens tab, the users are able to view their tokens (for retrieving or uploading data to the SYNERGY Platform through the SYNERGY APIs). By selecting the Generate new token, the user is asked to provide a name for the token that is to be generated and select the scope of the intended use that is to retrieve data from the platform's API, or to upload data to the platform's API, or both. Finally, the user is able to delete generated personal access tokens by selecting the Delete button located at the right side of each access token in the Access Tokens tab, as depicted in Figure 95.

ගී				19 🌘	John Doe 🗸
ம் Home	Edit User Details Not verified			K Back	✓ Save
Data Check-In	USER PROFILE	ACCESS TOKENS	ON-PRE	MISE RUNNER	
CC My Assets	Personal Access Token Generate a personal access token for quick access.	NAME demoAccessToken			
Arketplace		SCDES Critice			
Data Models		Cancel Gene	rate token		

Figure 95: Edit User Profile - Access Tokens – Generate Token

6.3.2 Register an On-Premise Execution Environment

The user may download the On-Premise Execution Environment tab as shown in Figure 96. In particular, the user needs to select the operating system of his/her physical machine in order to start downloading the appropriate version of the On-Premise Environment for Windows, MacOS, Linux.





ගී						ø 🔅	John Do
G Home	Edit User Details Not verified					< Back	🗸 Sav
م a Check-in	USER PROFILE	ACCESS TOKENS			ON-PREMISE RUNNER		
88 y Assets	On-Premise Runner Download on-premise numer.	Windows	MacOS	Linux			
oetplace alytics	Registered Runners Register a runner to run DCJ and analyric workflows on-premise.			R	egister new runner		
⊡0 Models							

Figure 96: Edit User Profile - On-Premise Execution Environment

However, a prerequisite step to execute the On-Premise Execution Environment is to register it by selecting the "Register a new runner" button which then enables a text box asking to provide a name for the Execution Environment. As soon as the name is inserted correctly, the corresponding URL and code for the registered On-Premise Execution Environment will appear allowing the execution of the On-Premise Execution Environment, as shown in Figure 97. The generated URL and code need to be inserted during the installation phase on the On-Premise Execution Environment on the user's physical machine.

ගී					🦉 🐞 John Doe
۵	Edit User Details Not verified				K Back Save
Home Data Check-In	USER PROFILE	ACCESS TOKENS		ON-PREMISE RUNNER	
OD My Assets	On-Premise Runner Download on-premise runner.		é	\$	
A Marketplace		Windows	MacOS	Linux	
≪ Analytics	Registered Runners Register a runner to run DCJ and analyric workflows on-premise.	Make sure to copy the token now and register the runner u	sing the url below, before the token e	xpires. You won't be able to see it again!	
Data Models		https://synergy-bigdata.eu ex2.public.ey.JpZCI8ijg3NjC0YTc5LTdjZDEtNDU02S1hYWQyL J2WSicmd5U.WjpZ2Rind6EuZXUILCJ1c2VjSW0(OUslimInd Q6NDg6NTUu0TQ4WU9U-bgrm0103J8/pwkd8wfx0TeMp	CI6IjIwMjEtMDYtMjhUMTQ6Mzg6NTU	JuOTQ4WilsImV4cCl6ljlwMjEtMDYtMjhUMT	

Figure 97: Edit User Profile - On-Premise Execution Environment - Register a New Runner





6.4 Wallet Manager

An organization's manager may view the contracts of his/her organization by navigating to the Contracts view of the Marketplace, as shown in Figure 98. In the case that the organization does not have a linked wallet, the manager may import an existing wallet, or create a new one as described in the subsequent sections (Section 6.4.1 and Section 6.4.2, respectively). It needs to be noted that, if data asset providers have already initiated requests for data assets (that do not belong to their organization) but the organization manager has not set up a wallet yet, they will not be able to proceed in buying a data asset, until a wallet is generated in the Contracts tab.

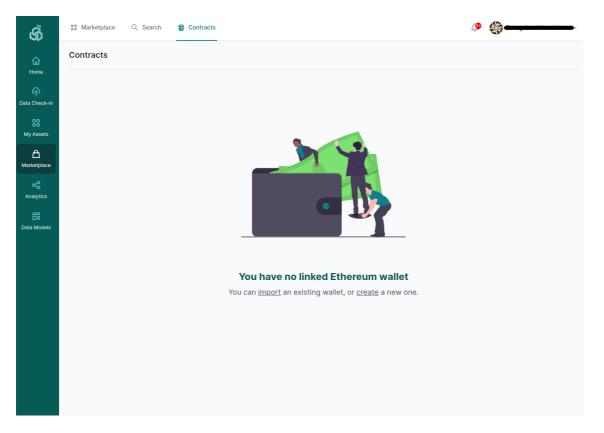


Figure 98: Contracts List

6.4.1 Import an Existing Wallet

The organization's manager may import an existing wallet by entering the private key and the wallet password, as depicted in Figure 99.





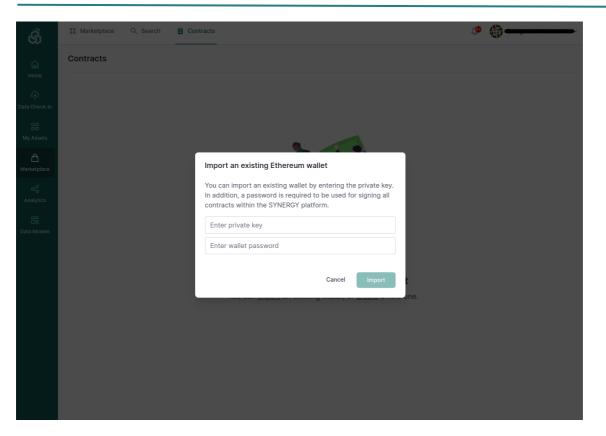


Figure 99: Import an Existing Wallet

6.4.2 Create a New Wallet

An organization that does not have a wallet can create one by selecting the import link in the Contracts view of the Marketplace, as shown in Figure 98. A wallet password needs to be entered accordingly to encrypt the organization's wallet, as Figure 100 depicts. It needs to be noted that this password will be used for signing all contracts within the SYNERGY Platform. By selecting the Create button, the platform starts generating and encrypting the wallet as Figure 101 shows.





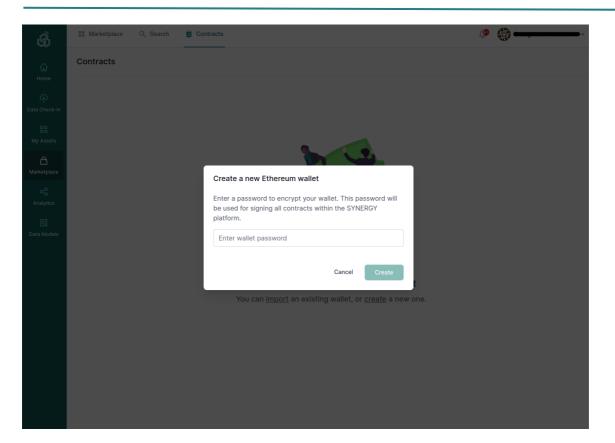


Figure 100: Create a New Wallet

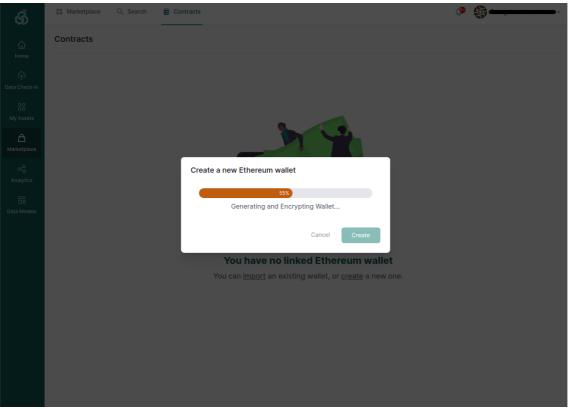


Figure 101: Create a New Wallet - Generating and Encrypting Wallet





7 Platform Integration and Support Activities

This section focuses on the integration approach followed by the technical partners to deliver the beta release of the SYNERGY Integrated Platform, as well as on the support channels that are in place to acquaint the overall SYNERGY consortium with the platform functionalities.

7.1 Integration Approach

As described in Section 2, the SYNERGY Platform brings together an extensive number of components, services and technologies that have been integrated at the Cloud deployment and the Server On-Premise Environments in accordance with the SYNERGY architecture (described in the SYNERGY Deliverable D2.6) and the integration plan (defined in the SYNERGY Deliverable D3.3). In order to proactively meet the software delivery challenges that a complex project as SYNERGY brings, an agile integration approach based on DevOps and GitOps has been put into place:

- I. **Release design and planning** that bring together the involved development teams, the WP3 and WP4 leaders, and the Technical Coordinator to discuss the priorities, the identification of dependencies and the main milestones per development cycle.
- II. Collaborative software development and testing with each development cycle (for new features) typically lasting 3 weeks and providing a new minor release of the SYNERGY Platform. It involves the involved development teams in: (a) coding, (b) running unit tests, (c) committing code to the respective component repo, (d) reviewing the code changes and providing comments to the pull requests, (e) merging the new features and fixes included in the code, (f) running appropriate integration tests.
- III. Deployment in staging environment that ensures that any minor release of the SYNERGY Platform is tested by the involved development teams before it is shipped to production. Depending on its criticality, any error or bug noticed may be prioritised to be immediately fixed or put in the backlog for the next minor release in (II).
- IV. Deployment in production environment that occurs per major release (i.e. on M18 for the beta release, on M24 for Release 1.00 and on M36 for Release 2.00) and on selected minor releases (for bug fixes or for prioritised features that need to be shipped earlier than the next major release). Experimentation on the production environment is available to all SYNERGY partners.





In this context, the integration activities of the SYNERGY Platform rely on a set of state-of-the art techniques and open-source tools to ensure collaborative and continuous planning, development and deployment cycles for each release as depicted in the following table.

Integration-related Activity	Tool
Source code versioning and issue tracking	Github
Automated build and testing	Github Actions
Error Tracking	Sentry
Deployment	Docker, Kubernetes
GitOps	Flux

Taking into consideration the current status of the beta release of the SYNERGY Platform and the components interrelations, the integration plan that shall be followed for M24 is depicted in the following figures.

		Beta Release		Release 1.00	Release 2.00 신다 (A 本)
		Q2 - 2021	Q3 - 2021	Q4 - 2021	(Q1-Q4) 2022
	Data Handling Manager	UI Development, Integration	Stress Testing	DCJ templates (Cloning)	Improvements, Bug fixing
Data	Matching Prediction Engine	UI Development, Integration	Stress Testing	Additional Algorithms Training	Improvements, Bug fixing
Collection Services	Data Ingestion Service	Integration, End-to-end Testing	Stress Testing	Improved execution handling	Improvements, Bug fixing
Bundle	Mapping & Transformation Service	Integration, End-to-end Testing 🔍	Stress Testing	Improved execution handling	Improvements, Bug fixing
	Cleaning Service	Integration, End-to-end Testing	Stress Testing	Improved execution handling	Improvements, Bug fixing
Data Security	Anonymisation Service	Integration, End-to-end Testing	Stress Testing	Improved execution handling	Improvements, Bug fixing
Services Bundle	Encryption Engine	Secure key exchange, Integration	Performance Testing	Impr. exec. handling, Key revocation	Investigat. of Searchable Encryption
	Access Policy Engine	UI Development, Integration	Stress Testing	Additional properties	Improvements, Bug fixing
	Data & Al Marketplace	UI Development, Integration	Data Assets (Algorithms/Mo	dels, Results) in Marketplace	Improvements, Bug fixing
Data Sharing Services Bundle	Contract Lifecycle Manager	UI Development, Integration	Multi-party Data	a Asset Contracts	Improvements, Bug fixing
Bunale	Remuneration Engine	Integration, End-to-end Testing	Frictionless Payment	s R	emuneration Payments
	Wallet Manager	Back-end Development	Support in OPE	Stress Testing	Improvements, Bug fixing
Data Matchmaking	Query Builder	UI Development, Integration	Search in Data Assets (Algorithms	/Models, Results) beyond Datasets	Improvements, Bug fixing
Services Bundle	Matchmaking Engine	UI Development	Integration, Performance Testing	Additional Recomm. Algorithms	Improvements, Bug fixing

Figure 102: SYNERGY Integration Plan on M18-Part I





		Beta Release		Release 1.00 気1 冷	Release 2.00
		Q2 - 2021	Q3 - 2021	Q4 - 2021	(Q1-Q4) 2022
	Analytics Workbench	UI Development, Integration	Additional Execu	tion Frameworks	Configuration of SMC in UI
Data	Visualization & Reporting Engine	UI Development, Integration	Additional Built-in Chart	Additional Built-in Charts, Export/Save Diagrams	
Analytics Services	Data Manipulation Service	Integration, End-to-end Testing	Additional Data Manipulation Functions		Improvements, Bug fixing
Bundle	Analytics Execution	Integration, End-to-end Testing	Additional Basic & Pre-trained Models Secure Multi-pr		party Computations (SMC) Execution
	Service Secure Results Export Service	Integration, End-to-end Testing	Stress Testing	Improvements, Bug fixing	Improvements, Bug fixing
Data	Master Controller	Integration, End-to-end Testing	Stress Testing Ed	lge OPE Orchestration	Support for SMC
Governance Services Bundle	Data Lineage Service	-		Back-end Development	UI Development, Integration
Bullule	CIM Manager	UI Development, Integration	Stress Testing	Improvements, Bug fixing	Improvements, Bug fixing
	Resources Orchestrator	Integration, End-to-end Testing	Stress Testing	SEP Improvements, Bug fixing	Improvements, Bug fixing
Platform Management Services	Notifications Engine	UI Development, Integration	Stress Testing	Mail notifications	Improvements, Bug fixing
	Security, Auth. & Author. Engine	Secure Key Exchange	Stress Testing	Penetration Testing	Invest. of Remote Attest., Bug fixing
Bundle	Platform Analytics Engine	•		Back-end Development	UI Development, Integration
	API Gateway	UI Development, Integration	Stress Testing	Improved failure management	Investigation of OPE Data Retrieval

Figure 103: SYNERGY Integration Plan on M18-Part II

7.2 Platform Availability

The SYNERGY Platform is deployed in its production environment at:

https://www.synergy-bigdata.eu/

Since the SYNERGY Platform allows organisation-based access, the organisation manager needs to register his/her personal profile and then register the organisation. Taking into consideration that the SYNERGY Platform is addressed to stakeholders of the electricity data value chain, the platform administrator moderates the organisation's registration and may grant or deny access to the platform. If an organisation is approved, then the organisation's manager can invite additional members to join the organisation, which receive an invitation code via email.

Demo accounts may become available upon request at: <u>admin@synergy-bigdata.eu</u>

7.3 Platform Support Mechanisms

The beta release of the SYNERGY Integration Platform has been deployed and is available to the SYNERGY consortium for experimentation and testing. In order to provide efficient support to the demo partners in WP8 and the energy application developers in WP5-WP7, the following support mechanisms shall be leveraged:



SYNERGY

- Online training sessions aiming at providing walkthroughs and guidance of the SYNERGY Platform functionalities to the demo partners and the application developers. The first platform demo event is scheduled for July 7th, 2021 in which the SYNERGY Platform will be launched to the whole SYNERGY Consortium.
- **Direct support channel in Slack** under a "SYNERGY-Support" workspace in which dedicated channels have been created:
 - *#platform-app-support* intended for the energy apps developers to request support and guidance on the use of SYNERGY Platform by the technical team.
 - *#platform-demo-support* intended for the demo partners to request support and guidance on the use of SYNERGY Platform by the technical team.
 - *#platform-deployment* intended for the technical team to announce maintenance times and any redeployments that occur on the SYNERGY Platform, to the SYNERGY Consortium.
- Issue tracking in Github under a dedicated organization entitled "SYNERGY Project". A "platform-support" private repo has been created and includes pre-defined templates: (a) Bug report to collect all the bug information that will help the technical team to reproduce and solve the issue encountered (as depicted in the following figure); (b) Feature report to collect any enhancements and ideas that can improve the user experience and can be considered by the technical team for the next releases of the Platform. In order to properly manage, prioritise and track any problems/issues until they have been successfully resolved, all partners will be invited to join the SYNERGY Project in Github. It needs to be noted that issue reporting via email will not be allowed.





 Code O Issues Netline equest Discussions Actions Projects Security Minights Settings 	Search or jump to	Pull requests issues Marketplace Explore		¢ +• (€••
Iteration Statement Concert Iteration Iteration Iteration Iteration Iteration The Iteration of Iterat	SYNERGY-Big-Data-P	Platform / platform-support Private	③ Unwatch → 1	Star 0
Create a report to help us improve. If this doesn't lock right, choose a different type. If the If the If the <	<> Code 🛈 Issues 👔	Pull requests 🖓 Discussions 💿 Actions 🔟 Projects 🕕 Security 🗠 Insights 🛞 Settings		
Create a report to help us improve. If this doesn't look right, choose a different type. Fittle Fit	lssue: B	ug report	2	ŵ
Intel Image: Control of the serve of	Create a repo	ort to help us improve. If this doesn't lock right, choose a different type.	No one—assign yourself	
## Summary Projects Image: Second Secon		Title		\$3 1
Summarize in concise manner the bug which you encountered in the SYNERGY Platform, explaining the current behavior (what actually happens when you use the SYNERGY Platform) versus the expected behavior (what you expected to happen). ## Steps to reproduce Describe step-by-step how we can reproduce the issue. This part needs to be filled in in detail, including any relevant screenshots and logs - plases use (a) numbering, (b) screenshots, and (c) appropriate code blocks (^) to format console output, logs, and code, as it's very hard to read otherwise. 1. Go to '' 2. Click on '' 3. Scroll down to '' 4. See error Helpful resources Helpful resourc		Write Preview H B $I \coloneqq \heartsuit \mathscr{O} \coloneqq \boxminus \boxtimes \boxtimes \boxdot$	Projects	tộ:
actually happens when you use the SYNERGY Platform) versus the expected behavior (what you expected to happen). ## Steps to reproduce Describe step-by-top how we can reproduce the issue. This part needs to be filled in in detail, including any relevant screensthots and logs - please use: (a) numbering, (b) screenschots, and (c) appropriate code blocks (C) to format console output, logs, and code, asit's very hard to read otherwise. 1. Go to '' 2. Click on '' 3. Scroll down to '' 4. See error			None yet	
### Steps to reproduce Linked pull requests Describe text-by-step how we can reproduce the issue. This part needs to be filled in in detail, including any relevant Linked pull requests screenshots and logs - please use: (a) numbering. (b) screenshots, and (c) appropriate code blocks (") to format console output, Linked pull requests 1. Go to '-' Sccreenshot on '-' Sccreenshot on '-' 2. Click on '' None yet 3. Scroll down to '' Helpful resources Scee error Screenshot content is to independent in the scenario of the				¢3
logs, and code, as it's very hard to read otherwise. Successfully merging a pull request may close this sole. 1. Go to '' Successfully merging a pull request may close this sole. 2. Click on '' None yet 3. Social clown to '' Helpful resources Helpful resources Helpful resources			No milestone	
1. Go to '' 2. Click on '' 3. Scroll down to '' 4. See error Helpful resources Strittle Computed Strikelines				
3. Scroll down to '' 4. See error Helpful resources Gittab Computer Science as				je this
4. See error Heipful resources			None yet	
Sithub Computer Suidelines				
GitHub Community Guidelines				
Attach files by dragging & dropping, selecting or pasting them.	A	Attach files by dragging & dropping, selecting or pasting them.	GitHub Community Guidelines	

Figure 104: Bug report in Github





8 Conclusions & Next Steps

In this deliverable, D3.4 "SYNERGY Integrated Platform – Beta Release", a thorough description of the different functionalities offered by the SYNERGY Platform, is provided. In particular, the different workflows of the core platform functionalities (i.e., data check-in, data search and acquisition, data analytics) and the additional platform functionalities (CIM manager, edit organisation profile, edit user profile, wallet management) that are supported by the SYNERGY Platform, are described thoroughly as user journeys for data asset providers and consumers.

Towards this end, this deliverable documented the beta release of the SYNERGY Platform, including an overall description of the SYNERGY Platform described in Section 2, as well as all the supported functionalities that were developed under WP3 and WP4 towards the delivery of this beta release. In particular, the main user journeys of the developed functionalities include:

- The Data Check-in User Journey which describes thoroughly the steps from the data checkin job creation to its configuration and execution (including data harvester, mapper, cleaner, anonymiser, and encryption), and additionally the profiling of the resulting data assets.
- The Data Search and Acquisition User Journey which describes the functionalities that allow users (both data asset providers and consumers) to browse data assets (i.e. datasets for this release) within the SYNERGY Marketplace, and prepare data contracts between data asset providers and consumers in order to retrieve such assets from the SYNERGY Platform, according to their preferences.
- The Data Analytics User Journey which describes the functionalities for designing, configuring, executing data analytics pipelines, such that the platform's users gain valuable insights for their own and acquired data (by visualizing the appropriate results or retrieving them according to their terms).
- The additional functionalities available in the SYNERGY Platform (i.e., CIM manager, the organisation profiles, the user profiles, and the wallet manager).

This document not only provides the aforementioned user journeys for the supported data services bundles and the exposed APIs, but it also provides updates on the previous release (D3.3 "SYNERGY Integrated Platform, Alpha mock-up Release") including certain considerations, assumptions, and restrictions that occurred during the development.



SYNERGY

The integration plan (as defined in D3.3 "SYNERGY Integrated Platform – Alpha, Mock-ups Release") remains valid, driving two major back-end development releases (to be reported on M22 in D3.5 "Data Collection, Security, Storage, Governance & Management Services Bundles – Release 1.00", D4.3 "SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles – Release 1.00"), and the next integrated platform release (to be reported on M24 with D3.6 "SYNERGY Integrated Platform & Open APIs – Release 1.00").

The deployment of the different components and services included in the current release followed a solid integration plan to ensure that their interrelations and dependencies are properly reflected and their integration is appropriately prioritised in order to provide the most added value possible to the electricity data value chain stakeholders through quick-wins. The components and services that fall under the context of the current release were deployed in the SYNERGY Cloud Platform and its Secure Experimentation Playgrounds (SEP), and the On-Premise Environments (OPE), as described in detail in D2.6 "SYNERGY Framework Architecture including functional, technical and communication specifications v1". Finally, the secure transfer of data across the platform's different layers (i.e., Cloud, SEP, OPE) is deployed as mentioned in the previous integration plan (defined in D3.3 "SYNERGY Integrated Platform – Alpha, Mock-ups Release").

The current version of the SYNERGY Integrated Platform will be available to SYNERGY stakeholders in the beginning of M19 for early assessment and feedback that will be collected during the development activities of the various SYNERGY energy apps in WP5-WP7, the demonstration activities in WP8, and the living lab activities in WP9. Based on these, necessary enhancements and updates will be taken into consideration and potentially introduced in the upcoming versions of the platform.





9 References

SYNERGY Consortium. (2020). Description of Action

- SYNERGY Consortium. (2020). SYNERGY D2.1 "End-user and Business requirements analysis for big data-driven innovative energy services and ecosystems v1"
- SYNERGY Consortium. (2020). SYNERGY D2.6 "SYNERGY Framework Architecture including functional, technical and communication specifications v1"
- SYNERGY Consortium. (2021). SYNERGY D3.1 "SYNERGY Common Information Model"
- 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 D4.1 " SYNERGY Data Analytics, Sharing & Matchmaking Services Bundles Beta Release"
- ICARUS Project (2021). https://www.icarus2020.aero/
- CYBELE Project (2021). https://www.cybele-project.eu/
- UNICORN Project (2021). http://unicorn-project.eu/
- MATILDA Project (2021). https://www.matilda-5g.eu/

