



D3.3 SYNERGY Integrated Platform – Alpha, Mock-ups Release



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



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

Deliverable n°:	D3.3
Deliverable name:	SYNERGY Integrated Platform – Alpha, Mock-ups Release
Version:	1.0
Release date:	31/03/2021
Dissemination level:	Public
Status:	Submitted
Authors:	MAG, Suite5, UBITECH, KBZ, ETRA, ICCS, FVH, CUE, COBRA, CIRCE, CAV, IPTO, KRK, KONCAR, ENES



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

Document history:

Version	Date of issue	Content and changes	Edited by
0.1	18/01/2021	Template version for DEM deliverables	Suite5
0.2	08/02/2021	Initial D3.3 Table of Contents	MAG
0.3	01/03/2021	Initial contributions to Sections 2, 3, 4	MAG, Suite5, UBITECH
0.4	12/03/2021	Updated contributions to Sections 2, 3, 4, 5	MAG, Suite5, UBITECH, KBIZ
0.5	16/03/2021	Draft section 6	MAG, Suite5
0.6	17/03/2021	Draft circulated for internal review	MAG
0.7	29/03/2021	Updates based on the internal review feedback	MAG, Suite5, UBITECH
0.8	30/03/2021	Updated version cross-checking alignment with D3.2, D4.1	MAG, Suite5
1.0	31/03/2021	Final version to be submitted to the EC	MAG (Nikos Achilleopoulos)

Peer reviewed by:

Partner	Reviewer
COBRA	Miguel Crespo Chacon
CAV	Ida Johansson

Deliverable beneficiaries:

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

Table of contents

Executive summary.....	5
1 Introduction	6
1.1 Purpose of the document	6
1.2 Scope of the document	6
1.3 Structure of the document	7
2 Data Check-in User Journey.....	9
2.1 Create a New Data Check-in Job	9
2.2 Configure a Data Check-in Job	10
2.2.1 Data Ingestion Configuration	11
2.2.1.1 File Upload Method	11
2.2.1.2 Application's Own (External) API Method	12
2.2.1.3 Platform's (Internal) API Method	15
2.2.1.4 Platform's (Internal) Kafka PubSub Method	16
2.2.1.5 Application's Own (External) Kafka PubSub Method	18
2.2.2 Pre-processing Rules Definition	19
2.2.2.1 Mapping Configuration	19
2.2.2.2 Cleaning Configuration	22
2.2.2.3 Anonymisation Configuration	25
2.2.2.4 Encryption Configuration	27
2.2.3 Define New Data Asset Profile	28
2.3 Execute a Data Check-in Job	32
2.3.1 View all Data Check-in Jobs	32
2.3.2 Pre-processing Rules Execution	33
2.3.2.1 Mapping Execution	34
2.3.2.2 Cleaning Execution	35
2.3.2.3 Anonymisation Execution	37
2.3.2.4 Encryption Execution	38
3 Data Search and Acquisition User Journey	40
3.1 Navigate to the SYNERGY Marketplace	40
3.2 Search for Data Assets	44
3.3 Data Asset Acquisition	47
3.3.1 Request for Data Assets (Data Asset Consumer Perspective)	48
3.3.2 Review a Data Asset Request (Data Asset Provider Perspective)	49

3.3.3	Prepare a draft contract (Data Asset Provider Perspective)	50
3.3.4	Review a draft contract (Data Asset Consumer Perspective)	52
3.3.5	Negotiate a draft contract (Data Asset Consumer Perspective)	54
3.3.6	Review a revised contract (Data Asset Provider Perspective)	55
3.3.7	Settle a finalized contract (Data Asset Provider Perspective)	56
3.3.8	View all organization's Contracts	57
3.4	View My Assets	58
3.5	Create a Retrieval Query	59
3.5.1	Create Retrieval Query through API	60
3.5.2	Create Retrieval Query for Streaming Data	62
3.5.3	View all Retrieval Queries	64
4	Data Analytics User Journey	66
4.1	View all Analytics Pipelines	66
4.2	Create an Analytics Pipeline	67
4.3	Configure an Analytics Pipeline	68
4.3.1	Graph View	69
4.3.2	Table View	70
4.3.3	Results View	71
4.3.4	Schedule Execution	72
4.3.5	Execution History	73
4.4	Visualization Results of an Analytics Pipeline	74
5	Additional Platform Functionalities	76
5.1	Common Information Model (CIM) Manager	76
5.2	Edit Organisation Profile	77
5.3	Edit User Profile	79
6	Integration Plan	81
7	Conclusions.....	85
8	References.....	87

Abbreviations and Acronyms

Acronym	Description
AI	Artificial Intelligence
API	Application Programming Interfaces
CIM	Common Information Model
CSV	Comma-Separated Values
DL	Deep Learning
DoA	Description of Action (annex I of the Grant Agreement)
JSON	JavaScript Object Notation
ML	Machine Learning
OPE	On-Premise Environment
SASL	Simple Authentication and Security Layer
SEP	Secure Experimentation Playground
TSV	Tab-Separated Values
XML	Extensible Markup Language
WP	Work Package

Executive summary

The present document D3.3 presents the alpha, mock-ups release of the SYNERGY Integrated Platform that brings together 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, that are developed under WP3 “End-to-end Interoperable Big Data Management Platform” and WP4 “Big Data Analytics and Data Sharing Mechanisms”.

D3.3 aims to navigate the technical partners, as well as the demo partners, to the user experience that data asset providers and data asset consumers will acquire in the SYNERGY Integrated Platform through high-fidelity mock-ups. Seventy one (71) mock-ups have been designed in detail to capture the core functionalities that are offered in three core user journeys, as well as through additional platform functionalities: (a) the data check-in user journey, describing how data asset providers will make available their data assets in the SYNERGY Platform (with a clear separation of concerns between the design and execution time); b) the data search and acquisition user journey which allows data asset consumers to search, explore and acquire in a trustful manner data assets, through the SYNERGY Data & AI Marketplace; and c) the data analytics user journey for both the data asset providers and data asset consumers, including the data analytics pipeline design and execution that aims to provide valuable insights into their own and acquired data assets.

In addition, the integration plan of the SYNERGY Platform has been devised taking into consideration the interrelations among the different components and services, the SYNERGY use cases and requirements (from D2.1), the SYNERGY architecture (defined in D2.6) and the prioritisation that had to be enforced to ensure that the beta release of the SYNERGY Platform will already bring significant added value to the electricity data value chain stakeholders.

The alpha release of the SYNERGY Platform has been extensively discussed among the technical partners and the early feedback of the demo partners over the mock-ups has been already collected (during the SYNERGY Plenary Meeting that was held on March 2nd-3rd, 2021) and addressed. Such an early release naturally builds on the back-end developments that have been delivered in the different Data Services Bundles in WP3 and WP4 and will drive the integration activities in T3.5 towards the beta release of the SYNERGY Platform, that is expected on M18.

1 Introduction

1.1 Purpose of the document

This deliverable constitutes a description of the initial efforts expended, and the corresponding results obtained regarding the SYNERGY Integrated Platform Alpha Version comprising the functional, and graphical mock-ups. In particular, this deliverable reports the procedures that were followed, and the results of the task T3.5 “Platform and Services Bundles Continuous Integration and SYNERGY Open APIs Delivery” in bringing together 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, developed under WP3 “End-to-end Interoperable Big Data Management Platform” and WP4 “Big Data Analytics and Data Sharing Mechanisms”.

This deliverable essentially documents the alpha release of the SYNERGY Integrated Platform describing step-by-step the core functionality through the high-fidelity user interfaces (mockups) and is expected to guide the integration activities in WP3 and WP4 towards the SYNERGY Integrated Platform Beta Release on M18 (that shall be documented in the SYNERGY Deliverable D3.4).

1.2 Scope of the document

D3.3 aims to present and describe thoroughly the user experience of data asset providers and data asset consumers in the SYNERGY Integrated Platform. In particular, the scope of this document is:

- To present the set of core functionalities that are divided into three distinct core workflows, as well as additional platform functionalities. Specifically, a complete description of the core functionalities is provided and presented separately for each workflow: a) the data check-in user journey, describing the process that is essentially followed by data asset providers for data uploading; b) the data search and acquisition user journey which includes the available functionalities that are available to data asset consumers for data search, exploration and acquisition; and c) the data analytics user

journey for both the data asset providers and data asset consumers, including the analytics pipeline design and execution in order to perform data analysis and visualisations. In addition, a description of the additional platform functionalities is provided to supplement the range of functionalities offered by the SYNERGY Integrated Platform.

- To present the integration and release plan of the SYNERGY platform by identifying eventual dependencies between the functionalities and components that are indirectly documented in this deliverable (but are also presented in detail in the SYNERGY Deliverables D2.6, D3.2 and D4.1).

This initial Alpha version of the SYNERGY Integrated Platform serves as the foundation for the integration activities of the beta release of the different Data Services Bundles that have been developed in WP3 and WP4 as documented in the SYNERGY Deliverables D3.2 and D4.1. It has been developed taking into consideration the technical requirements and the 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. The alpha release of the SYNERGY Integrated Platform has also inspired discussions among the technical partners and the demo partners with regard to the expected functionalities of the SYNERGY Integrated Platform’s beta release, while potential refinements or adjustments that emerged were taken into consideration.

1.3 Structure of the document

The structure of the document is organized as follows:

- Section 2 presents the data check-in user journey. In particular, a complete workflow for the data check-in workflow is provided to guide energy data value chain stakeholders, that serve as data asset providers, through the data check-in functionalities during the design and execution phases.
- Section 3 provides a thorough description of the data search and acquisition workflow. This workflow encompasses the steps that energy data value chain stakeholders, acting as data consumers, may follow in order to search for data assets in the SYNERGY data & AI marketplace and purchase them.

- Section 4 describes the data analytics user journey. It reveals the capabilities that an energy data value chain consumer has, regarding the analysis on their own and acquired data in order to gain valuable insights. More specifically, in Section 4 the workflow to design and configure an analytics pipeline composed of different blocks (e.g., data preparation, learning algorithms) to unveil such insights is presented.
- Section 5 provides a description of the workflow of additional functionalities that the SYNERGY Platform enables to different actors/users.
- Section 6 documents the integration plan for future releases of the SYNERGY Integrated Platform.
- Section 7 concludes this deliverable D3.3 “SYNERGY Integrated Platform – Alpha, Mock-ups Release”, providing a brief overview of the next steps towards the SYNERGY Integrated Platform beta release.
- Section 8 includes the list of relevant references.

2 Data Check-in User Journey

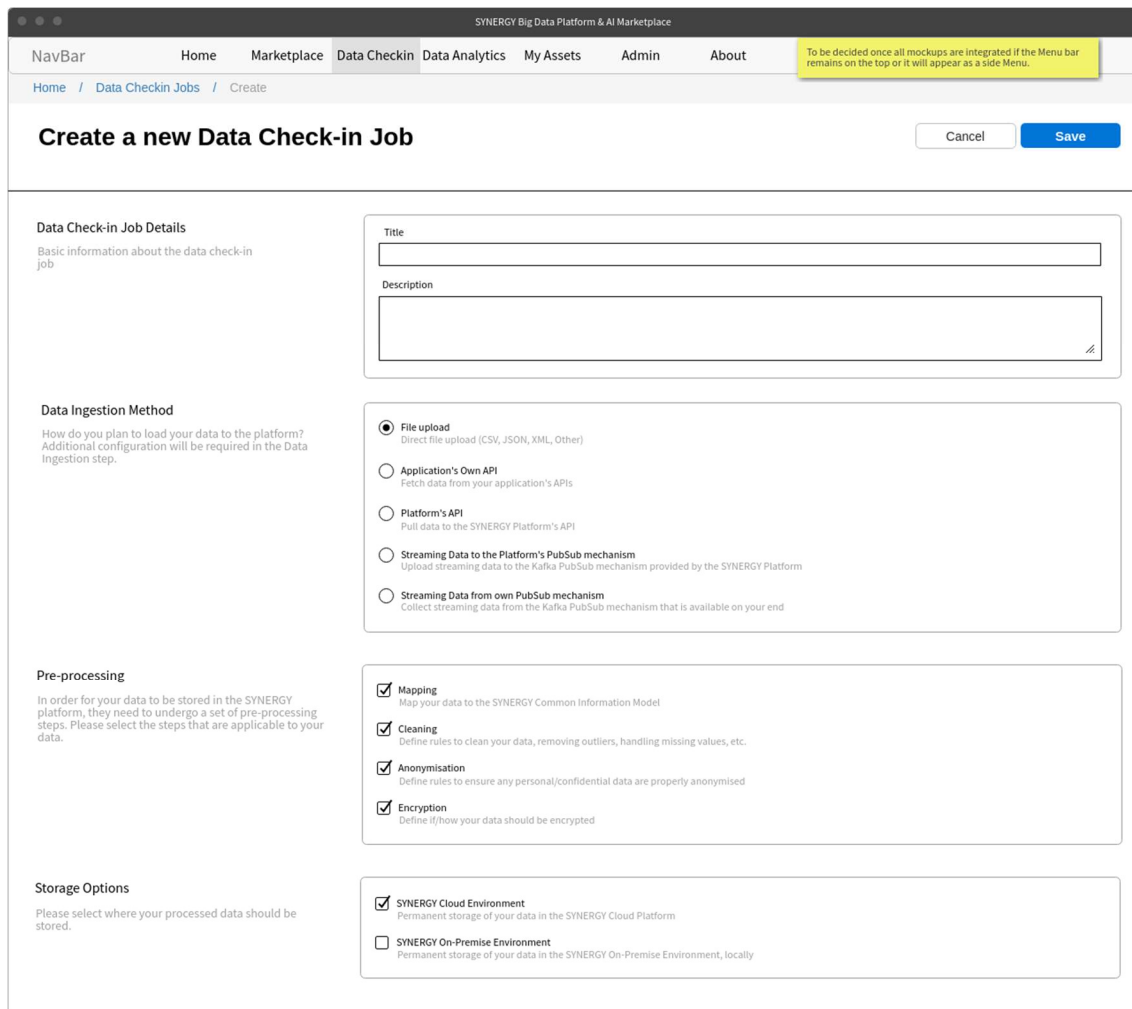
In this section, the data check-in user journey including the various steps that the data asset provider needs to follow in order to load data on the SYNERGY Platform is presented. In particular, during this workflow the data asset provider starts by creating a new data check-in job, configures the data ingestion process that is responsible to load the data on the SYNERGY Platform, defines the pre-processing rules that are to be applied on the data asset, defines the data asset's metadata and licensing details and finally runs the data check-in job execution to start loading the data to the SYNERGY Platform.

In general, it should be emphasized that there is a distinct separation between the creation and configuration (design) phase of the data check-in job as described in sections 2.1 and 2.2, and the actual execution of the configured data check-in job by the SYNERGY Platform as described in section 2.3.

2.1 Create a New Data Check-in Job

The data check-in workflow initiates with 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 Data Check-in option located at the top navigation bar of the SYNERGY Big Data Platform & AI Marketplace as depicted in Figure 1. 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. Following that, the data asset provider needs to select the data ingestion method which determines how the data will be loaded on the SYNERGY Platform. The available options for data ingestion are through: a) file upload which allows data asset providers to upload files in different formats (i.e., CSV, JSON, XML, and other formats); b) application's own (external) API, where data can be fetched from APIs owned by the data asset providers; c) platform's API, where data can be pulled to the SYNERGY Platform API; d) streaming data that are published to the Kafka PubSub mechanism provided by the SYNERGY Platform; and e) streaming data collection by subscribing to the Kafka PubSub mechanism owned by the data asset provider. Another important configuration that has to be set during the creation of a new data check-in job, is to select the pre-processing rules to be applied on the loaded data. In particular, the data asset provider may select the applicable pre-processing rules (i.e., Mapping, Cleaning, Anonymisation, and Encryption) to be applied on the data, with Mapping being a precondition for the rest of the

steps to be enabled. The final step for the creation of the data check-in job, is to select the environment where the data should be stored. For permanent storage of the processed data to the SYNERGY Cloud Platform, data asset providers may select SYNERGY Cloud Environment, while for permanent storage of the processed data locally to an on-premise environment, they may select the SYNERGY On-Premise Environment option. By selecting the save button at the top of the screen, the data check-in job is saved, and the user is directed to a new configuration page depending on the data ingestion method selected.



The screenshot shows the SYNERGY Big Data Platform & AI Marketplace interface. The top navigation bar includes links for Home, Marketplace, Data Checkin, Data Analytics, My Assets, Admin, and About. A yellow tooltip indicates that the menu bar will be integrated into the top or side menu once all mockups are integrated. The main content area is titled 'Create a new Data Check-in Job' and includes a 'Cancel' button and a 'Save' button. The form is divided into four sections: 'Data Check-in Job Details', 'Data Ingestion Method', 'Pre-processing', and 'Storage Options'. The 'Data Check-in Job Details' section includes fields for 'Title' and 'Description'. The 'Data Ingestion Method' section includes radio buttons for 'File upload', 'Application's Own API', 'Platform's API', 'Streaming Data to the Platform's PubSub mechanism', and 'Streaming Data from own PubSub mechanism'. The 'Pre-processing' section includes checkboxes for 'Mapping', 'Cleaning', 'Anonymisation', and 'Encryption'. The 'Storage Options' section includes checkboxes for 'SYNERGY Cloud Environment' and 'SYNERGY On-Premise Environment'.

Figure 1: Data Check-in – Create New Check-in Job

2.2 Configure a Data Check-in Job

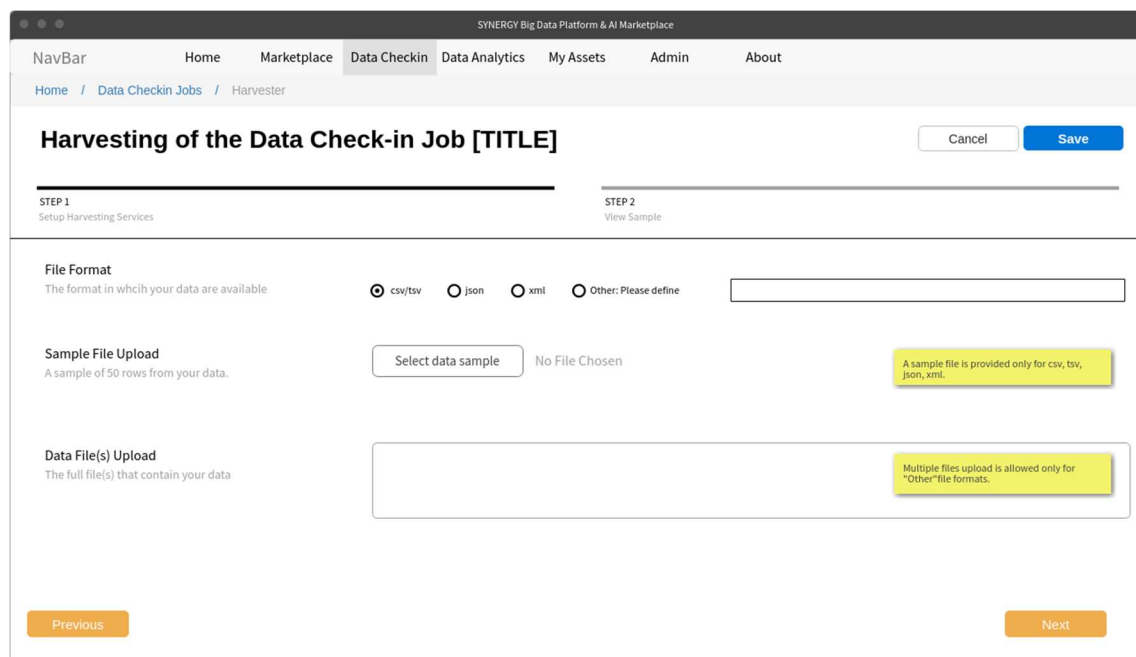
Data asset providers are able to load their data assets to the SYNERGY Platform using different methods that are described thoroughly in the subsequent subsections. Starting from the creation of a new data check-in job as described in section 2.1, a data asset provider is able to

configure among others, the source, periodicity, error handling strategy, and the execution environment of the data check-in job (i.e., SYNERGY Cloud Environment or SYNERGY On-Premise Environment). In this section, the data check-in job configuration workflow, as well as the different parameters that need to be configured, are described.

2.2.1 Data Ingestion Configuration

2.2.1.1 File Upload Method

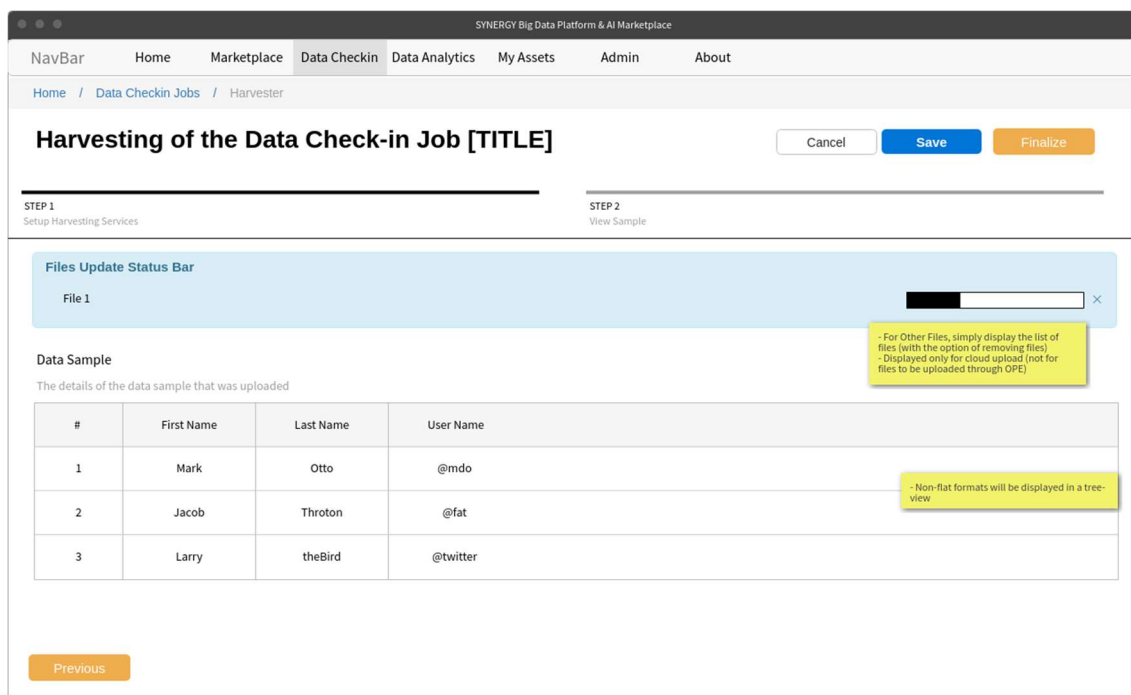
If the data asset provider has selected the data upload option, the Data Ingestion configuration page appears as depicted in Figure 2. In this configuration page, the file format of the data that are to be uploaded needs to be selected. Once the data asset provider has selected the data file format, a sample file including 50 entries (rows) from the whole data asset (applies to CSV, TSV, JSON, and XML files only), should be uploaded (separately from the actual data). It needs to be noted that if the file is not to be uploaded directly in the cloud, the data asset provider may alternatively provide the location of the file (where it is stored locally and it will be retrieved from the On-Premise Environment). Following the data sample upload, 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 3.



The screenshot shows the SYNERGY Big Data Platform & AI Marketplace interface. The top navigation bar includes links for Home, Marketplace, Data Checkin, Data Analytics, My Assets, Admin, and About. The breadcrumb trail is Home / Data Checkin Jobs / Harvester. The main title is 'Harvesting of the Data Check-in Job [TITLE]' with 'Cancel' and 'Save' buttons. The interface is divided into two steps: STEP 1 (Setup Harvesting Services) and STEP 2 (View Sample). Under STEP 1, there are three sections: 'File Format' with radio buttons for csv/tsv (selected), json, xml, and Other: Please define; 'Sample File Upload' with a 'Select data sample' button and 'No File Chosen' status; and 'Data File(s) Upload' with a file upload area. There are also two yellow informational boxes: 'A sample file is provided only for csv, tsv, json, xml.' and 'Multiple files upload is allowed only for "Other" file formats.' At the bottom, there are 'Previous' and 'Next' buttons.

Figure 2: Data Check-in – File Upload Method - Configure Harvester (Step 1)

During this step, the data asset provider may view the rows and columns of the data sample uploaded, in order to review the structure of the data that are to be uploaded. In case the files are directly uploaded in the cloud, the data asset provider is able to view a list of the uploaded files along with the uploading progress, while an option of removing files will be available as well. By selecting the Finalize button, the data asset provider confirms that the data sample is correct and final, and thus no further changes on the data sample are allowed.



Harvesting of the Data Check-in Job [TITLE] Cancel Save Finalize

STEP 1 Setup Harvesting Services STEP 2 View Sample

Files Update Status Bar

File 1 ✕

Data Sample

The details of the data sample that was uploaded

#	First Name	Last Name	User Name
1	Mark	Otto	@mdo
2	Jacob	Throton	@fat
3	Larry	theBird	@twitter

Previous

- For Other Files, simply display the list of files (with the option of removing files).
- Displayed only for cloud upload (not for files to be uploaded through OPE)

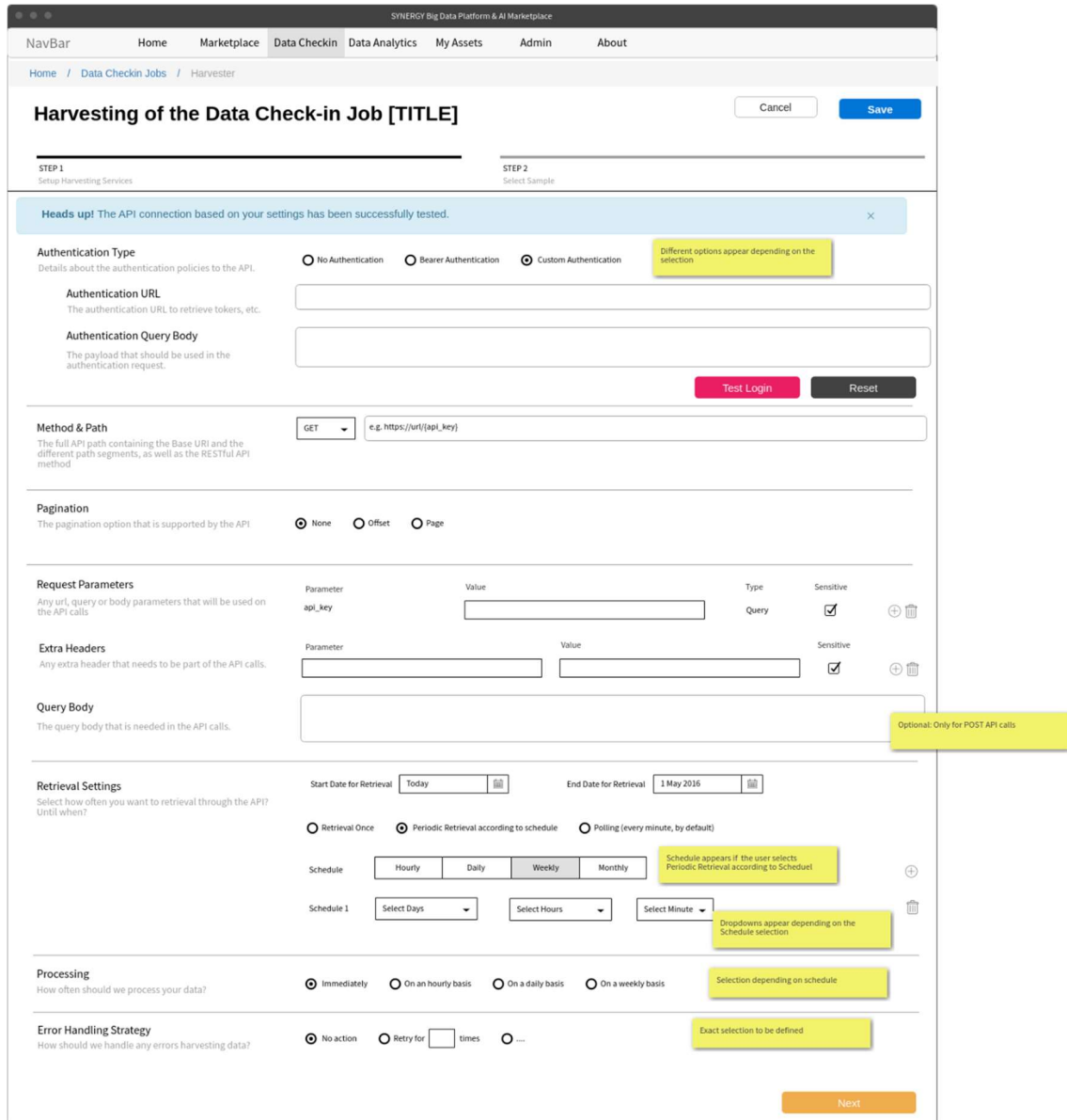
- Non-flat formats will be displayed in a tree-view

Figure 3: Data Check-in - File Upload Method - Finalize Harvester (Step 2)

2.2.1.2 Application's Own (External) API Method

In case the data asset provider selected to ingest data from an own (external) API that is exposed by his/her organisation's systems, an appropriate configuration page will appear as shown in Figure 4. Initially, the data asset provider is asked to select an option regarding the authentication type (i.e., No Authentication, Bearer Authentication, and Custom Authentication) and 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, while the data asset provider has the option to test or reset the authentication policies inserted, by selecting the button Test Login or Reset, respectively). Then, the full API path including the base URL and the query body of the request, along with the appropriate method (i.e., GET, POST, PUT), need to be inserted. 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. The data asset provider may also add any request parameters such as URL, query or body parameters, and extra headers that can be used for the API call. 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 through the API. 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).



Harvesting of the Data Check-in Job [TITLE]

Cancel Save

STEP 1
Setup Harvesting Services

STEP 2
Select Sample

Heads up! The API connection based on your settings has been successfully tested.

Authentication Type
Details about the authentication policies to the API.

☐ No Authentication ☐ Bearer Authentication ☒ Custom Authentication

Authentication URL
The authentication URL to retrieve tokens, etc.

Authentication Query Body
The payload that should be used in the authentication request.

Test Login Reset

Method & Path
The full API path containing the Base URI and the different path segments, as well as the RESTful API method

GET e.g. https://url/api_key

Pagination
The pagination option that is supported by the API

☒ None ☐ Offset ☐ Page

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

Parameter	Value	Type	Sensitive
api_key		Query	<input checked="" type="checkbox"/>

Extra Headers
Any extra header that needs to be part of the API calls.

Parameter	Value	Sensitive
		<input checked="" type="checkbox"/>

Query Body
The query body that is needed in the API calls.

Retrieval Settings
Select how often you want to retrieval through the API? Until when?

Start Date for Retrieval Today End Date for Retrieval 1 May 2016

☐ Retrieval Once ☒ Periodic Retrieval according to schedule ☐ Polling (every minute, by default)

Schedule Hourly Daily Weekly Monthly

Schedule 1 Select Days Select Hours Select Minute

Processing
How often should we process your data?

☒ Immediately ☐ On an hourly basis ☐ On a daily basis ☐ On a weekly basis

Error Handling Strategy
How should we handle any errors harvesting data?

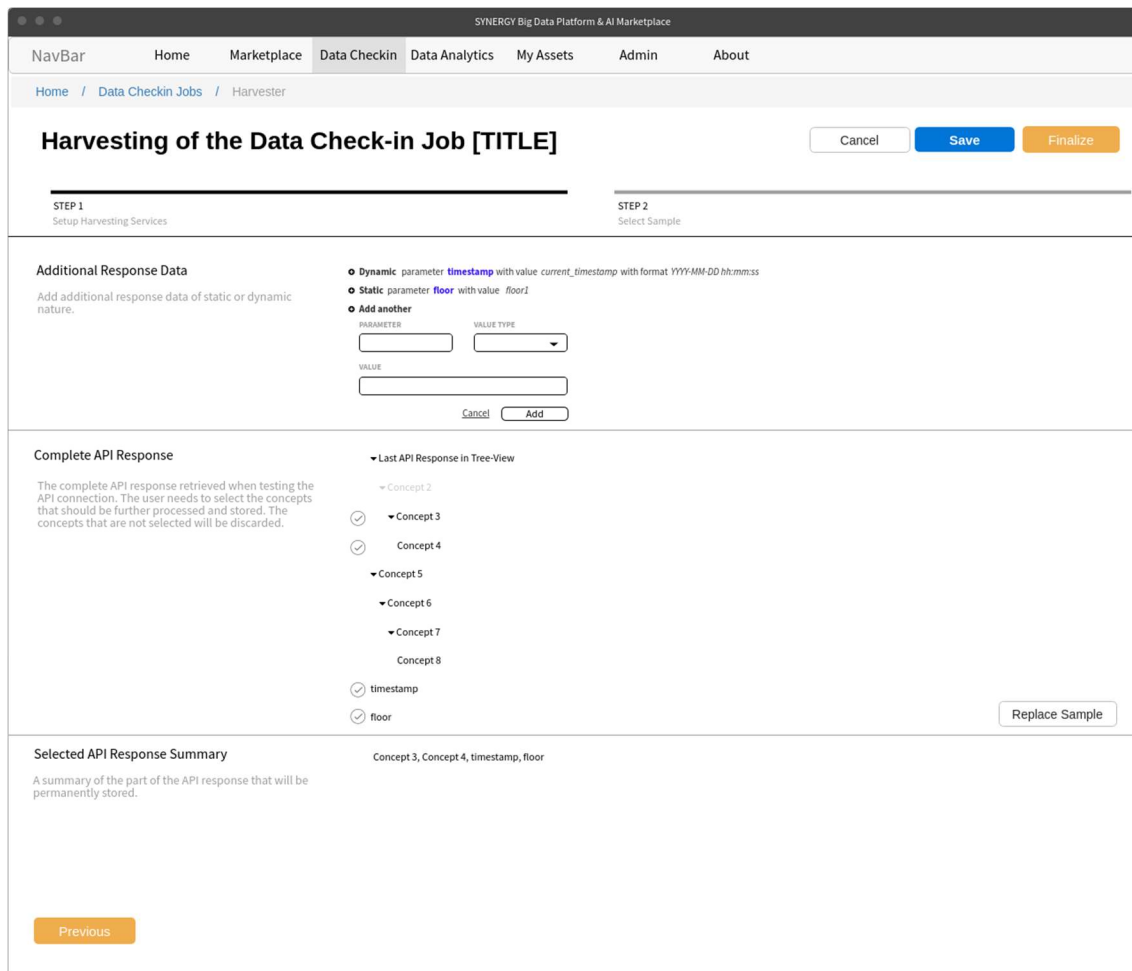
☒ No action ☐ Retry for times ☐ ...

Next

Figure 4: Data Check-in - Application's Own API Method - Configure Harvester

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 pooling 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. The option for the Error Handling Strategy determines the way that an error is handled. Currently there are two options available: no action, and setting a particular number of retries in case of error. By selecting the Next button, the SYNERGY platform will make a call to the API endpoint according to the configuration set. If it is successful, the data asset provider comes across the Finalize Harvester step as depicted in Figure 5. During this step, the data asset provider may add additional response data, associated with a value and its value type (static or dynamic), that will be stored along with each API response. 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.



The screenshot shows the SYNERGY Big Data Platform & AI Marketplace interface. The top navigation bar includes links for Home, Marketplace, Data Check-in, Data Analytics, My Assets, Admin, and About. The breadcrumb trail is Home / Data Check-in Jobs / Harvester. The main heading is 'Harvesting of the Data Check-in Job [TITLE]' with buttons for Cancel, Save, and Finalize. The interface is divided into two steps: STEP 1 (Setup Harvesting Services) and STEP 2 (Select Sample). STEP 2 is the active step, showing options for 'Additional Response Data' (Dynamic, Static, or Add another) and a 'Complete API Response' section with a tree view of concepts (Concept 2 through Concept 8, timestamp, floor) and a 'Replace Sample' button. At the bottom, there is a 'Selected API Response Summary' section and a 'Previous' button.

Figure 5: Data Check-in - Application's Own API Method - Finalize Harvester

2.2.1.3 Platform's (Internal) API Method

If 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 6 depicts. Similarly, as in other data ingestion methods, 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 fetched, by the SYNERGY Platform. In addition to this, the data asset provider needs to upload a sample file containing up to 50 entries (rows) from the data that will be sent to the SYNERGY Platform's API.

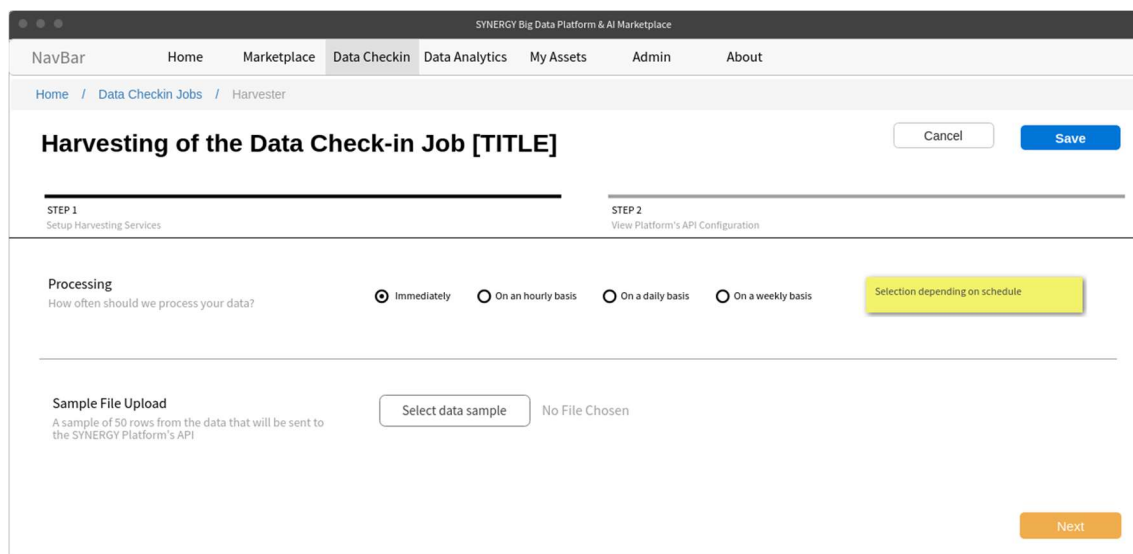


Figure 6: Data Check-in - Platform's API Method - Configure Harvester

During the second step of the Platform's API data harvesting method that is shown in Figure 7, the data asset provider is able to see an auto generated API endpoint in the SYNERGY platform. However, to use this generated API, the user should be authenticated. Hence, instructions on how to use the POST endpoint are provided accordingly. Finally, the data asset provider needs to review the details of the configuration and select Finalize to proceed to the next steps of the Data Check-in workflow.

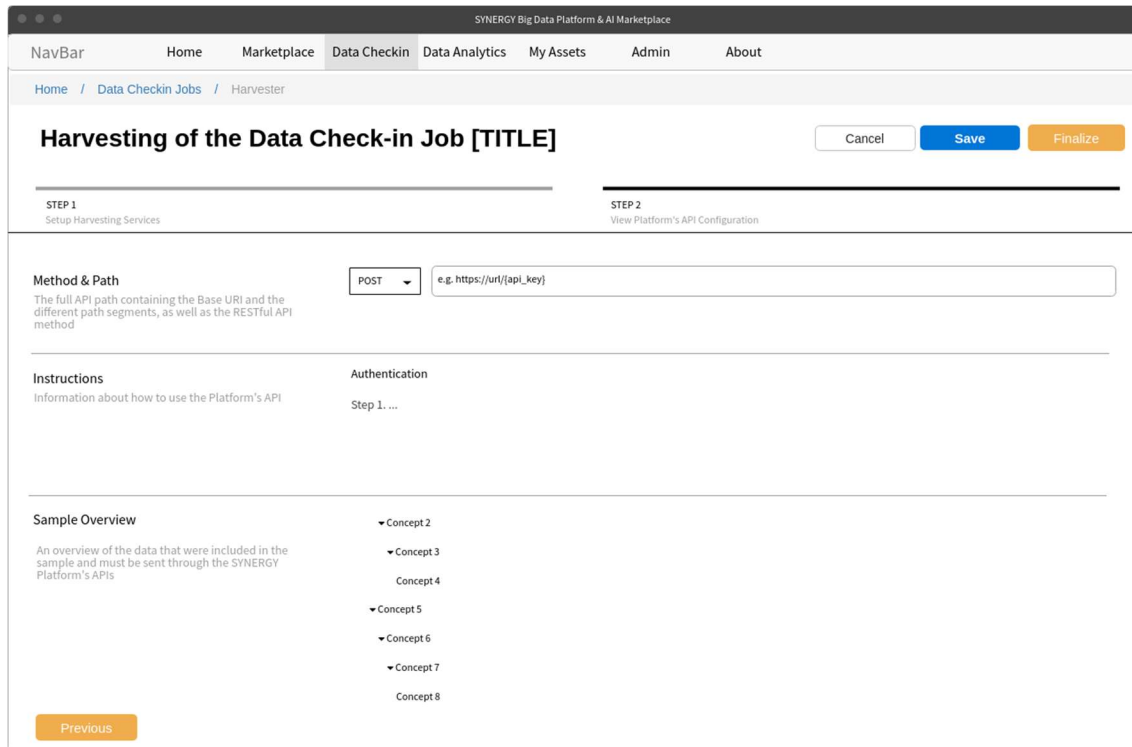
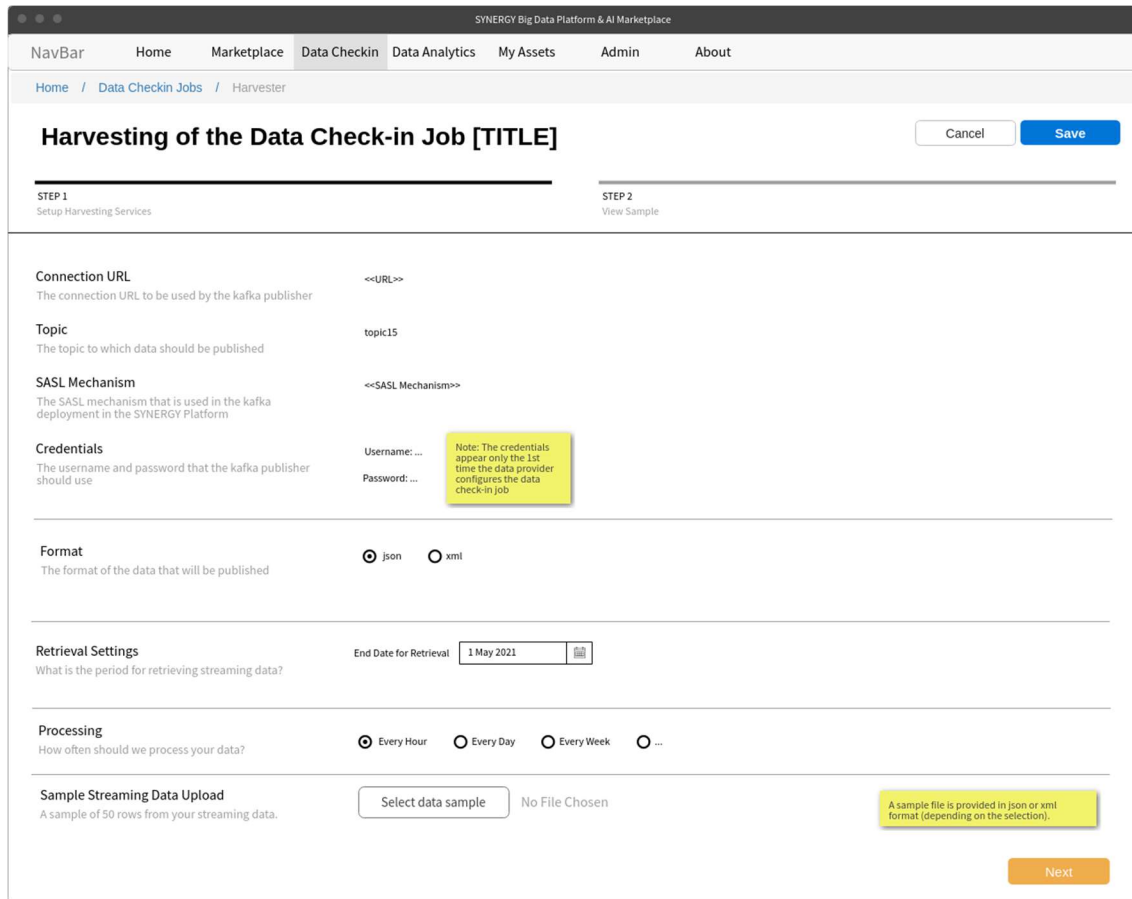


Figure 7: Data Check-in - Platform's API Method - Finalize Harvester

2.2.1.4 Platform's (Internal) Kafka PubSub Method

If the data asset provider selected the harvesting of the data through the Platform's Kafka PubSub Mechanism, the corresponding configuration page is shown as Figure 8 depicts. During this configuration, the data asset provider views 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. Then, the data asset provider should select the format of the data to be published (i.e., JSON, or XML), the retrieval end date and the periodicity of the data processing in order to define when the streaming data retrieval should stop by the SYNERGY platform, and how often the streaming data should be processed. 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 View Sample 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 9.



SYNERGY Big Data Platform & AI Marketplace

NavBar Home Marketplace Data Checkin Data Analytics My Assets Admin About

Home / Data Checkin Jobs / Harvester

Harvesting of the Data Check-in Job [TITLE]

Cancel Save

STEP 1 Setup Harvesting Services

STEP 2 View Sample

Connection URL
The connection URL to be used by the kafka publisher <<URL>>

Topic
The topic to which data should be published topic15

SASL Mechanism
The SASL mechanism that is used in the kafka deployment in the SYNERGY Platform <<SASL Mechanism>>

Credentials
The username and password that the kafka publisher should use
Username: ... Password: ...
Note: The credentials appear only the 1st time the data provider configures the data check-in job

Format
The format of the data that will be published ☒ json ☐ xml

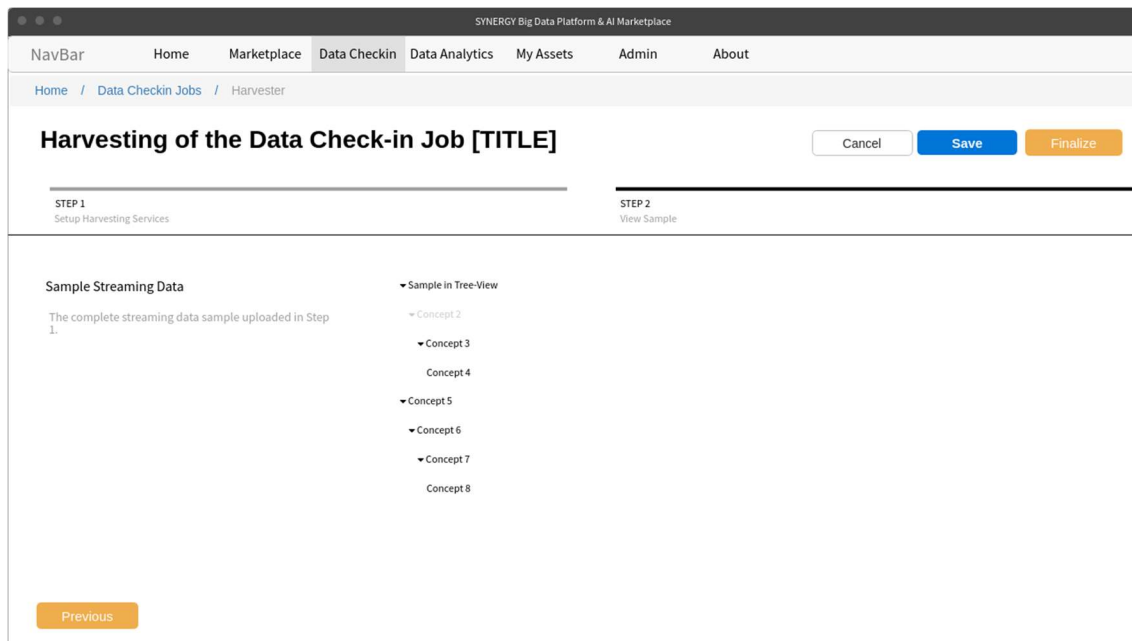
Retrieval Settings
What is the period for retrieving streaming data? End Date for Retrieval 1 May 2021

Processing
How often should we process your data? ☒ Every Hour ☐ Every Day ☐ Every Week ☐ ...

Sample Streaming Data Upload
A sample of 50 rows from your streaming data. Select data sample No File Chosen
A sample file is provided in json or xml format (depending on the selection).

Next

Figure 8: Data Check-in - Platform's Kafka Method - Configure Harvester (Step 1)



SYNERGY Big Data Platform & AI Marketplace

NavBar Home Marketplace Data Checkin Data Analytics My Assets Admin About

Home / Data Checkin Jobs / Harvester

Harvesting of the Data Check-in Job [TITLE]

Cancel Save Finalize

STEP 1 Setup Harvesting Services

STEP 2 View Sample

Sample Streaming Data
The complete streaming data sample uploaded in Step 1.

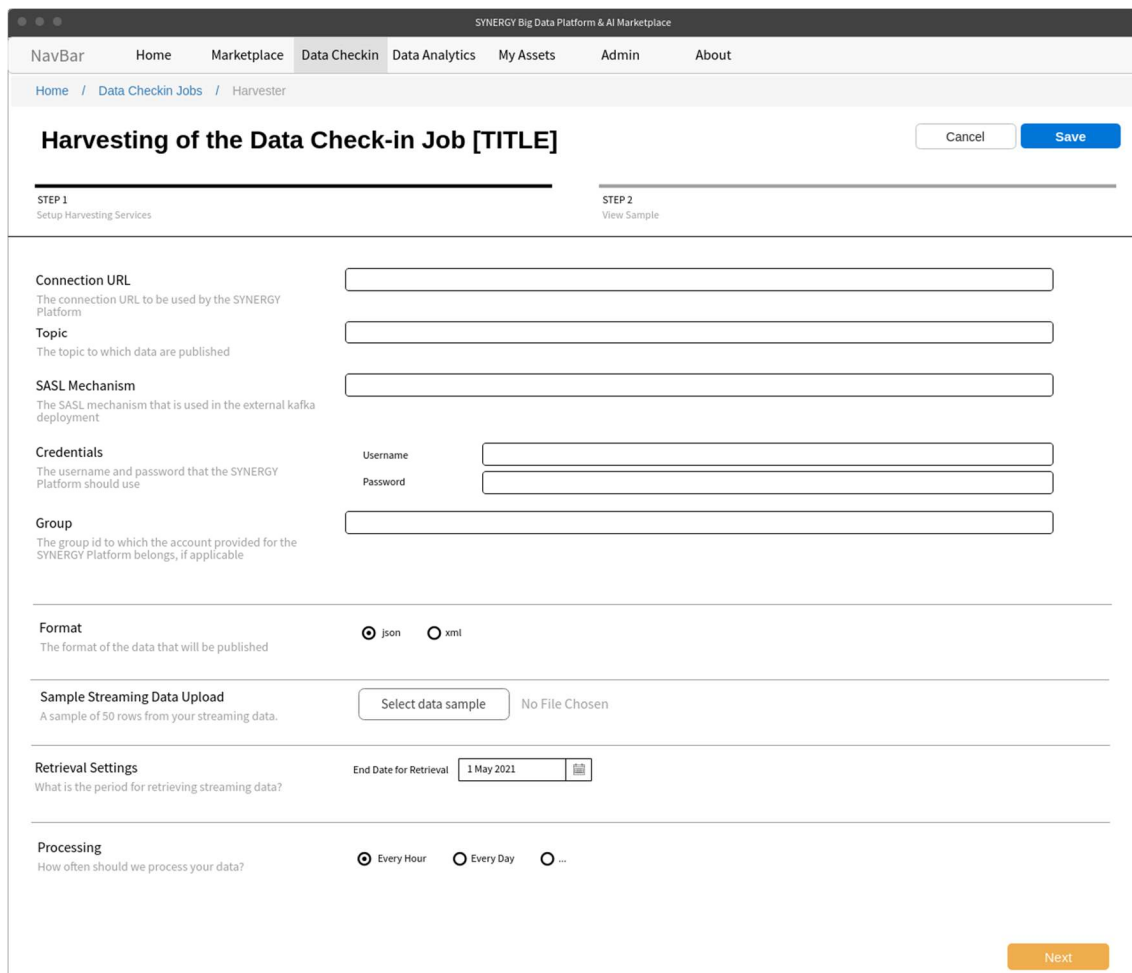
- Sample in Tree-View
 - Concept 2
 - Concept 3
 - Concept 4
 - Concept 5
 - Concept 6
 - Concept 7
 - Concept 8

Previous

Figure 9: Data Check-in - Platform's Kafka Method - Finalize Harvester (Step 2)

2.2.1.5 Application's Own (External) Kafka PubSub Method

In case the data asset provider selected to use his/her organisation's (external) Kafka PubSub mechanism for subscribing the SYNERGY Platform to the already published streaming data, the configuration page as shown in Figure 10 is presented. Initially, the data asset provider needs to 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. The next step of the configuration is 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. The retrieval end date and the processing periodicity need to be defined as well.



SYNERGY Big Data Platform & AI Marketplace

NavBar Home Marketplace Data Checkin Data Analytics My Assets Admin About

Home / Data Checkin Jobs / Harvester

Harvesting of the Data Check-in Job [TITLE]

Cancel Save

STEP 1 Setup Harvesting Services

STEP 2 View Sample

Connection URL
The connection URL to be used by the SYNERGY Platform

Topic
The topic to which data are published

SASL Mechanism
The SASL mechanism that is used in the external kafka deployment

Credentials
The username and password that the SYNERGY Platform should use

Username

Password

Group
The group id to which the account provided for the SYNERGY Platform belongs, if applicable

Format
The format of the data that will be published

☒ json ☐ xml

Sample Streaming Data Upload
A sample of 50 rows from your streaming data.

Select data sample No File Chosen

Retrieval Settings
What is the period for retrieving streaming data?

End Date for Retrieval 1 May 2021

Processing
How often should we process your data?

☒ Every Hour ☐ Every Day ☐ ...

Next

Figure 10: Data Check-in – Application's Own Kafka Method - Configure Harvester

By selecting the Next button, the data asset provider needs to review and select the concepts to be further processed and stored in the SYNERGY Platform (since there might be data

published in the specific topic that the data asset provider does not want the SYNERGY platform to eventually store). Finally, as Figure 11 depicts, the data asset provider may review the selected concepts from the Selected Response Summary section and select the Finalize button to complete the configuration of external streaming data ingestion.

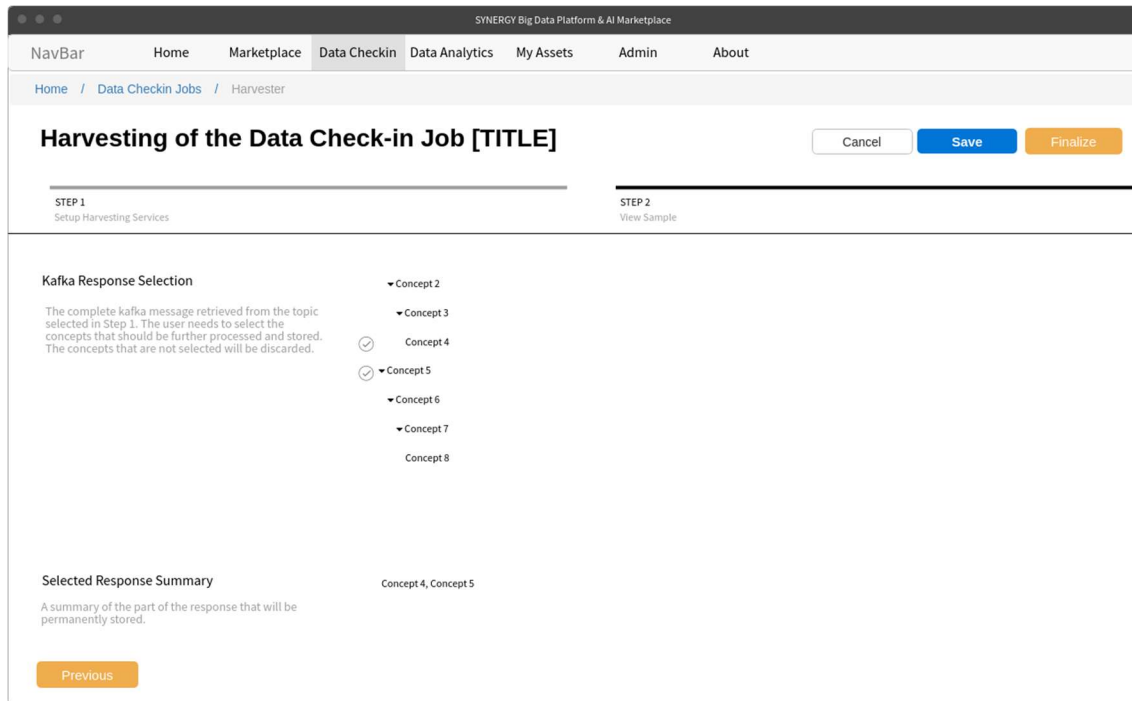


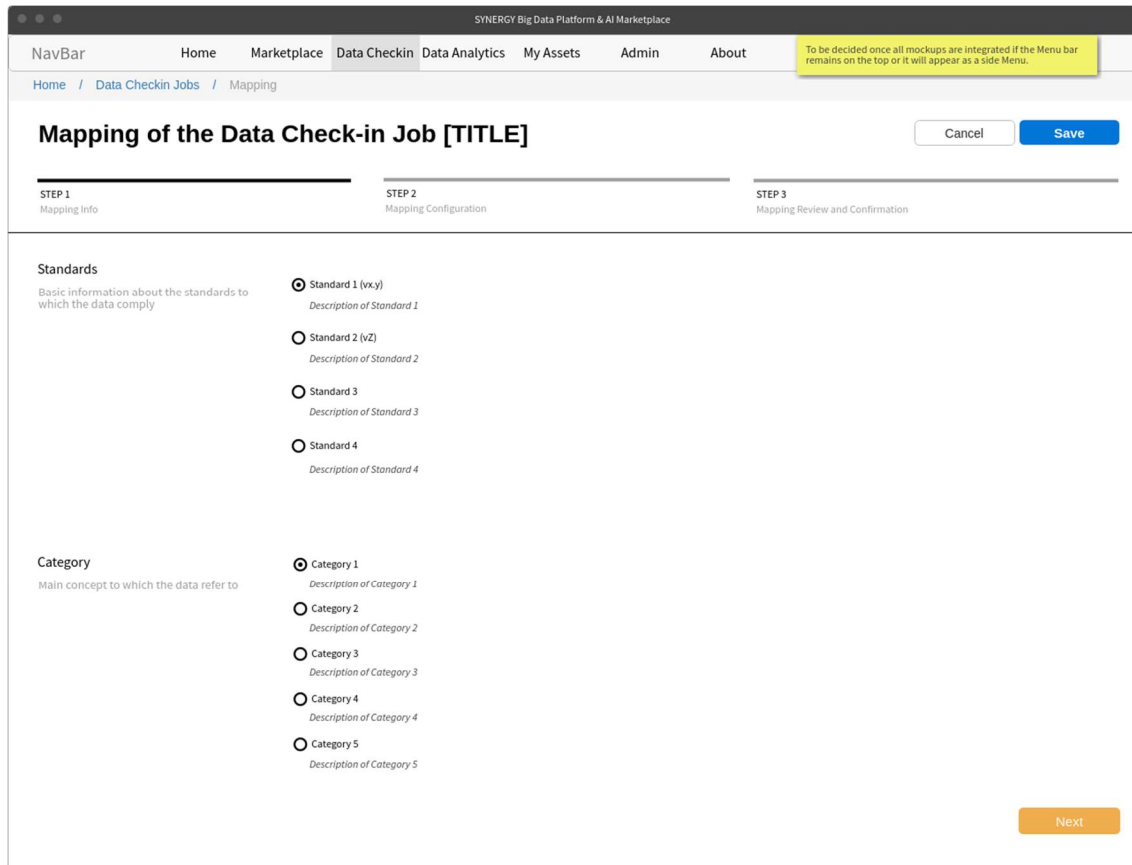
Figure 11: Data Check-in – Application’s Own Kafka Method - Finalize Harvester

2.2.2 Pre-processing Rules Definition

During the design and creation of a new check-in job, the data asset providers are asked to select some pre-processing steps ranging from Mapping to Cleaning, Anonymisation, and Encryption (which are optional steps), that are to be applied on the data upon their ingestion. This section describes the workflow that a data asset provider needs to follow in order to configure these pre-processing steps, prior to their actual execution which is described in Section 2.3.

2.2.2.1 Mapping Configuration

The mapping configuration for a data check-in job 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 data comply, if applicable, and the main category in which the data refer to, as depicted in Figure 12.



SYNERGY Big Data Platform & AI Marketplace

NavBar Home Marketplace Data Checkin Data Analytics My Assets Admin About

Home / Data Checkin Jobs / Mapping

Mapping of the Data Check-in Job [TITLE] Cancel Save

STEP 1 Mapping Info STEP 2 Mapping Configuration STEP 3 Mapping Review and Confirmation

Standards
Basic information about the standards to which the data comply

- ☒ Standard 1 (vx.y)
Description of Standard 1
- ☐ Standard 2 (vZ)
Description of Standard 2
- ☐ Standard 3
Description of Standard 3
- ☐ Standard 4
Description of Standard 4

Category
Main concept to which the data refer to

- ☒ Category 1
Description of Category 1
- ☐ Category 2
Description of Category 2
- ☐ Category 3
Description of Category 3
- ☐ Category 4
Description of Category 4
- ☐ Category 5
Description of Category 5

Next

Figure 12: Mapping - Provide Mapping Info

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) in the Mapping Playground, as depicted in Figure 13. In case there is no corresponding mapping or the mapping provided is wrong, the data asset provider can search for a certain concept under the Data Model Details section, or even propose a new one in case the CIM has not already foreseen it. Once the concept to which a field/concept in the source data is to be mapped is identified, it needs to be copied to the Mapping Playground section using a drag n’ drop functionality to the appropriate field/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, e.g. for a numeric field, the applicable measurement unit needs to be provided while for a datetime field, the datetime format and the applicable time-zone. In the Mapping Playground section, the data asset provider is able to filter the view of the concepts by different categories (i.e., Predicted Mappings, Unidentified Mappings, Corrected Mappings, Invalid Mappings, and Selected Mappings). It needs to be noted that during this step, data type

casting according to the CIM provisions is also performed (e.g. from integer to double, from string to datetime, etc.).

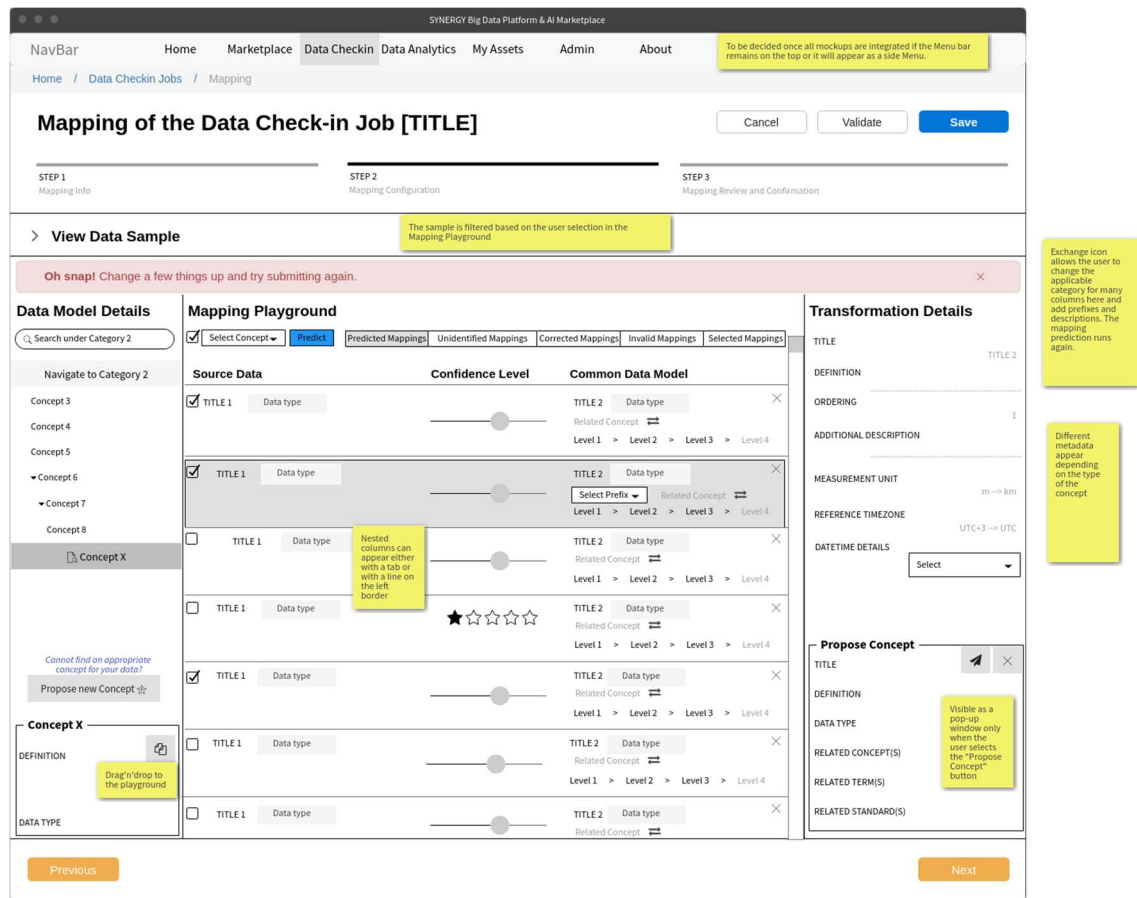
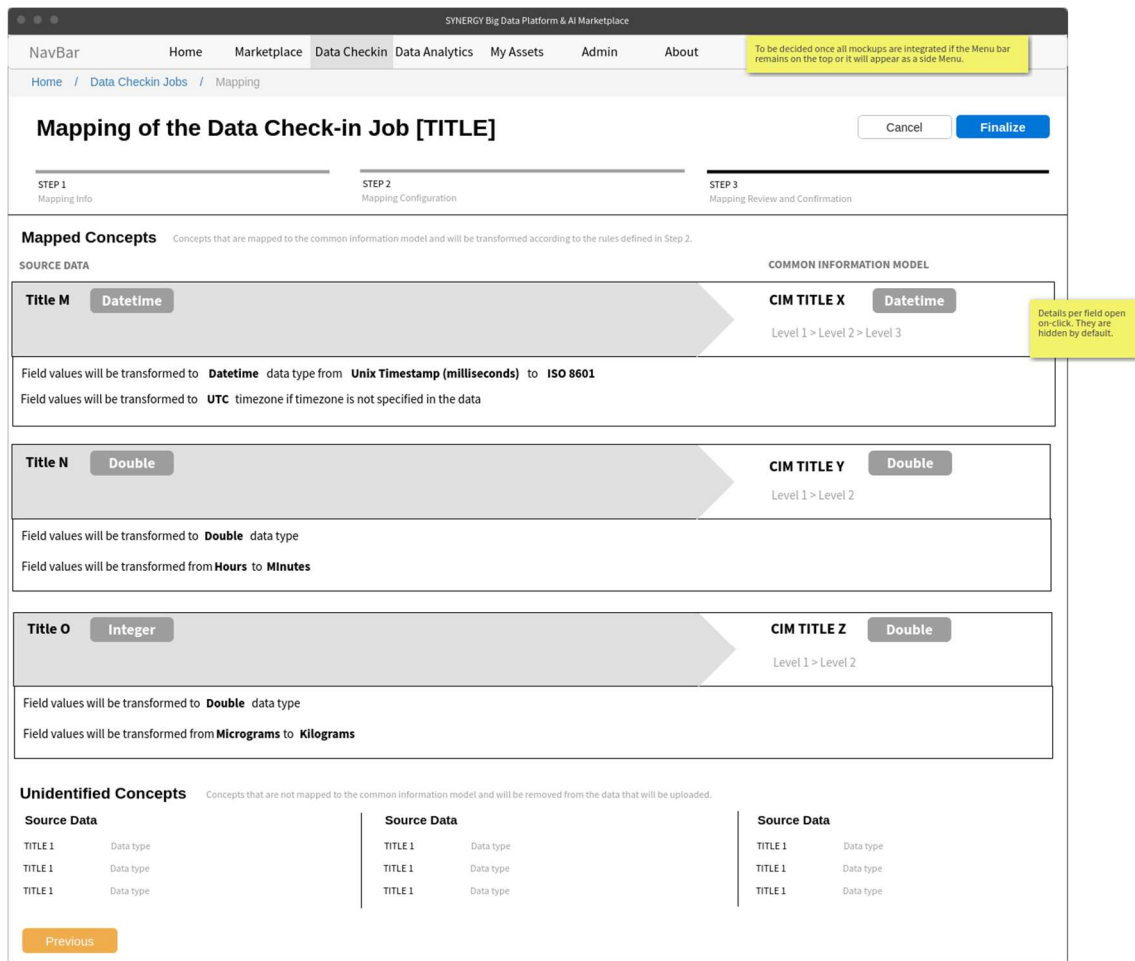


Figure 13: Mapping - Configure Mapping

Once the data asset provider is satisfied with the mapping configuration, by clicking the Next button, the Mapping Review and Confirmation page is loaded. During this step, the data asset provider is able to see a summary of the concepts that are mapped to the common information model and will be transformed according to the rules provided during the previous step, as it can be seen in Figure 14. In particular, the user can see an overview of the mapped concepts, as well as details related to each concept (per field) by clicking on the appropriate concept row. 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 common information model, and thus they will be excluded from the mapped data that will be uploaded and proceed to further processing.



SYNERGY Big Data Platform & AI Marketplace

NavBar Home Marketplace Data Checkin Data Analytics My Assets Admin About

Home / Data Checkin Jobs / Mapping

Mapping of the Data Check-in Job [TITLE] Cancel Finalize

STEP 1 Mapping Info STEP 2 Mapping Configuration STEP 3 Mapping Review and Confirmation

Mapped Concepts Concepts that are mapped to the common information model and will be transformed according to the rules defined in Step 2.

SOURCE DATA **COMMON INFORMATION MODEL**

Title M Datetime **CIM TITLE X** Datetime
Level 1 > Level 2 > Level 3

Field values will be transformed to **Datetime** data type from **Unix Timestamp (milliseconds)** to **ISO 8601**
Field values will be transformed to **UTC** timezone if timezone is not specified in the data

Title N Double **CIM TITLE Y** Double
Level 1 > Level 2

Field values will be transformed to **Double** data type
Field values will be transformed from **Hours** to **Minutes**

Title O Integer **CIM TITLE Z** Double
Level 1 > Level 2

Field values will be transformed to **Double** data type
Field values will be transformed from **Micrograms** to **Kilograms**

Unidentified Concepts Concepts that are not mapped to the common information model and will be removed from the data that will be uploaded.

Source Data **Source Data** **Source Data**

TITLE 1 Data type
TITLE 1 Data type
TITLE 1 Data type

TITLE 1 Data type
TITLE 1 Data type
TITLE 1 Data type

TITLE 1 Data type
TITLE 1 Data type
TITLE 1 Data type

Previous

Details per field open on-click. They are hidden by default.

To be decided once all mockups are integrated if the Menu bar remains on the top or it will appear as a side Menu.

Figure 14: Mapping - Finalize Mapping

2.2.2.2 Cleaning Configuration

If the data asset provider has selected the cleaning pre-processing step during the creation of the data check-in job, the cleaning workflow of the data check-in job that is responsible to manipulate the data ingested into the SYNERGY Platform, in such a way that they are accurate, complete and consistent, needs to be configured. This workflow involves several features offered by the Cleaning Service described in Section 4.1.5 of D2.6 “SYNERGY Framework Architecture including functional, technical and communication specifications v1”. 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. Note 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 15, the data asset provider may define different cleaning rules and constraints based on the constraint type (i.e., allowed value ranges, uniqueness constraints, mandatory constraints, regular expression patterns, and outliers identification). In addition, the data asset provider can define a rule (i.e., drop, replace with previous value, replace with mean value, or replace with default value) in case where an entry is considered as an outlier.

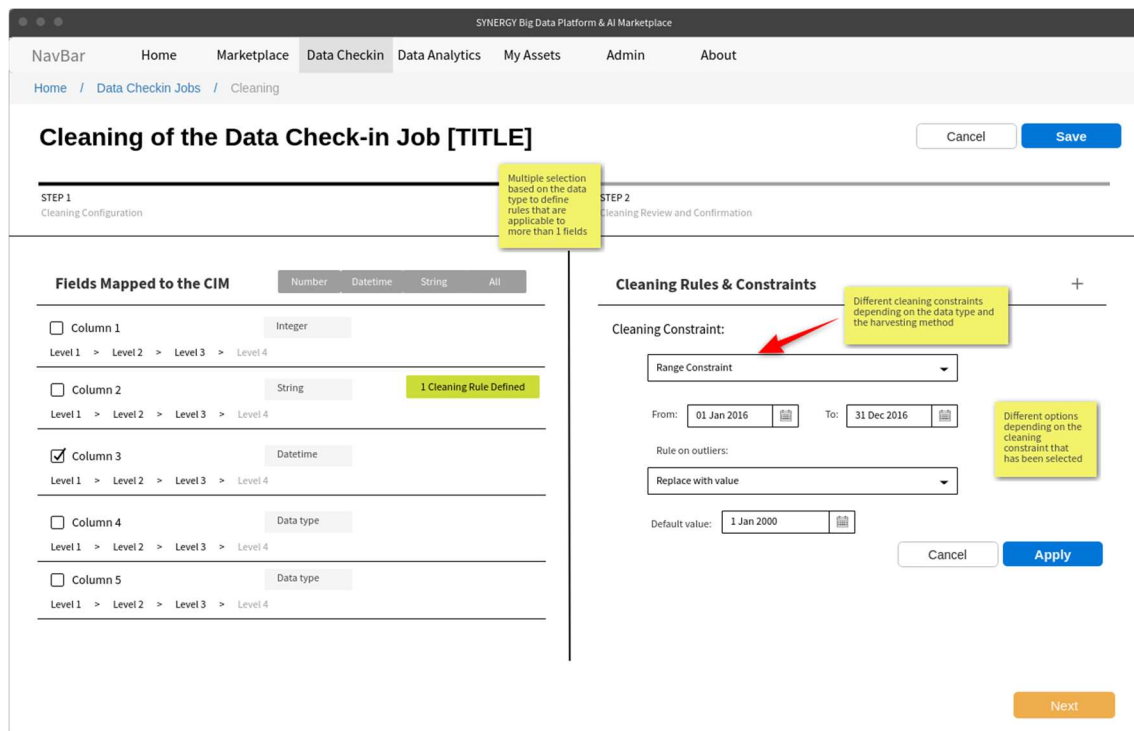


Figure 15: Cleaning - Configure Cleaning

Each time a cleaning rule is defined, the data asset provider can view what exactly will be applied in the data in simple language, as well as add a new rule, edit or remove an existing rule, or even change the order of the rules that are to be applied as shown in the right part of Figure 16.

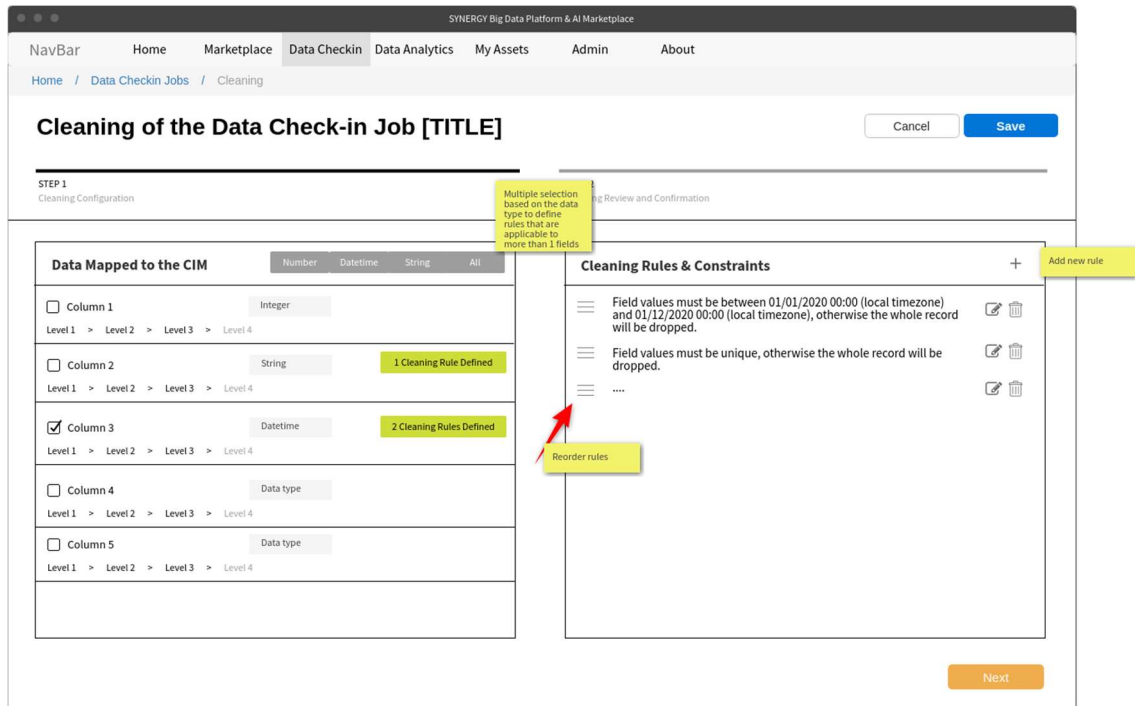


Figure 16: Cleaning - Configure Cleaning - View Rules

Once the data asset provider is satisfied with the cleaning configuration, by selecting the Next button, he/she is able to review the final cleaning rules that are defined for each field, as in Figure 17.

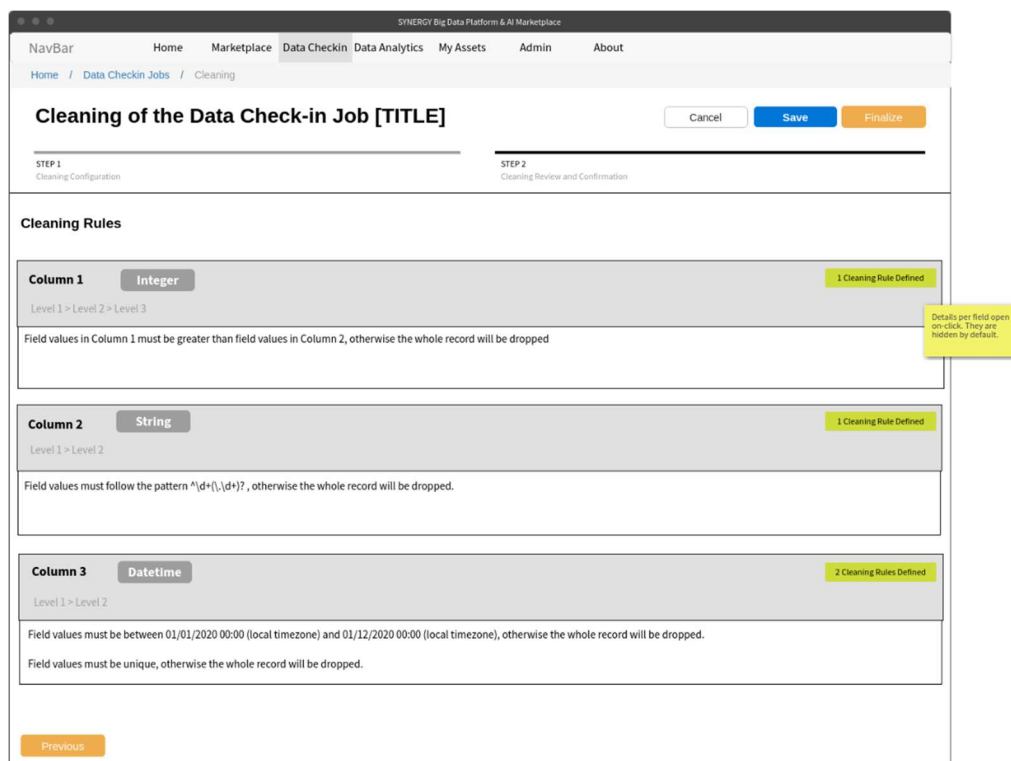
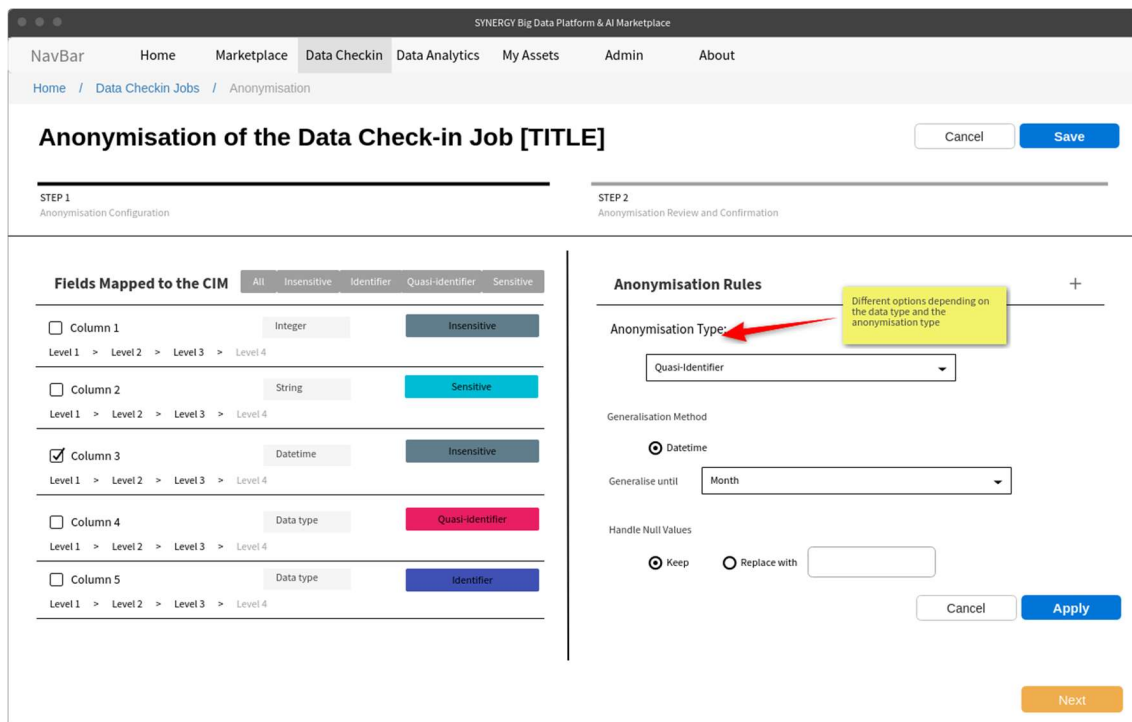


Figure 17: Cleaning - Finalize Cleaning

2.2.2.3 Anonymisation Configuration

Another optional pre-processing step of the data check-in job that is available to data asset providers is the Anonymisation task. 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 19. During this step, a data asset provider defines which of the fields that have been mapped to the CIM during the Mapping step are sensitive, quasi-identifiers and identifiers since by default all fields are considered as sensitive. 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).

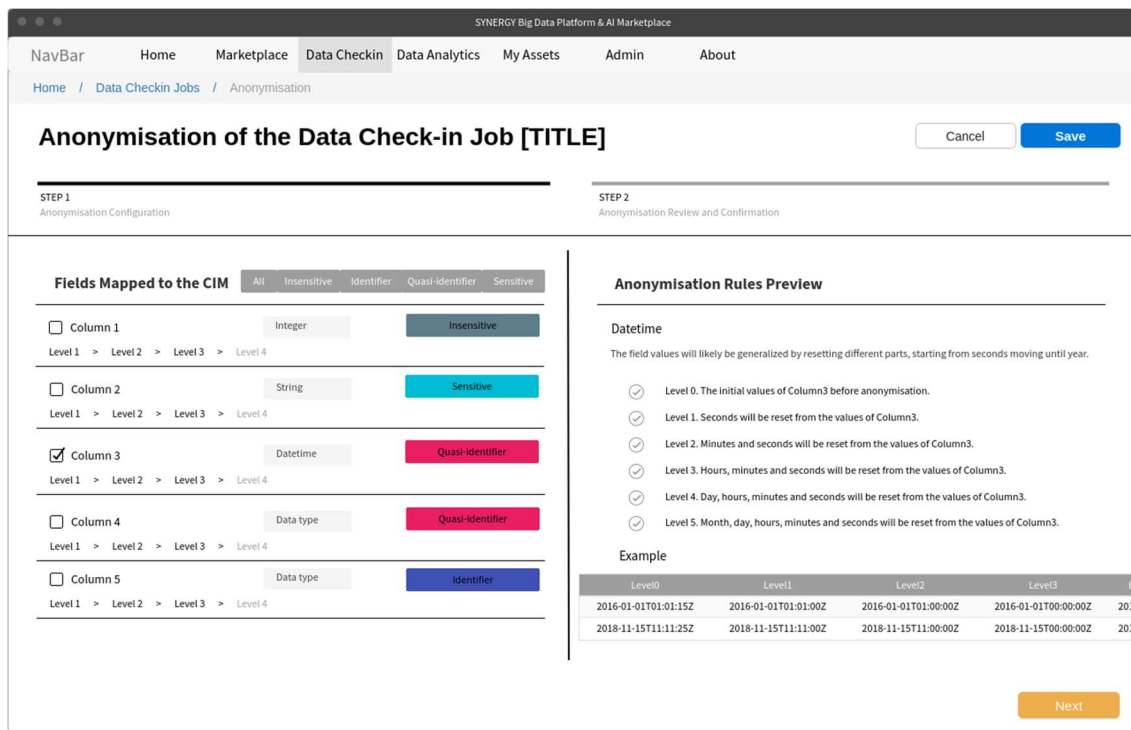


The screenshot displays the 'Anonymisation of the Data Check-in Job [TITLE]' configuration interface. It features a navigation bar at the top with links to Home, Marketplace, Data Checkin, Data Analytics, My Assets, Admin, and About. The main content area is divided into two steps: STEP 1 (Anonymisation Configuration) and STEP 2 (Anonymisation Review and Confirmation). In STEP 1, the 'Fields Mapped to the CIM' section lists five columns with their data types and selected anonymisation types. The 'Anonymisation Rules' section shows the configuration for a specific rule, including the anonymisation type, generalisation method, and null value handling. A red arrow points to the 'Anonymisation Type' dropdown menu, which is currently set to 'Quasi-Identifier'. A yellow callout box explains that different options are available depending on the data type and the anonymisation type. The 'Generalisation Method' is set to 'Datetime', and the 'Generalise until' is set to 'Month'. The 'Handle Null Values' section shows the 'Keep' option selected. The 'Apply' button is highlighted in blue.

Figure 19: Anonymisation – Configure Anonymisation

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 20. 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.



Anonymisation of the Data Check-in Job [TITLE]

Cancel Save

STEP 1
Anonymisation Configuration

Fields Mapped to the CIM

All Insensitive Identifier Quasi-identifier Sensitive

☐ Column 1 Integer Insensitive
Level 1 > Level 2 > Level 3 > Level 4

☐ Column 2 String Sensitive
Level 1 > Level 2 > Level 3 > Level 4

☒ Column 3 Datetime Quasi-identifier
Level 1 > Level 2 > Level 3 > Level 4

☐ Column 4 Data type Quasi-identifier
Level 1 > Level 2 > Level 3 > Level 4

☐ Column 5 Data type Identifier
Level 1 > Level 2 > Level 3 > Level 4

STEP 2
Anonymisation Review and Confirmation

Anonymisation Rules Preview

Datetime
The field values will likely be generalized by resetting different parts, starting from seconds moving until year.

- ☒ Level 0. The initial values of Column3 before anonymisation.
- ☒ Level 1. Seconds will be reset from the values of Column3.
- ☒ Level 2. Minutes and seconds will be reset from the values of Column3.
- ☒ Level 3. Hours, minutes and seconds will be reset from the values of Column3.
- ☒ Level 4. Day, hours, minutes and seconds will be reset from the values of Column3.
- ☒ Level 5. Month, day, hours, minutes and seconds will be reset from the values of Column3.

Example

Level0	Level1	Level2	Level3	
2016-01-01T01:01:15Z	2016-01-01T01:01:00Z	2016-01-01T01:00:00Z	2016-01-01T00:00:00Z	20:
2018-11-15T11:11:25Z	2018-11-15T11:11:00Z	2018-11-15T11:00:00Z	2018-11-15T00:00:00Z	20:

Next

Figure 20: Anonymisation - View Anonymisation Rules

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 21). The data asset provider also needs to 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.

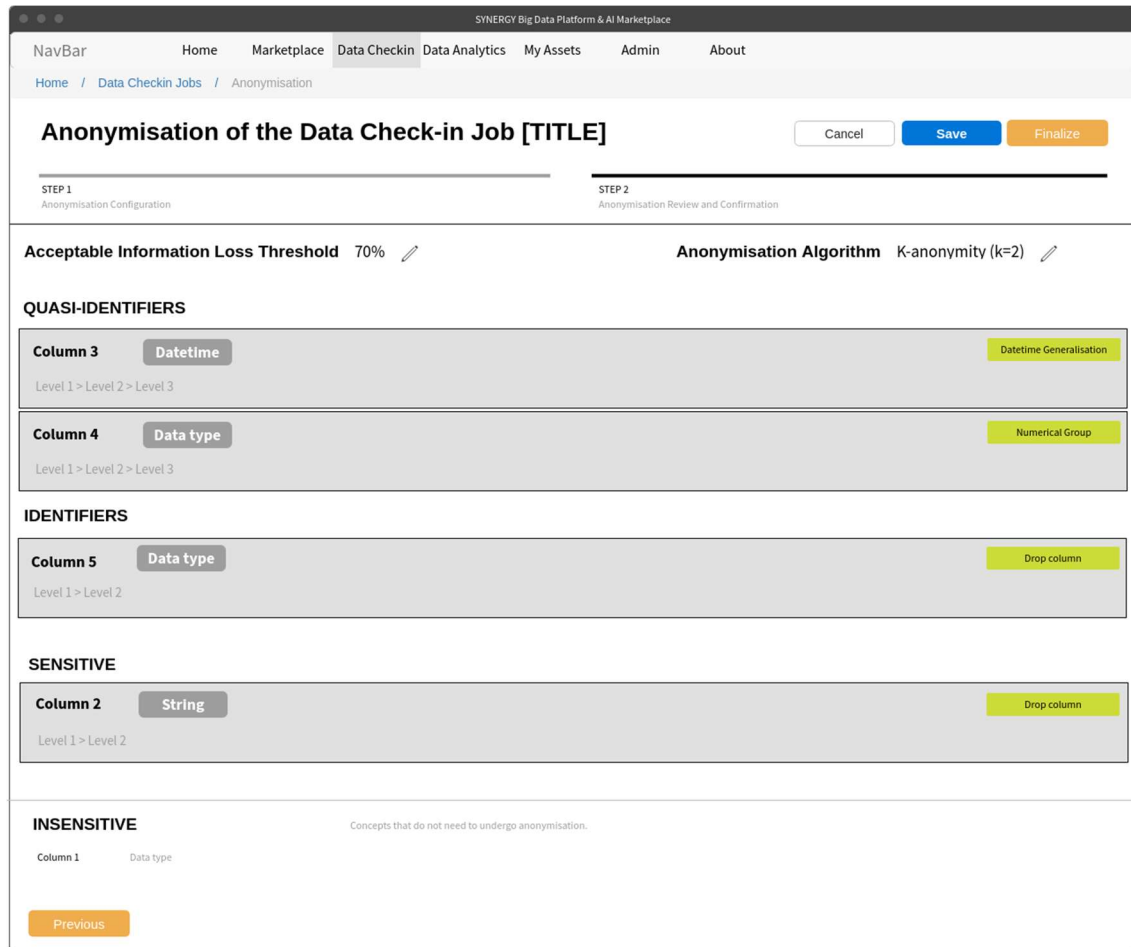


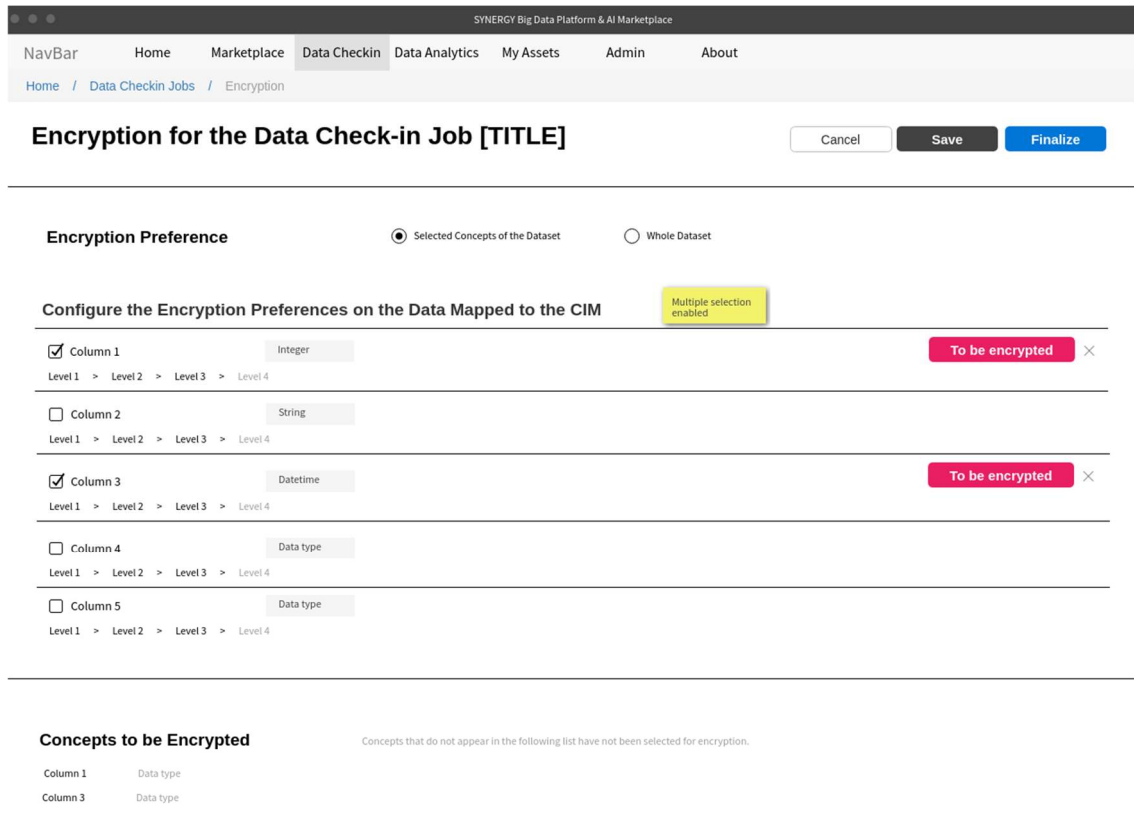
Figure 21: Anonymisation - Finalize Anonymisation

2.2.2.4 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 have installed locally) to eliminate the risk of unauthorized data access or leakage. As Section 4.2.2 of D2.6 describes, data asset providers are able to (optionally) set encryption parameters to be applied either on the whole dataset, or on a selected part of the dataset (i.e., the contents that correspond to specific dataset's fields/columns to be encrypted, whereas other fields/columns may remain unencrypted). 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 pre-processing steps (as in Figure 1).

As depicted in Figure 18, the data asset provider is able to either select the concepts of the dataset that are to be encrypted or select to encrypt the whole dataset. If the data asset provider

chose to encrypt a part of the dataset, the selected concepts will appear in the list Concepts to be Encrypted at the bottom part of the page.



The screenshot shows the 'Encryption for the Data Check-in Job [TITLE]' page in the SYNERGY Big Data Platform & AI Marketplace. The page has a navigation bar with links: NavBar, Home, Marketplace, Data Checkin (active), Data Analytics, My Assets, Admin, and About. Below the navigation bar is a breadcrumb trail: Home / Data Checkin Jobs / Encryption.

The main content area is titled 'Encryption for the Data Check-in Job [TITLE]' and includes buttons for 'Cancel', 'Save', and 'Finalize'. Below this is the 'Encryption Preference' section, which has two radio buttons: 'Selected Concepts of the Dataset' (selected) and 'Whole Dataset'.

The 'Configure the Encryption Preferences on the Data Mapped to the CIM' section is highlighted with a yellow box labeled 'Multiple selection enabled'. It contains a list of columns with checkboxes and data types:

- ☒ Column 1 (Integer) - To be encrypted
- ☐ Column 2 (String)
- ☒ Column 3 (Datetime) - To be encrypted
- ☐ Column 4 (Data type)
- ☐ Column 5 (Data type)

Each column has a breadcrumb trail: Level 1 > Level 2 > Level 3 > Level 4.

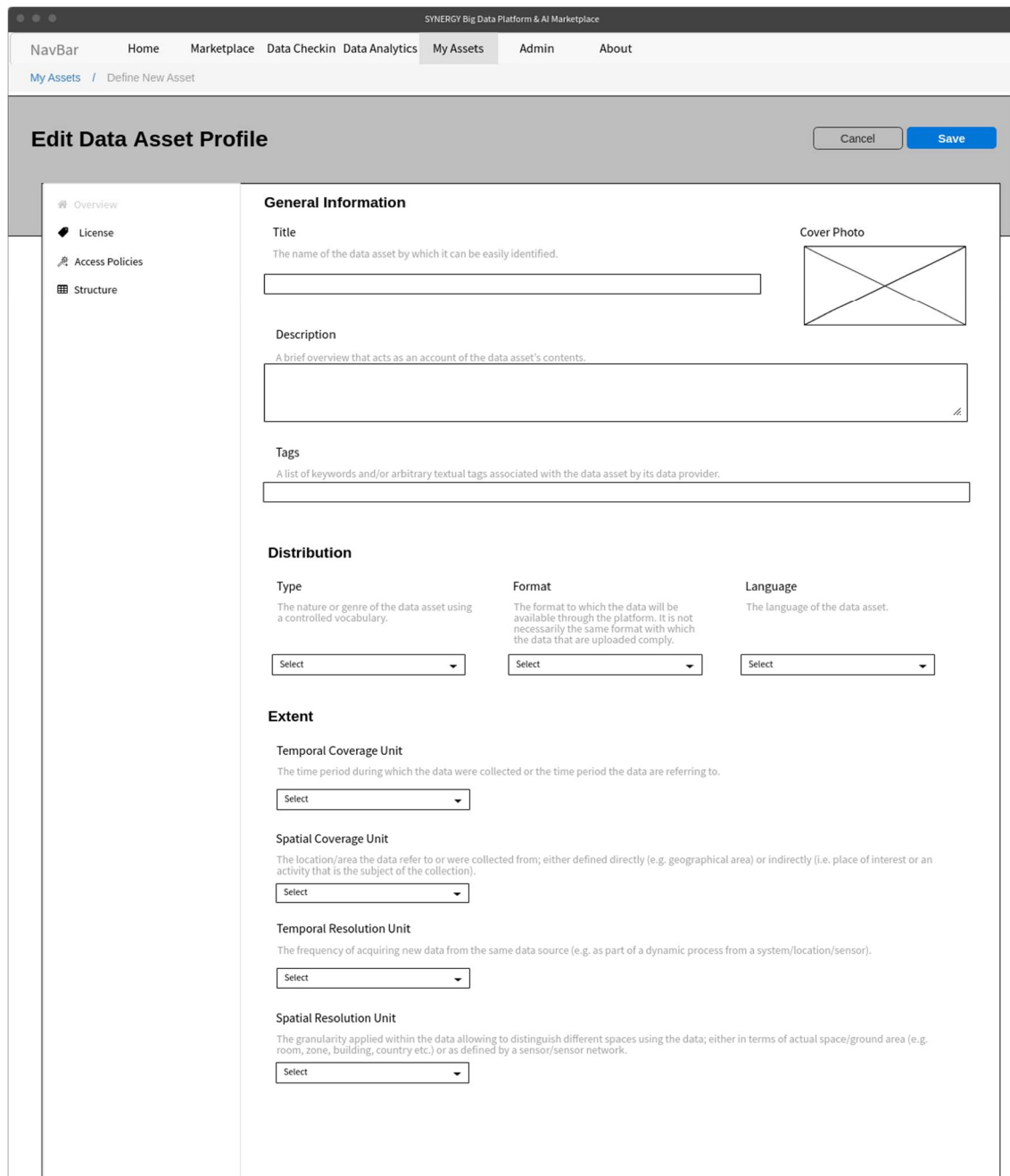
At the bottom, the 'Concepts to be Encrypted' section shows a list of concepts: Column 1 (Data type) and Column 3 (Data type). A note states: 'Concepts that do not appear in the following list have not been selected for encryption.'

Figure 18: Encryption - Configure Encryption

2.2.3 Define New Data Asset Profile

Since the outcome of the data check-in job is stored as a data asset in the SYNERGY Platform, its detailed profile needs to be defined by the data provider, according to the SYNERGY Metadata Schema that is presented in D4.1. A data asset provider thus needs to navigate to My Assets on the main navigation bar. As depicted in Figure 22, the data asset provider needs to provide a suitable title for the data asset, a brief description regarding the data asset's contents, and optionally a cover photo which is related to the data asset. In addition, the data asset provider needs to add tags (i.e., list of keywords, or arbitrary textual tags) that are associated with a particular data asset. The type which describes the nature/genre 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.



SYNERGY Big Data Platform & AI Marketplace

NavBar Home Marketplace Data Checkin Data Analytics **My Assets** Admin About

My Assets / Define New Asset

Edit Data Asset Profile

Cancel Save

General Information

Title
The name of the data asset by which it can be easily identified.

Description
A brief overview that acts as an account of the data asset's contents.

Tags
A list of keywords and/or arbitrary textual tags associated with the data asset by its data provider.

Cover Photo

Distribution

Type
The nature or genre of the data asset using a controlled vocabulary.

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.

Language
The language of the data asset.

Extent

Temporal Coverage Unit
The time period during which the data were collected or the time period the data are referring to.

Spatial Coverage Unit
The location/area the data refer to or were collected from; either defined directly (e.g. geographical area) or indirectly (i.e. place of interest or an activity that is the subject of the collection).

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).

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.

Overview

License

Access Policies

Structure

Figure 22: Define New Data Asset Profile - Overview

Then, the data asset provider needs to set the access 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 that: a) Allows other organisations to access the data asset if the access policies are satisfied and there is an active contract (Private Access); b) Denies the access of the data asset to external organisations as it is only intended for

use within the organisation that owns it (Confidential Access); and c) Allows the access of the data asset to any organisation without requiring any access policies to be satisfied or the existence of a data contract (Public Access). If the access level is set to Private Access, then the data asset provider needs to define the related access policies by setting the applicable strategy (Allow-all or Deny-all) and adding appropriate exceptions. Such exceptions can be added, edited, or removed as depicted in Figure 23. To add exceptions, a data asset provider needs to select a user property (i.e., organisation country, organisation type, etc.), a condition (i.e., equal, not equal, etc.), and the value of the property. In addition, the data asset provider can combine two or more exceptions by linking them with a logical operator (i.e. and, or).

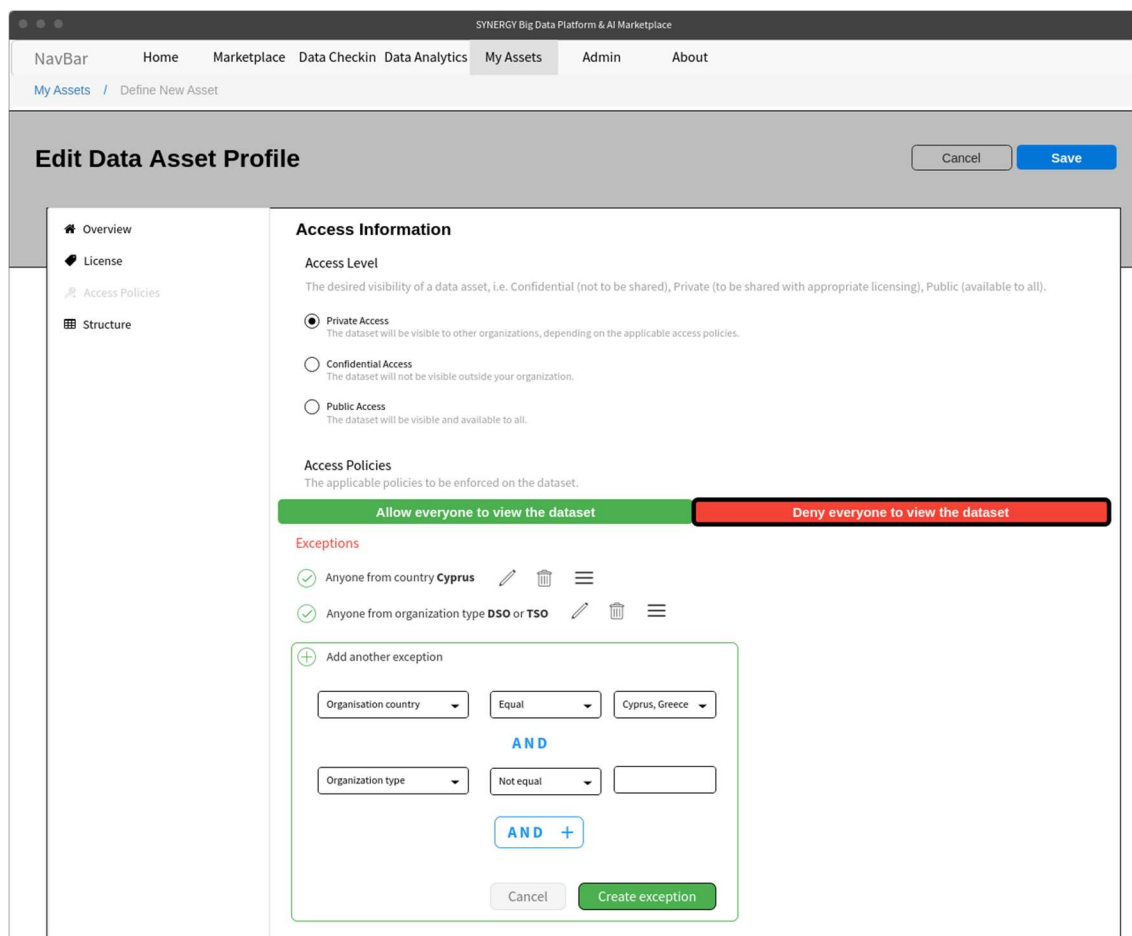
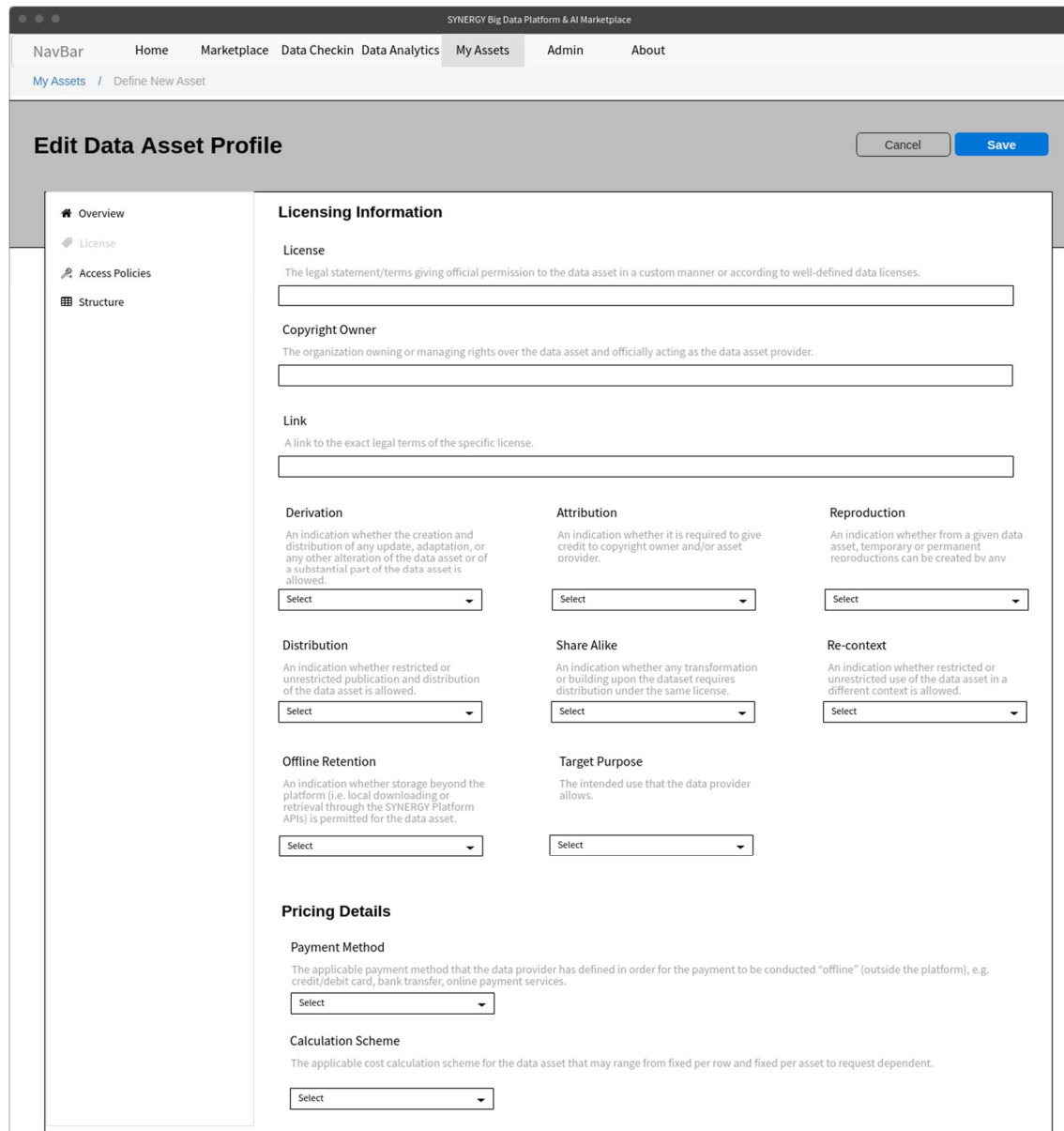


Figure 23: Define New Data Asset Profile - Access Policies

Following the definition of access policies, the data asset provider needs to insert the licensing information, as depicted in Figure 24. In this sub-menu, the data asset provider needs to define the license (ranging from open data licenses to a custom license), the Copyright Owner (that is the organisation owning or managing rights over the data asset in case it is not the organisation that has uploaded the data to the SYNERGY Platform) and a link to the license. Depending on

the license selection, different aspects need to be clarified, i.e. Derivation, Attribution, Reproduction, Distribution, Share Alike, Re-context, Offline Retention, and Target Purpose. Then, the data asset provider needs to select the payment method and the calculation scheme, representing the applicable method of payment to be conducted outside the platform (e.g., by credit/debit card, bank transfer, or using online payment services), and the cost calculation scheme for the data asset (e.g., fixed per row, fixed per asset) and the related cost, respectively.



Edit Data Asset Profile

Licensing Information

License
The legal statement/terms giving official permission to the data asset in a custom manner or according to well-defined data licenses.

Copyright Owner
The organization owning or managing rights over the data asset and officially acting as the data asset provider.

Link
A link to the exact legal terms of the specific license.

Derivation
An indication whether the creation and distribution of any update, adaptation, or any other alteration of the data asset or of a substantial part of the data asset is allowed.

Attribution
An indication whether it is required to give credit to copyright owner and/or asset provider.

Reproduction
An indication whether from a given data asset, temporary or permanent reproductions can be created by any

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 upon the dataset requires distribution under the same license.

Re-context
An indication whether restricted or unrestricted use of the data asset in a different context is allowed.

Offline Retention
An indication whether storage beyond the platform (i.e. local downloading or retrieval through the SYNERGY Platform APIs) is permitted for the data asset.

Target Purpose
The intended use that the data provider allows.

Pricing Details

Payment Method
The applicable payment method that the data provider has defined in order for the payment to be conducted "offline" (outside the platform), e.g. credit/debit card, bank transfer, online payment services.

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.

Figure 24: Define New Data Asset Profile - Licensing Information

The data asset profile is finalized by reviewing the data structure of the data asset provided, as depicted in Figure 25. In particular, the data asset provider may view the fields to which

transformation, cleaning, anonymisation and encryptions rules were applied. More details regarding each field can be found by clicking on each field accordingly.

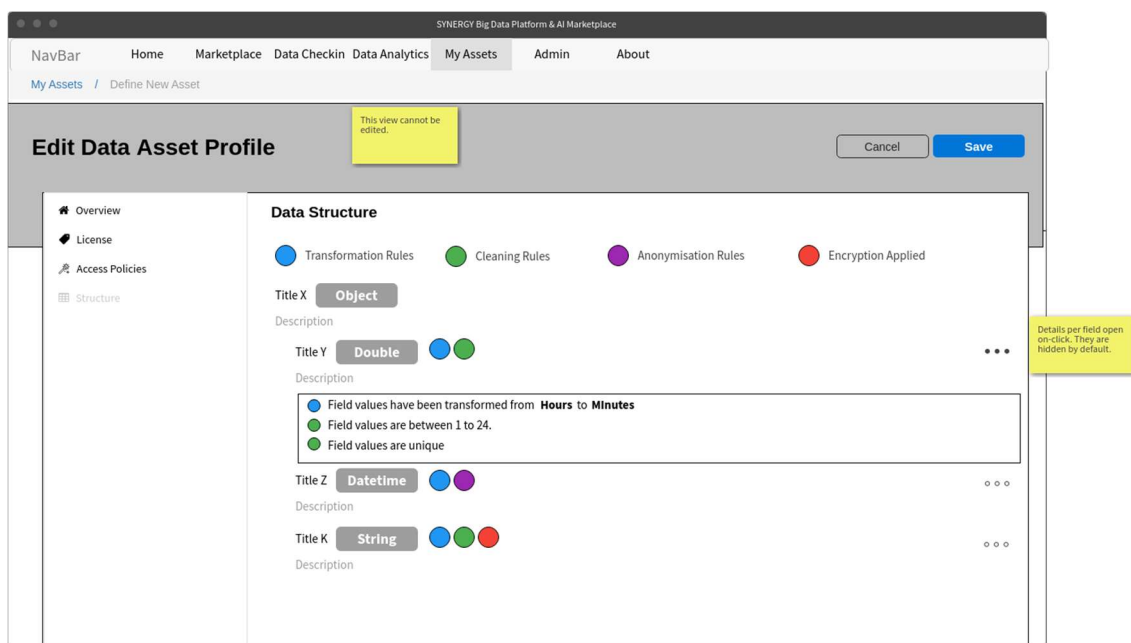


Figure 25: Define New Data Asset Profile - Data Structure

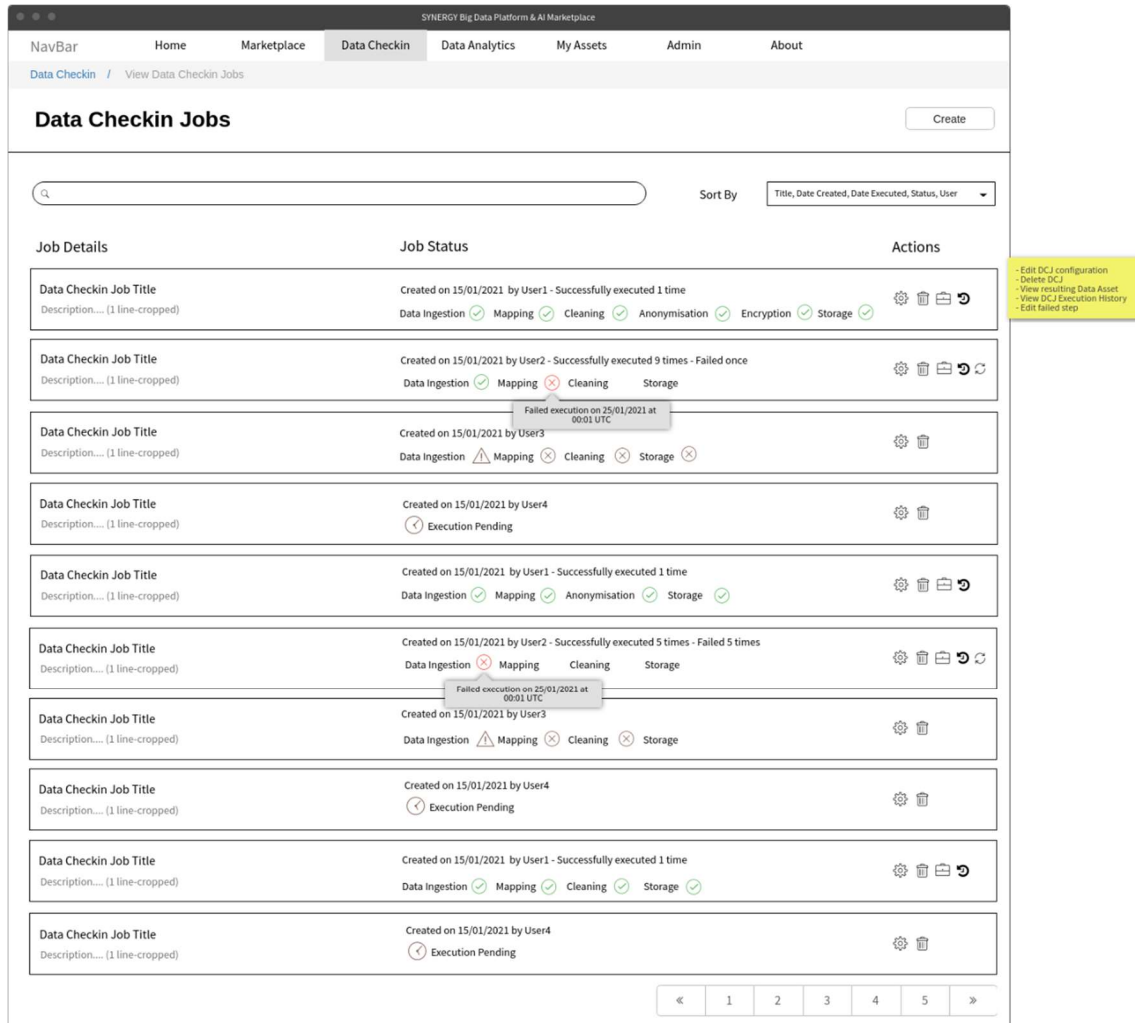
Obviously, the data asset provider is able to edit the metadata (with the exception of the Data Structure) at any point through the “My Assets” View (described in section 3.4).

2.3 Execute a Data Check-in Job

2.3.1 View all Data Check-in Jobs

Once the data asset provider has completed the configuration for the data asset check-in job, he/she may view the summary list of all data check-in jobs as shown in Figure 26. 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. Data check-in job search and sorting is also available at the top of the page. 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 for each job. 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., Data Ingestion, Mapping, Cleaning, Anonymisation, Encryption, Storage) 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. Next to the Job Status column, the available Actions are depicted depending on the current status of a data check-in job, namely edit the configuration, delete a data check-in job, view the resulting data asset, view the data check-in job execution history, and edit the configuration for a failed step.



Job Details	Job Status	Actions
Data Checkin Job Title Description.... (1 line-cropped)	Created on 15/01/2021 by User1 - Successfully executed 1 time Data Ingestion ✓ Mapping ✓ Cleaning ✓ Anonymisation ✓ Encryption ✓ Storage ✓	⚙️ 🗑️ 📄 🕒
Data Checkin Job Title Description.... (1 line-cropped)	Created on 15/01/2021 by User2 - Successfully executed 9 times - Failed once Data Ingestion ✓ Mapping ✗ Cleaning ✗ Storage ✗	⚙️ 🗑️ 📄 🕒
Data Checkin Job Title Description.... (1 line-cropped)	Created on 15/01/2021 by User3 Data Ingestion ⚠️ Mapping ✗ Cleaning ✗ Storage ✗	⚙️ 🗑️
Data Checkin Job Title Description.... (1 line-cropped)	Created on 15/01/2021 by User4 ⌚ Execution Pending	⚙️ 🗑️
Data Checkin Job Title Description.... (1 line-cropped)	Created on 15/01/2021 by User1 - Successfully executed 1 time Data Ingestion ✓ Mapping ✓ Anonymisation ✓ Storage ✓	⚙️ 🗑️ 📄 🕒
Data Checkin Job Title Description.... (1 line-cropped)	Created on 15/01/2021 by User2 - Successfully executed 5 times - Failed 5 times Data Ingestion ✗ Mapping ✗ Cleaning ✗ Storage ✗	⚙️ 🗑️ 📄 🕒
Data Checkin Job Title Description.... (1 line-cropped)	Created on 15/01/2021 by User3 Data Ingestion ⚠️ Mapping ✗ Cleaning ✗ Storage ✗	⚙️ 🗑️
Data Checkin Job Title Description.... (1 line-cropped)	Created on 15/01/2021 by User4 ⌚ Execution Pending	⚙️ 🗑️
Data Checkin Job Title Description.... (1 line-cropped)	Created on 15/01/2021 by User1 - Successfully executed 1 time Data Ingestion ✓ Mapping ✓ Cleaning ✓ Storage ✓	⚙️ 🗑️ 📄 🕒
Data Checkin Job Title Description.... (1 line-cropped)	Created on 15/01/2021 by User4 ⌚ Execution Pending	⚙️ 🗑️

Figure 26: View All Data Check-in Jobs

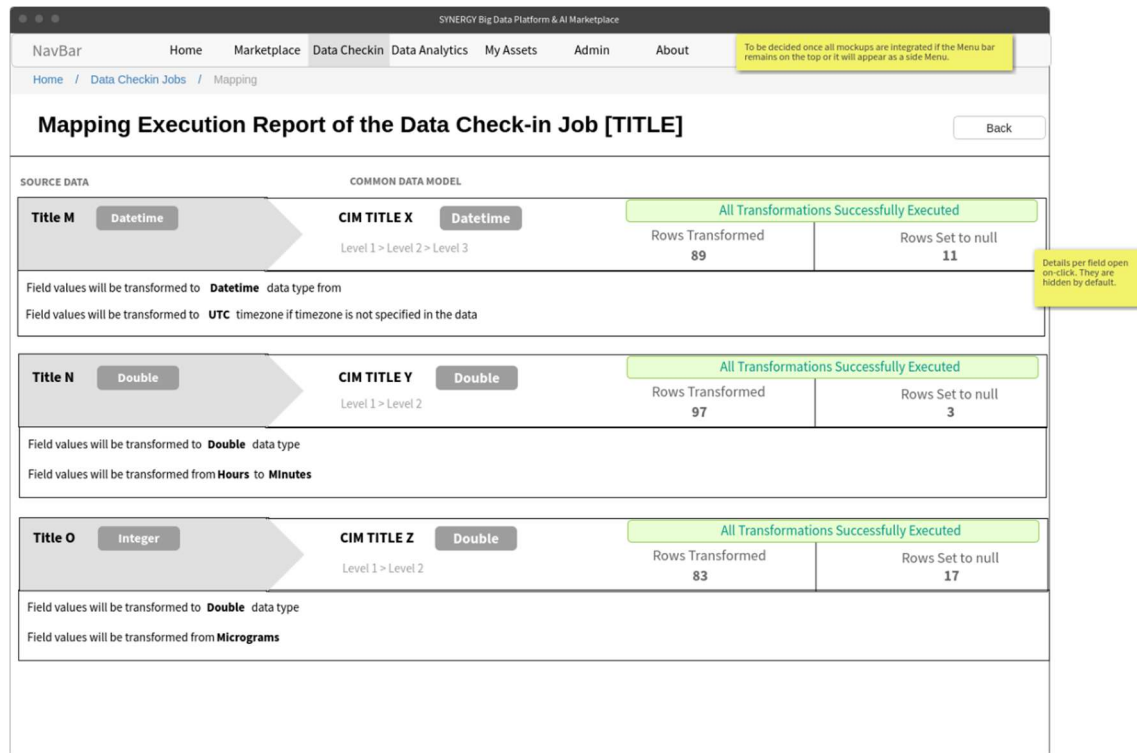
2.3.2 Pre-processing Rules Execution

As mentioned already, the pre-processing steps configuration (at design time) has a distinct separation with 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 pre-processing steps (i.e., Mapping, Cleaning, Anonymisation, and Encryption) can be only executed if each pre-processing configuration has been finalised.

2.3.2.1 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 27.



SOURCE DATA	COMMON DATA MODEL	TRANSFORMATIONS
Title M Datetime Field values will be transformed to Datetime data type from Field values will be transformed to UTC timezone if timezone is not specified in the data	CIM TITLE X Datetime Level 1 > Level 2 > Level 3	All Transformations Successfully Executed Rows Transformed: 89 Rows Set to null: 11
Title N Double Field values will be transformed to Double data type Field values will be transformed from Hours to Minutes	CIM TITLE Y Double Level 1 > Level 2	All Transformations Successfully Executed Rows Transformed: 97 Rows Set to null: 3
Title O Integer Field values will be transformed to Double data type Field values will be transformed from Micrograms	CIM TITLE Z Double Level 1 > Level 2	All Transformations Successfully Executed Rows Transformed: 83 Rows Set to null: 17

Figure 27: Mapping - Mapping Execution Report – Successful

Otherwise, if the Mapping step failed (e.g. there was a mistake in the transformation rules defined), the data asset provider can view in which field there was a problem, in order to revise the transformations that have been defined (as shown in Figure 28). It needs to be noted that the Mapping step may have run successfully for n times and failed in the n+1 time, which is why exact number of rows are provided per execution.

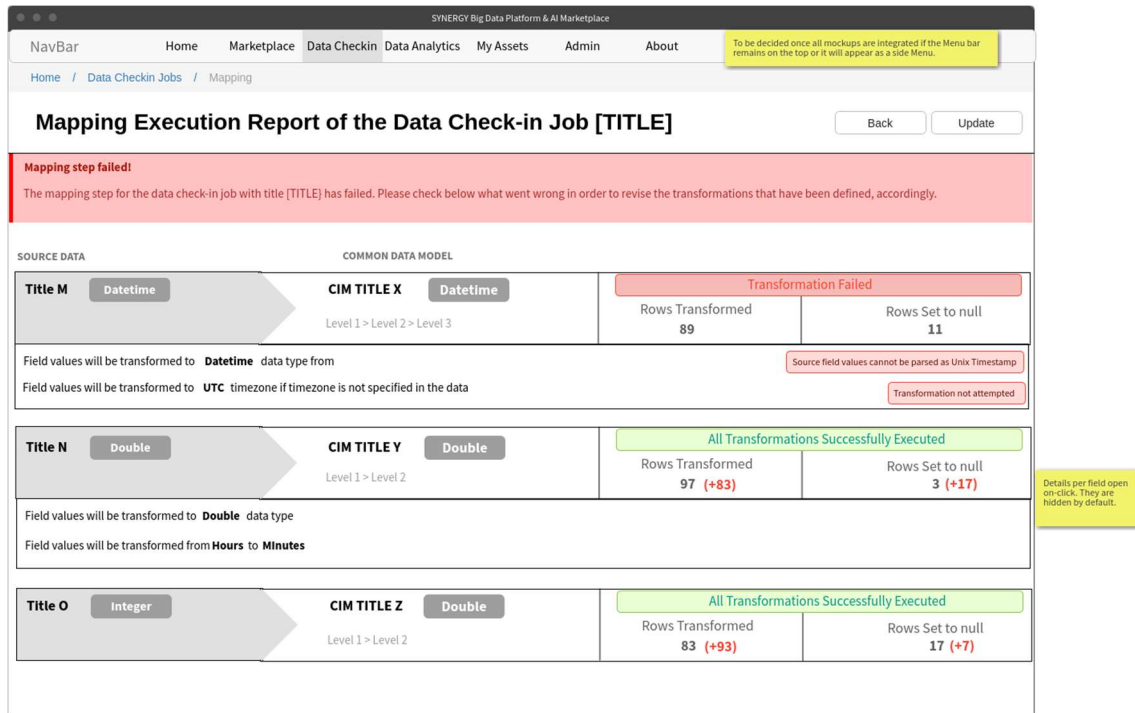


Figure 28: Mapping - Mapping Execution Report – Failed

2.3.2.2 Cleaning Execution

If the Cleaning step has been successfully executed, the data asset provider may view what cleaning constraints were met and resulted into transformations per field in the source data, as shown in Figure 29.

Otherwise, if the Cleaning step failed (e.g. there was a mistake when defining the cleaning rules, or if any of the field values of a column do not follow the defined pattern), the data asset provider can view in which field there was a problem, in order to revise the cleaning rules and constraints that have been defined (as shown in Figure 30).

Cleaning Report of the Data Check-in Job [TITLE]	
<div> <div>Home</div> <div>Marketplace</div> <div>Data Checkin</div> <div>Data Analytics</div> <div>My Assets</div> <div>Admin</div> <div>About</div> </div> <div>Home / Data Checkin Jobs / Cleaning</div>	
<div> <div>1 Cleaning Rule Successfully Executed</div> <div>Dropped Rows/Records</div> <div>11</div> </div>	
<div> <div>Field values in Column 1 must be greater than field values in Column 2, otherwise the whole record will be dropped</div> </div>	
<div> <div>1 Cleaning Rule Successfully Executed</div> <div>Dropped Rows/Records</div> <div>4</div> </div>	
<div> <div>Field values must follow the pattern <code>^\d+(\.\d+)?</code>, otherwise the whole record will be dropped.</div> </div>	
<div> <div>2 Cleaning Rules Successfully Executed</div> <div>Dropped Rows/Records</div> <div>30</div> </div>	
<div> <div>Field values must be between 01/01/2020 00:00 (local timezone) and 01/12/2020 00:00 (local timezone), otherwise the whole record will be dropped.</div> <div>Field values must be unique, otherwise the whole record will be dropped.</div> </div>	

Figure 29: Cleaning - Cleaning Report – Successful

Cleaning Report of the Data Check-in Job [TITLE]	
<div> <div>Home</div> <div>Marketplace</div> <div>Data Checkin</div> <div>Data Analytics</div> <div>My Assets</div> <div>Admin</div> <div>About</div> </div> <div>Home / Data Checkin Jobs / Cleaning</div>	
<div> <div>1 Cleaning Rule Successfully Executed</div> <div>Dropped Rows/Records</div> <div>11</div> </div>	
<div> <div>Field values in Column 1 must be greater than field values in Column 2, otherwise the whole record will be dropped</div> </div>	
<div> <div>1 Cleaning Rule Failed</div> <div>Dropped Rows/Records</div> <div>None of the field values follows the defined pattern.</div> </div>	
<div> <div>Field values must follow the pattern <code>^\d+(\.\d+)?</code>, otherwise the whole record will be dropped.</div> </div>	
<div> <div>2 Cleaning Rules Successfully Executed</div> <div>Dropped Rows/Records</div> <div>30</div> </div>	
<div> <div>Field values must be between 01/01/2020 00:00 (local timezone) and 01/12/2020 00:00 (local timezone), otherwise the whole record will be dropped.</div> <div>Field values must be unique, otherwise the whole record will be dropped.</div> </div>	

Figure 30: Cleaning - Cleaning Report – Failed

2.3.2.3 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 successful as Figure 31 depicts.

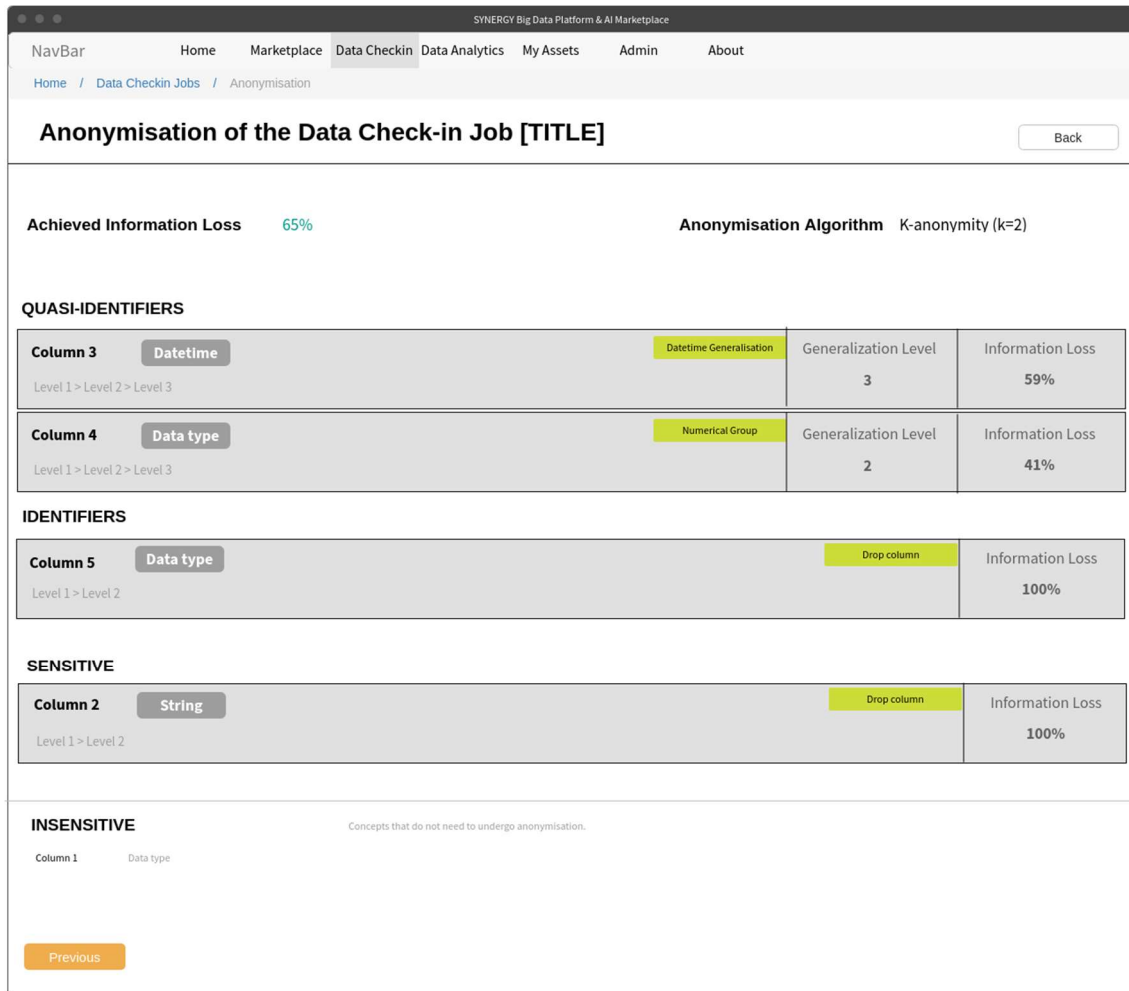


Figure 31: Anonymisation - Anonymisation Report – Successful

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 32 suggests.

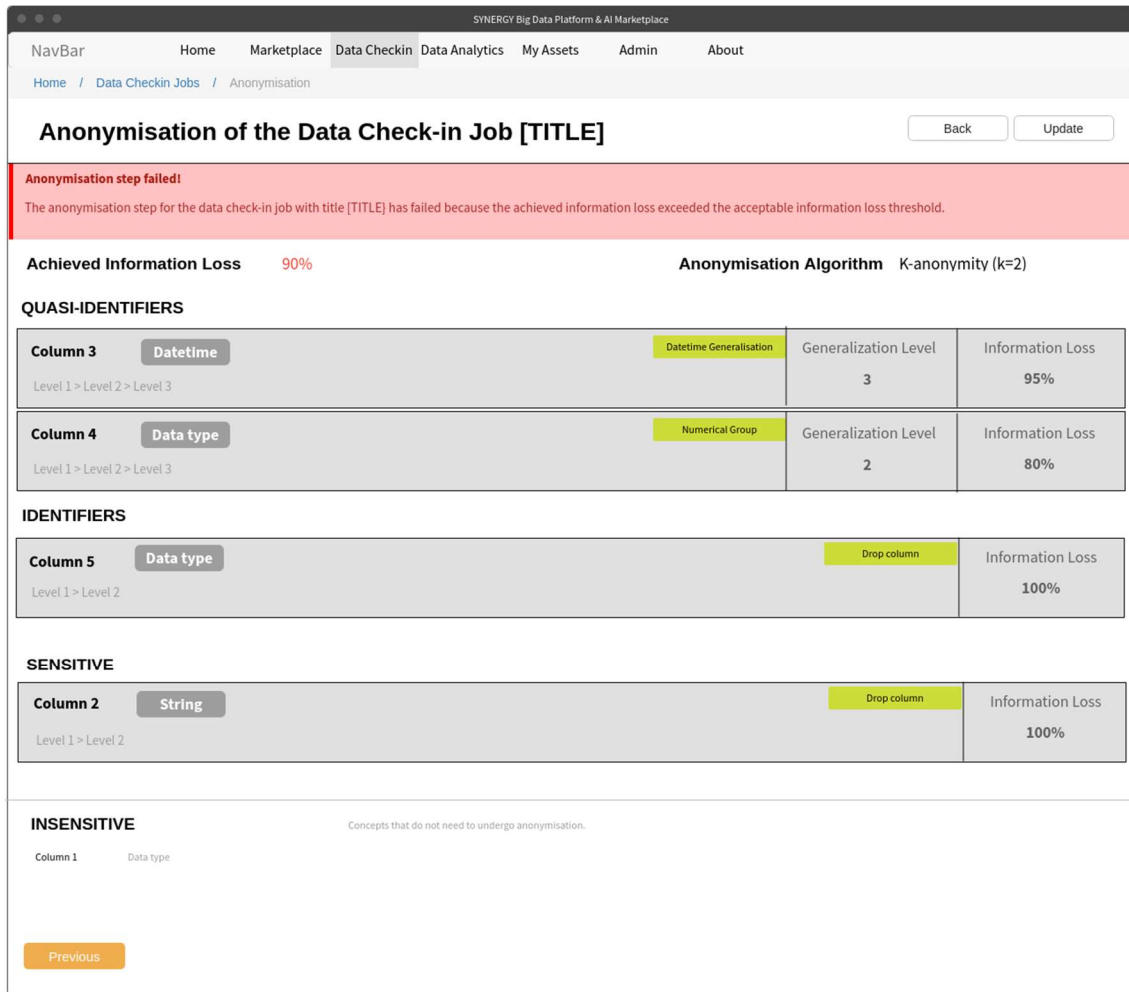


Figure 32: Anonymisation - Anonymisation Report - Failed

2.3.2.4 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 (either to the selected fields or to the whole dataset, according to the configuration provided), as depicted in Figure 33.

SYNERGY Big Data Platform & AI Marketplace

NavBar
Home
Marketplace
Data Checkin
Data Analytics
My Assets
Admin
About

Home / Data Checkin Jobs / Encryption

Encryption Report of the Data Check-in Job [TITLE]
Back

Encryption Preferences

<div> Column 1 Integer </div> <div> Level 1 > Level 2 > Level 3 </div>	<div> Successfully Encrypted </div>
<div> Column 2 String </div> <div> Level 1 > Level 2 </div>	<div> No encryption applied </div>
<div> Column 3 Datetime </div> <div> Level 1 > Level 2 </div>	<div> Successfully Encrypted </div>
<div> Column 4 String </div> <div> Level 1 > Level 2 </div>	<div> No encryption applied </div>
<div> Column 5 Datetime </div> <div> Level 1 > Level 2 </div>	<div> No encryption applied </div>

Figure 33: Encryption - Finalize Encryption

3 Data Search and Acquisition User Journey

The SYNERGY Platform allows data asset consumers to search and acquire existing data assets that belong to other electricity data value chain stakeholders in the SYNERGY Data & AI Marketplace. In order to acquire a data asset, the data asset provider and consumer need to reach an agreement on the data sharing terms that is prepared based on the contract preparation process followed in the SYNERGY Platform. Once a data asset is acquired through an active contract, it is available for use by its consumer 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.

3.1 Navigate to the SYNERGY Marketplace

Data asset consumers access the SYNERGY Data & AI Marketplace where they may navigate all the data assets that they are eligible to view (upon satisfying the access policies set by their respective providers). As depicted in Figure 34, data asset consumers view highlight info regarding each data asset such as the type of data asset (i.e., dataset, algorithm, analytics results), 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. The SYNERGY Marketplace gives the potential to data asset consumers to search for a particular data asset, pin data assets to favourites, add data assets to cart for acquisition, and to view the status of the contracts created for a particular data asset.

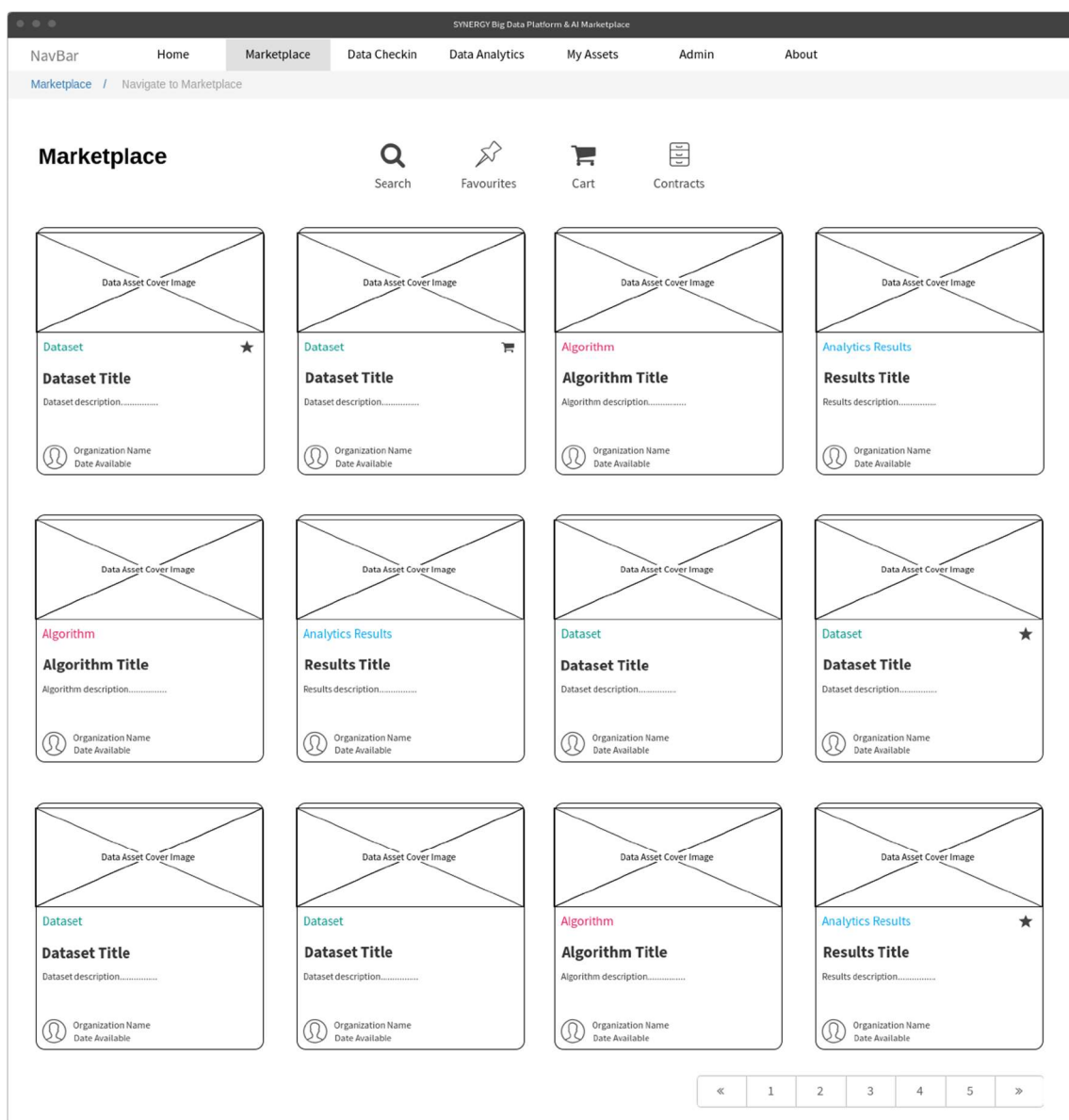


Figure 34: SYNERGY Data & AI Marketplace

When a data asset consumer selects a data asset from the Marketplace, an overview of this data asset will appear in a slideover, in the right side of the page. The overview of the data asset includes additional information related to the data asset such as its volume, accessibility, temporal coverage, license, date created, and the date of last update of that data asset, as shown in Figure 35. The data asset consumer may select to view more details for the specific data asset, mark it as favourite or add it to the cart (that belongs to the specific organisation).

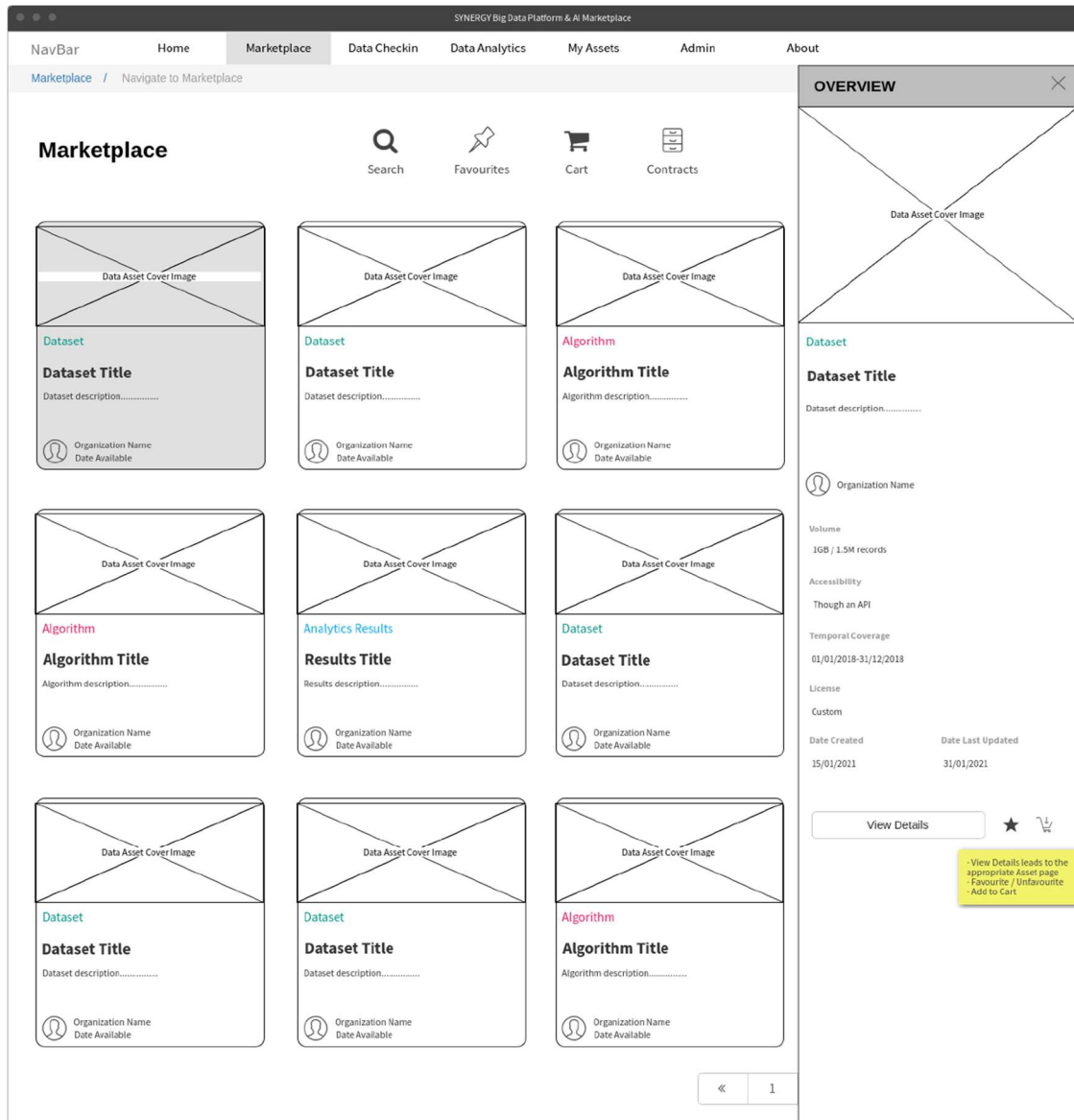
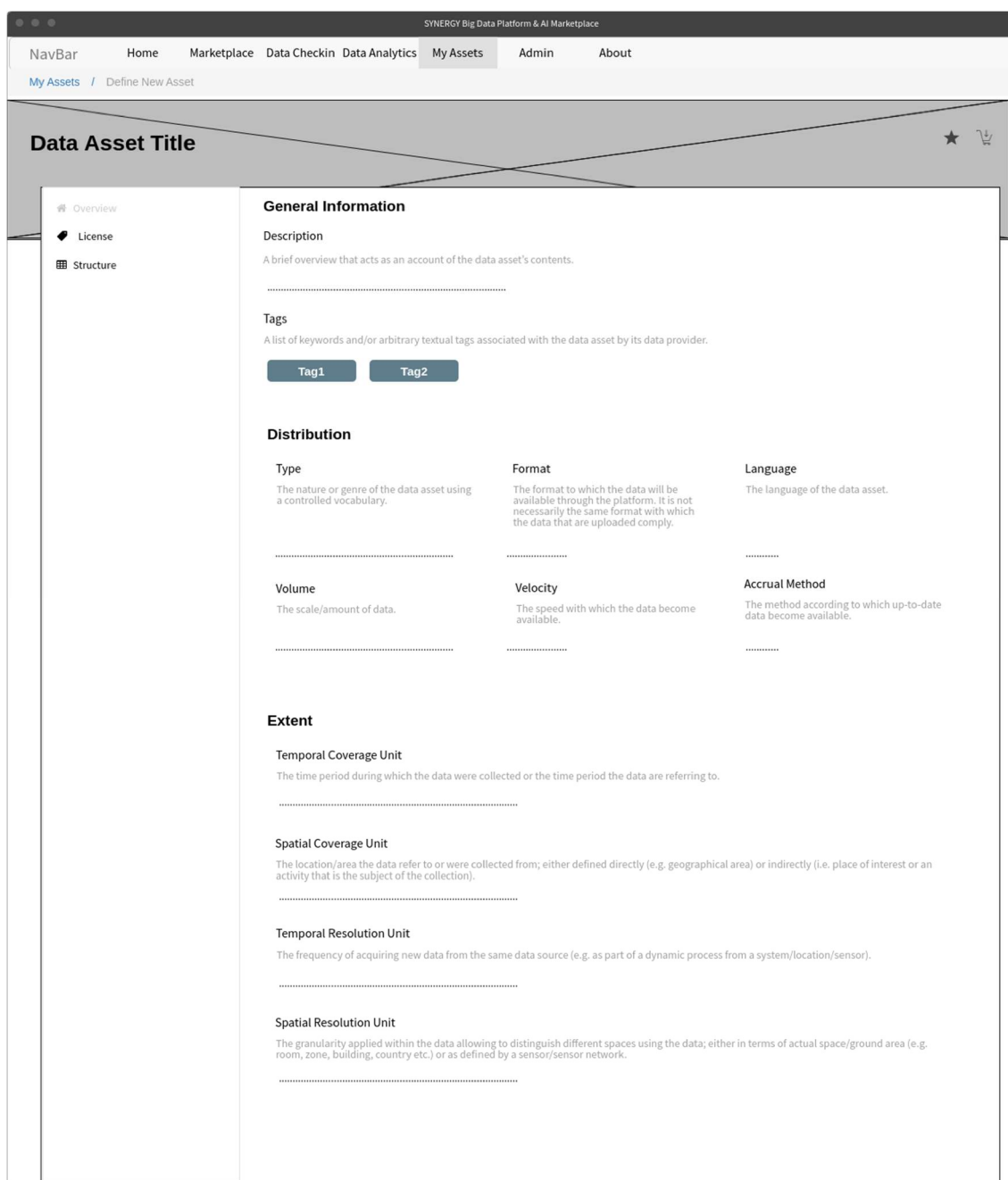


Figure 35: SYNERGY Data & AI Marketplace – Quick Glimpse at a Data Asset

If a data asset consumer has opted to view more details for a data asset, the corresponding Asset profile page appears and includes more insights on the data asset (i.e., overview, license information, data structure), as Figure 36 depicts.



SYNERGY Big Data Platform & AI Marketplace

NavBar Home Marketplace Data Checkin Data Analytics My Assets Admin About

My Assets / Define New Asset

Data Asset Title

Overview License Structure

General Information

Description
A brief overview that acts as an account of the data asset's contents.

Tags
A list of keywords and/or arbitrary textual tags associated with the data asset by its data provider.

Tag1 Tag2

Distribution

Type The nature or genre of the data asset using a controlled vocabulary.	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.	Language The language of the data asset.
Volume The scale/amount of data.	Velocity The speed with which the data become available.	Accrual Method The method according to which up-to-date data become available.

Extent

Temporal Coverage Unit
The time period during which the data were collected or the time period the data are referring to.

Spatial Coverage Unit
The location/area the data refer to or were collected from; either defined directly (e.g. geographical area) or indirectly (i.e. place of interest or an activity that is the subject of the collection).

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).

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.

Figure 36: SYNERGY Data & AI Marketplace – Data Asset Profile

If a data asset provider has added a data asset (of any type) to the favourites list, the star icon will change to dark grey colour, indicating that the selected data asset will be included in the list of favourite data assets. At any moment, the data asset provider may resort to the list of favourite data assets, by clicking the Favourites button in the Marketplace page, as shown in Figure 37. The Favourites List page provides a tabular view for easy access to the data assets' information that the data asset provider has pinned as favourites.

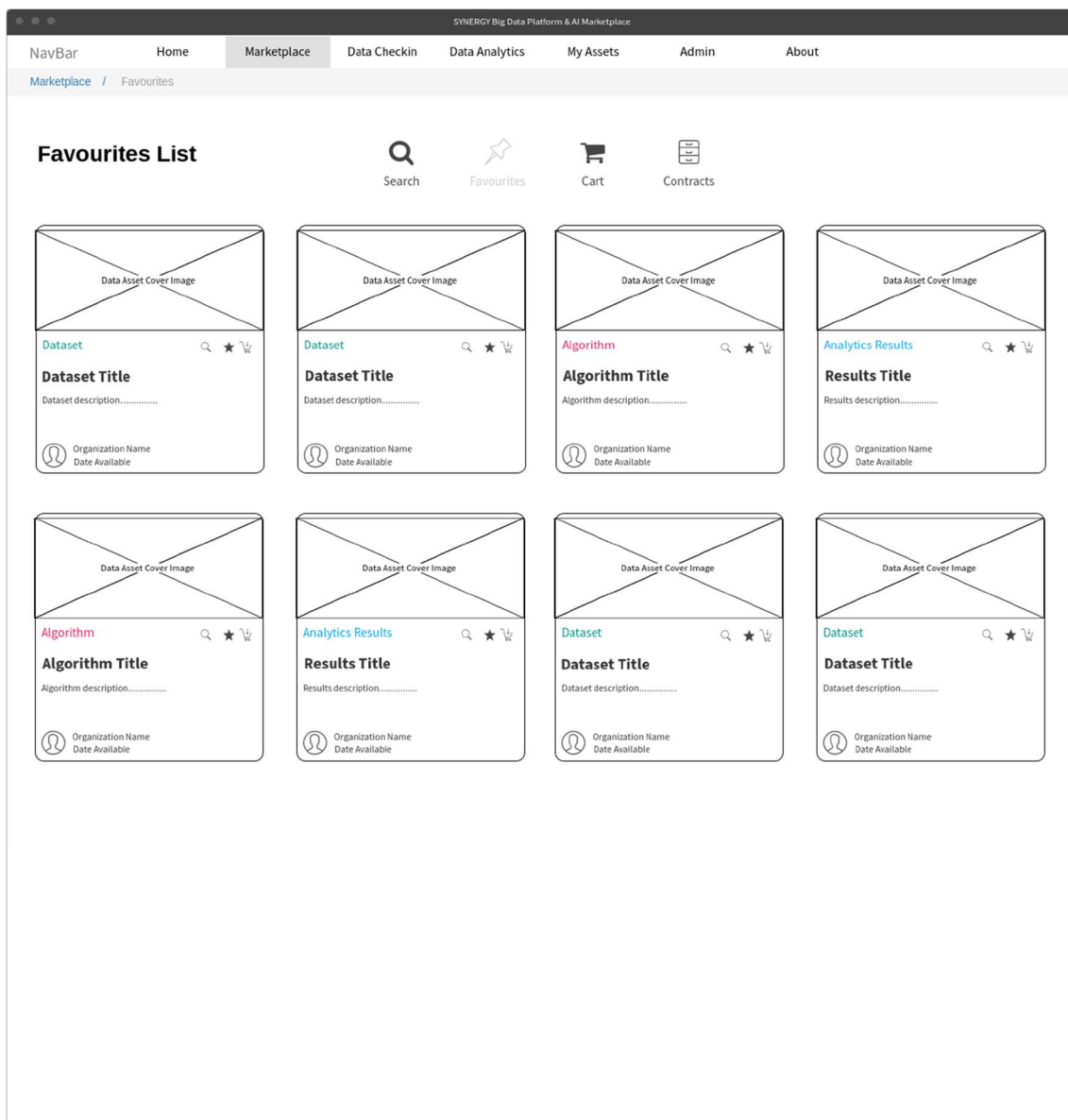


Figure 37: SYNERGY Data & AI Marketplace – Favourites List

3.2 Search for Data Assets

Data asset consumers are able to search for a data asset of interest by selecting the Search button in the Marketplace page, and the Search page is loaded as shown in Figure 38. Within this view, the data asset consumer may search for a particular keyword. Through a faceted search functionality, the data asset consumer may apply filters to the results, e.g. in terms of domain, category, type, accessibility, format, and language of the data asset. By applying any of these filters, the data asset results will appear accordingly, and the data asset consumer may sort them by different ways (i.e., relevance, title, date created, and organisation). At any

moment, the data asset consumer can view the filters that have been defined and navigate to the recommendations for data assets of interest. In case the search did not return any results, the recommendations may not only take the form of recommended data assets, but also of recommended organisations that might potentially have the data asset of interest.

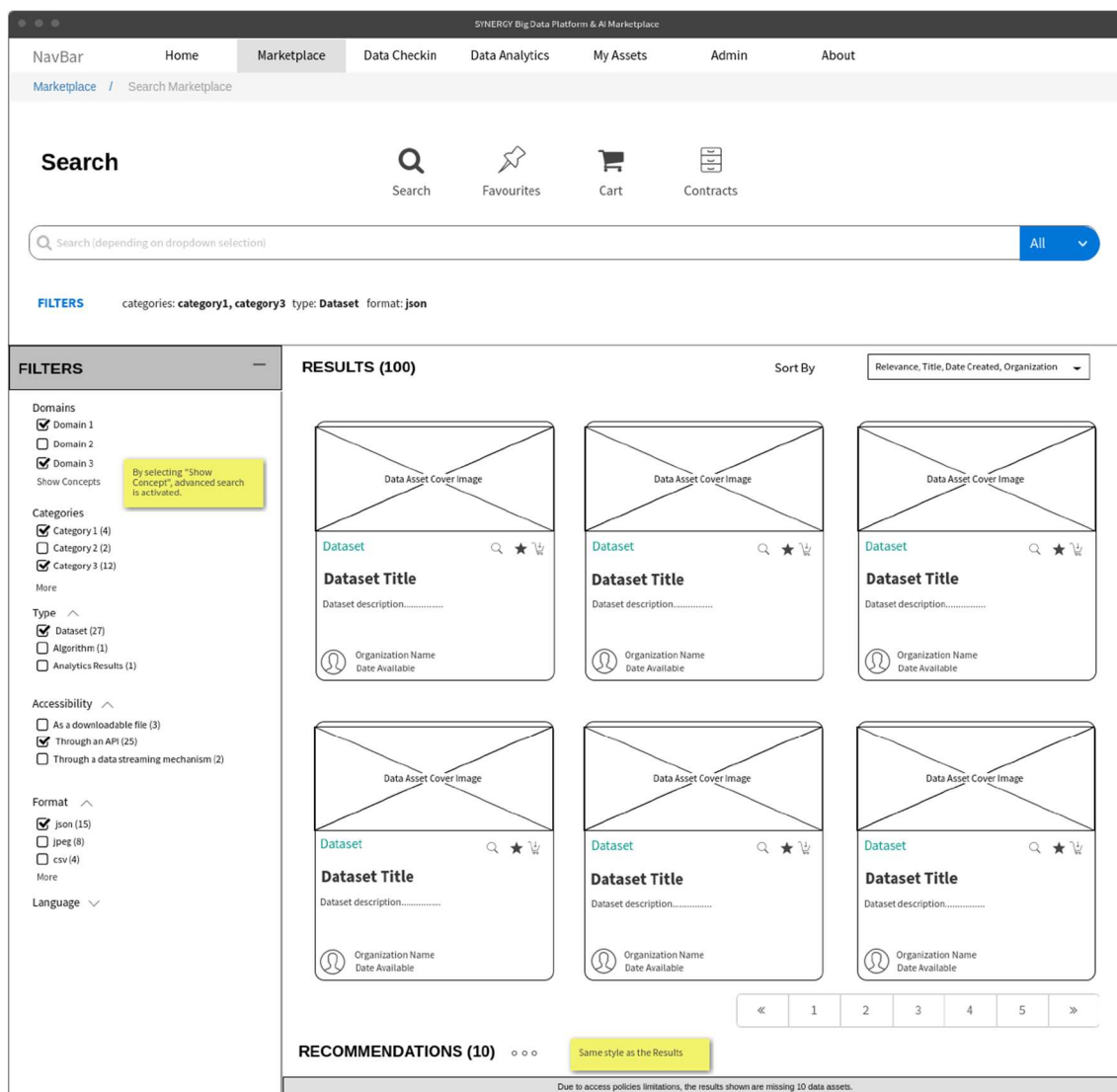
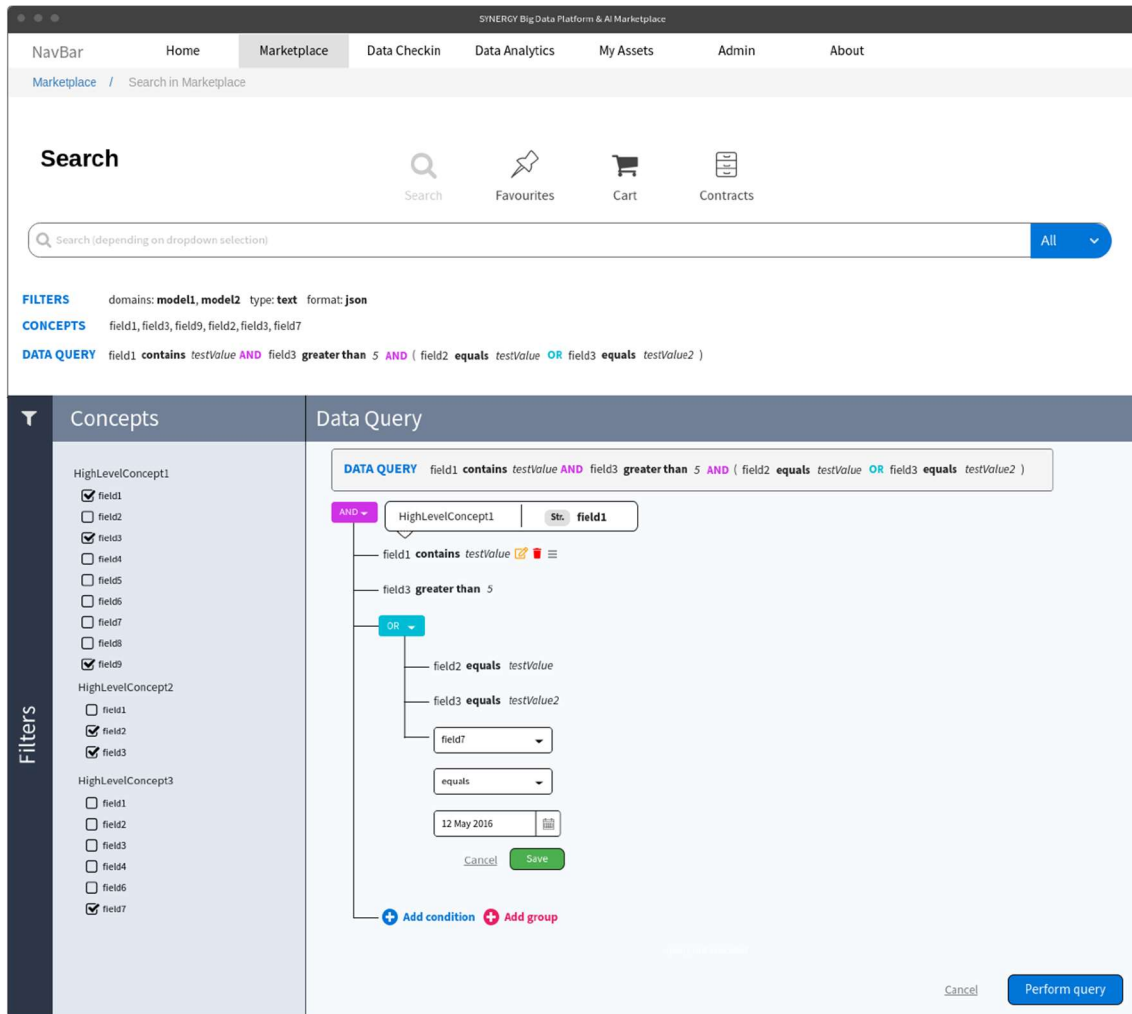


Figure 38: SYNERGY Data & AI Marketplace – Faceted Search

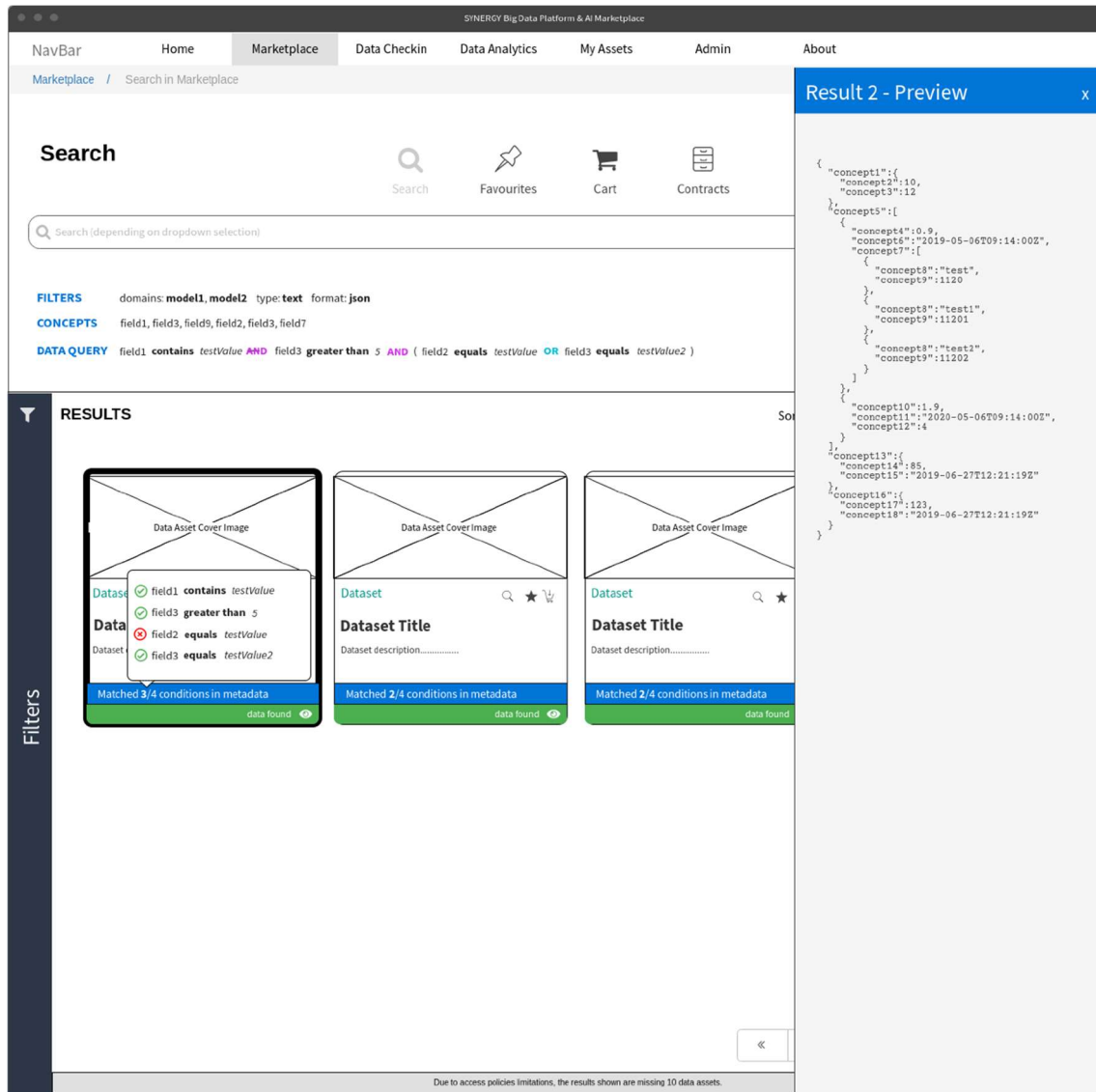
In case the data asset consumer has selected to see the concepts of the Common Information Model in the Domain filter, the advanced search is activated, as displayed in Figure 39. During the advanced search step, the data asset consumer may select the concepts and fields of interest from the SYNERGY Common Information Model, and optionally add as many conditions as needed. Individual conditions for specific fields can be applied and connected in groups for complex conditions. Once the data asset consumer has expressed all the desired conditions, he/she may perform a query that will fetch the appropriate results.



The screenshot displays the SYNERGY Big Data Platform & AI Marketplace interface. The top navigation bar includes links for Home, Marketplace, Data Checkin, Data Analytics, My Assets, Admin, and About. The Marketplace section is active, showing a search bar and icons for Search, Favourites, Cart, and Contracts. Below the search bar, there are filters and concepts. The filters section shows domains: model1, model2, type: text, format: json. The concepts section lists field1, field3, field9, field2, field3, field7. The data query section shows a query: field1 contains testValue AND field3 greater than 5 AND (field2 equals testValue OR field3 equals testValue2). The query builder on the right allows users to construct queries using concepts and fields. It shows a tree structure with conditions like field1 contains testValue, field3 greater than 5, and field2 equals testValue. The interface also includes a 'Perform query' button.

Figure 39: SYNERGY Data & AI Marketplace – Advanced Search Definition

Upon executing such a data query, the results that potentially match the expressed conditions are displayed as depicted in Figure 40. In each result, the number of conditions (from the data query) that are satisfied is identified while a clear indication whether data have been found in the data asset is also present. By selecting a specific data asset, a preview of the data asset is provided (only for unencrypted data).



The screenshot displays the SYNERGY Data & AI Marketplace interface. The top navigation bar includes links for Home, Marketplace, Data Checkin, Data Analytics, My Assets, Admin, and About. Below the navigation bar, there is a search bar and a sidebar with filters. The main area shows search results for datasets, each with a cover image, title, description, and a status bar indicating how many conditions in the metadata were matched. A right-hand panel shows a JSON preview of the search results.

Search Bar: Search (depending on dropdown selection)

Filters: domains: model1, model2 type: text format: json

CONCEPTS: field1, field3, field9, field2, field3, field7

DATA QUERY: field1 contains testValue AND field3 greater than 5 AND (field2 equals testValue OR field3 equals testValue2)

RESULTS: Three dataset cards are shown, each with a cover image, title, description, and a status bar indicating how many conditions in the metadata were matched.

JSON Preview:

```

{
  "concept1": {
    "concept2": 10,
    "concept3": 12
  },
  "concept5": {
    "concept4": 0.9,
    "concept6": "2019-05-06T09:14:00Z",
    "concept7": {
      "concept8": "test",
      "concept9": 1120
    },
    "concept8": "test1",
    "concept9": 11201
  },
  "concept9": {
    "concept3": "test3",
    "concept9": 11202
  }
},
{
  "concept10": 1.9,
  "concept11": "2020-05-06T09:14:00Z",
  "concept12": 4
},
{
  "concept13": {
    "concept14": 85,
    "concept15": "2019-06-27T12:21:19Z"
  },
  "concept16": {
    "concept17": 123,
    "concept18": "2019-06-27T12:21:19Z"
  }
}

```

Figure 40: SYNERGY Data & AI Marketplace - Advanced Search Results

3.3 Data Asset Acquisition

The data asset consumer may add a data asset (of any type) to the cart for acquisition and the cart icon will change to dark grey colour. The data asset consumer will be able to view the My Cart page, through the SYNERGY Data & AI Marketplace page, as shown in Figure 41. The cart lists all the data assets that were added to the cart at any moment by any member of the data asset consumer's organisation. The data asset consumer can either acquire all the data assets (that are listed in the cart) at once, or each data asset separately.

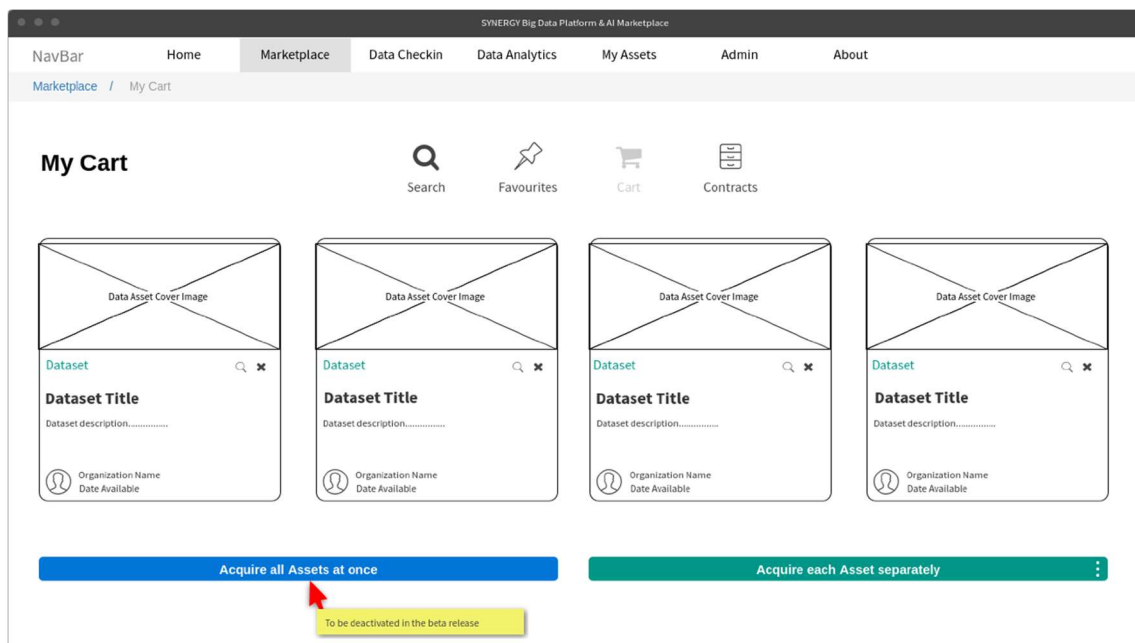


Figure 41: SYNERGY Data & AI Marketplace – My Cart

3.3.1 Request for Data Assets (Data Asset Consumer Perspective)

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 42. 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 a copy of the data asset locally. 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.

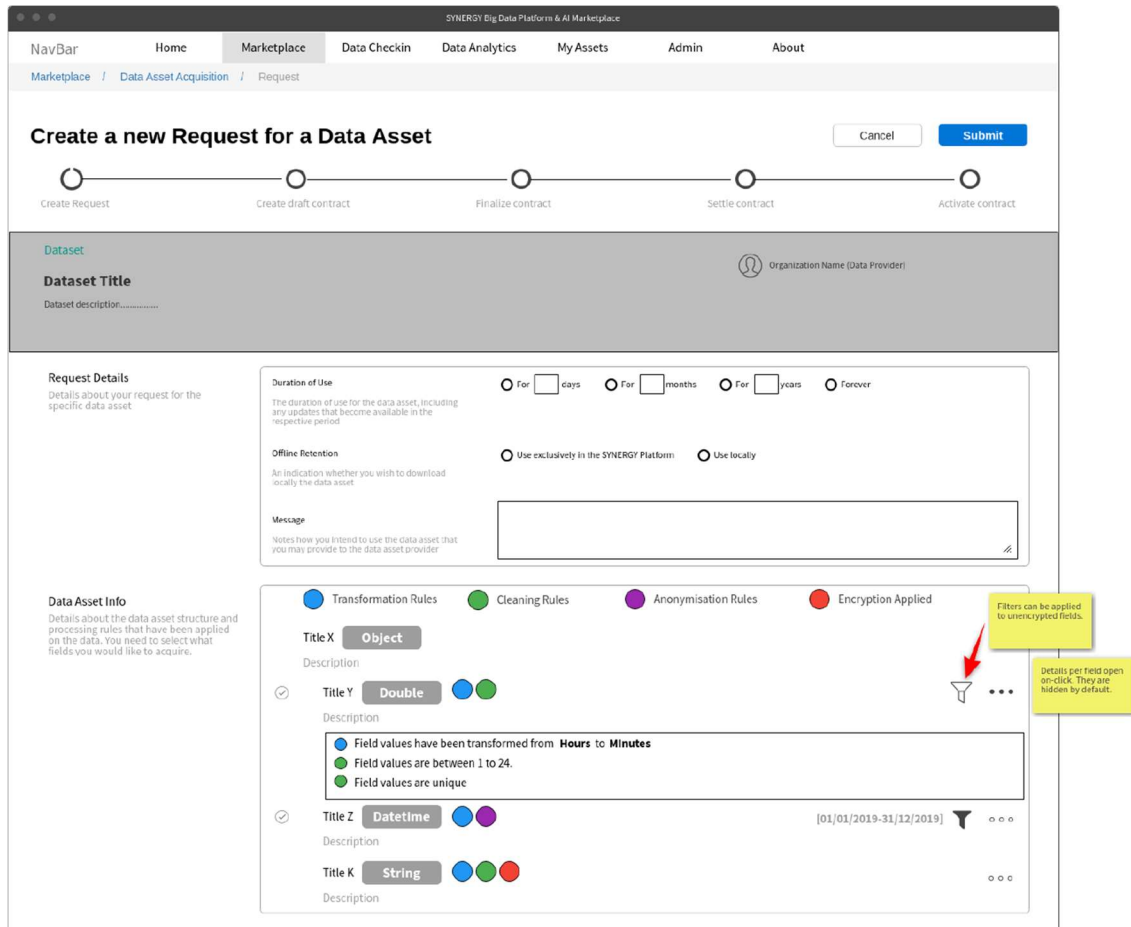


Figure 42: Create New Request for Data Asset Acquisition (Consumer Perspective)

3.3.2 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 43.

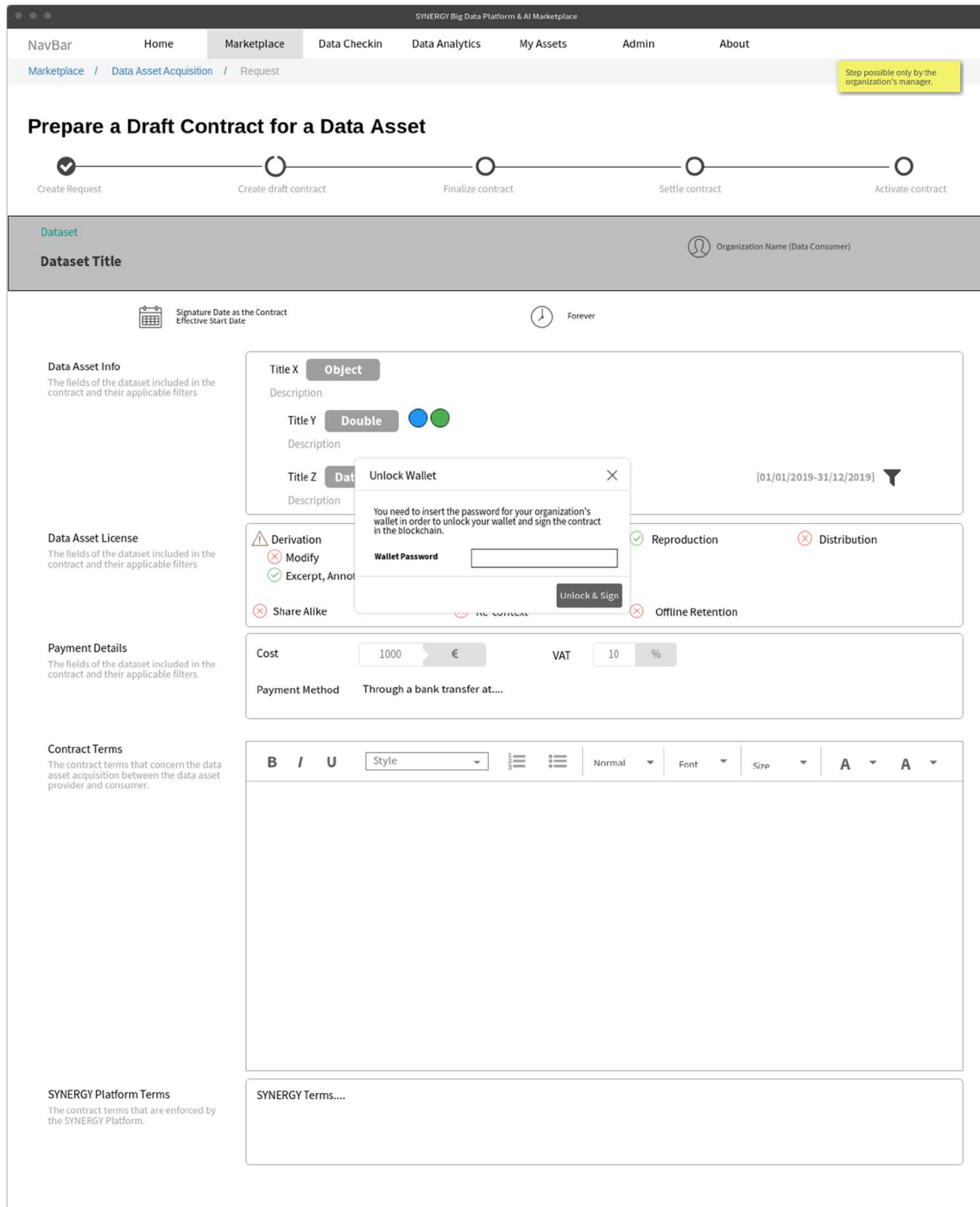
Figure 43: Review Request for Data Asset Acquisition by the Organisation's Manager (Provider Perspective)

3.3.3 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 44. 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.

- Step possible only by the organization's manager.
- Sign -> unlock wallet -> mine (write in blockchain)

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, as shown in Figure 45.



SYNERGY Big Data Platform & AI Marketplace

NavBar Home Marketplace Data Checkin Data Analytics My Assets Admin About

Marketplace / Data Asset Acquisition / Request

Step possible only by the organization's manager.

Prepare a Draft Contract for a Data Asset

Create Request Create draft contract Finalize contract Settle contract Activate contract

Dataset

Dataset Title Organization Name (Data Consumer)

Signature Date as the Contract Effective Start Date Forever

Data Asset Info
The fields of the dataset included in the contract and their applicable filters

Title X Object
Description
Title Y Double
Description
Title Z Data
Description

Data Asset License
The fields of the dataset included in the contract and their applicable filters

Derivation
Modify
Excerpt, Annotation
Share Alike

Reproduction
Distribution
Offline Retention

Payment Details
The fields of the dataset included in the contract and their applicable filters

Cost 1000 € VAT 10 %
Payment Method Through a bank transfer at...

Contract Terms
The contract terms that concern the data asset acquisition between the data asset provider and consumer.

SYNERGY Platform Terms
The contract terms that are enforced by the SYNERGY Platform.

SYNERGY Terms....

Unlock Wallet

You need to insert the password for your organization's wallet in order to unlock your wallet and sign the contract in the blockchain.

Wallet Password

Unlock & Sign

Figure 45: Unlock Wallet for Data Asset Acquisition by the Organisation's Manager (Provider Perspective)

3.3.4 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 46. 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.

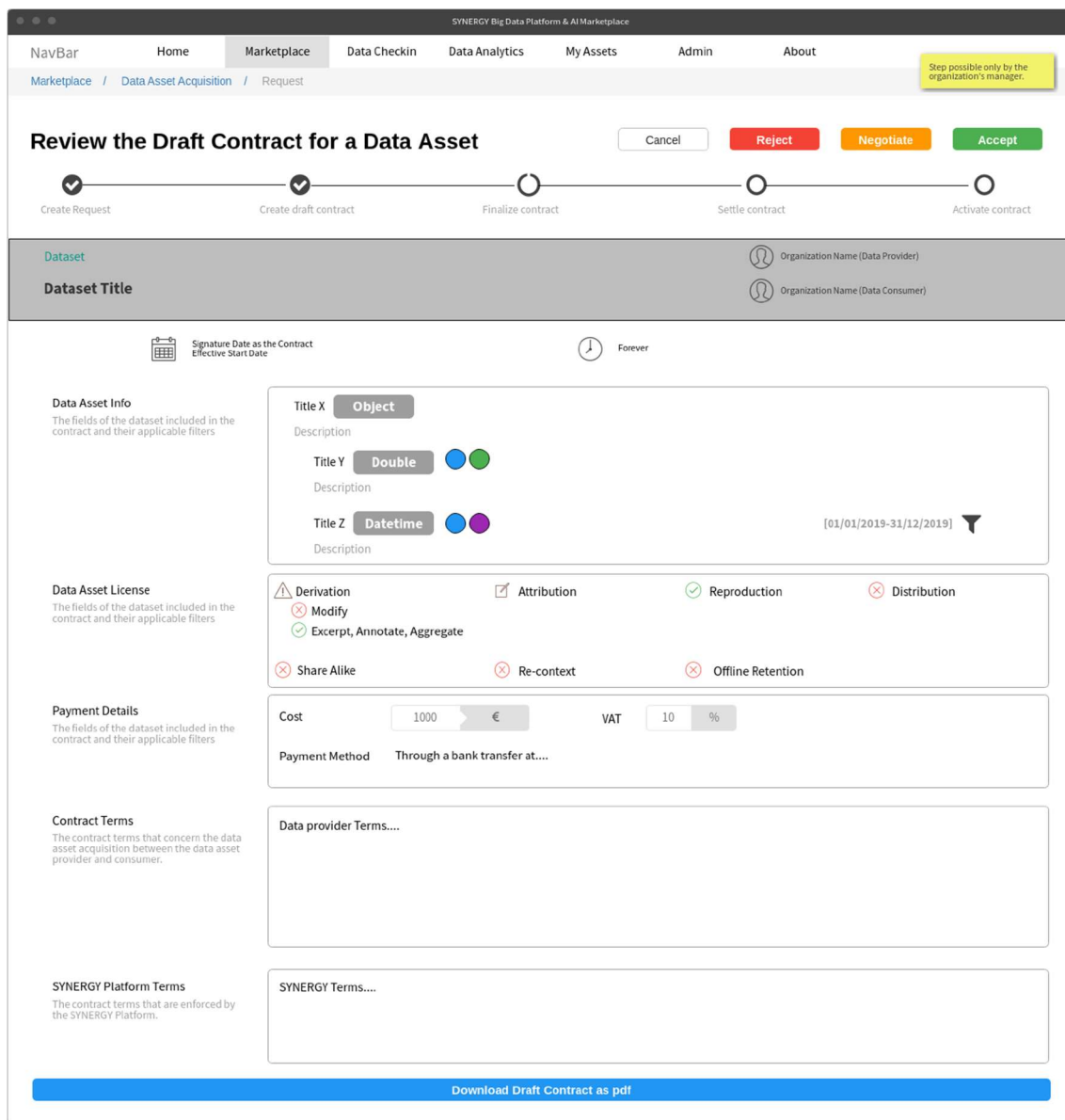


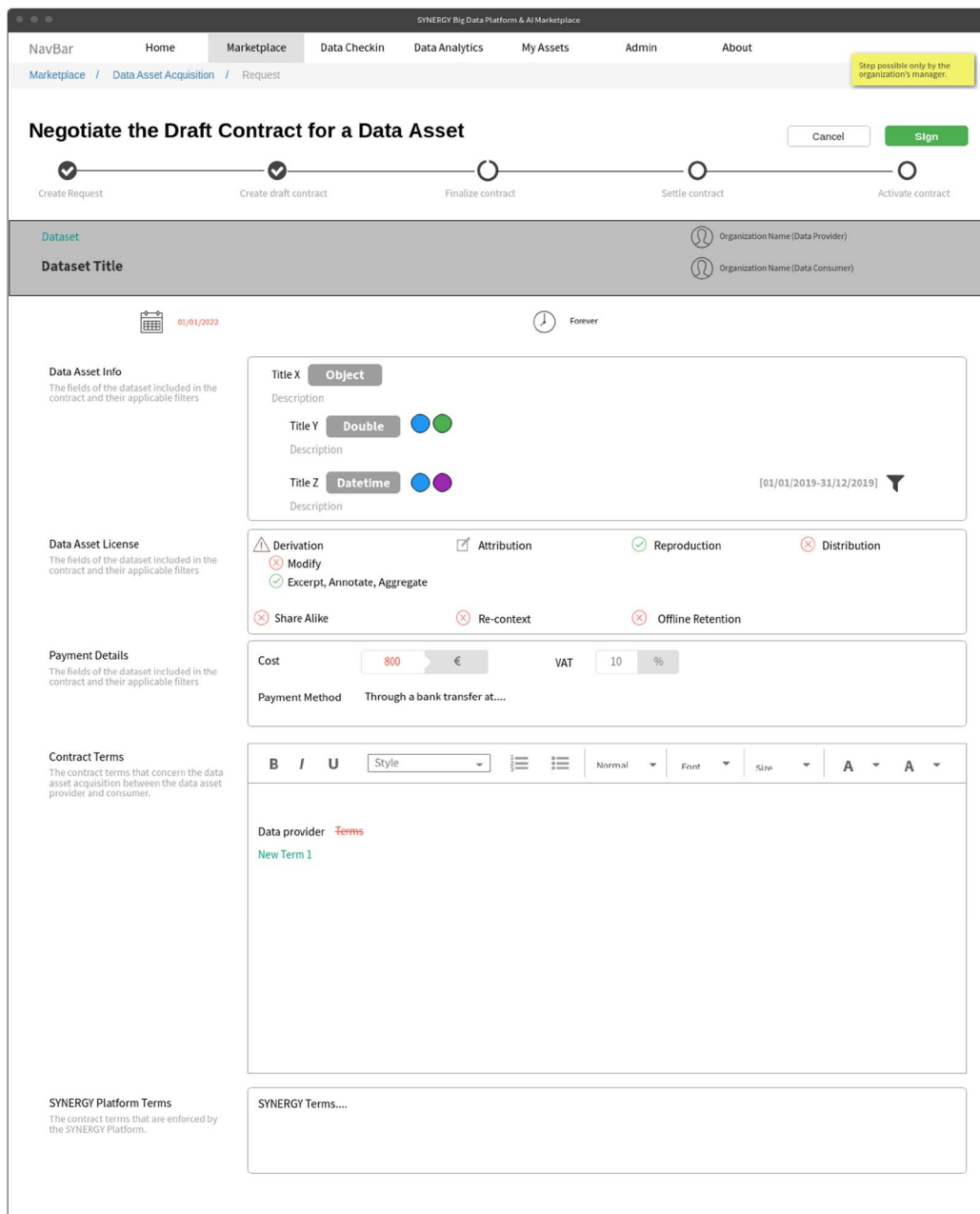
Figure 46: Review Draft Contract for Data Asset Acquisition (Consumer Perspective)

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. The process continues as described in section 3.3.7.

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.

3.3.5 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 48), 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.

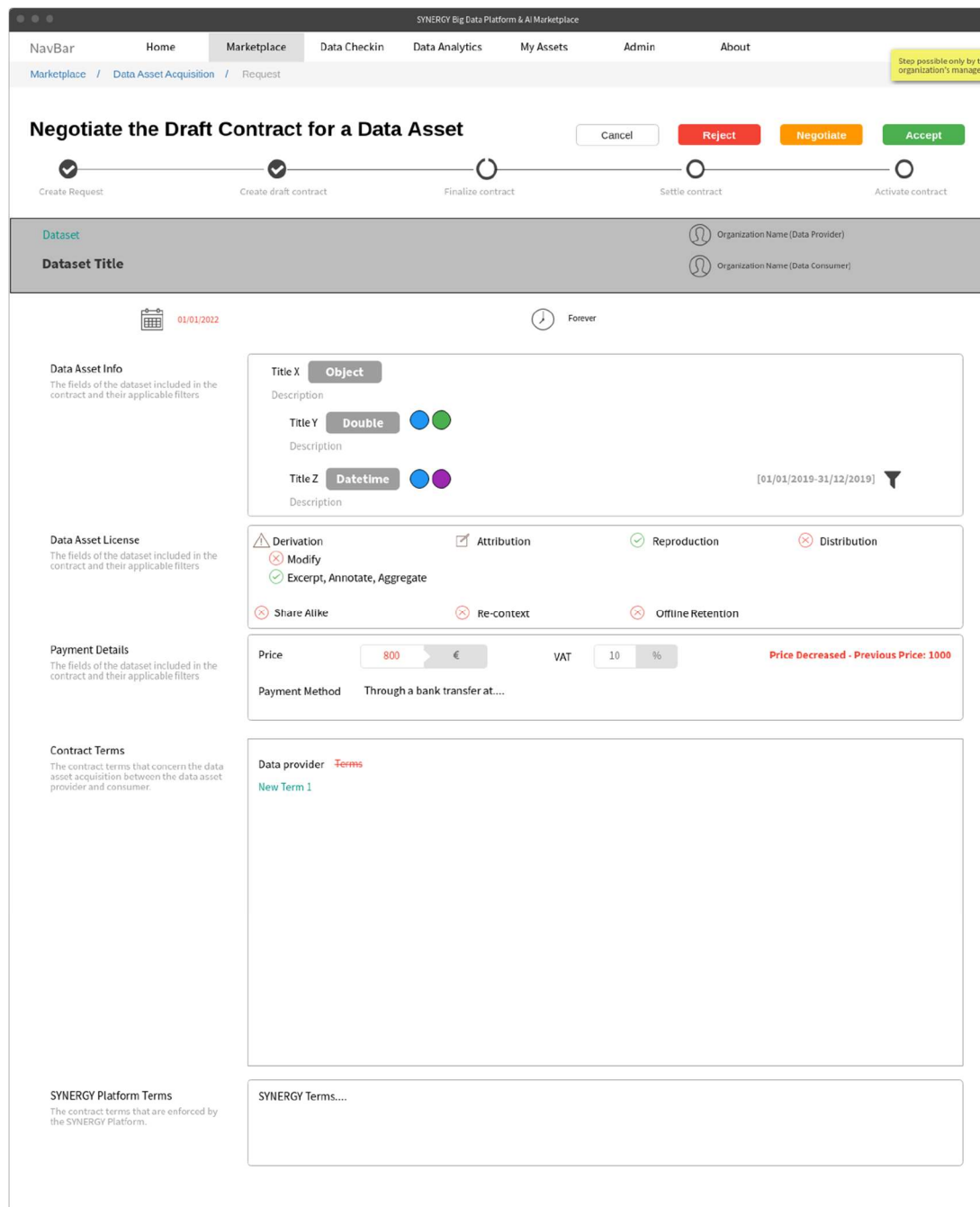


The mock-up shows a web application titled "SYNERGY Big Data Platform & AI Marketplace". The navigation bar includes "NavBar", "Home", "Marketplace", "Data Checkin", "Data Analytics", "My Assets", "Admin", and "About". A breadcrumb trail shows "Marketplace / Data Asset Acquisition / Request". A yellow notification box states: "Step possible only by the organization's manager." The main heading is "Negotiate the Draft Contract for a Data Asset". Below it is a progress bar with five steps: "Create Request" (checked), "Create draft contract" (checked), "Finalize contract" (active), "Settle contract", and "Activate contract". "Cancel" and "Sign" buttons are on the right. The interface is divided into sections for "Dataset", "Data Asset Info", "Data Asset License", "Payment Details", "Contract Terms", and "SYNERGY Platform Terms". The "Dataset" section shows "Dataset Title" and "Organization Name (Data Provider/Consumer)". "Data Asset Info" includes fields for Title X (Object), Title Y (Double), and Title Z (Datetime) with descriptions and a date range filter. "Data Asset License" shows a list of license terms with checkboxes for Derivation, Attribution, Reproduction, Distribution, Modify, Excerpt, Annotate, Aggregate, Share Alike, Re-context, and Offline Retention. "Payment Details" includes "Cost" (800 €), "VAT" (10 %), and "Payment Method" (Through a bank transfer at...). "Contract Terms" features a rich text editor with a toolbar and a list of terms including "Data provider Terms" and "New Term 1". "SYNERGY Platform Terms" shows a list of terms starting with "SYNERGY Terms....".

Figure 47: Negotiate Draft Contract for Data Asset Acquisition (Consumer Perspective)

3.3.6 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 48. 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.



SYNERGY Big Data Platform & AI Marketplace

NavBar: Home, Marketplace, Data Checkin, Data Analytics, My Assets, Admin, About

Marketplace / Data Asset Acquisition / Request

Negotiate the Draft Contract for a Data Asset

Cancel, Reject, Negotiate, Accept

Progress: Create Request (✓), Create draft contract (✓), Finalize contract (○), Settle contract (○), Activate contract (○)

Dataset: Dataset Title

Organization Name (Data Provider), Organization Name (Data Consumer)

01/01/2022, Forever

Data Asset Info
The fields of the dataset included in the contract and their applicable filters

Title X: Object
Description
Title Y: Double
Description
Title Z: Datetime
Description [01/01/2019-31/12/2019]

Data Asset License
The fields of the dataset included in the contract and their applicable filters

Derivation (⚠), Attribution (✓), Reproduction (✓), Distribution (✗)
Modify (✗), Excerpt, Annotate, Aggregate (✓)
Share Alike (✗), Re-context (✗), Offline Retention (✗)

Payment Details
The fields of the dataset included in the contract and their applicable filters

Price: 800 €, VAT: 10 %, Price Decreased - Previous Price: 1000
Payment Method: Through a bank transfer at...

Contract Terms
The contract terms that concern the data asset acquisition between the data asset provider and consumer.

Data provider: Terms
New Term 1

SYNERGY Platform Terms
The contract terms that are enforced by the SYNERGY Platform.

SYNERGY Terms...

Figure 48: Review Updated Draft Contract for Data Asset Acquisition (Provider Perspective)

It needs to be noted that this procedure described in sections 3.3.4, 3.3.5 and 3.3.6 may iterate until both the data asset provider and data asset consumer are satisfied with the contract for the data asset acquisition.

3.3.7 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 49.

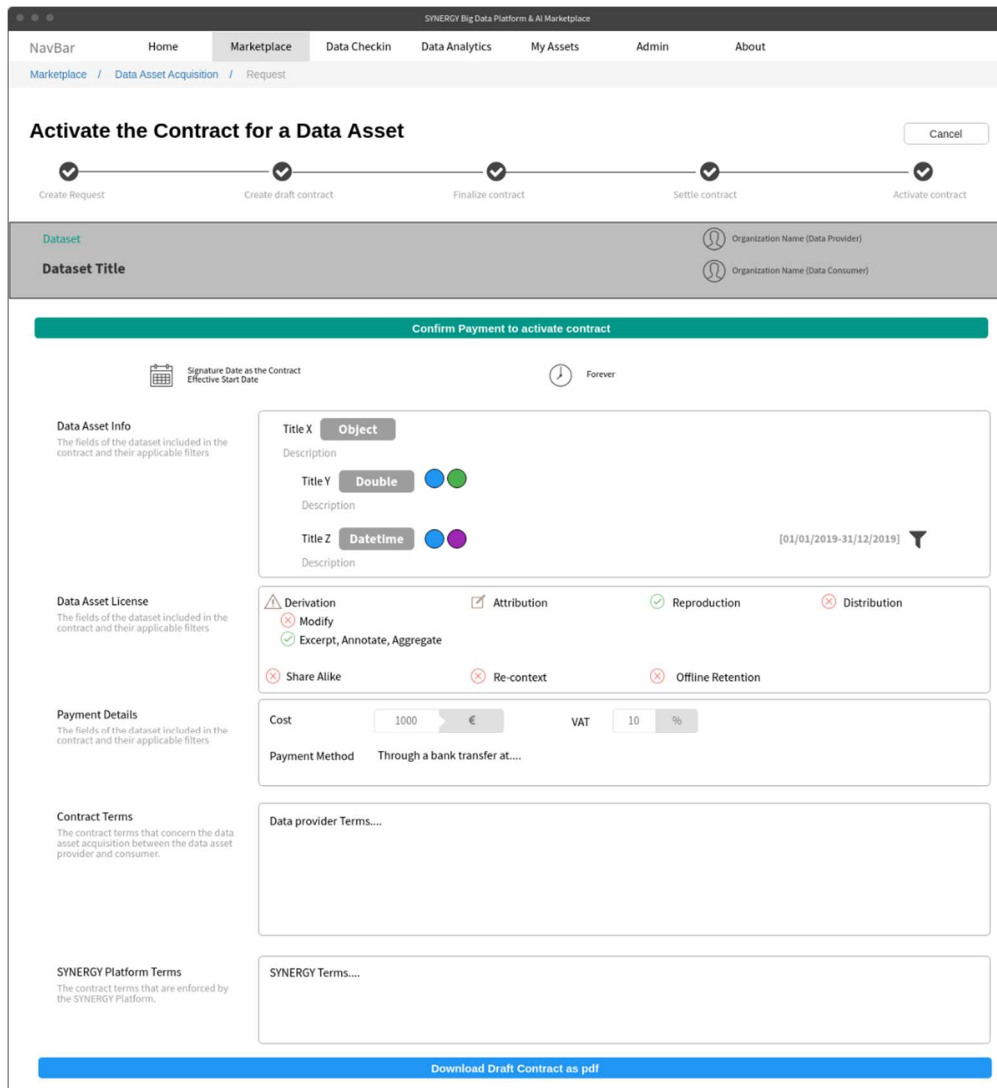
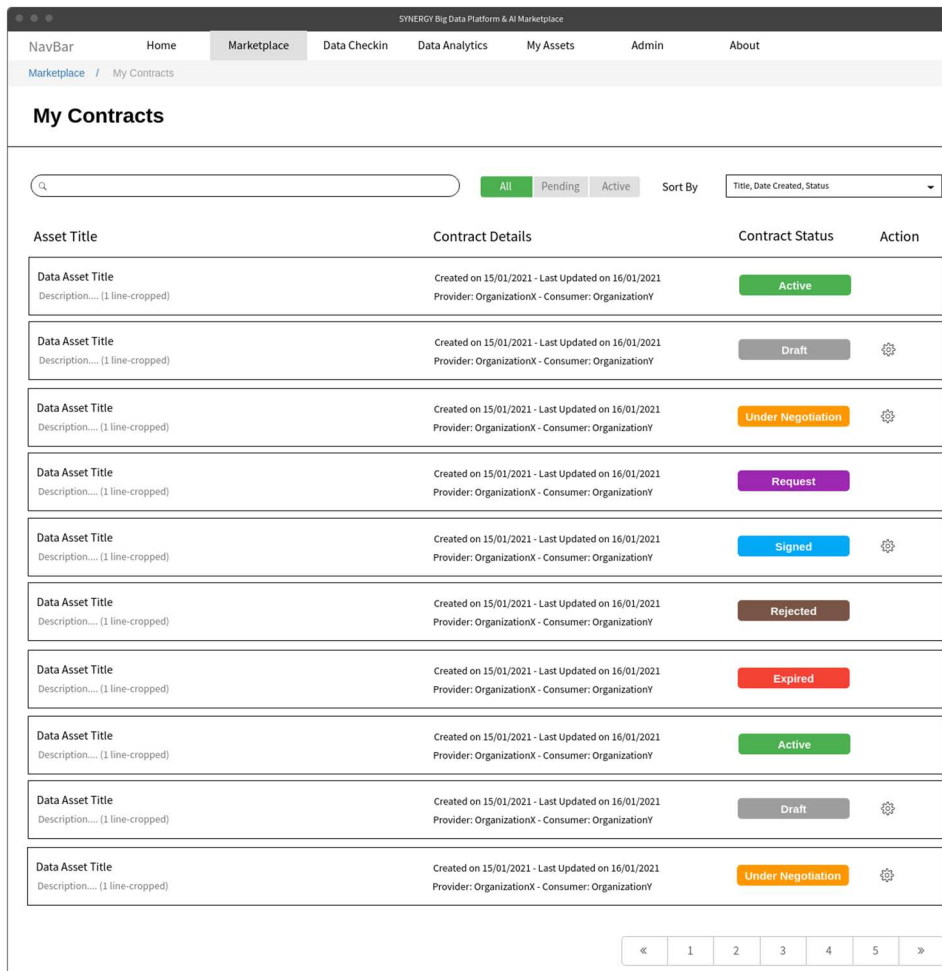


Figure 49: Activate Draft Contract for Data Asset Acquisition (Provider Perspective)

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).

3.3.8 View all organization's Contracts

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 50. 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 the functionality for action (if available).



Asset Title	Contract Details	Contract Status	Action
Data Asset Title Description.... (1 line-cropped)	Created on 15/01/2021 - Last Updated on 16/01/2021 Provider: OrganizationX - Consumer: OrganizationY	Active	
Data Asset Title Description.... (1 line-cropped)	Created on 15/01/2021 - Last Updated on 16/01/2021 Provider: OrganizationX - Consumer: OrganizationY	Draft	
Data Asset Title Description.... (1 line-cropped)	Created on 15/01/2021 - Last Updated on 16/01/2021 Provider: OrganizationX - Consumer: OrganizationY	Under Negotiation	
Data Asset Title Description.... (1 line-cropped)	Created on 15/01/2021 - Last Updated on 16/01/2021 Provider: OrganizationX - Consumer: OrganizationY	Request	
Data Asset Title Description.... (1 line-cropped)	Created on 15/01/2021 - Last Updated on 16/01/2021 Provider: OrganizationX - Consumer: OrganizationY	Signed	
Data Asset Title Description.... (1 line-cropped)	Created on 15/01/2021 - Last Updated on 16/01/2021 Provider: OrganizationX - Consumer: OrganizationY	Rejected	
Data Asset Title Description.... (1 line-cropped)	Created on 15/01/2021 - Last Updated on 16/01/2021 Provider: OrganizationX - Consumer: OrganizationY	Expired	
Data Asset Title Description.... (1 line-cropped)	Created on 15/01/2021 - Last Updated on 16/01/2021 Provider: OrganizationX - Consumer: OrganizationY	Active	
Data Asset Title Description.... (1 line-cropped)	Created on 15/01/2021 - Last Updated on 16/01/2021 Provider: OrganizationX - Consumer: OrganizationY	Draft	
Data Asset Title Description.... (1 line-cropped)	Created on 15/01/2021 - Last Updated on 16/01/2021 Provider: OrganizationX - Consumer: OrganizationY	Under Negotiation	

Figure 50: SYNERGY Data & AI Marketplace – Contracts List

3.4 View My Assets

The members of an organization may view a list of their owned or acquired data assets in a tabular view in the SYNERGY Platform, as depicted in Figure 51. They are also able to filter the assets based on their type, and appropriately sort them. Depending on whether it is a data asset they own or have purchased, additional functionalities are available, such as to view more details about the selected data asset, to edit the details of the selected data asset (available only if the data asset is owned by the specific organisation), see the status available for analytics (i.e. whether the data asset has been transferred to the organisation’s Secure Experimentation Playground), add to results (in order to create a new retrieval query as described in section 3.5), visualize data, subscribe to streaming data (available only if the data asset is acquired by the user), and finally to delete the asset (available only if the data asset is owned by the user). In case where the contract of an acquired data asset has expired, an appropriate tag will be visible at the data asset’s card, and that data asset will not be available to be used anymore by the specific organisation.

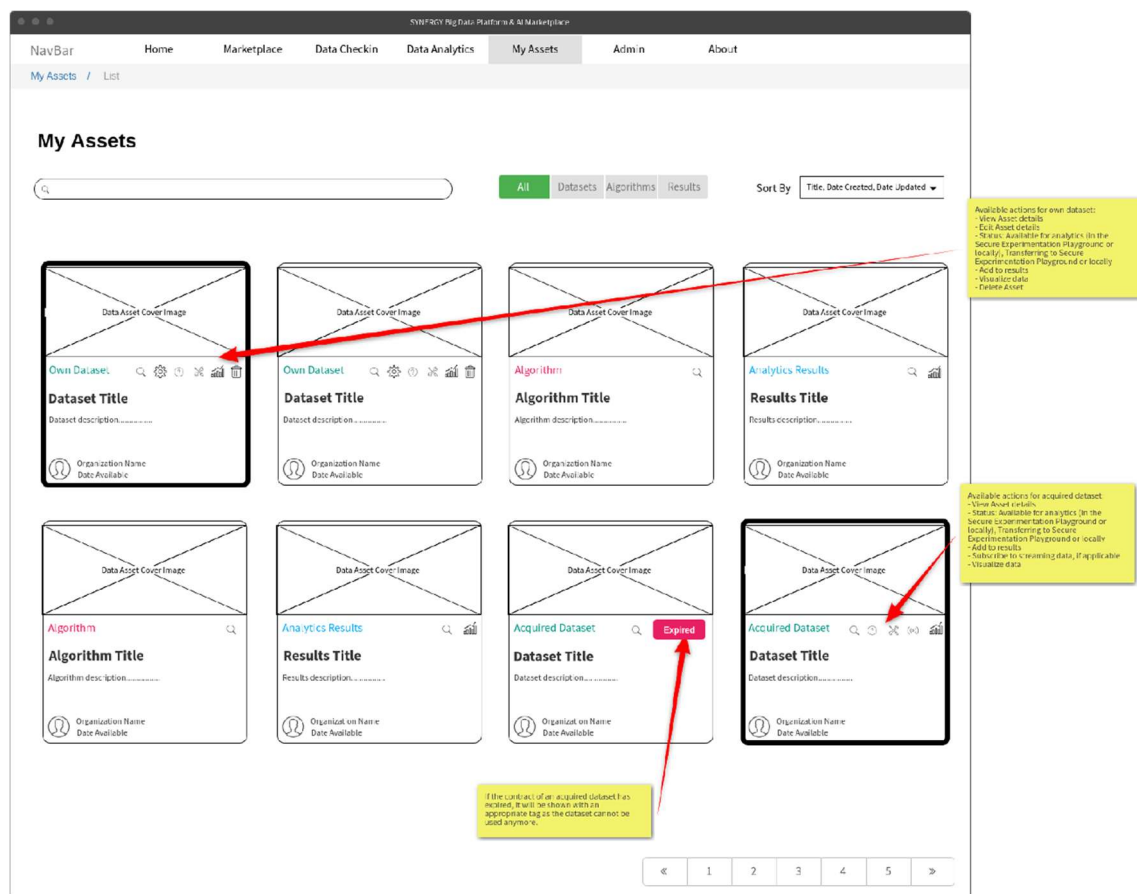


Figure 51: My Assets

3.5 Create a Retrieval Query

As shown in Figure 52, the members of an organization are able to create a retrieval query through API (icon highlighted in a blue circle) or a retrieval query for data streaming (icon highlighted in a red circle) through “My Assets” in the SYNERGY Platform. The procedure for configuring each method is described in the next subsections accordingly. It needs to be clarified that retrieval queries are enabled only for unencrypted data that are stored in the SYNERGY Cloud Platform.

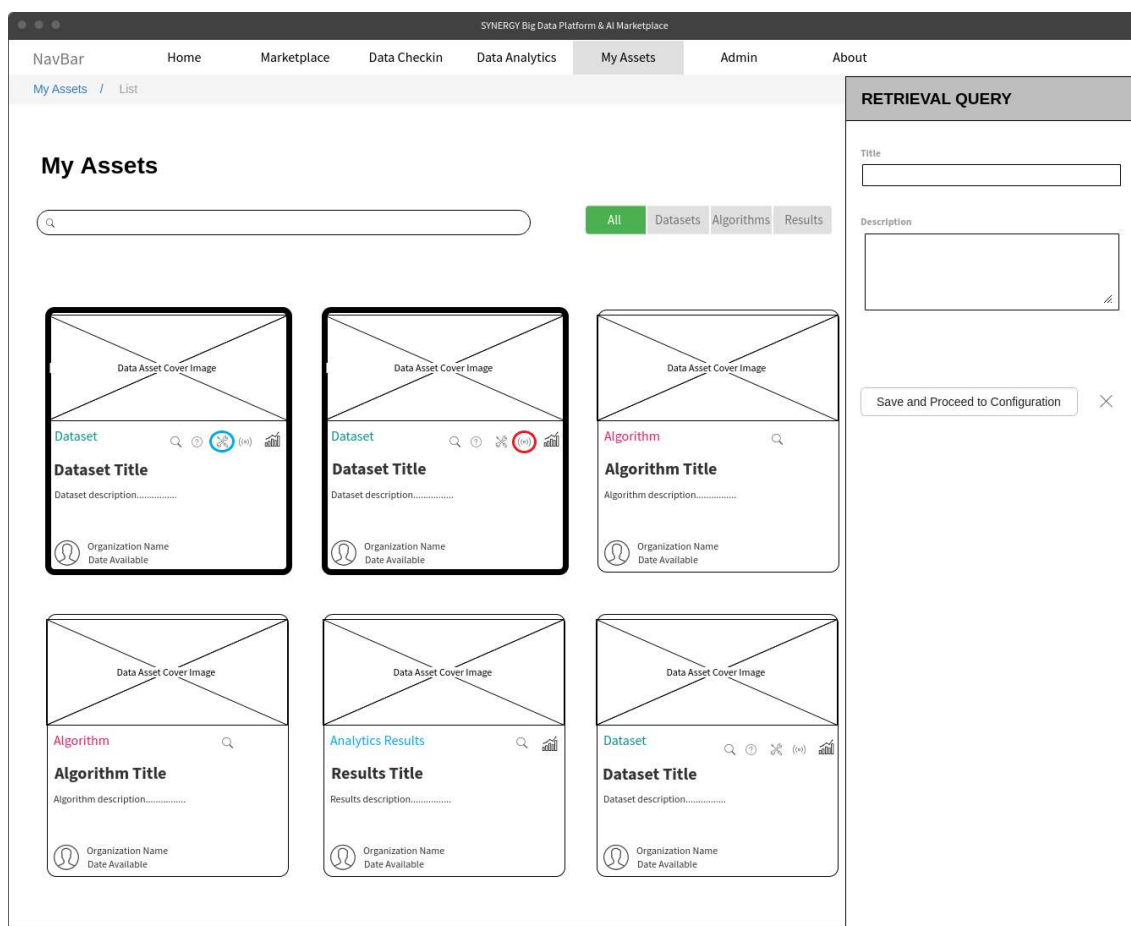


Figure 52: My Assets – Create a Retrieval Query

In order to create a retrieval query through API, the retrieval query basic information (i.e., title, description) needs to be filled in, to the right section that appears as a slideover.

3.5.1 Create Retrieval Query through API

In the Retrieval Configuration page through API as shown in Figure 53, the Retrieval Settings need 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 datasets were selected in the “My Assets” page, 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.

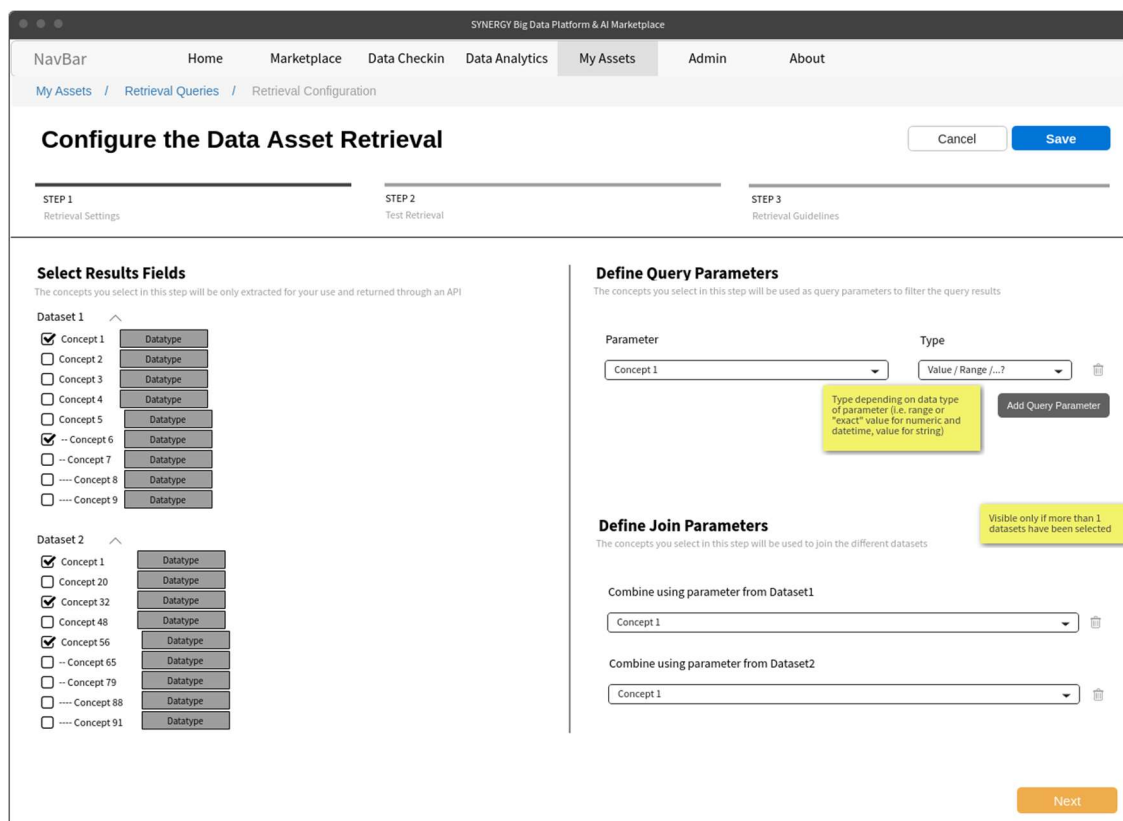


Figure 53: Data Asset (API) Retrieval Configuration – Retrieval Settings

Once the retrieval configuration is ready, the configured retrieval can be tested as shown in Figure 54. At the left side of this page, the users shall be able to edit the body with indicative values 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.

The last step of the retrieval configuration contains information on how to acquire the retrieval results through the SYNERGY APIs as depicted in Figure 55. To this end, instructions regarding authentication, 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)

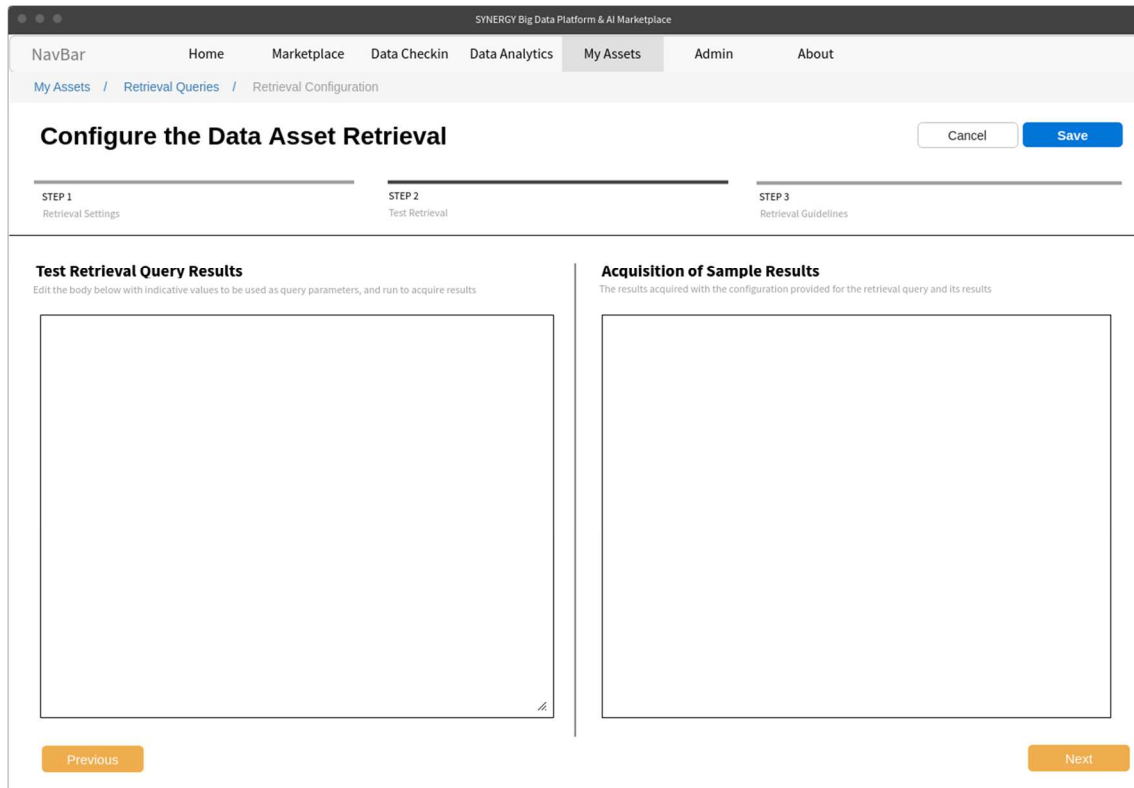
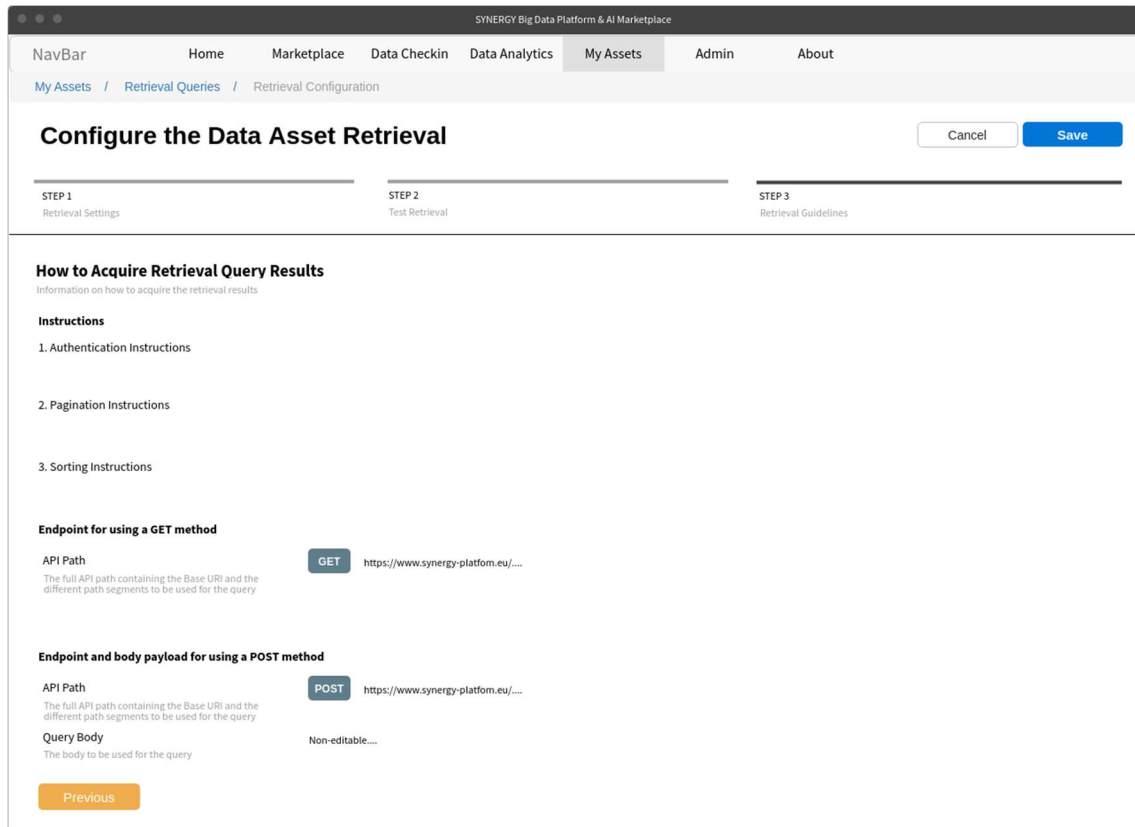


Figure 54: Data Asset (API) Retrieval Configuration - Test Retrieval



SYNERGY Big Data Platform & AI Marketplace

NavBar Home Marketplace Data Checkin Data Analytics My Assets Admin About

My Assets / Retrieval Queries / Retrieval Configuration

Configure the Data Asset Retrieval

Cancel Save

STEP 1 Retrieval Settings STEP 2 Test Retrieval STEP 3 Retrieval Guidelines

How to Acquire Retrieval Query Results

Information on how to acquire the retrieval results

Instructions

1. Authentication Instructions
2. Pagination Instructions
3. Sorting Instructions

Endpoint for using a GET method

API Path
The full API path containing the Base URI and the different path segments to be used for the query

GET <https://www.synergy-platform.eu/...>

Endpoint and body payload for using a POST method

API Path
The full API path containing the Base URI and the different path segments to be used for the query

POST <https://www.synergy-platform.eu/...>

Query Body
The body to be used for the query

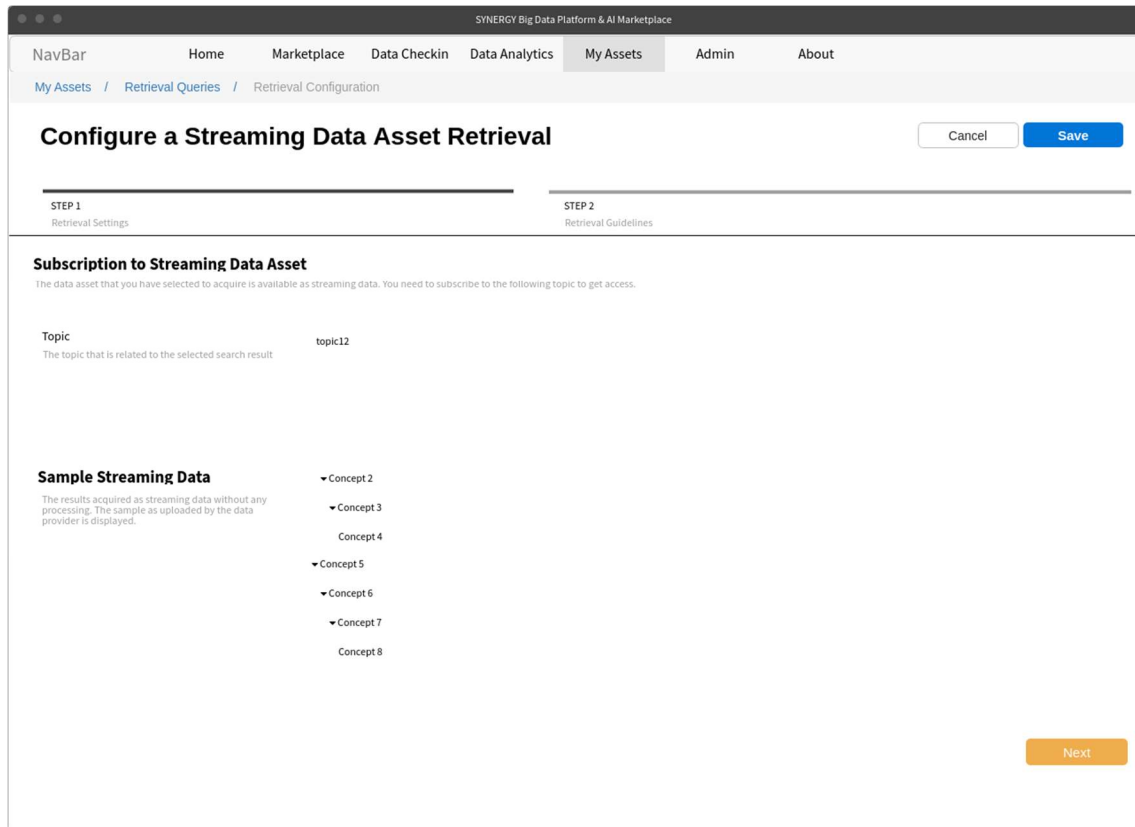
Non-editable....

Previous

Figure 55: Data Asset (API) Retrieval Configuration - Retrieval Guidelines

3.5.2 Create Retrieval Query for Streaming Data

In the case of streaming data (collected in the SYNERGY PubSub mechanism), the data asset consumers shall subscribe to the topics listed for the selected data asset, as shown in Figure 56. A sample of the streaming data without any processing as uploaded by the data asset provider is also displayed.



SYNERGY Big Data Platform & AI Marketplace

NavBar Home Marketplace Data Checkin Data Analytics My Assets Admin About

My Assets / Retrieval Queries / Retrieval Configuration

Configure a Streaming Data Asset Retrieval

Cancel Save

STEP 1 Retrieval Settings

STEP 2 Retrieval Guidelines

Subscription to Streaming Data Asset

The data asset that you have selected to acquire is available as streaming data. You need to subscribe to the following topic to get access.

Topic

The topic that is related to the selected search result

topic12

Sample Streaming Data

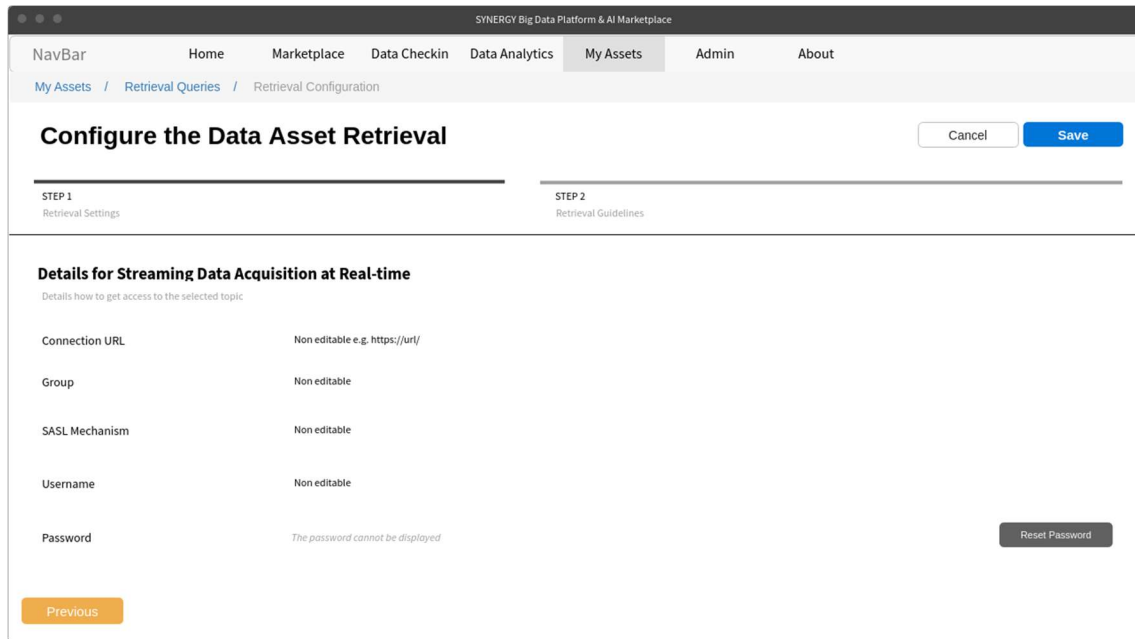
The results acquired as streaming data without any processing. The sample as uploaded by the data provider is displayed.

- ▼ Concept 2
- ▼ Concept 3
- Concept 4
- ▼ Concept 5
- ▼ Concept 6
- ▼ Concept 7
- Concept 8

Next

Figure 56: Data Asset (Streaming Data) Retrieval Configuration – Retrieval Settings

Then, as depicted in Figure 57, specific guidelines are provided in the form of the connection URL to the SYNERGY Cloud PubSub mechanism, the group, the SASL Mechanism, as well as the username and password of the data asset consumer's account (that has been created and is subscribed to the above topic).



SYNERGY Big Data Platform & AI Marketplace

NavBar Home Marketplace Data Checkin Data Analytics My Assets Admin About

My Assets / Retrieval Queries / Retrieval Configuration

Configure the Data Asset Retrieval

Cancel Save

STEP 1 Retrieval Settings

STEP 2 Retrieval Guidelines

Details for Streaming Data Acquisition at Real-time

Details how to get access to the selected topic

Connection URL	Non editable e.g. https://url/
Group	Non editable
SASL Mechanism	Non editable
Username	Non editable
Password	The password cannot be displayed

Reset Password

Previous

Figure 57: Data Asset (Streaming Data) Retrieval Configuration – Retrieval Guidelines

It needs to be noted that the acquisition of the streaming data as described in this section implies that the raw data are retrieved without any processing by the SYNERGY Platform.

3.5.3 View all Retrieval Queries

A member of an organization can access the retrieval queries that have been created for the specific organization as shown in Figure 58. Within this page he/she may view more details regarding each retrieval query such as its title, a short description, the date of creation, the datasets that were selected for the query, the status of the query, and additional actions such as edit or remove the retrieval query. Note that, in case where the access policies of a dataset (that is included in a retrieval query) has changed, the user will be appropriately notified, including more information about the changes occurred. The same applies to the case where the retrieval query definition is incomplete.

SYNERGY Big Data Platform & AI Marketplace		
NavBar	Home	Marketplace
Data Checkin	Data Analytics	My Assets
Admin	About	
My Assets / Retrieval Queries		
Retrieval Queries		
<input type="text"/>	Sort By	Title, Date Created
Retrieval Query Title	Query Status	Actions
Retrieval Query Title Description..... (1 line-cropped)	Created on 15/01/2021 by User1 Dataset1 & Dataset2	
Retrieval Query Title Description..... (1 line-cropped)	Created on 15/01/2021 by User2 Dataset4 & Dataset5	
Retrieval Query Title Description..... (1 line-cropped)	Created on 15/01/2021 by User3 Incomplete Query Definition	
Retrieval Query Title Description..... (1 line-cropped)	Created on 15/01/2021 by User4 Dataset7	
Retrieval Query Title Description..... (1 line-cropped)	Created on 15/01/2021 by User1 Dataset10	
Retrieval Query Title Description..... (1 line-cropped)	Created on 15/01/2021 by User1 Incomplete Query Definition	
Retrieval Query Title Description..... (1 line-cropped)	Created on 15/01/2021 by User1 - Executed 5 times Dataset 10&11	
Retrieval Query Title Description..... (1 line-cropped)	Created on 15/01/2021 by User1 Incomplete Query Definition	
Retrieval Query Title Description..... (1 line-cropped)	Created on 15/01/2021 by User1 - Executed 100 times Dataset 100&101	
Retrieval Query Title Description..... (1 line-cropped)	Created on 15/01/2021 - Executed 1 time Dataset 10001	
<div> << 1 2 3 4 5 >> </div>		

Figure 58: My Assets – View Retrieval Queries

4 Data Analytics User Journey

The SYNERGY Platform allows data asset consumers to run data analytics over their own data assets (that have been checked in in the SYNERGY Platform as discussed in Section 2), or in conjunction with data assets that they have acquired from other electricity data value chain stakeholders (according to the Data Search & Acquisition User Journey described in Section 3). Similarly, as in the case of Data Check-in workflow, there is a distinct separation between the design of a data analytics pipeline (described in sections 4.2 and 4.3.1-4.3.4) and its execution either in the organisation's Secure Experimentation Playground in the cloud, or on-premise as described in section 4.3.5 and 4.4. 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.

4.1 View all Analytics Pipelines

The data asset consumers may see all the analytics pipelines they have defined as depicted in Figure 60, where the details, the status and the available actions for each analytics pipeline is displayed. A data asset consumer 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 user (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. Details about the successful and failed executions that are highlighted with appropriate icons, will be provided through tooltips. Finally, a data asset consumer 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.









































SYNERGY Big Data Platform & AI Marketplace		
NavBar	Home	Marketplace
Data Checkin	Data Analytics	My Assets
Admin	About	
Analytics Workbench / Pipeline Designer		
Analytics Pipelines		Create
<input type="text"/>		Sort By Title, Date Created, Date Executed, Status, User
Pipeline Details	Pipeline Status	Actions
Pipeline Title Description..... (1 line-cropped)	Created on 15/01/2021 by User1 - Executed 1 time Latest execution status ✓	   
Pipeline Title Description..... (1 line-cropped)	Created on 15/01/2021 by User2 - Executed 10 times Latest executions status ✗ ✓ ✓ ✓ ✗	   
Pipeline Title Description..... (1 line-cropped)	Created on 15/01/2021 by User3 ⚠ Incomplete Workflow Definition	   
Pipeline Title Description..... (1 line-cropped)	Created on 15/01/2021 by User4 ⌚ Execution Pending	   
Pipeline Title Description..... (1 line-cropped)	Created on 15/01/2021 by User1 - Executed 20 times Latest executions status ✗ ✓ ✓ ✓ ✗	   
Pipeline Title Description..... (1 line-cropped)	Created on 15/01/2021 by User1 ⚠ Incomplete Workflow Definition	   
Pipeline Title Description..... (1 line-cropped)	Created on 15/01/2021 by User1 - Executed 5 times Latest executions status ✗ ✓ ✓ ✓ ✗	   
Pipeline Title Description..... (1 line-cropped)	Created on 15/01/2021 by User1 ⌚ Execution Pending	   
Pipeline Title Description..... (1 line-cropped)	Created on 15/01/2021 by User1 - Executed 100 times Latest executions status ✗ ✓ ✓ ✓ ✗	   
Pipeline Title Description..... (1 line-cropped)	Created on 15/01/2021 - Executed 1 time Latest execution status ✓	   
		« 1 2 3 4 5 »

Figure 60: Analytics Pipeline - View all Analytics Pipelines

4.2 Create an Analytics Pipeline

Data asset consumers may create a new analytics pipeline through the Data Analytics page, as shown in Figure 61. Initially, they need 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, they need to select the execution framework of their analytics pipeline, that is the framework (i.e., Apache Spark engine, Python environment, or other) that will be used for the data processing tasks included in the pipeline. In addition, data asset consumers need to select the location

where the analytics pipeline job will be executed (i.e., Cloud Execution, On-Premise Execution, Edge Execution). 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, Edge Execution) using a Python environment. By clicking on the Save button, the Analytics Workflow Design page is loaded accordingly as described in section 4.3.

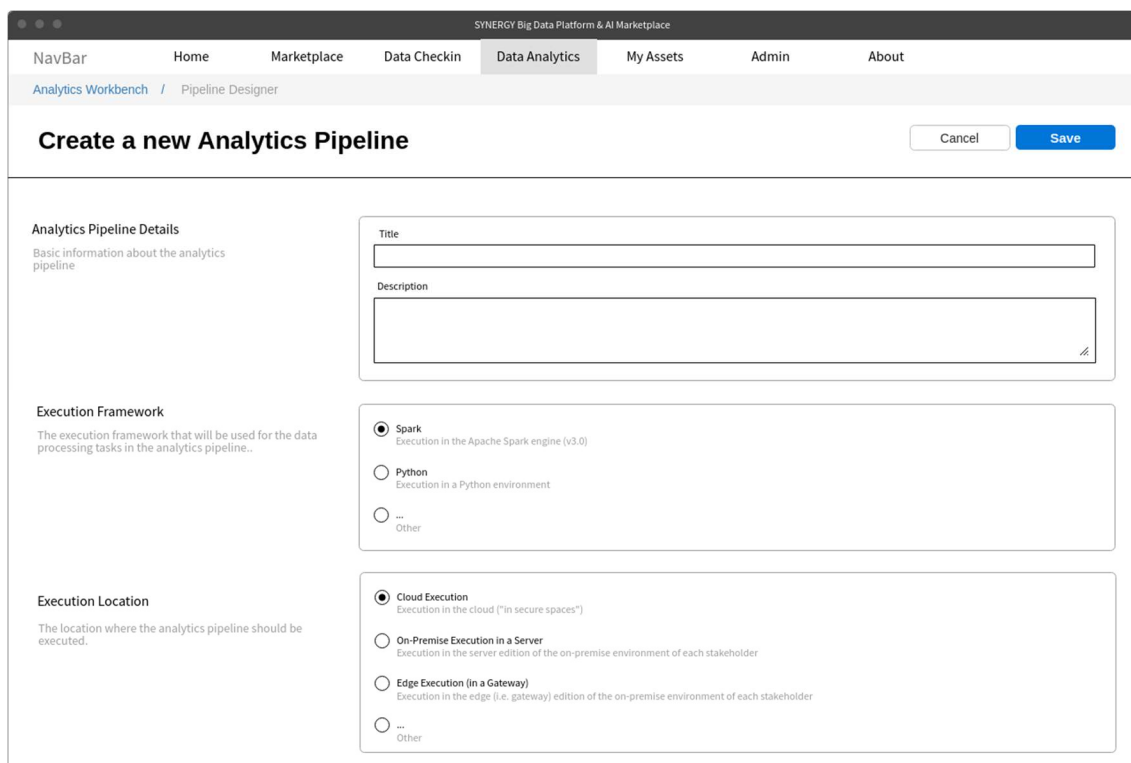


Figure 61: Analytics Pipeline - Create a new Analytics Pipeline

4.3 Configure an Analytics Pipeline

Through the Analytics Workbench, the data asset consumers are able to appropriately configure an analytics pipeline, by defining the necessary input data, the data manipulation functions and the machine learning (ML) algorithms that are to be applied on the input data, and the output data options. Such a configuration is done in a user-friendly way where the data asset consumers can build the whole analytics pipeline in a visual manner, in the form of easily configurable blocks that are appropriately connected to form the final data analysis pipeline. Indicatively, in the input block configuration the exact dataset needs to be defined while in the output configuration, it needs to be clarified how the data will be obtained, e.g. through a visualization,

through the SYNERGY Platform’s APIs. The data asset consumers are able to use a wide range of basic algorithms developed in different frameworks, as well as pre-trained algorithms to solve specific energy problems.

4.3.1 Graph View

Through the Graph view, data asset consumers are able to visually design an analytics pipeline as shown in Figure 62. By searching and filtering in the Library, the data asset consumer 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 consumer can add them in the main Graph View by selecting the Add button. 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 user configures the block connections in the right sidebar.

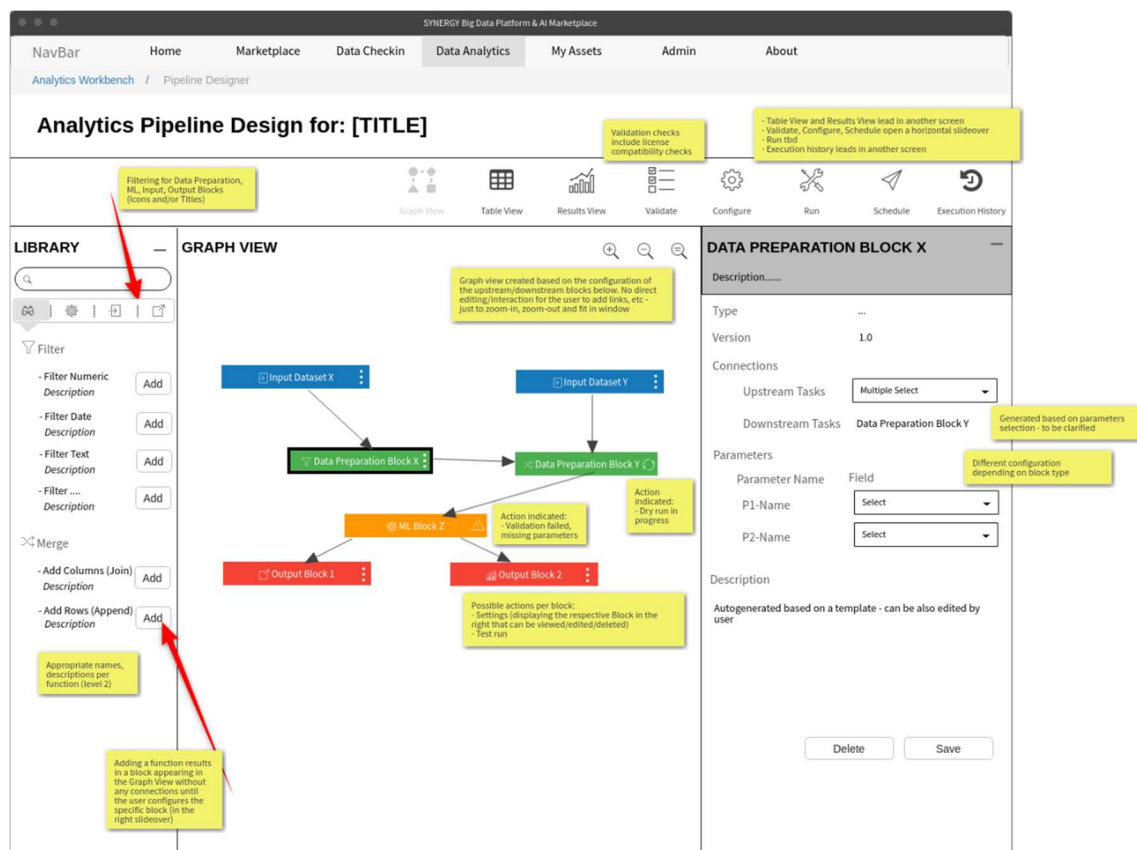


Figure 62: Analytics Pipeline - Graph View

Practically, by selecting a block, the data asset consumer 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. As a result of the block definition by the data asset consumer, an autogenerated description (that can be edited) of the whole block will appear. By selecting the Save button, a dry run will be executed for the configured block in order to retrieve the structure of the expected output, and a related status icon will appear in the right side of the block in the Graph View. 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. In order for the data asset consumer to view the results of the test run, he/she needs to transition to the Table view that is described in section 4.3.2.

4.3.2 Table View

In the Table view, data asset providers obtain an actual overview of a data sample, as depicted in Figure 63. In the left slideover, a data asset provider may view the blocks that constitute the pipeline, while by selecting the overflow menu icon (vertical ellipsis) of each block, reveals two available options for the block: a) Settings, which loads the right slideover menu with the block configuration as in the Graph View; and b) Test run, which runs the pipeline until the specific block for a data sample. By selecting a block from the Overview slideover menu, the exact workflow is highlighted, and results of the test run appear in the Data View. Within this 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.

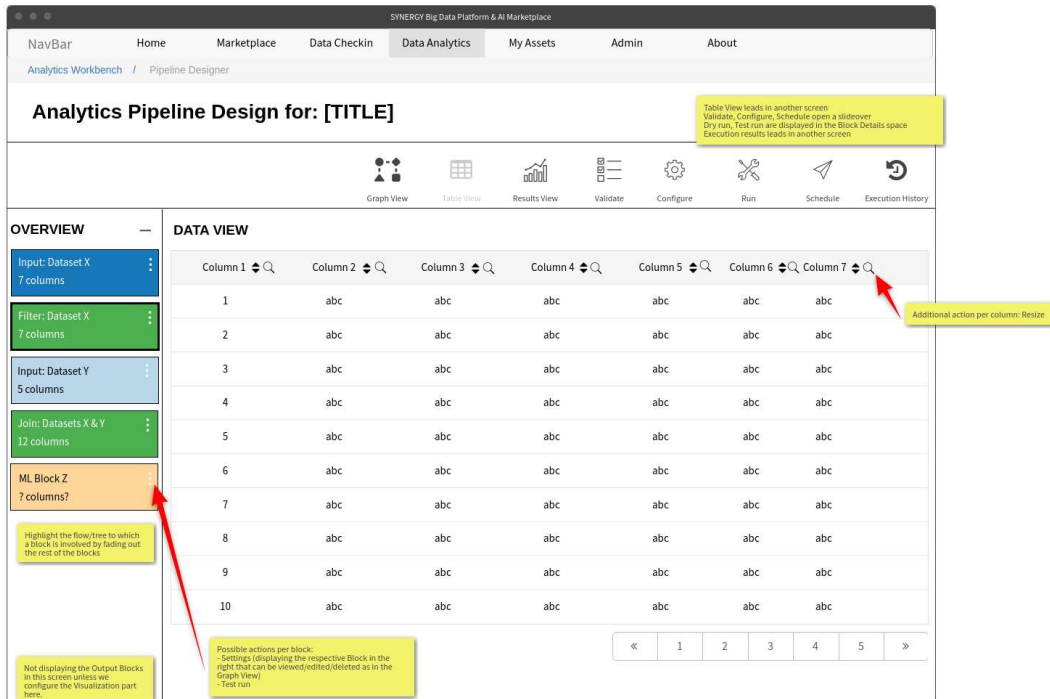


Figure 63: Analytics Pipeline - Table View

4.3.3 Results View

Once the data asset consumer is satisfied with the data analytics pipeline he/she has defined, the desired visualization of the results also needs to be configured, as shown in Figure 64. In particular, the data asset consumer may select the type of diagram that needs to be visualized from the Diagrams left slideover menu. This menu provides several types of diagrams for visualization such as line graph, scatter plot, bar chart, and others. Following that, the data asset consumer needs to configure the parameters that are to be visualized, from the Diagram Configuration right slideover menu, by selecting the parameters for each axis according to the type of the diagram. In addition, the data asset consumer may add a title, legends and axis labels for the diagram.

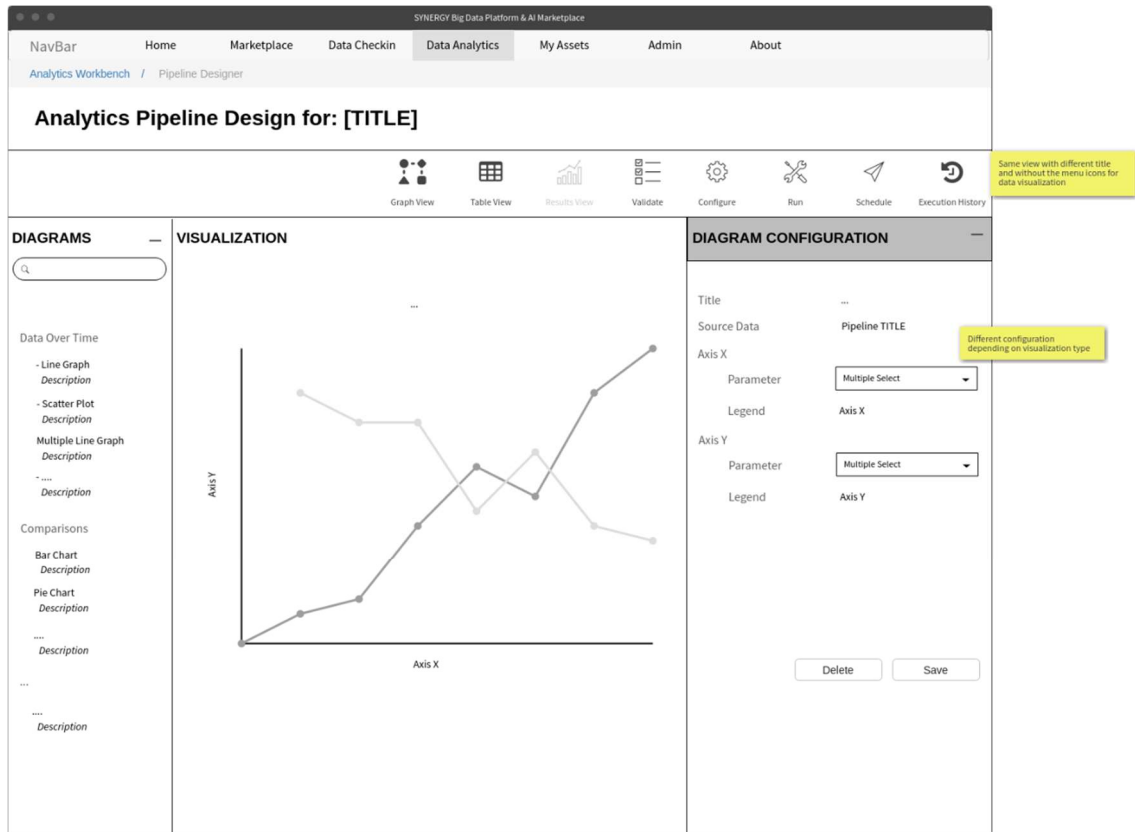


Figure 64: Analytics Pipeline – Results View

4.3.4 Schedule Execution

Apart from the blocks workflow and the results that need to be configured in an analytics pipeline, the data asset consumer may configure the scheduling of the pipeline execution, by selecting the Schedule option from the Analytics Workbench bar, as shown in Figure 65. A slideover to define an appropriate schedule is displayed, and the user can select either to run the pipeline execution by creating a schedule, or by defining an event-triggered mechanism. When a schedule is selected, the execution period (i.e., start date, end date) needs to be defined, as well as the details for each schedule (possibly multiple schedules) that includes the frequency of the execution (i.e., hourly, daily, weekly, monthly). When an event-triggered mechanism is selected, the data asset consumer may select to trigger the pipeline execution by different

conditions, such as when new data are available, or when new data are sent via an API call, and others.

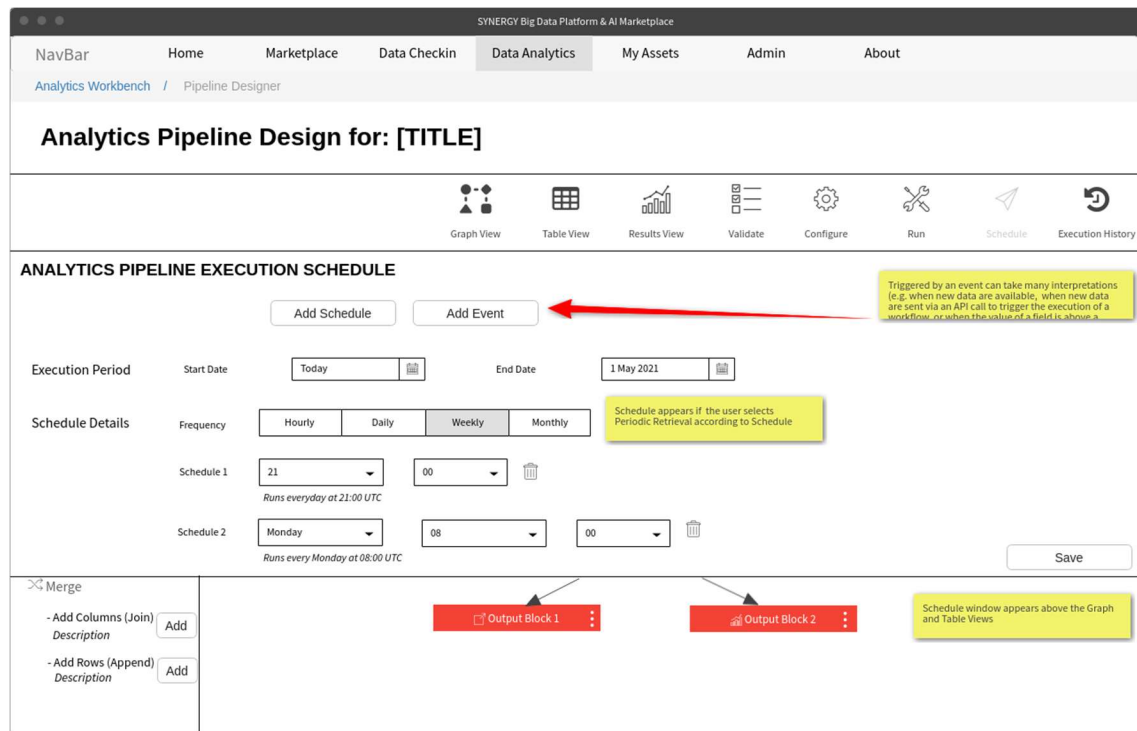


Figure 65: Analytics Pipeline - Schedule Execution

4.3.5 Execution History

By selecting the Execution History option from the Analytics Workbench bar, a new screen including the execution logs will appear as shown in Figure 66. The data asset consumer is able to 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 consumer 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 of the specific execution 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.

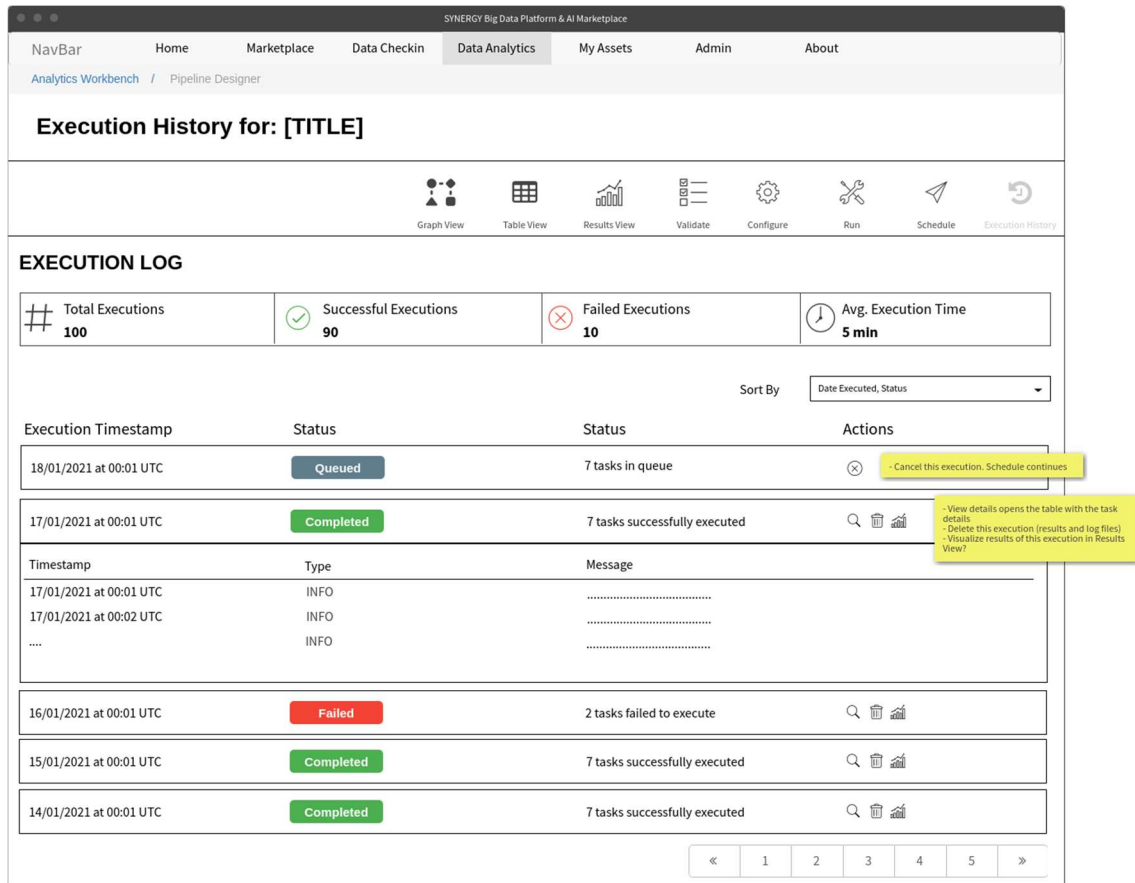


Figure 66: Analytics Pipeline - Execution History

4.4 Visualization Results of an Analytics Pipeline

Once an analytics pipeline has successfully been executed, the data asset consumer may view the results as shown in Figure 67. 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.

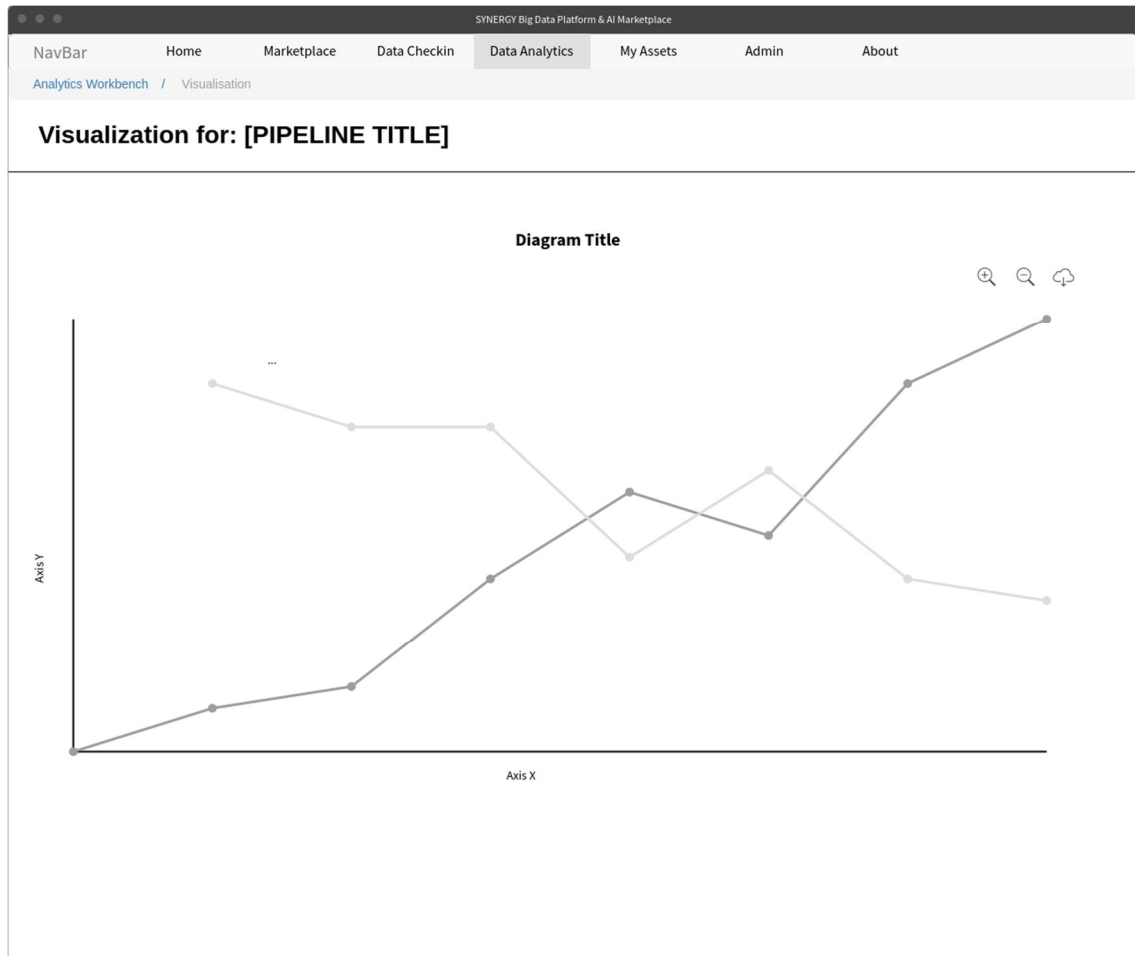


Figure 67: Analytics Pipeline – View Results of an Analytics Pipeline

5 Additional Platform Functionalities

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) in order to be able to perform different functionalities, e.g. edit user and organisation profiles, as well as the common information model lifecycle management for the CIM administrators.

5.1 Common Information Model (CIM) Manager

The CIM Manager is a functionality that is enabled for the platform’s administrator in order to manage the lifecycle of the Common Information Model. As it can be seen from Figure 67, an administrator is able to search for concepts (among active, all, deprecated, or proposed) from the left slideover menu.

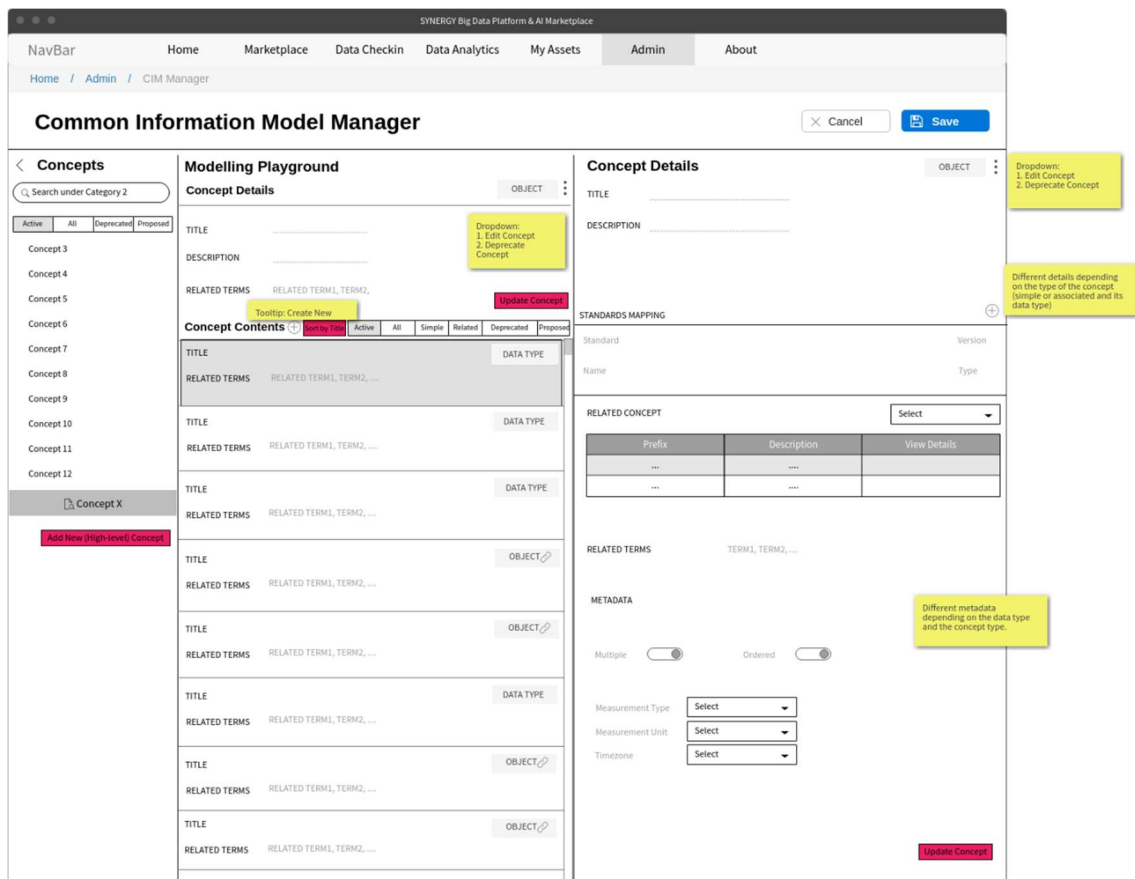
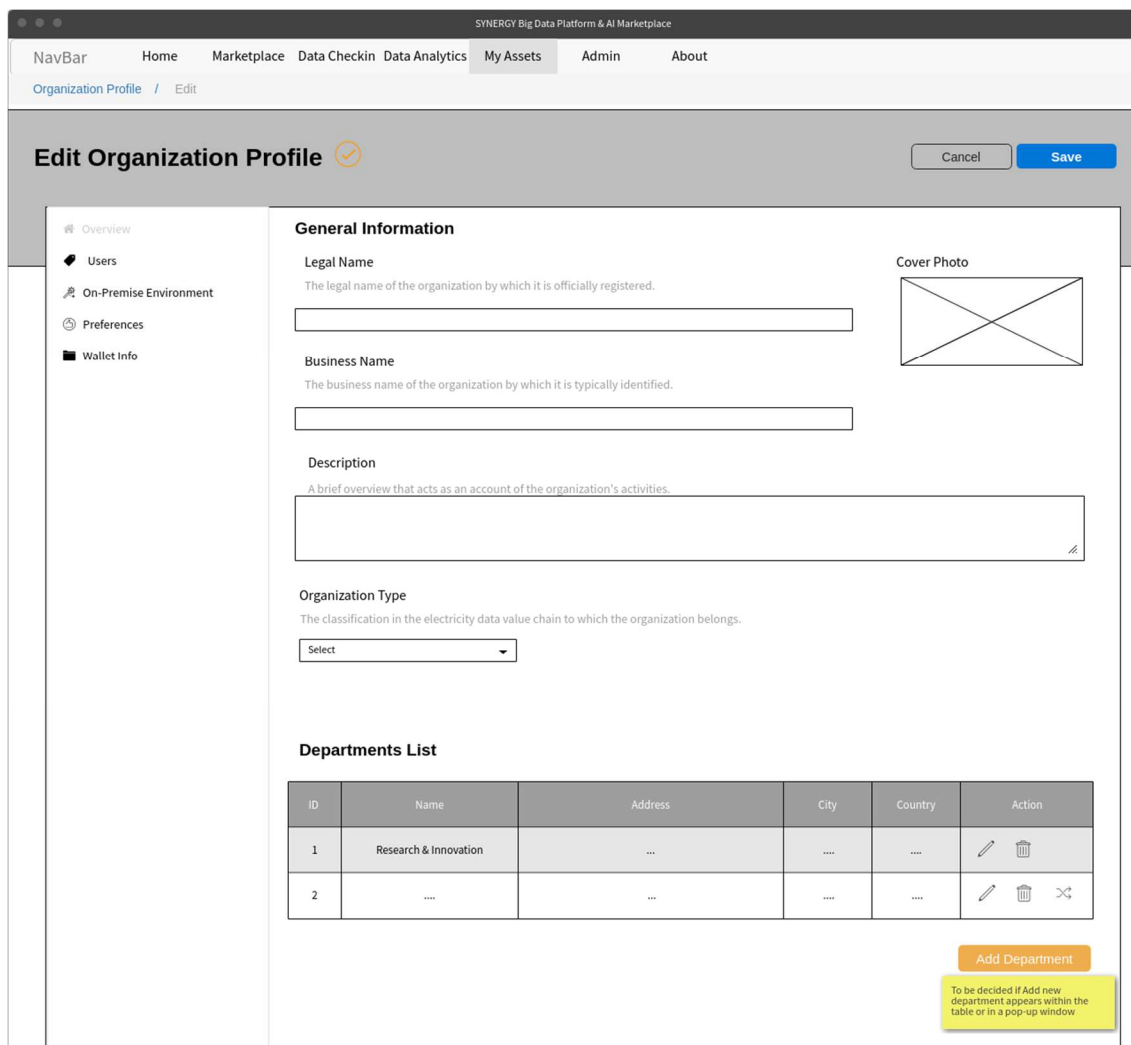


Figure 67: Common Information Model (CIM) Manager

By selecting a concept, he/she may view the concept details and the concept's fields under the Modelling Playground. 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 administrator is able to edit or deprecate a selected concept. Furthermore, by selecting a concept, a right slideover menu containing further details regarding the concept also appears, such as its title, description, mapping to different standards, related concepts and terms. Depending on the field's datatype, the CIM administrator needs to add different complementary metadata. Finally, the concept is updated by clicking the Update Concept button, 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 concept available in the Mapping stage (described in section 2.2.2.1). It needs to be noted that any proposed concepts by the data asset providers (as depicted in Figure 13) are also visible in the CIM Manager in order to allow the CIM administrator to take appropriate action.

5.2 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 68. 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.



SYNERGY Big Data Platform & AI Marketplace

NavBar Home Marketplace Data Checkin Data Analytics My Assets Admin About

Organization Profile / Edit

Edit Organization Profile

Cancel Save

- Overview
- Users
- On-Premise Environment
- Preferences
- Wallet Info


General Information

Legal Name
The legal name of the organization by which it is officially registered.






Business Name
The business name of the organization by which it is typically identified.

Description
A brief overview that acts as an account of the organization's activities.

Organization Type
The classification in the electricity data value chain to which the organization belongs.

Select 

Departments List

ID	Name	Address	City	Country	Action
1	Research & Innovation	 
2	  

Add Department

To be decided if Add new department appears within the table or in a pop-up window

Figure 68: 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 69.

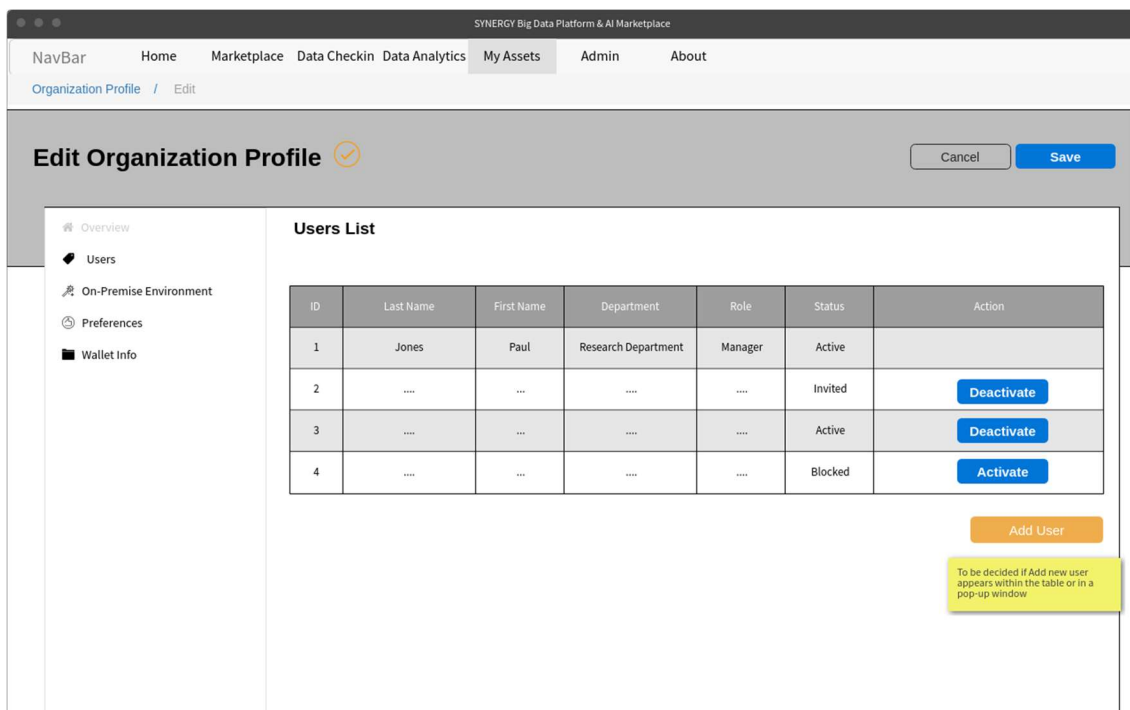


Figure 69: Edit Organisation Profile - Users List

5.3 Edit User Profile

The organisation’s members may edit their individual user profile, by navigating to the Edit User Profile page as shown in Figure 70. In particular, they need to fill some basic information (i.e., last name, first name, photograph), and information regarding their account (i.e., email, username). In addition to these, the users are able to change their current password to a new one. By selecting the Tokens tab, the users are able to view their tokens (to retrieve or upload data to the SYNERGY Platform through the SYNERGY APIs), to generate new ones, or even to delete existing tokens they own, as depicted in Figure 71.

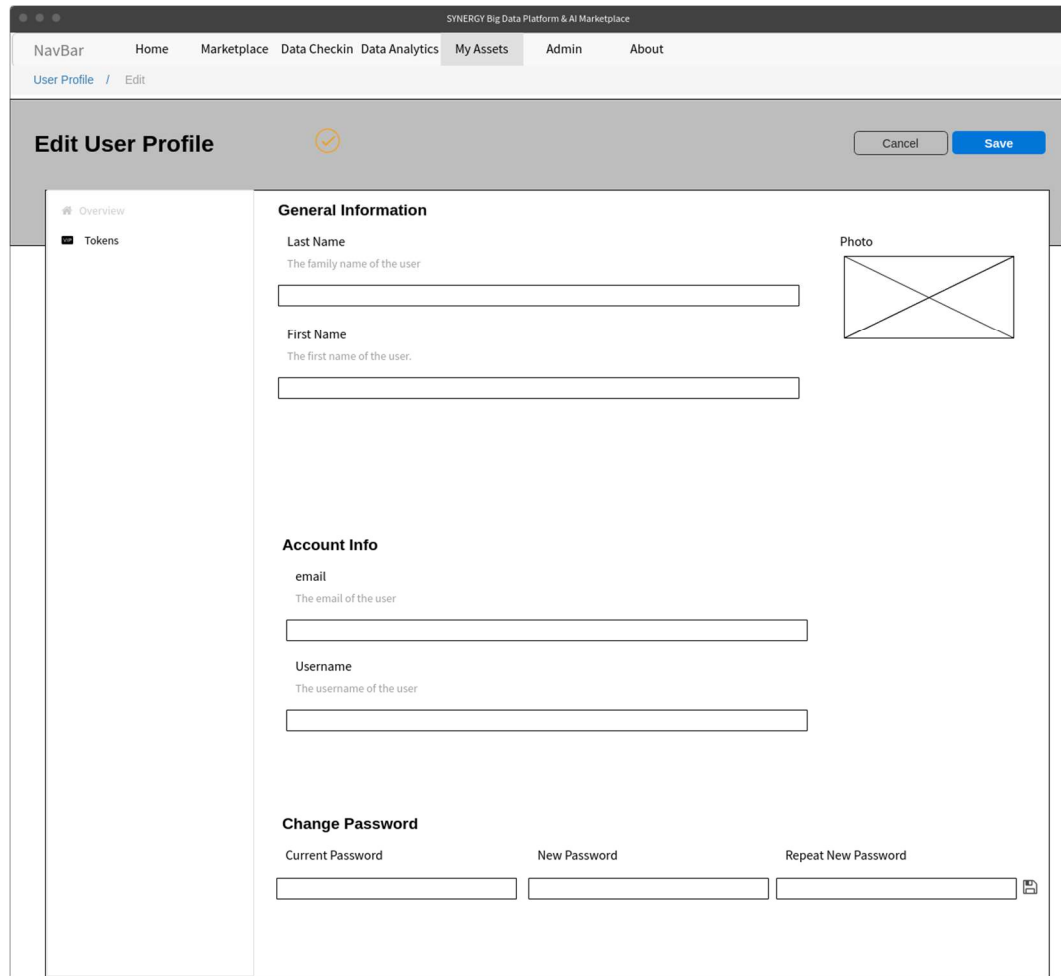


Figure 70: Edit User Profile - Overview

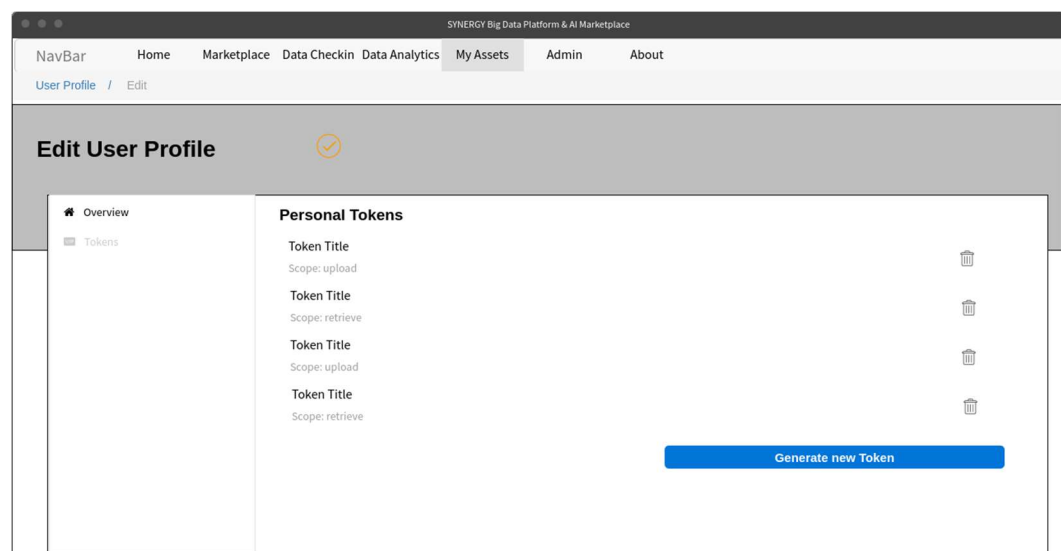


Figure 71: Edit User Profile - Tokens

6 Integration Plan

Taking into consideration the different user journeys described in sections 2-4, the core workflows that were described in detail in the SYNERGY Deliverable D2.6 (concerning the data check-in workflow, the data search and sharing workflow, the data analytics workflow) and the development activities across the different services bundles (as documented in D3.2 and D4.1), the interactions among the different SYNERGY platform components have been identified as depicted in the following table. It needs to be highlighted that 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 Data Storage Services Bundle communicates with all components and services in the SYNERGY Platform and has thus not been included in the following table.

Table 1: SYNERGY Components Interrelations

Component	Related Components with which there is interaction/integration
Data Collection Services Bundle	
Data Handling Manager	Matching Prediction Engine, Access Policy Engine, Master Controller
Matching Prediction Engine	Mapping & Transformation Service, CIM Manager
Data Ingestion Service	Master Controller
Mapping & Transformation Service	Master Controller
Cleaning Service	Master Controller
Data Security Services Bundle	
Anonymisation Service	Master Controller
Encryption Engine	Master Controller, Wallet Manager
Access Policy Engine	Data Handling Manager, Data & AI Marketplace, Analytics Workbench, API Gateway
Data Sharing Services Bundle	
Data & AI Marketplace	Query Builder, Matchmaking Engine, Contract Lifecycle Manager
Contract Lifecycle Manager	Remuneration Engine, Notifications Engine, Master Controller
Remuneration Engine	Wallet Manager
Wallet Manager	Encryption Engine, Remuneration Engine
Data Matchmaking Services Bundle	
Query Builder	Data & AI Marketplace, Matchmaking Engine, API Gateway
Matchmaking Engine	Query Builder, Data & AI Marketplace
Data Analytics Services Bundle	
Analytics Workbench	Visualization & Reporting Engine, Data & AI Marketplace, API Gateway
Visualization & Reporting Engine	Service Secure Results Export Service
Data Manipulation Service	Master Controller
Analytics Execution	Master Controller
Service Secure Results Export Service	Master Controller
Data Governance Services Bundle	

Component	Related Components with which there is interaction/integration
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
Data Lineage Service	Master Controller
CIM Manager	Matching Prediction Engine, Query Builder
Platform Management Services Bundle	
Resources Orchestrator	Master Controller
Notifications Engine	Master Controller, Contract Lifecycle Manager
Security, Authentication & Authorisation Engine	All Components & Services
Platform Analytics Engine	Data Handling Manager, Data & AI Marketplace, Contract Lifecycle Manager
API Gateway	Access Policy Engine, Contract Lifecycle Manager, Query Builder, Service Secure Results Export Service

The functionalities of the different services bundles have been carefully planned across the different components to ensure that they are aligned and appropriately prioritised to deliver a beta version that brings the most added value possible to the different electricity data value chain stakeholders. As already mentioned, the details of the back-end development towards the beta release of all SYNERGY Data Services Bundles are documented in detail in D3.2 and D4.1, yet their overall integration plan has been collaboratively created taking into consideration the present alpha, mock-up release. The details of the integration plan highlight all milestones, and the alignment between the core features of the different components as depicted in Figures 72-73.

It needs to be noted that the most critical functionalities that cross-cut different components are stressed in the SYNERGY Integration Plan, which shall be revised once the SYNERGY Integrated Platform in its beta version is released on M18 (as it will be documented in D3.4).

		Beta Release		Release 1.00	Release 2.00
		Q2 - 2021	Q3 - 2021	Q4 - 2021	(Q1-Q4) 2022
Data Collection Services Bundle	Data Handling Manager	UI Development, Integration	Stress Testing	DCJ templates (Cloning)	Improvements, Bug fixing
	Matching Prediction Engine	UI Development, Integration	Stress Testing	Additional Algorithms Training	Improvements, Bug fixing
	Data Ingestion Service	Integration, End-to-end Testing	Stress Testing	Improved execution handling	Improvements, Bug fixing
	Mapping & Transformation Service	Integration, End-to-end Testing	Stress Testing	Improved execution handling	Improvements, Bug fixing
Data Security Services Bundle	Cleaning Service	Integration, End-to-end Testing	Stress Testing	Improved execution handling	Improvements, Bug fixing
	Anonymisation Service	Integration, End-to-end Testing	Stress Testing	Improved execution handling	Improvements, Bug fixing
	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 Sharing Services Bundle	Data & AI Marketplace	UI Development, Integration	Data Assets (Algorithms/Models, Results) in Marketplace		Improvements, Bug fixing
	Contract Lifecycle Manager	UI Development, Integration	Multi-party Data Asset Contracts		Improvements, Bug fixing
	Remuneration Engine	Integration, End-to-end Testing	Frictionless Payments		Remuneration Payments
	Wallet Manager	Back-end Development	Support in OPE	Stress Testing	Improvements, Bug fixing
Data Matchmaking Services Bundle	Query Builder	UI Development, Integration	Search in Data Assets (Algorithms/Models, Results) beyond Datasets		Improvements, Bug fixing
	Matchmaking Engine	UI Development, Integration	Additional Recommendation Algorithms		Improvements, Bug fixing

Figure 72: Integration Plan – Part I

		Beta Release		Release 1.00	Release 2.00
		Q2 - 2021	Q3 - 2021	Q4 - 2021	(Q1-Q4) 2022
Data Analytics Services Bundle	Analytics Workbench	UI Development, Integration	Additional Execution Frameworks		Configuration of SMC in UI
	Visualization & Reporting Engine	UI Development, Integration	Additional Built-in Charts, Export/Save Diagrams		Improvements, Bug fixing
	Data Manipulation Service	Integration, End-to-end Testing	Additional Data Manipulation Functions		Improvements, Bug fixing
	Analytics Execution	Integration, End-to-end Testing	Additional Basic & Pre-trained Models		Secure Multi-party Computations (SMC) Execution
	Service Secure Results Export Service	Integration, End-to-end Testing	Stress Testing	Improvements, Bug fixing	Improvements, Bug fixing
Data Governance Services Bundle	Master Controller	Integration, End-to-end Testing	Stress Testing	Edge OPE Orchestration	
	Data Lineage Service	--	-	Back-end Development	UI Development, Integration
	CIM Manager	UI Development, Integration	Stress Testing	Improvements, Bug fixing	Improvements, Bug fixing
Platform Management Services Bundle	Resources Orchestrator	Integration, End-to-end Testing	Stress Testing	SEP Improvements, Bug fixing	Improvements, Bug fixing
	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
	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 73: Integration Plan – Part II

The different components and services are to be deployed in the SYNERGY Cloud Platform and its Secure Experimentation Playgrounds (SEP), and the On-Premise Environments (OPE)

according to the SYNERGY architecture (described in detail in D2.6). With regard to the On-Premise Environments, it needs to be clarified that there will not be any user interface, but a tray shall appear in the operating system (where the server OPE has been installed) when they are executed and may redirect the user to the Cloud Platform whenever needed. As part of the integration activities towards the beta release in M18, the secure transfer of data across the different layers (Cloud, SEP, OPE) is also currently developed and tested.



7 Conclusions

In this deliverable (D3.3 “SYNERGY Integrated Platform – Alpha, Mock-ups Release”), a thorough description of the SYNERGY Platform user journeys for configuring and executing the core platform functionalities (i.e., data check-in, data search and acquisition, data analytics), as well as some additional functionalities (i.e., CIM manager, edit organisation profile, edit user profile), was provided. Towards this end, this deliverable documented the alpha (high fidelity mock-up) version of the SYNERGY Platform across its basic functionalities in terms of workflows for data check-in, data search and acquisition, and data analytics, that users (i.e., data asset providers, data asset consumers) may configure and execute.

In particular, this deliverable (that consolidates inputs from both the WP3 and WP4 development activities) documented in detail:

- The data check-in user journey, describing the workflow that needs to be configured regarding the creation, configuration and execution of data check-in jobs, as well as the profiling of the resulting data assets.
- The data search and acquisition user journey which allows data asset consumers to browse within the SYNERGY Marketplace for data assets of interest, to prepare contracts in order to acquire data assets (i.e., datasets, algorithms, etc) by their respective providers, and to retrieve such assets from the SYNERGY Platform, according to their preferences.
- The data analytics user journey for both the data asset providers and data asset consumers, which offers the data analytics pipeline design and execution, in order to provide valuable insights for their own and acquired data (by visualizing the appropriate results or retrieving them according to their terms).
- The additional functionalities, which are available within the SYNERGY Platform (i.e., CIM manager, the organisation profiles and the user profiles).
- The integration and release plan of the integrated SYNERGY platform, which provides information regarding the components’ different dependencies, and high-level functionalities that cross-cut different components and shall be supported in future releases.

This alpha version of the SYNERGY Integrated Platform was discussed in different technical meetings in order to bring the demo partners and the energy apps to the same page along the expected functionalities. It shall also serve as the foundation for the front-end development and integration activities in WP3 and WP4 towards the beta release of the SYNERGY Platform that is expected in M18.



8 References

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

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

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"

