H2020-LC-SC3-EE-2019

EUROPEAN COMMISSION

European Climate, Infrastructure and Environment Executive Agency

Grant agreement no. 893857

frESCO

New business models for innovative energy services bundles for residential consumers

Project acronym	**frESCO**
Full title	New business models for innovative energy service bundles for residential consumers
Grant agreement number	893857
Programme	H2020-EU.3.3.1. - Reducing energy consumption and carbon footprint by smart and sustainable use
Topic identifier	LC-SC3-EE-13-2018-2019-2020 - Enabling next-generation of smart energy services valorising energy efficiency and flexibility at demand-side as energy resource
Call	H2020-LC-SC3-EE-2019
Funding scheme	IA – Innovation Action
Project duration	42 months (1 June 2020 – 30 November 2023)
Project adviser	Rebecca Kanellea - CINEA
Coordinator	CIRCE – Fundacion Circe Centro de Investigacion de Recursos y Consumos Energeticos
Consortium partners	CIRCE, S5, EI-JKU, CARTIF, UBITECH, UBE, KONCAR KET, KRK, COSMA, LCTE, VOLT, VERD, IOSA, RINA-C
Website	http://fresco-project.eu
Cordis	https://cordis.europa.eu/project/id/893857

DISCLAIMER OF WARRANTIES

This document has been prepared by frESCO project partners as an account of work carried out within the framework of the EC-GA contract no. 893857.

Neither Project Coordinator, nor any signatory party of frESCO Project Consortium Agreement, nor any person acting on behalf of any of them:

(a) makes any warranty or representation whatsoever, expressed or implied,

(i). with respect to the use of any information, apparatus, method, process, or similar item disclosed in this document, including merchantability and fitness for a particular purpose, or

(ii). that such use does not infringe on or interfere with privately owned rights, including any party's intellectual property, or

(iii). that this document is suitable to any particular user's circumstance; or

(b) assumes responsibility for any damages or other liability whatsoever (including any consequential damages, even if the Project Coordinator or any representative of a signatory party of the frESCO Project Consortium Agreement has been informed of the possibility of such damages) resulting from your selection or use of this document or any information, apparatus, method, process, or similar item disclosed in this document.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement n. 893857

Disclaimer: *The European Commission is not responsible for any use made of the information contained herein. The content does not necessarily reflect the opinion of the European Commission.*

Deliverable D4.4
frESCO Integrated Platform – Alpha, Mockups Release

Deliverable number	D4.4
Deliverable name	frESCO Integrated Platform – Alpha, Mockups Release
Lead beneficiary	Suite5
Description	The alpha version (with limited functionality) of the Integrated frESCO platform
WP	WP4
Related task(s)	Task 4.3. Core Big Data Ingestion, Curation and Management Services Task 4.4. Personal and Enterprise Big Data and Edge Analytics Algorithms Definition as a baseline for advanced energy service bundles Task 4.5. Data Assets Security, Encryption and Privacy Mechanisms
Type	Other
Dissemination level	Public
Delivery date	11.11.2021
Main author	Giorgos Papadopoulos (S5), Alexandros Tsitsanis (S5), Tasos Tsitsanis (S5)
Contributors	

Document history

Version	Date	Changes	Author
V1 – first draft	06.10.2021		S5
V1 – reviews	11.10.2021		UBITECH
V1 – reviews	15.10.2021		CIRCE
V1 – consolidated version	11.11.2021		S5
Final version	11.11.2021		S5
Final deliverable submission	12.11.2021		CIRCE

ABBREVIATIONS

Abbreviation	Name
API	Application Programming Interface
CA	Consortium Agreement
CIM	Common Information Model
CSV	Comma Separated Values
D	Deliverable
DER	Distributed Energy Resources
EC	European Commission
ESCO	Energy Service Company
ICT	Information and Communication Technology
GA	Grant Agreement
H2020	Horizon 2020 The EU Framework Programme for Research and Innovation
JSON	JavaScript Object Notation
PDF	Portable Document Format
PubSub	Publish-Subscribe
SASL	Simple Authentication and Security Layer
URL	Uniform Resource Locator
XML	Extensible Markup Language
WP	Work package

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Deliverable 4.4 presents the alpha, mock-ups release of the frESCO Integrated Platform, outlining also the modules that comprise the platform (Data Collection, Data Security and Storage, Data Analytics, Data Search, Platform Governance). Building upon the frESCO conceptual architecture, as presented in D2.5, in this document, all modules and their respective features are described to such an extent, in order to depict in detail, the range of functionalities that the frESCO Platform users will be enabled with.

D4.4 provides a detailed description of the frESCO Platform user flows for configuring and executing the core platform functionalities (i.e., data collection, data search and retrieval, data analytics), as well as different supplementary functionalities (i.e., User Profile Management, Organization Profile Management, Notification Alerts). To this purpose, nineteen (19) mock-ups have been designed, to illustrate and thoroughly describe the most crucial features that are offered in the three fundamental user flows, but also present the supplementary functionalities that will be offered to the frESCO Platform users. Indicatively, the user flows are briefly described as: (a) the data collection user flow, which outlines the process that data asset providers use to upload data; b) the data search and retrieval user flow, which describes the functionalities that the users of the platform will be offered, in order to search, explore, and retrieve data; and c) the data analytics user flow, which aims to provide valuable insights to the frESCO Platform users, on the data assets they own, with the deployment of appropriate pre-trained algorithms.

This deliverable introduces and describes gradually all the fundamental functionalities, that are included in the alpha release of the frESCO Integrated Platform, through the use of appropriate mock-ups, serving as a foundation towards the frESCO Integrated Platform Beta Release on M20 (January 2022), that will be documented in the frESCO Deliverable D4.6.

INTRODUCTION

1.1 Purpose and target group

The alpha, mock-ups release of the frESCO Integrated Platform is described in Deliverable 4.4, as well as the modules that constitute the platform (Data Collection, Data Security and Storage, Data Analytics, Data Search, Platform Governance). The extent of the functionalities that the frESCO Platform users will be able to access, is defined by describing all modules and their associated features.

D4.4 describes the frESCO Platform user flows for configuring and performing the fundamental platform functions (data collection, data search and retrieval, and data analytics), as well as various supplementary functionalities (User profile Management, Notification Alerts).

1.2 Scope of the document

D4.4 aims to delineate and provide proper guidance on the user flows that the users of the frESCO Platform, such as energy service beneficiaries and service providers will be enabled with. In particular, the scope of this document is:

- To define the core user flows, together with the various supplementary functionalities, that the frESCO Platform users will be enabled with. Every user flow and its associated features are detailed thoroughly, in the sections that follow.

- To establish this alpha version of the frESCO Integrated Platform as the basis for the integration activities of the platform's beta release of the various Data Services Bundles, as well as to describe the connections among the different modules that are presented in this deliverable, which shall also be extensively described in the respective deliverables (frESCO Deliverables 4.3 (M18), 4.5 (M18), 4.6 (M20)).

1.3 Structure of the document

The structure of the document is organized as follows:

Section 2 presents a descriptive overview of the frESCO Big Data Management Platform, the different modules that comprise the platform (Data Collection, Data Security and Storage, Data Analytics, Data Search and Platform Governance), together with the various features of each module.

Section 3 delineates the user's data collection flow. An extensive data collection workflow is offered, in order to guide the frESCO platform users through the data collection functions, during the design and implementation phases.

Section 4 describes the data analytics user flow. Pre-trained algorithms will be executed over data made available in the platform and once an analysis is completed, the users of the frESCO platform will receive useful insights via the visualization of the analytics results. Also, through appropriate mechanisms, users will be able to export or download these analytics results.

Section 5 provides a detailed overview of the data search and retrieval user flow. This workflow outlines the actions that users shall follow, in order to search for data assets in the frESCO Platform and retrieve them using appropriate functions.

Section 6 presents the various supplementary functionalities that will be offered to the users of the frESCO Platform.

Section 7 concludes this deliverable D4.4 "frESCO Integrated Platform – Alpha, Mock-ups Release", providing a brief summary of what has been presented in the different sections of the deliverable.

Section 8 includes the list of relevant references.

2 BIG DATA MANAGEMENT PLATFORM AND MODULES OVERVIEW

As part of the backbone for all integration activities in the frESCO project, the Big Data Management Platform poses an essential ICT framework to allow interoperable and secure data acquisition and processing, to enable the deployment of novel energy services. The platform will contribute both to the granular communication and data exchange between numerous sources, such as buildings, DER management systems, weather data sources and wholesale energy prices, as well as to the optimal mass ingestion and storage of big volumes of the corresponding data sets. Suitable techniques and methods for data importing, curation, and semantic harmonization will be utilized, to ensure high performance and ability to adapt to the requirements and needs of the value chain stakeholders. In addition, the Big Data Platform will also offer a big data analytics module, that essentially allows the preparation and execution of analytics algorithms, along with reporting of the respective results, utilizing a catalogue of pre-trained analytics models, to generate new insights and knowledge for all frESCO stakeholders. The Big data platform consists of the following main modules, which will be further analysed in the next sections and are illustrated in Figure 1, namely:

- Data Collection Module
- Data Security and Storage Module
- Data Analytics Module
- Data Search & Retrieval Module
- Platform Governance Module

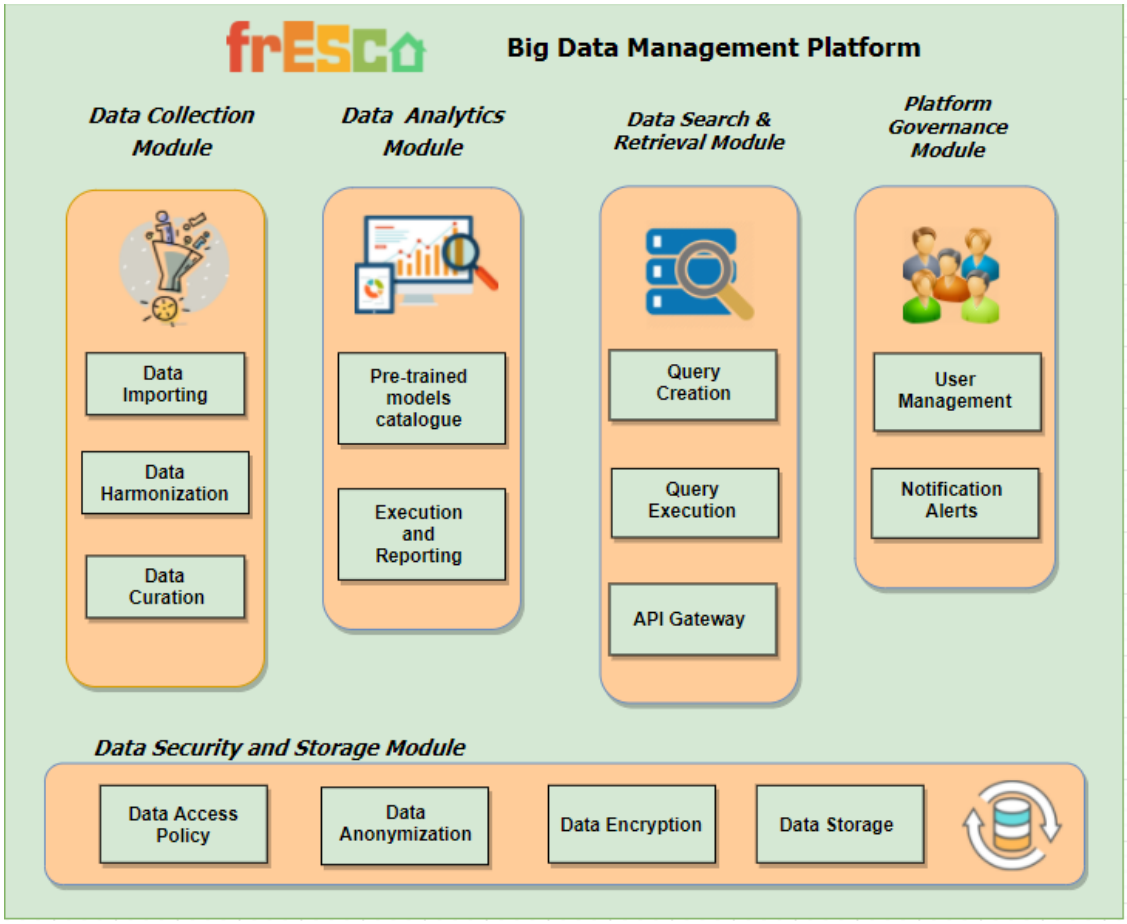

Figure 1 frESCO Big Data Management Platform and Modules Overview

In comparison to the deliverable D2.5 “Report on the FRESCO conceptual architecture”, where an initial representation of the Big Data Management Platform was provided, the present deliverable encapsulates several changes conducted in the different modules that are going to be implemented towards the release of the frESCO Platform. Briefly, the changes that were made are presented as follows:

- The Data Handling feature was removed from the Data Collection Module, as its functionalities can be addressed by the Data Importing feature.

- The API gateway feature was transferred to the Data Search & Retrieval Module, since it responds to API calls and conducts the retrieval of the data requested.

It is worth mentioning that some renaming on the different features of the frESCO modules was also performed, namely:

- Data Semantic Mapping feature was renamed to Data Harmonization.
- Data Storage and Indexing feature was renamed to Data Storage.
- Notifications feature was renamed to Notification Alerts.

Finally, the Platform Orchestration Module was renamed to Platform Governance Module.

2.1 Data Collection Module Overview

The Data Collection Module, responsible for the data ingestion process in the frESCO Big Data Management Platform, is considered as a fundamental part in the overall Big Data Management Platform configuration since it will provide a variety of functions and features, spanning data importing from various sources, and semantic harmonization of the data imported in the platform, together with data quality assurance for ensuring the completeness and integrity of the data to be later processed within the platform or provided to the overlay components of the architecture (consisting of the components and applications that will be responsible for the end-user services of the frESCO project). The key features involved in the Data Collection Module are presented in the next sub-sections.

2.1.1 Effective Handling of Data Management Processes

Through effective handling of data management processes, the insertion of data from different sources into the frESCO Big Data Management Platform is performed in a simple and straightforward way. The data asset providers, by specifying how, when and what data they own and are ready to share, shall utilize the platform, in order to:

- Configure the parameters of the data ingestion process, by defining their needs and preferences regarding the method that the data will be inserted in the platform (API, file uploading, streaming data ingestion).
- Perform proper harmonization actions, as the user interface of the frESCO platform will provide guidance to the users, in such a way that all the attributes of the inserted data will be harmonized based on the frESCO Common Information Model (CIM).
- Define the steps for data curation, safeguarding both the quality and value of the ingested data.
- Determine appropriate rules for effective anonymization of their data, in order to avoid sharing any potential “personal” information within their data.
- Decide whether data encryption should be performed over the data. The users will also have the option to upload their data without applying any encryption parameters.

2.1.2 Flexible Data Import Mechanisms and Methods

Offering multiple data importing alternatives, the user can perform the necessary steps to introduce the raw data into the frESCO Big Data Management Platform. In order to support the stakeholder needs and the various data forms, the frESCO platform allows the consumption of data through files, through APIs and through the usage of Publish/subscribe messaging mechanisms, a form of asynchronous service-to-service communication used in serverless and microservices architectures [Pub/Sub Messaging 2021]. Different steps and configuration options need to be followed for each one of the different data import methods. The platform typically supports cloud execution, but it may be extended to on-premise execution (in future releases) to accommodate the security needs of the stakeholders.
Data import activities are initiated by the user, who is authorized to perform the following actions:

- Import data from files: allowing for data retrieval from files in formats that can be processed (e.g., csv).

- Import data from APIs: allowing data retrieval from pilots' systems APIs, and from Open Data APIs (e.g., weather data, other local sources).
- Secure and reliable data upload: Utilizing secure and reliable mechanisms for uploading the data owned by any organisation towards the frESCO Big Data Management Platform infrastructure

The platform provides the means for flexible data importing into the frESCO Data Management Platform, based on the configuration performed by the user. The feature can handle various data import methods to support stakeholders needs, making it possible to consume data through APIs from the API Gateway. Each one of these methods has different configuration options, concerning scheduling (when data import should be performed through APIs), and authentication aspects.

This feature is triggered for execution on demand, based on the actions performed by the user. The imported data are stored in a temporary object storage for increased reliability and roll- back options in case of failure in any other subsequent feature and process performed in the module.

2.1.3 Enabling Data Interoperability through Semantic Harmonization into the Common Information Model (CIM)

As described in section 2.1.1, the users of the frESCO platform will be able to perform proper harmonization actions, so that all the attributes of the inserted data will be harmonized to the frESCO Common Information Model (CIM). Semantic harmonization is an important part of the data handling process and acts as an enabler for functionalities offered by other Data Management Platform components. Through this feature, the platform users will be able to:

- Match the imported data to the frESCO CIM by performing appropriate changes to the respective data attributes to match the appropriate CIM entity names.
- Perform the necessary harmonization actions on the imported data to comply with the frESCO CIM. With this feature, the data values are matched to the CIM measurement units and the respective data types, for example.

2.1.4 Effective Data Curation for Data Quality Assurance Enhancement

The Data Collection module also ensures that data imported into the frESCO Data Management Platform is correct and full, according to the data owner's experience. Data quality and reusability are improved by removing or correcting incomplete and inconsistent data, making insights extraction more reliable. In this regard, the Data Collection module will implement the necessary data curation techniques, such as simple value substitutions, reformatting, and duplicate removal, as well as more complex duties like outlier identification and substitution. In more detail, this feature will allow for:

- Execution of data curation rules: During the curation configuration process, the user specifies the constraints and limitations that the data ingested into the module may have, as well as the precise actions that must be taken if any of these constraints are breached. This configuration corresponds to a collection of curation rules that are carried out by executing the appropriate data curation procedures. Depending on the data type of each column/attribute, a variety of data validation methods are available and are complemented by two sorts of corrective measures: removing entries and changing values. A curation rule is created by combining a validation option with a corrective action for a specific attribute, and all curation rules are

applied to the changed data (coming from the semantic harmonization to the fresco CIM), resulting in curated data.

2.2 Data Security and Storage Module Overview

The Data Security and Storage Module's role is twofold. Firstly, it assumes the responsibility to address data security and privacy concerns of the Data asset providers, concerning the data that will be imported and handled in the frESCO Data Management Platform. For that reason, it offers several functions and features, that can be intuitively configured in the platform, and cover Data Access policy definitions, Anonymization and Encryption capabilities. Secondly, this module covers the need for reliable storage and indexing of the data, by providing resilience strategies and various indexing methods. The features of this module are described in the following sub-sections.

2.2.1 Easy Data Access Policy configuration

The Data Security and Storage module includes a data access policy feature that increases users' trust in the entire frESCO Data Management Platform by allowing them to set access rules that control access requests on their data within the platform in a flexible and straightforward way. For each data asset, permission policies can be defined to achieve the desired behavior. The Data Access Policy feature effectively ensures a suitable separation of concerns between policy definition and policy enforcement, guaranteeing that data can only be seen and consumed by eligible platform users. In more detail, this feature will allow for:

- Definition, configuration, and modification of access policies: Users can build sophisticated access policy rules for their assets using a graphical user interface. These criteria might be based on data properties as well as the requestor's characteristics. Policies can specify when access is permitted, and they can be coupled to construct complicated rules using Boolean logic. The rules are saved and can be simply modified by the Data Owner using the provided interface.
- Data access policies are enforced: When a request is made to access data that has been imported into the frESCO Data Management Platform, the access policies configured are applied. The Data Access Policy feature enables data access in a performant manner, affecting both the Data Search Module results and the data to be received through the Data Retrieval module.

2.2.2 Flexible Data Anonymization Methods

The Data Security and Storage Module provides a mechanism to protect data from inadvertent disclosure of personal or corporate confidential information. Prior to making the data available on the frESCO platform, anonymization actions that reflect into specified parts of the data (i.e., which attributes) are performed. frESCO platform users are provided with the tools they need to anonymize it, as well as any attributes in their data that they believe identifiable information might be included.

More specifically, this feature will enable:

- Attributes that require anonymisation: Data Anonymisation allows data asset providers to identify any data attributes that include sensitive or identifying personal information according to GDPR.
- Execution of data anonymization rules: Based on the type of the attribute that the data owner defines (personal, identifying, quasi-identifying) and the data type, the Data Anonymisation feature allows the data owner as a next step to define the anonymisation method that should be applied in straightforward way. Such methods include generalization approaches to create arithmetic intervals or categories for numeric attributes, and masking approaches for string attributes.

2.2.3 Flexible Data Encryption Methods

The Data Security and Storage Module provides data asset providers with data encryption capabilities as additional mechanisms to protect their data and prevent the risk of unauthorised data access and/or data leakage of any kind. To accommodate this functionality, the execution of data processing steps can be offered to the frESCO platform users also on-premise, with the utilization of an executable file that runs locally, and allows the encryption of the available data, before being uploaded to the frESCO Data Management Platform. Based on the evaluation of users' requirements, the implementation of the on-premise functionality, will be considered in future releases of the frESCO Data Management Platform and will not be part of the beta release that is expected on M20.

The Data Encryption mechanism allows for modular control over the data contents that should be encrypted by allowing for flexibility in the definition of the encryption configuration to be used for the data transfer. The engine, in particular, allows the data source to encrypt the entire data asset and allows for:

- Execution of data encryption rules: The Data Encryption mechanism provides symmetric encryption techniques and is responsible for both the production of the encryption key and the actual encryption of the underlying data, allowing the data owner to prevent unauthorized access to the data.
- Data decryption: The Data Encryption mechanism offers the services required to decrypt data so that it can be accessed by the frESCO platform users who have been granted access. The system ensures that the decryption key is available so that the underlying information may be decrypted and accessed.

2.2.4 Applicable Data Storage and Indexing

Storing a plethora of data, along with their associated metadata, and the jobs-related data in a secure and reliable manner is provided by the Data Security and Storage Module. All storage and indexing tools used in the context of frESCO are taken under consideration and are mainly pertaining to:

- Storage of data: Storing the different data (i.e., data assets, analytics models, analytics results) along with their metadata, in order to be available for all the components and services in the frESCO platform.
- Storage of log data: Storing log-related information for the frESCO Data Management platform operation and usage, users' and organisations' data and all administrative information required for the smooth operations of the frESCO Platform.

- Storage of the CIM: Storing the frESCO Common Information Model in its different versions along with its associated entities and attributes.

Depending on the type of information that is to be stored in the frESCO Platform and the way it will be retrieved, different storage and indexing tools are foreseen to accommodate the varying needs.

2.3 Data Analytics Module Overview

The Data Analytics Module allows users in the frESCO platform Infrastructure to run analytics over their own data. It provides a catalogue of pre-trained analytics models that can be selected based on the user needs, and by executing and visualizing their results, leverage the added value that data analytics can bring through the offerings of the frESCO models.

2.3.1 Propagation of pre-trained model's catalogue

This feature pre-trains and populates the analytics catalogue of the core frESCO platform with selected data analytics algorithms that are prioritized by the frESCO demo cases and demonstrators. The pre-trained data analytics in frESCO are divided into two categories: I personal data analytics (for example, consumer energy behaviour, comfort preferences, and flexibility), and (ii) enterprise/industrial data analytics (indicatively involving Building Energy analytics, Forecasting and Flexibility analytics). It is worth mentioning that if the requirement for appropriate algorithmic models to be executed on the edge is identified, it will be examined in future releases of the frESCO platform. In more detail this feature will allow for:

- Catalogue of Personal Data Analytics: Data asset providers will be able to leverage a number of pre-trained personal data analytics models that span a variety of comfort, energy, and flexibility characteristics. Because energy management optimization at the end-user level should not affect their comfort and well-being, pre-trained analytics methods will wisely embed user preferences, translating them into user comfort models and providing this perspective in flexibility profiling and energy management optimization, to ensure that end-user comfort levels are always maintained.
- Catalogue of Industrial Data Analytics: Users of the frESCO platform will have access to pre-trained models that provide analytics capabilities, and target specific energy domain problems, such as short-term demand forecasting, generation forecasting, and DER flexibility analytics.

2.3.2 Intuitive Analytics Reporting

The Data Analytics Services Module includes a simple mechanism that allows the frESCO platform users to visually obtain insights of the execution of the pre-trained models, by verifying the generated visualizations. In more detail this feature will allow for:

- Built-in visualization and reporting: The Analytics Execution and Reporting mechanism includes predefined charts for various analytics processes that are fed with data to create useful visualisations.
- Saving and exporting visualisations and reports: All assets created with the Analytics Reporting mechanism can be saved as static objects (i.e., image) (in which case a configuration template is created and used to render the updated results every time it is accessed).

2.4 Data Search Module Overview

The Data Search Module constitutes an essential part of the Big Data Management Platform of frESCO, enabling the users of the platform to search and discover data that can be proven useful, determine and define which of these available data are of importance and eventually, have a clear and thorough view on the provided results. The key features of the Data Search Module are described in the following sub-sections.

2.4.1 Dynamic Query Creation

Data asset providers will be provided with all necessary methods for data discovery and exploration as a result of this feature. They can conduct searches for data that are relevant to their needs, explore the results, and dig further into the data to select candidates for retrieval. In more detail this feature will allow for:

- User-friendly data search that supports keyword-based and metadata-based searches: The Query Creation feature enables query development utilizing both a flexible free-text search and filtering on the data assets' information to give users more flexibility in how they search for and discover data. Query Creation, in particular, allows users to search for assets based on metadata, content, and meta-information about those items.
- Search for Analytics Reports: The Query Creation feature allows for a more straightforward data search method that supports different types of assets, such as analytics results.

2.4.2 Adjustable Query Execution processes

The customizable Query Execution feature builds on the data discoverability and exploration capabilities of the Query Creation feature by providing a built-in way to execute the query created, providing a brief overview of the results that will be generated. As a result, the user can quickly determine whether the findings satisfy his or her expectations and make any necessary adjustments or corrections. More specifically, this feature will enable:

- Query execution and results: The Query Execution feature converts the query configuration provided by the data creator into a query that can be performed in an optimal way in the platform, and results that match the query are returned, processed, and provided to the user.

2.4.3 Configuration of APIs for Data Retrieval

The configuration of APIs for data retrieval is an important aspect of the module because it acts as a single point of entry for apps (both frESCO and third-party) to retrieve data from the frESCO Platform, that they're allowed to access, as well as analytics findings. All API calls are accepted by the API gateway, which then retrieves and aggregates data from the different services required to reply to the call and returns the appropriate result. More specifically, this feature will enable:

- Retrieval of data and analytics results: Using the frESCO Open APIs, authorised apps can configure the retrieval of data from a single dataset, or from analytics findings. The use of filters represented by API request parameters and the selection of particular attributes of the data needed, are mechanisms for fine-tuning the received results. The API Gateway generates

a unique identifier, gives endpoint usage instructions, and includes a test API feature for quickly checking what results they return.

- Failure management: The API Gateway evaluates whether a partial answer or an error is preferable to be returned to the authorised frESCO application and any authorised 3rd party application, depending on the service that may be unresponsive or unavailable.
- Management of API keys: Based on the API keys created and kept in the platform, the API Gateway, in conjunction with the User Management mechanism, applies authorisation and access control to all data.

2.5 Platform Governance Module Overview

The Platform Governance Module in the platform of frESCO is of crucial importance, as it establishes the means and processes for secure and reliable registration of the users to the platform. Through appropriate authentication and safeguarding processes, the users are granted access to the data they are eligible to use and are offered the capability to receive analytics results, as well. Also, the users will be informed for the data import or data analysis occurring to the platform, according to their preferences, via platform notifications. The key features of the Platform Governance Module are described in the following sub-sections.

2.5.1 Secure and Reliable User Management

The User Management feature is a critical component of the module since it provides several features and layers of security, such as identity checks and reliable registration of and authentication of platform users. Authenticated users are only given access to data that they are authorized to use, according to the data access policy rules defined. More specifically, this feature will enable:

- Identity information for users on the frESCO platform: The User Management functions as an identity provider, creating and managing identity information for people who are qualified to use the frESCO platform.
- Authentication and authorisation services to components and services: The User Management feature, provides authentication and authorisation services to limit the access to authorized users, according to the data access policy rules.

2.5.2 Dynamic Personalized Notifications

Users of the frESCO platform will be able to be notified about certain events that occur on the platform and are relevant to them, based on the choices they have made. At key milestones, real-time information on the progress of ongoing import jobs and/or data analysis jobs will be updated. Notifications will be provided both within the frESCO platform and via email to ensure that relevant information is received by the appropriate stakeholders as soon as possible. More specifically, this feature will enable:

- Delivery of notifications for various events in the frESCO Platform: The Notifications feature generates and delivers relevant notifications for the progress of a data import job's execution (regarding successful completion each time it is executed, or details for any failure), the progress of a data upload execution (regarding successful completion each time it is executed, or details for any failure), and the progress of a data analysis execution (regarding successful completion each time it is executed).

- Management of notifications: The users can examine and act on the various notifications or delete them as needed. In addition, users can customize which notifications they want to receive based on their personal preferences.

3 DATA COLLECTION USER FLOW

The Data Collection Module is responsible for the data ingestion process in the frESCO Big Data Management Platform. It is considered as a fundamental part in the overall Big Data Management Platform configuration that will offer a wide range of features and functions, such as:

- Data importing from various sources;
- Data handling and semantic harmonization of the data imported in the platform;
- Data quality assurance, establishing the completeness and integrity of the data to be later processed within the platform Data Security and Storage Module overview.
- Data Anonymization to prevent accidental disclosure of personal or corporate information.
- Data Encryption that prevents the possibility of unauthorised access and any data leakage

In this section, the procedure that has to be adopted by the data asset provider, in order to load data on the frESCO Big Data Management Platform is presented. Initially, the data asset provider will have to initiate the workflow with a new data collection job creation and establish the process for the data ingestion, so that the data can be loaded on the frESCO Big Data Management Platform. Afterwards, appropriate pre-processing rules have to be defined with regard to the data asset, also registering the asset's metadata. After conducting the aforementioned procedure, the data asset provider performs the execution of the data collection job, leading to the storage of the data to the frESCO Platform. The different steps that need to be followed for the ingestion of data asset on the Big Data Management Platform are, further and explicitly, described in the following sections.

3.1 Data Collection Job Creation

The data collection workflow initiates with the creation and configuration of a new data collection job. Navigating through the frESCO Big Data Management platform, the data asset provider will be offered with the possibility to create a data collection job, by selecting the Data Collection option, at the side navigation bar, as depicted in Figure 2.

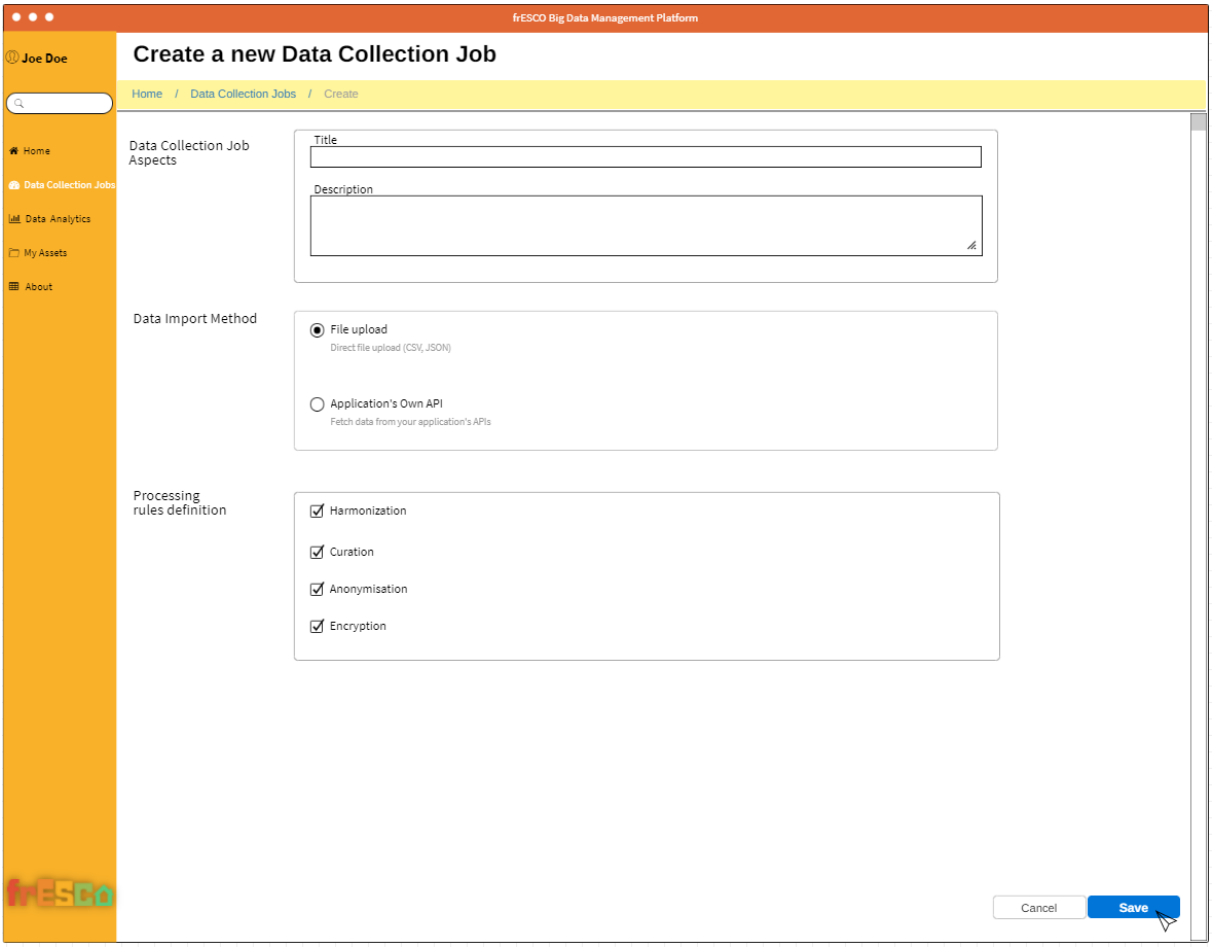

Figure 2 Data Collection Job Creation

The data asset provider will be requested to fill in some initial information about the data collection job, such as title and a brief description, in the Data Collection Job Details section. Subsequently, the data import mechanisms and methods will have to be defined, so that the data can be imported on the frESCO Big Data Management platform. Various data import methods can be handled, in order to support different needs, making it possible to upload data as files, and through APIs.

Additionally, an integral part for the creation of a data collection job is the definition of appropriate processing rules (i.e., Harmonization, Curation, Anonymisation, and Encryption) that need to be applied on the data to be stored on the platform, from the data asset provider. However, Harmonization stands as a prerequisite for the implementation of the different steps.

The data collection job is saved once the users make use of the “Save” button, at the right bottom corner of the screen and the users will be redirected to a new page, in accordance with the method applied for the data import.

3.2 Data Collection Job Configuration

In the following sections, the different means and methods offered to the frESCO platform users, in order to upload their data assets to the frESCO Big Data Management platform are presented. As anticipated from the previous section, after creating a data collection job, the data asset providers can handle different aspects of the data collection job.

3.2.1 Data Ingestion Configuration

3.2.1.1 Uploading Files

When data asset providers choose to upload their data utilizing files, they will be directed to the Data Import page. After landing on this page, the users will be requested to select the data file format of the data that are meant to be uploaded, as well as the matching file. In addition to the actual file, also a sample file is requested from the platform, allowing a quick verification from the user. This sample file should not exceed 50 rows of the entire data asset, and it applies to all file formats (csv, json, etc.,). Once the sample file is uploaded, the data asset providers will be offered with the opportunity to review the data sample particulars, by selecting the "Save&Proceed" button, as depicted in Figure 3.

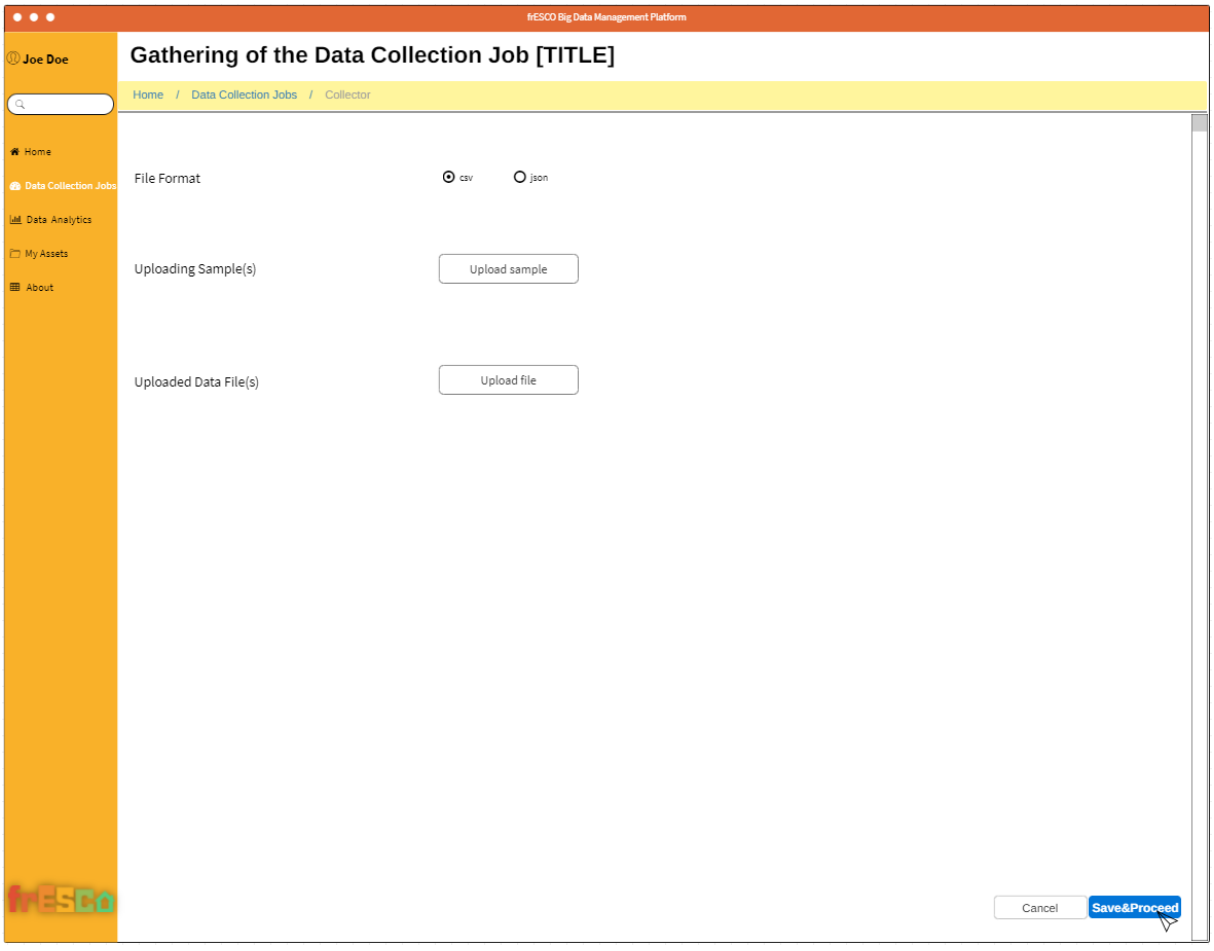

Figure 3 Uploading Files in the frESCO Platform

In the next step, the users can verify the list of the uploaded files and information on the uploading progress. They will also be presented with an option to remove files at their own volition, as depicted in Figure 4 .

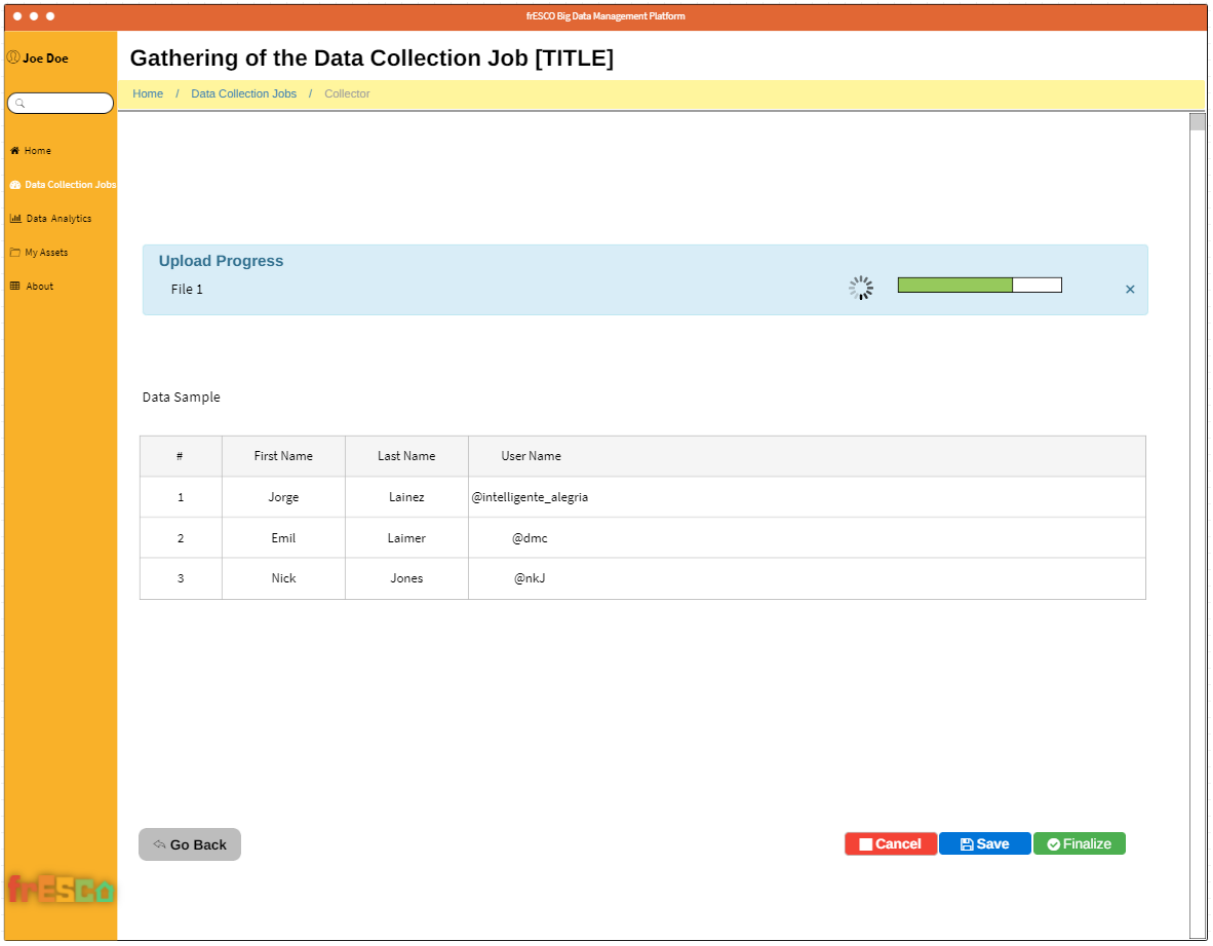

Figure 4 Sampling of the uploaded files

3.2.1.2 Application's own API method

Platform users that keep data in their organizations systems and expose them through APIs, can choose the option to import that data through the API retrieval functionality offered by the frESCO platform, utilizing the configuration page as depicted in Figure 5. Determining the type of authentication and providing the extra needs in terms of tokens and message bodies as required by the selection, is performed by the platform user. As next steps, the user needs to provide the full API URL, identifying the specific method to be used (i.e., GET, POST) and the query body of the request, if necessary.

Equally important during the configuration of the data import via API retrieval, is to identify the scheduling parameters when the API calls will take place. The platform user is offered by the platform with options to select the time period, with start and stop dates, as well the frequency of retrieval, once or periodically (i.e., hourly, daily, weekly, or monthly) based on a schedule that can be defined. A Polling option is also offered to the user, which by default is set to import data into the platform every minute.

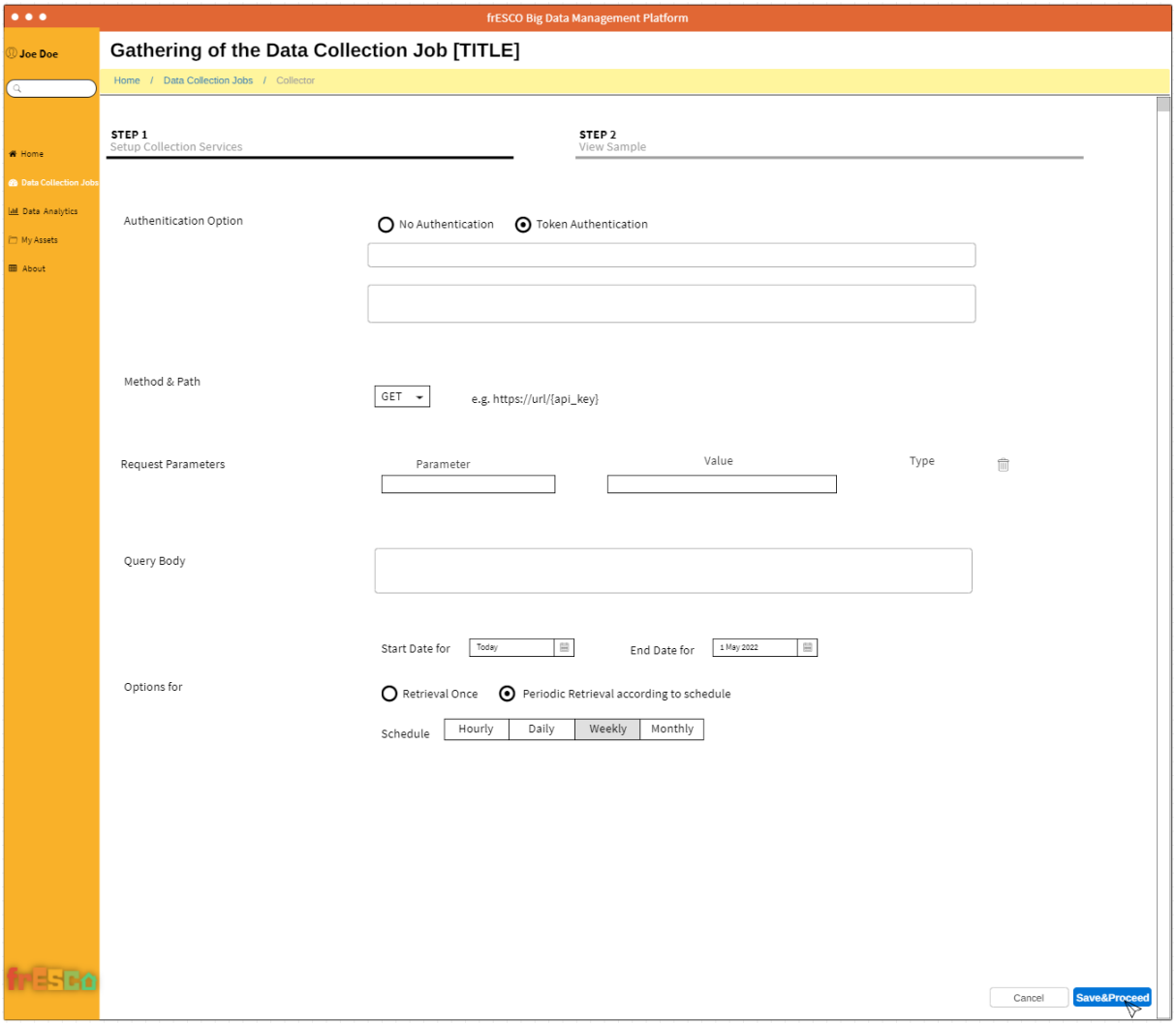

Figure 5 Publishing data through the application's own API mechanism

In the next step, as depicted in Figure 6, the frESCO platform initiates the API call to the previously defined endpoint. If the call is successful, the platform user is presented with the results of the API call and is able to select the entities that will be processed in the next steps. A summary of the structure that will be stored, including selected entities is offered to the user, to quickly check the data that will be stored in the platform. In case of unsuccessful API calls, then the user is properly informed about the problem along with the server response error message, if available.

At any time, the platform user can save the configuration by selecting the “Save” button and conclude the process by selecting the Finalize step, if no more changes are needed.

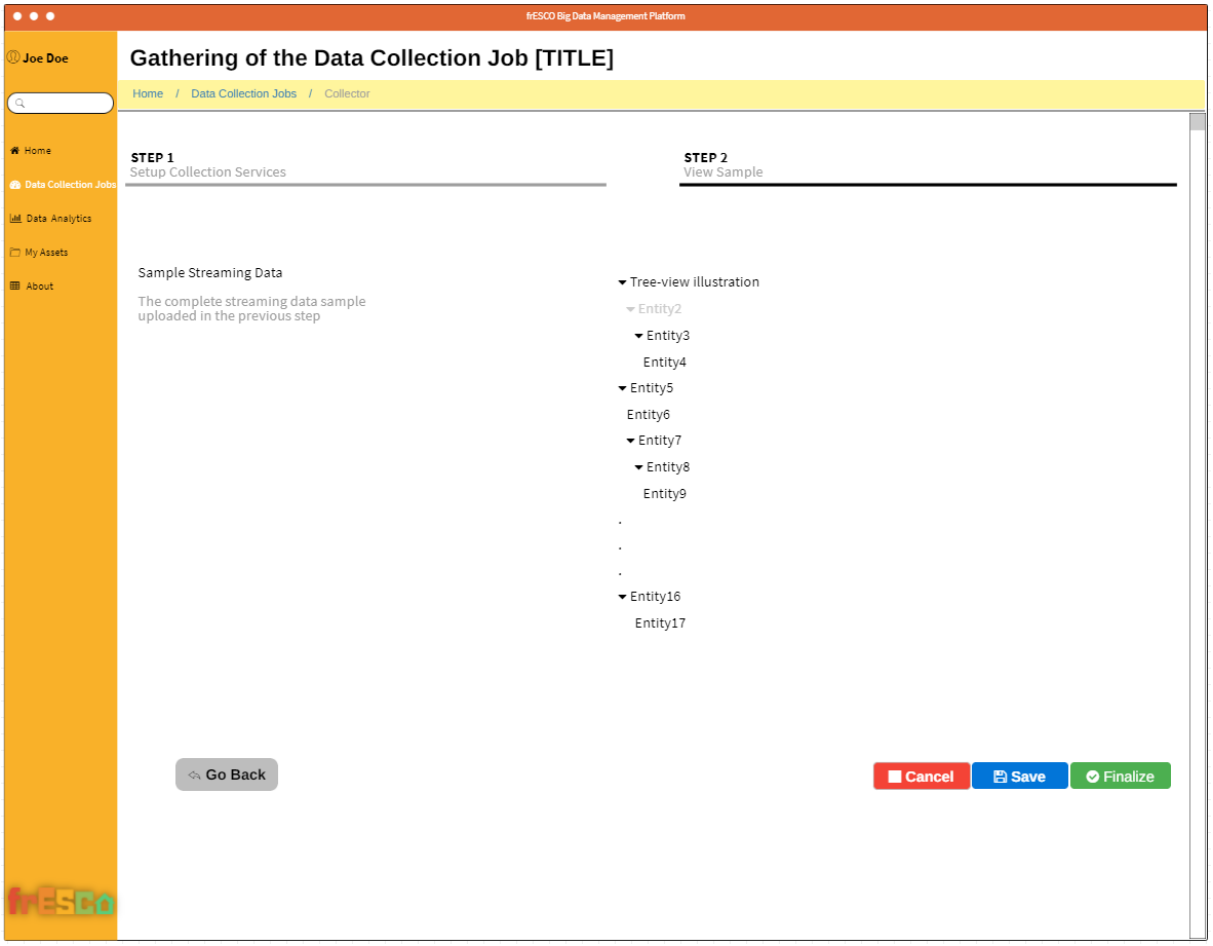

Figure 6 Sampling of the published data via the application's own API mechanism

3.3 Processing rules

The data asset providers, in order to design and create a new collection job, are requested to define a series of several pre-processing steps, spanning from Harmonization to Curation, Anonymisation, and Encryption (which are optional steps), that are to be applied on the data upon their ingestion. This section outlines the phases towards the effective configuration of these pre-processing steps, in preparation for their actual execution.

3.3.1 Harmonization

A data collection job's harmonization setup is broken down into two stages: Harmonization Info and Configuration. The data asset providers must first specify the data's compliance with particular standards, if applicable, as well as the entities of the appropriate category to which the data refer.

The predictions made by the frESCO Big Data Management Platform based on how the source data can be harmonized to the frESCO Common Information Model (CIM) in the Harmonization Playground are conveyed to the data asset providers. If no related harmonization exists or the harmonization provided is incorrect, the users will have the option of searching the Data Model Details section for a matching entity.

Once the entity to which an attribute/entity in the source data should be harmonized has been determined, data asset providers shall drag and drop it to the relevant attribute/entity in the Harmonization Playground area, as displayed in the Figure 7. The data owner can utilize the "Transformation Details" section, to provide any additional information needed per attribute to conclude the process.

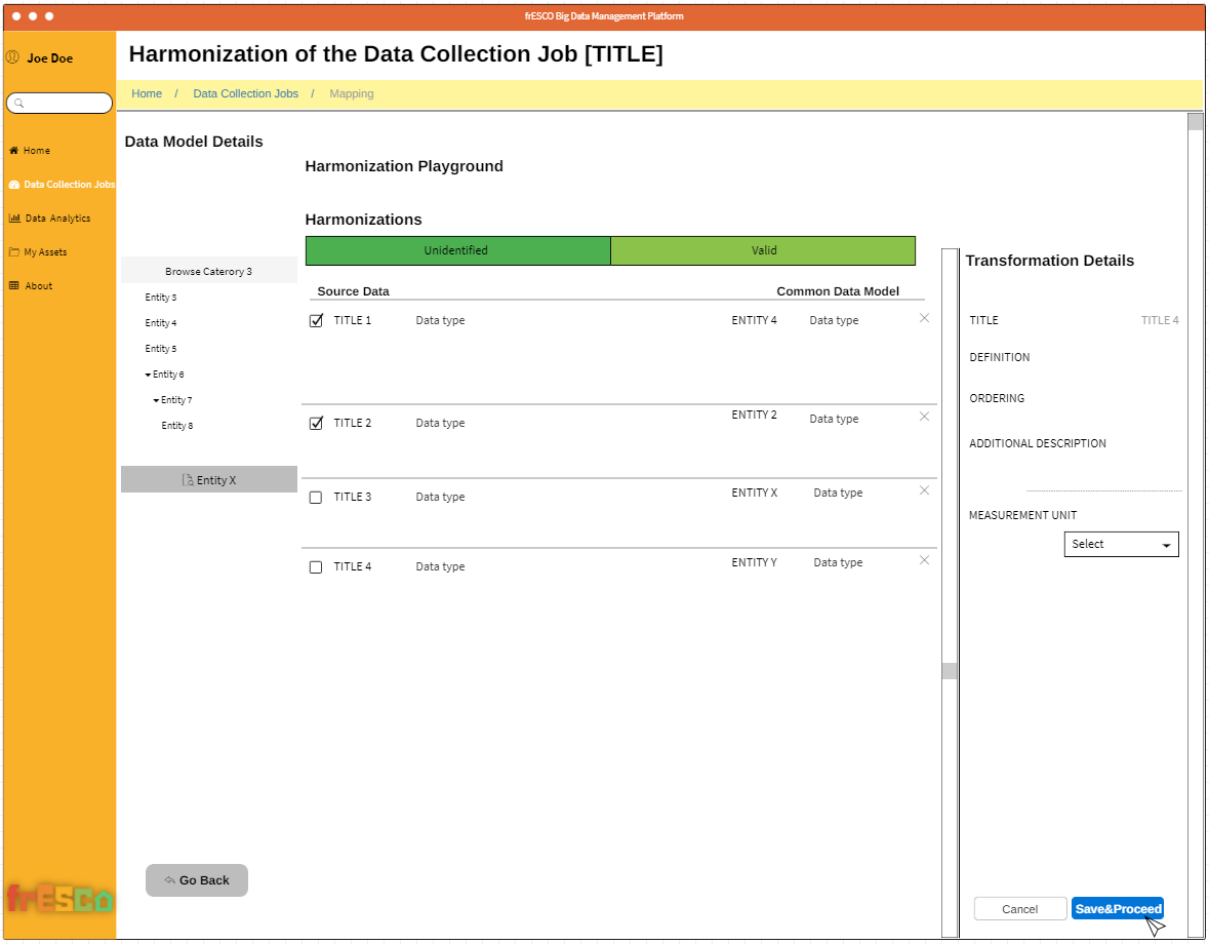

Figure 7 Harmonization Configuration

3.3.2 Curation

The data that are ingested into the frESCO Big Data Management Platform need to be handled appropriately, safeguarding their accuracy, completeness and consistency. Through the data curation feature of the frESCO Platform, necessary curation rules are introduced that hinge upon every attribute of the respective data asset and their individual data types, so to ensure the storage of a high-quality data asset. In order to initiate the workflow of data curation, the frESCO platform users need to indicate the relevant attributes/columns that specific curation rules and constraints will be applied to. After concluding the curation configuration, the users are able to conclude the final set of curation rules defined, by clicking on the "Save&Proceed" button, as shown in Figure 8.

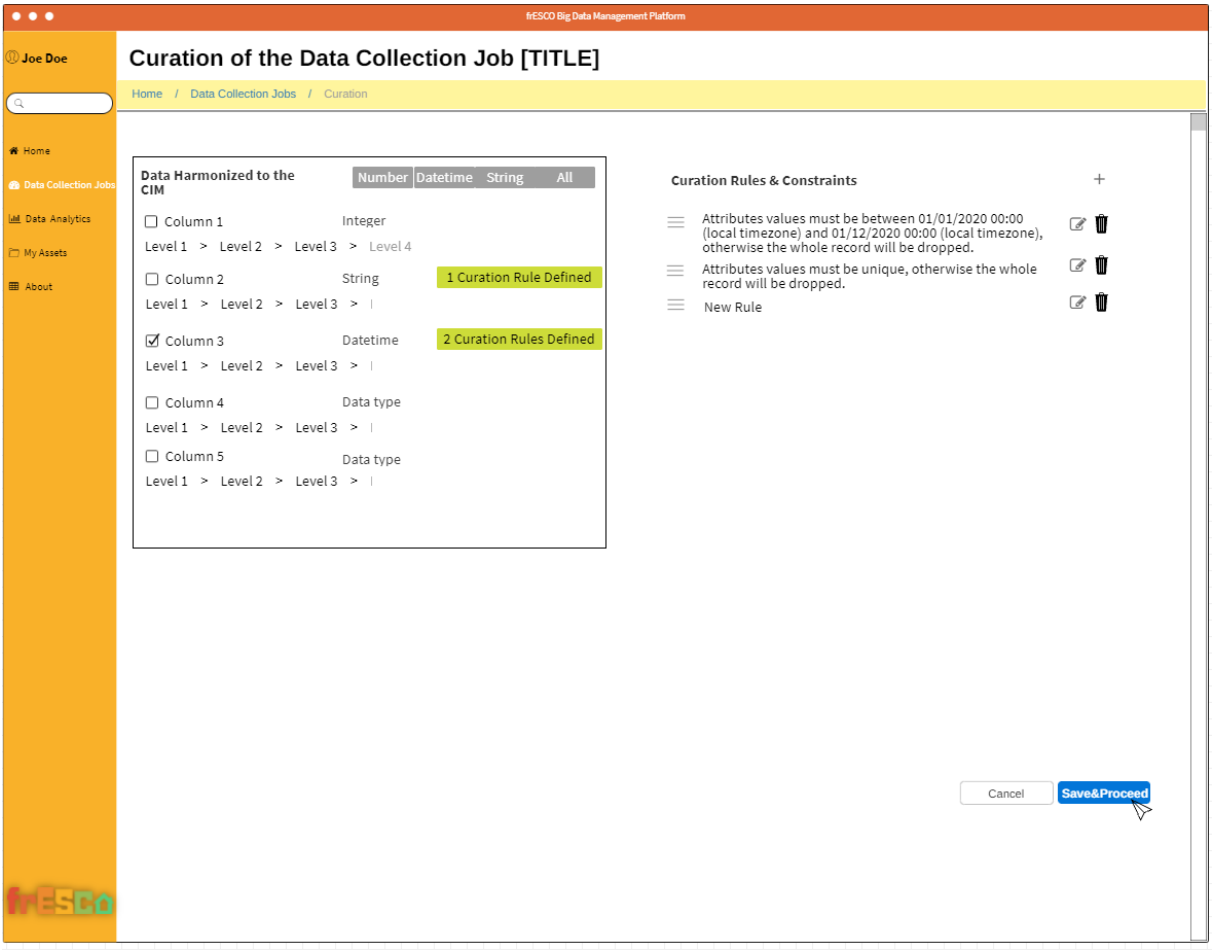

Figure 8 Curation Configuration

In the case that the curation execution has been performed successfully, the curation constraints applied for the transformations conducted on the data per attribute in the source data will be presented to the data asset providers. If the Curation Execution failed, the users will be notified accordingly for the attribute(s) that an issue occurred, so that they can modify the previously defined curation rules and constraints.

3.3.3 Anonymization

Anonymization is another feature that is made available to data asset providers as part of the pre-processing phase. Anonymization actions, that reflect into specific parts of the data (i.e., which attributes), are performed prior to making the data available within the frESCO platform. frESCO platform users are informed about potentially sensitive information within their data and provided with appropriate functionalities to anonymise this information, as well as any attribute of their data that they consider as containing any identifying information.

In more detail, the data asset providers can define the attributes (harmonized already in the Common Information Model) that are sensitive, quasi-identifiers and identifiers since by default all attributes are considered as neutral. Different anonymization methods with individual configuration options are introduced to the users, depending on the data type and the anonymization type they have selected

for an attribute. Finally, it is of crucial importance for the users of the platform to establish the acceptable information loss threshold and specific parameters of the selected anonymisation algorithm.

They, once finished with the anonymization configuration, are able to click on the "Save&Proceed" button, as depicted in Figure 9.

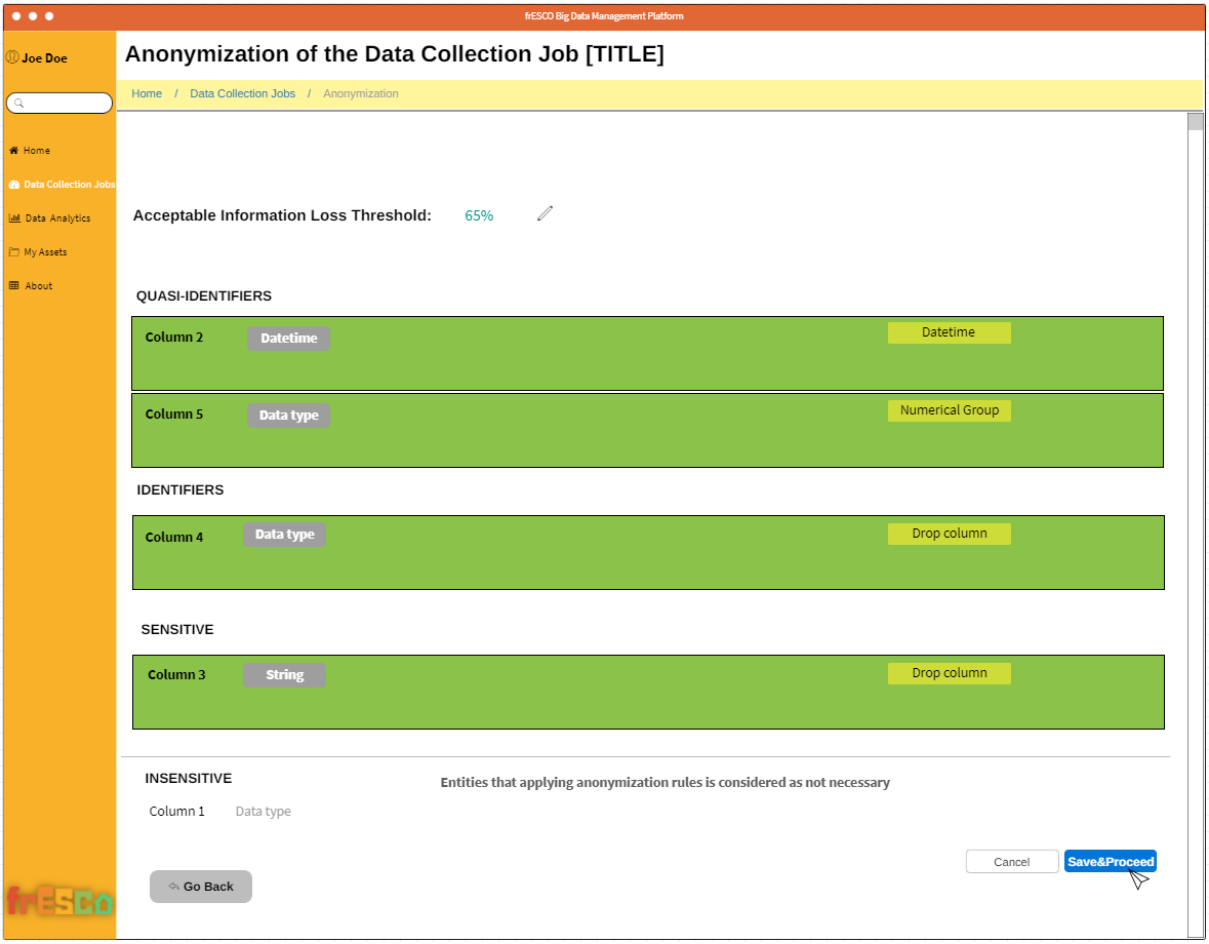

Figure 9 Anonymization Configuration

In the case that the anonymization execution has been performed successfully, the resulting data loss is below the acceptable information loss threshold that has been established during the configuration. Otherwise, if the resulting data loss is above the acceptable information loss threshold that has been established during the configuration, the Anonymisation Execution is considered as failed.

3.3.4 Encryption

The Data Security and Storage Module offers data encryption functionalities to data asset providers, as additional mechanisms to safeguard their data and eliminate the possibility of unauthorised data access and/or data leakage of any type. Specific data are therefore encrypted in the frESCO Data Management Platform, according to the data asset provider's needs and requirements. The Data Encryption feature provides flexibility in the definition of the encryption configuration to be applied, enabling modular control over the data contents that should be encrypted.

In particular, data asset owners are able to define encryption specifications not only for the total of the dataset, but also for specific parts of it, as depicted in Figure 10. The data encryption feature is optional; therefore, the users are asked for the inclusion of the feature during the creation of a data collection job.

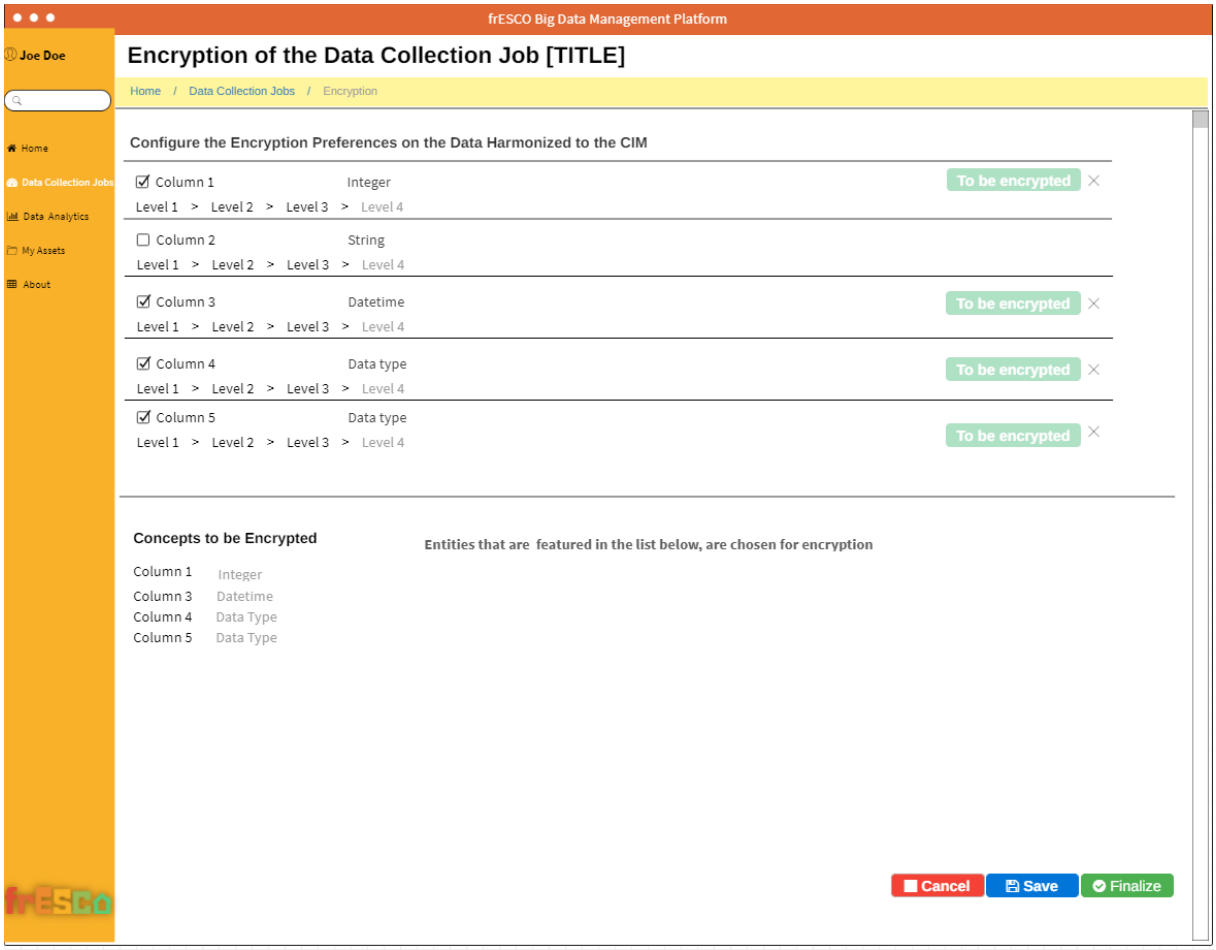

Figure 10 Encryption Configuration

Data asset providers can choose to either encrypt the whole data asset or specific entities of it. If they choose the latter option, the entities they picked will appear in the list Entities to be Encrypted, located at the bottom of the page. If the execution is performed successfully or unsuccessfully, the data asset providers are notified accordingly.

4 DATA ANALYTICS USER FLOW

The users of the frESCO platform, having selected from the catalogue of pre-trained analytics models the one to be executed over their own data, will be provided with the results of the execution, to leverage the added value that data analytics can bring through the offerings of the frESCO models.

4.1 Results Visualization

Once an analytics model is selected, the data asset providers will be presented with the results of the model application in a pre-configured graph chart, as illustrated in Figure 11. The users of the platform shall have the possibility to export the results or download them in a .pdf format, among others.

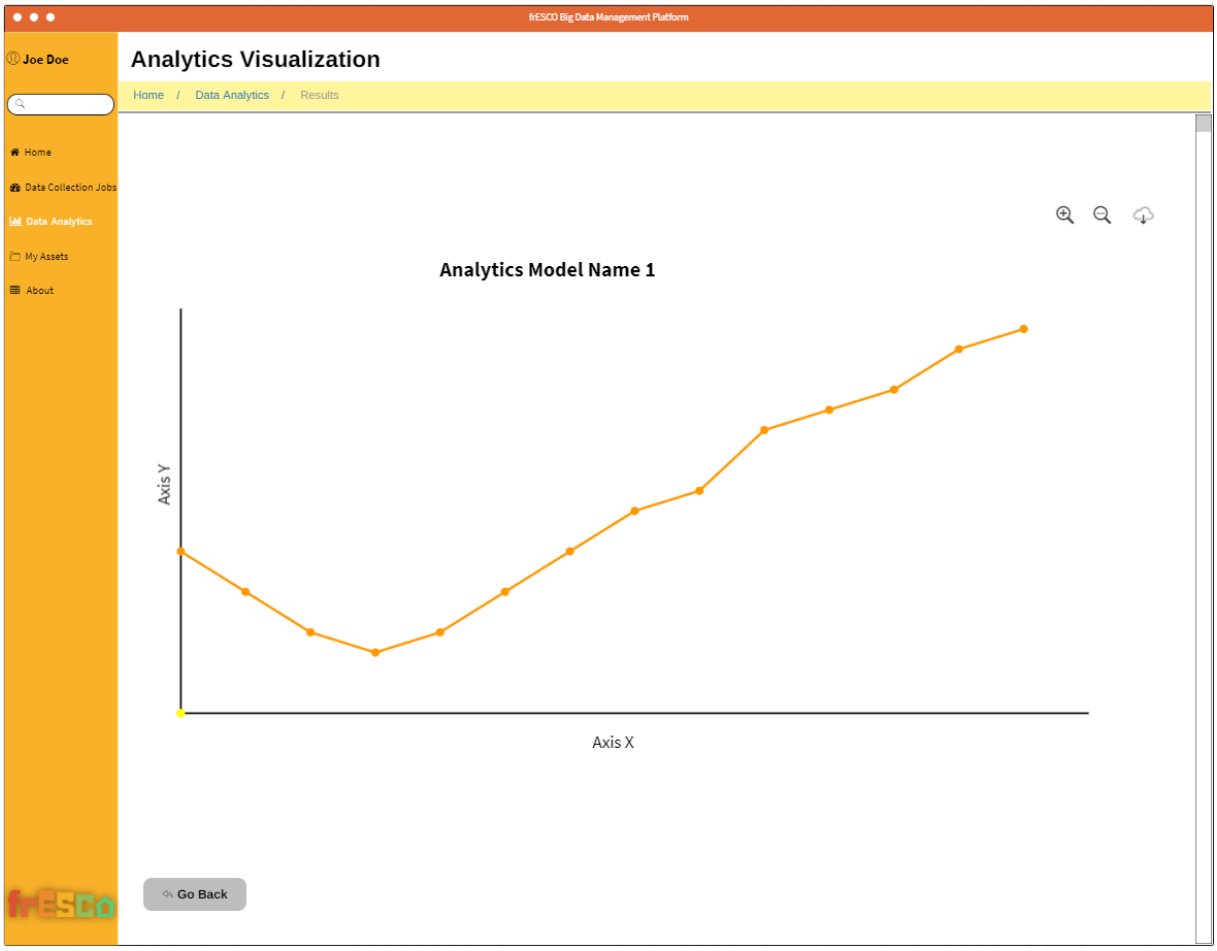

Figure 11 Algorithmic Model Application Results

5 DATA SEARCH AND RETRIEVAL USER FLOW

The Data Search Module constitutes an essential part in the Big Data Management Platform of frESCO, enabling the users of the platform to search and discover data that can be proven useful, determine and define which of these available data are of importance and eventually, have a clear and thorough view on the provided results.

5.1 Assets overview

The users of the frESCO Platform will have the possibility to overview the data assets they own, listed in the respective tab of the platform, as shown in Figure 12. Appropriate filtering tools will be made available to them, so that they are able to filter the data assets determined by their data type. Additionally, after the filtering has been performed, the users of the platform may be introduced with complementary functions, in order to:

- Review the details of a specific data asset
- Edit the details of a specific data asset
- Delete a specific data asset

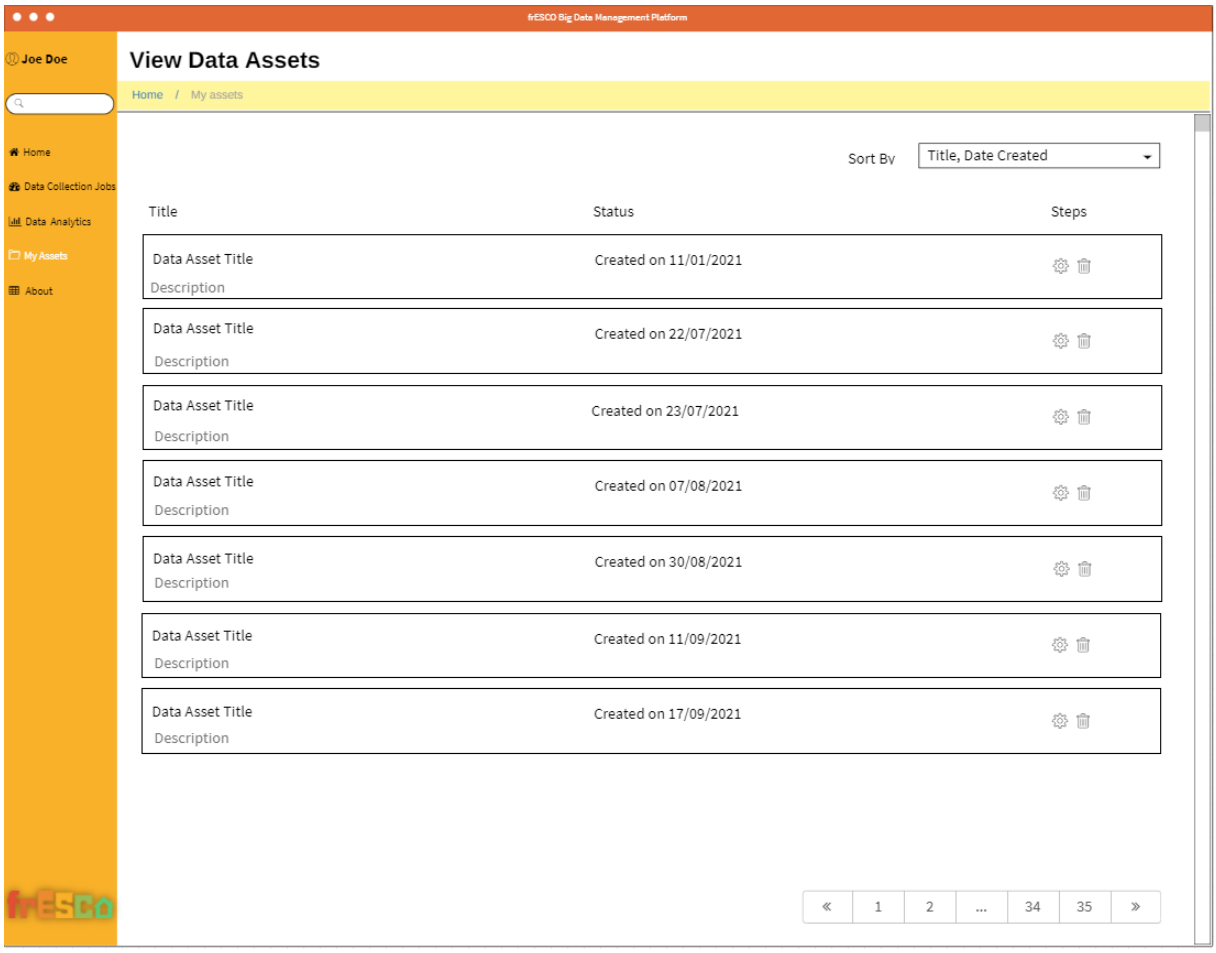

Figure 12 Synopsis of the data assets

5.2 Data Asset Profile Definition

Any result deriving from a data collection job is stored as a data asset in the frESCO Big Data Management Platform, thus a detailed profile of the data asset needs to be defined by the data providers. The users of the platform, after navigating to My Assets on the side navigation bar, will be requested to issue the data asset title and describe what is included in the data asset, as shown in Figure 13.

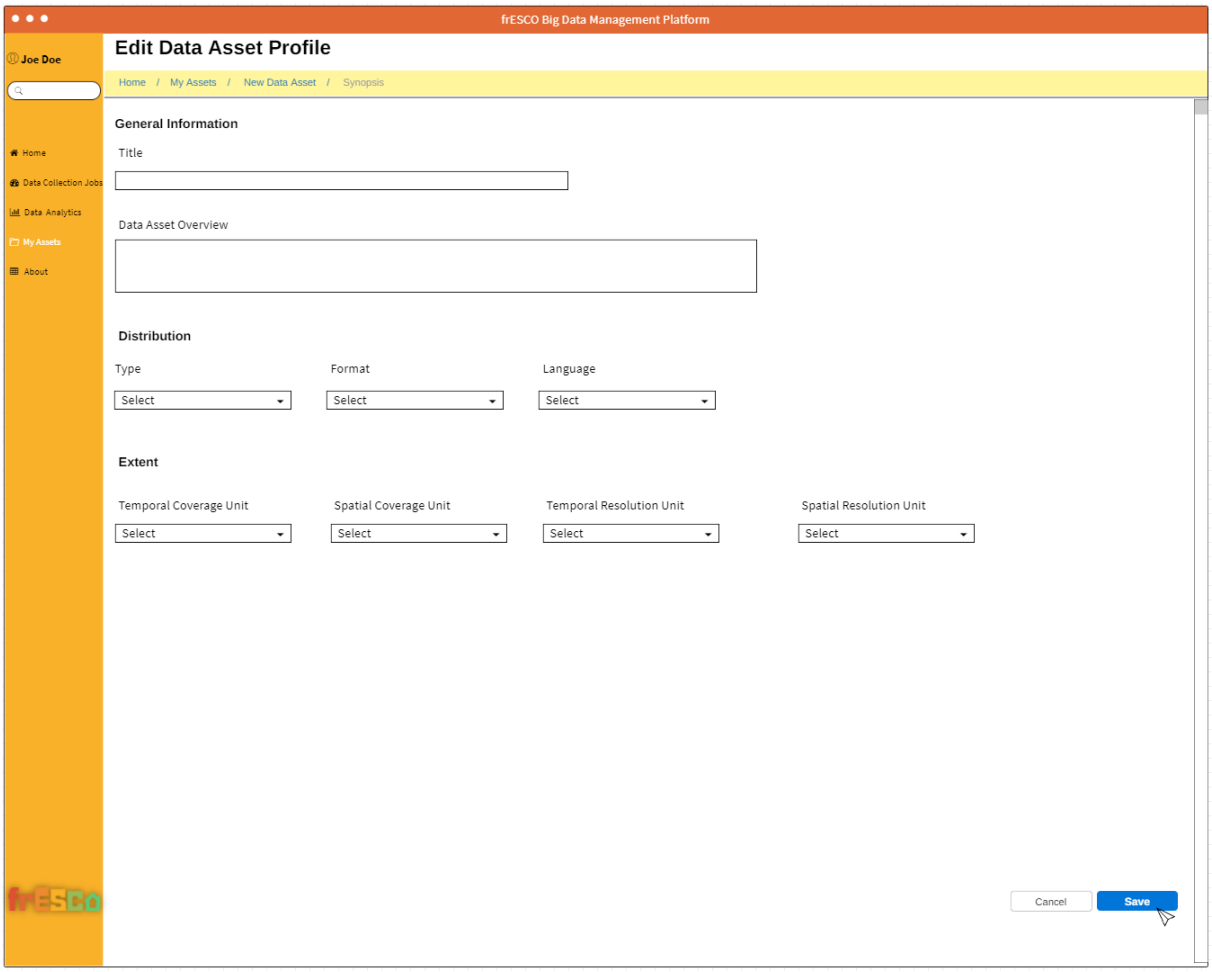

Figure 13 New Data Asset Creation

Also, the data asset providers through a series of drop-down lists will have the possibility to delineate the type of the data asset, the format of the data asset to which the data will be available, the language of the data asset, as well as specifics about the temporal coverage, spatial coverage, temporal resolution, and spatial resolution units of the data.

In the next figure, the frESCO platform users will have to establish the visibility levels along with respective access policies of the corresponding data assets. By establishing the visibility levels, the data asset provider is enabled to choose whether the data asset is for: i) Exclusive access (Access on the data asset is granted only if access policies are satisfied) and ii) Universal access (Free access to the data asset, no requirement for access policies to be met). If the access level is set to Exclusive, appropriate access policies need to be defined, together with the relevant strategy (Allow-all) and the inclusion of particular exceptions. 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, etc.), and the value of the property, as shown in Figure 14.

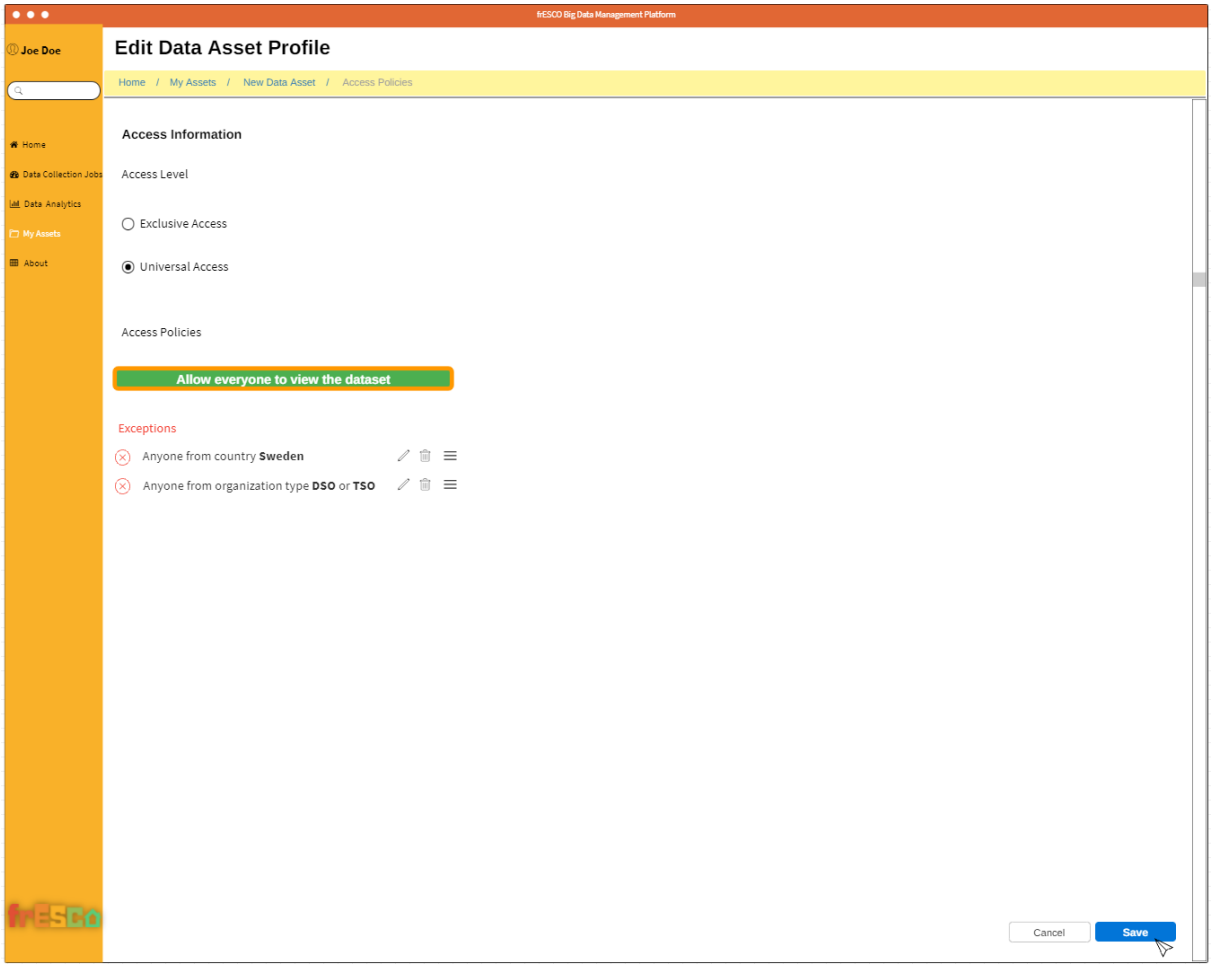

Figure 14 Establishing Data Access Policies

Afterwards, the data asset providers are required to define the license (ranging from open data licenses to a custom license), and the IPR 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 frESCO Platform).

5.3 Query Creation

A key element of the Data Search Module, providing data asset providers the necessary methods for data discoverability and exploration, that enables data sharing operations. The users of the platform can search for data that are interesting for their needs, browse through the results and investigate the data in more detail to identify useful candidates for retrieval, as depicted in Figure 15.

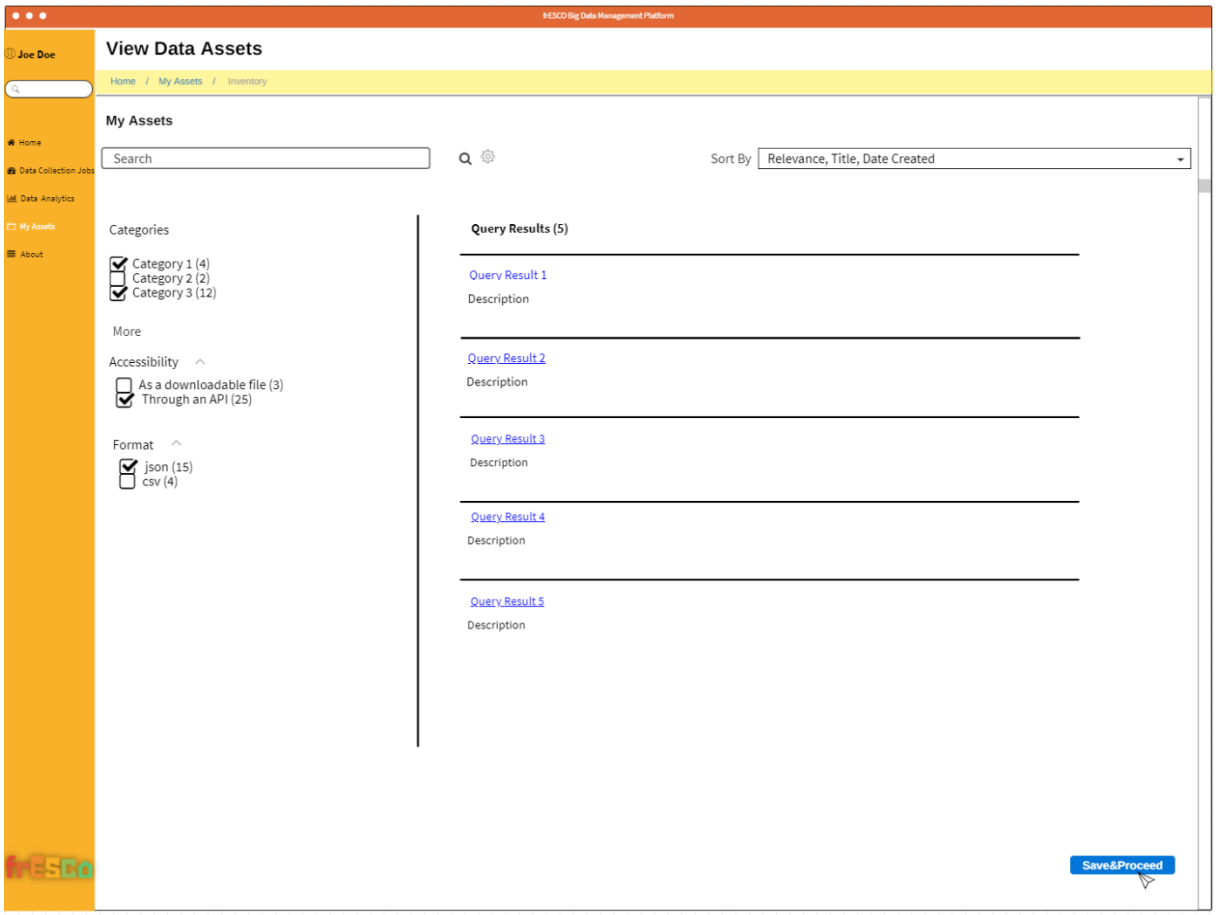

Figure 15 Retrieval Query Creation

5.3.1 Retrieval Query Creation through API

The Retrieval Settings must be specified in the Retrieval Configuration page through API, by selecting the entities that a user wants to retrieve from each selected data asset, as shown in Figure 16. Furthermore, the entities that will be used as query parameters to filter the query results must be defined (with different options depending on the data type of the entity). The configured retrieval can be tested once the retrieval configuration is complete. Users will be able to amend the body with suggested values for query parameters in order to filter the data asset appropriately, while a sample preview of the result will be displayed also.

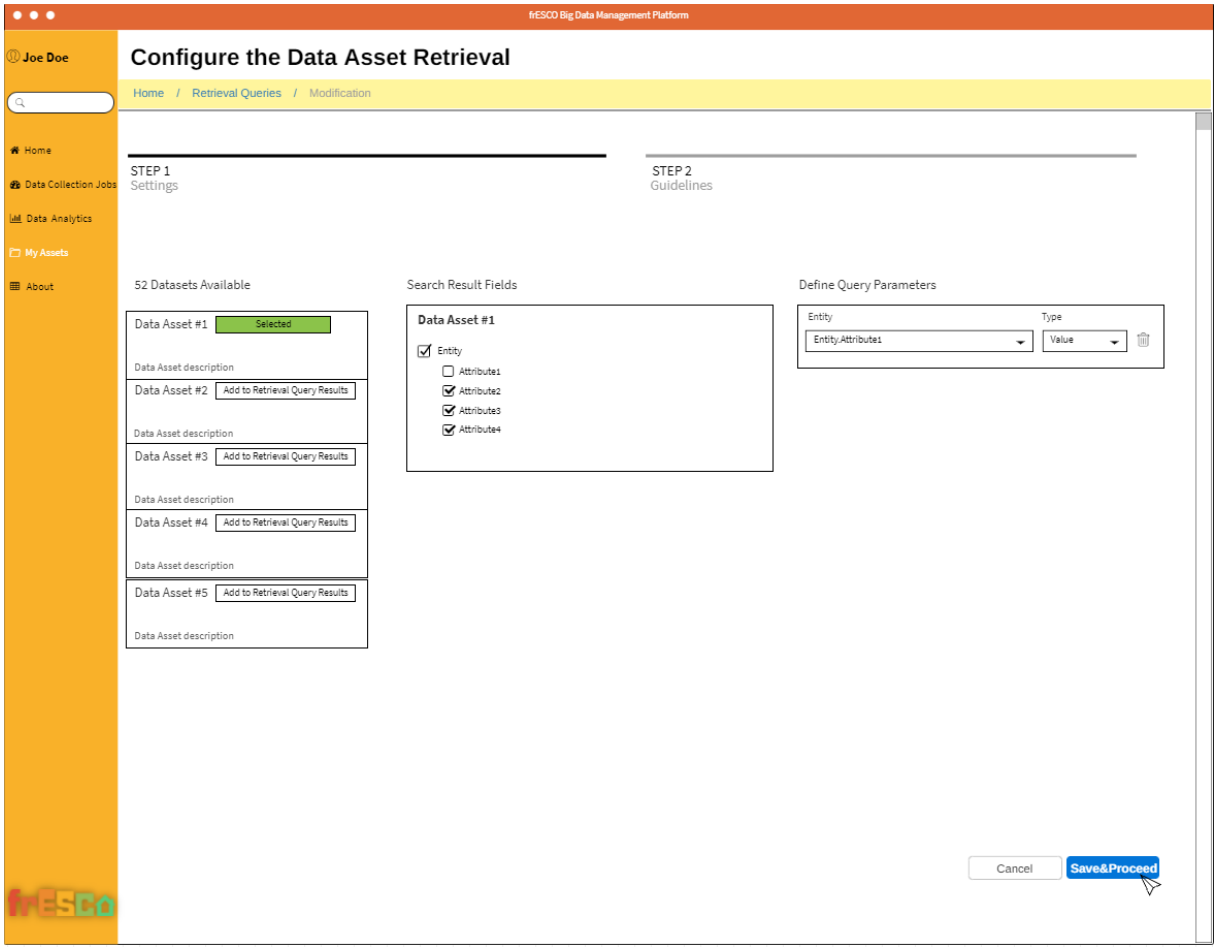

Figure 16 API Retrieval Query Configuration

As depicted in Figure 17, the final step of the retrieval configuration comprises information on how to obtain the retrieval results using the frESCO APIs. To this end, instructions regarding authentication, are provided in addition to the endpoints (i.e., for GET and POST methods) including the full API paths.

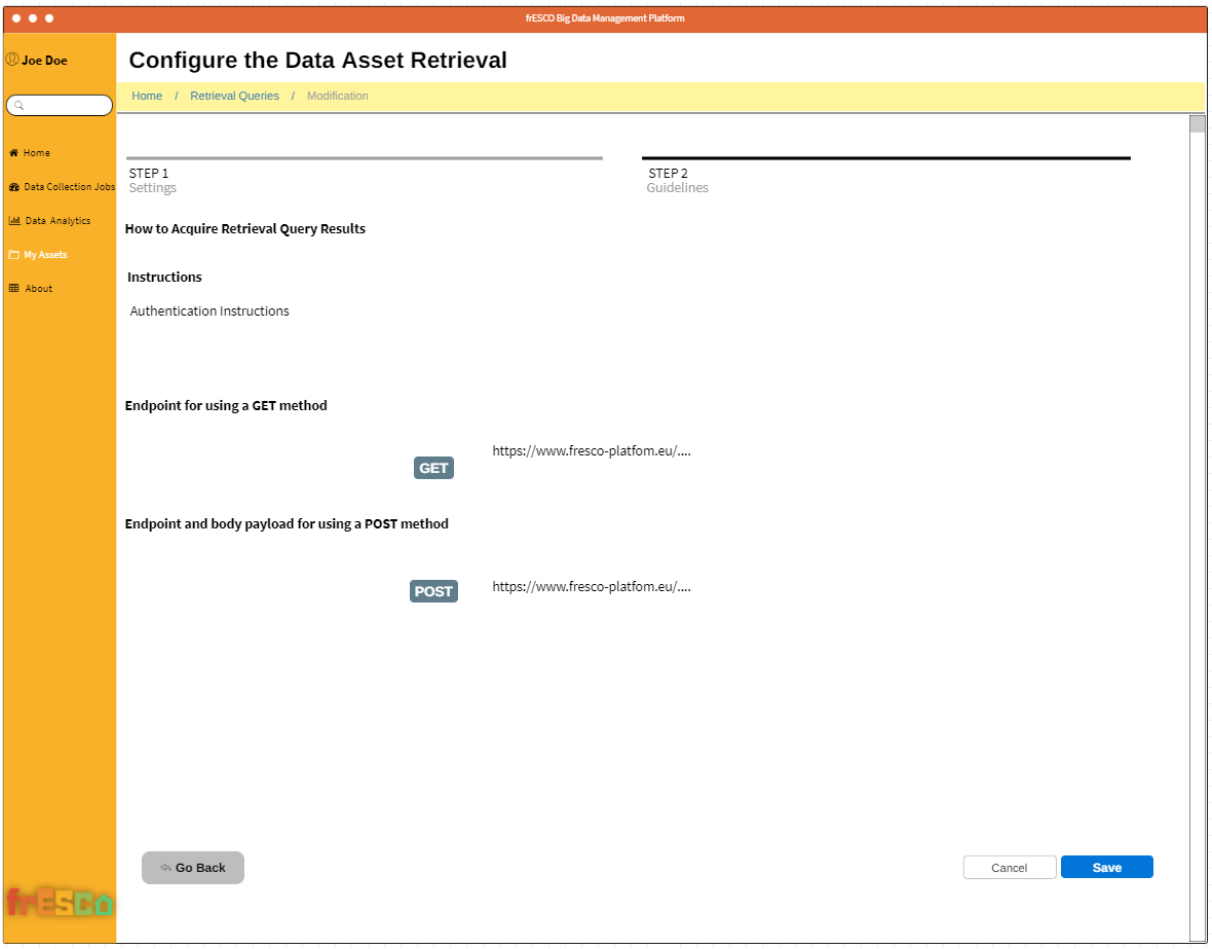

Figure 17 Finalization of the API Retrieval

6 SUPPLEMENTARY PLATFORM FUNCTIONALITIES

Different sorts of users (platform administrators, data asset consumers, data asset suppliers, and others) have access to additional features on the frESCO Platform, such as editing user profiles, as well as receiving notification alerts on the status of data import jobs (executed, failed to execute, and pending) along with the respective functionalities to handle them effectively.

6.1 User Profile Management

Members of each organization that has access on the frESCO Platform, can change their own user profiles by navigating to the "Home" where they find the Edit User Profile page, as shown in Figure 18.

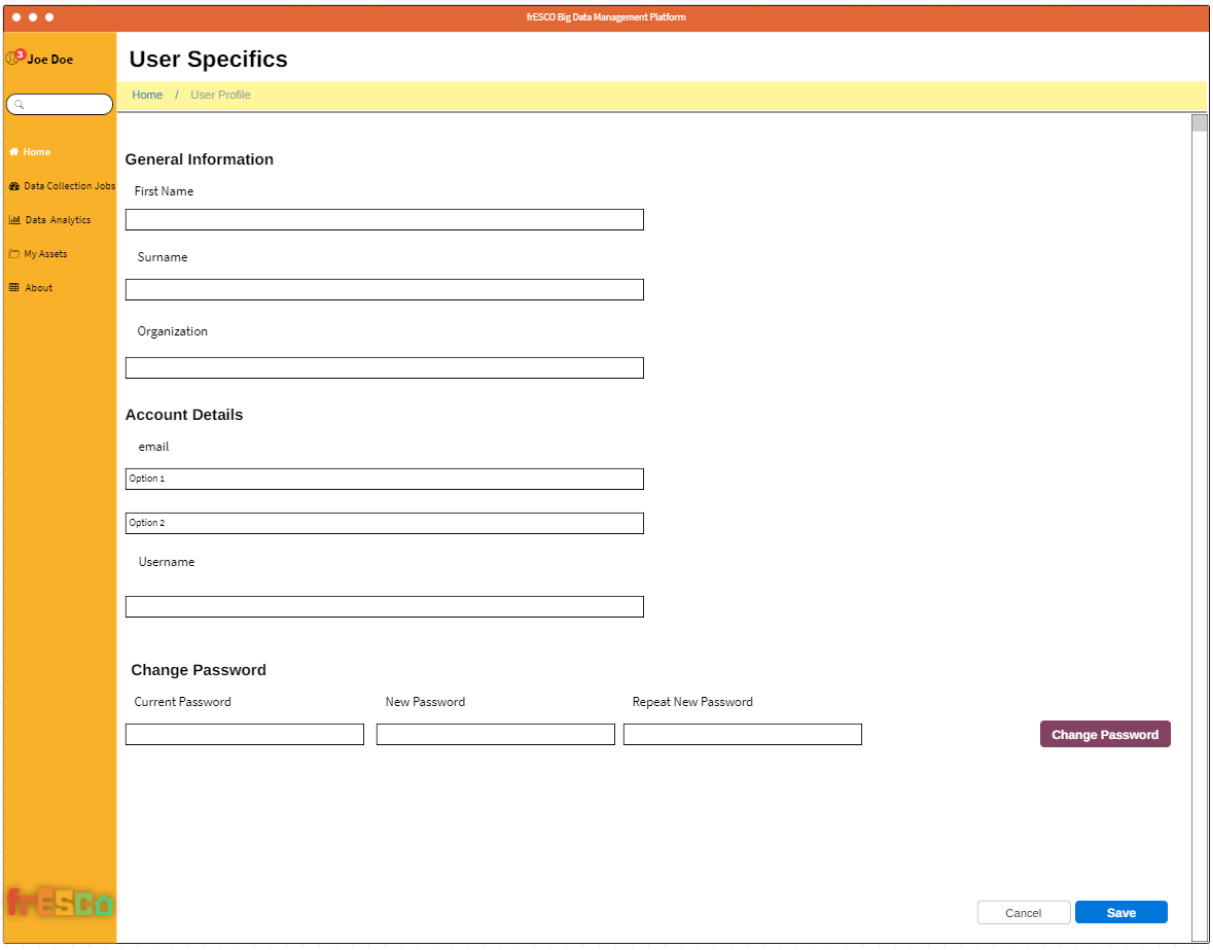

Figure 18 User Profile Creation

They must fill out some basic information (such as their surname, first name, etc.,) as well as several account details (i.e., email, username). In addition, users shall have the option to change their existing password to a new one.

6.2 Notification Alerts

As described in section 2.5.2, the users of the frESCO platform will be presented with the possibility to receive adequate notification alerts in regard to the progress of their data import jobs and the respective status of them, as depicted in Figure 19. Also, through appropriate functions, the users of the platform will be able to define what they want to do with the notifications they receive (e.g., examine, delete).

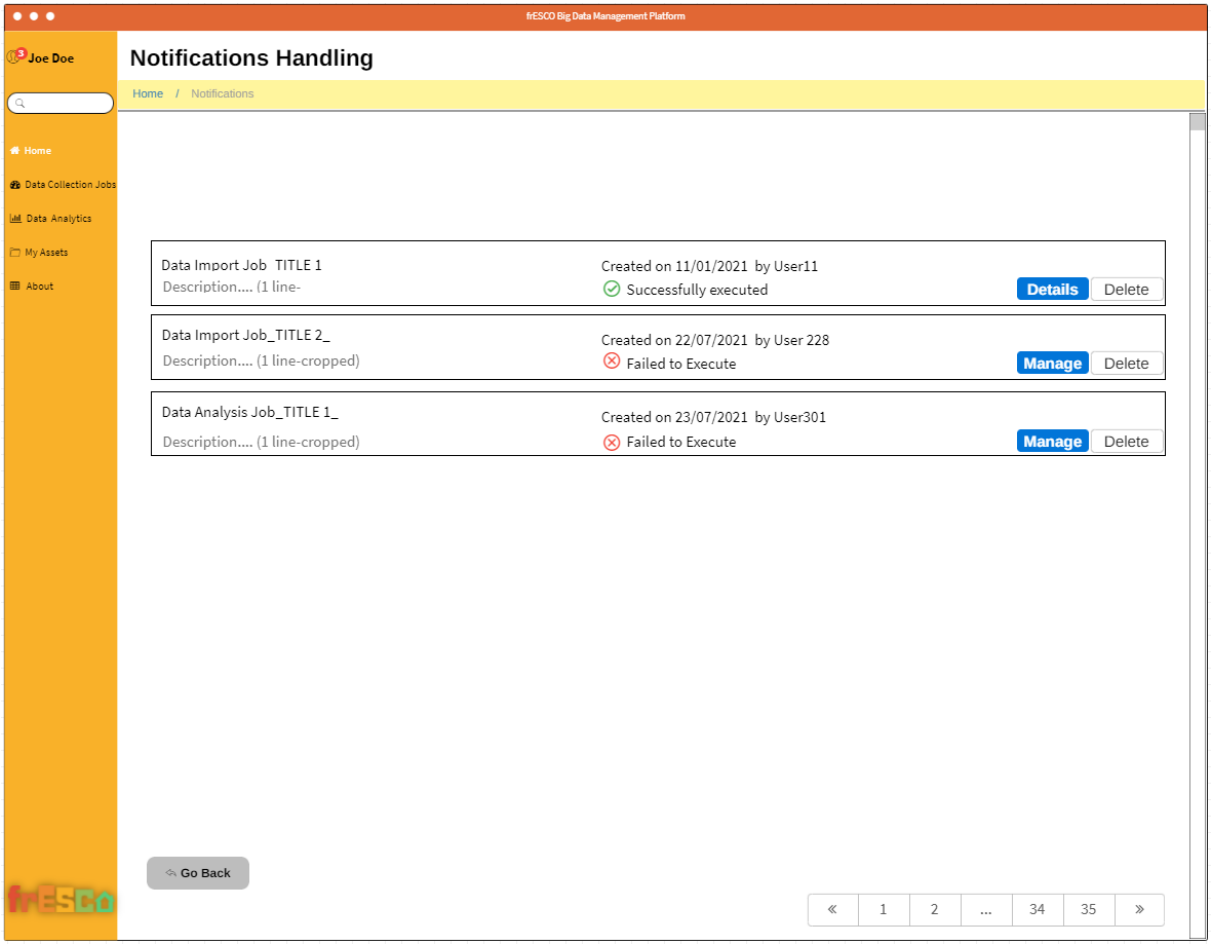

Figure 19 Notification Alerts

7 CONCLUSIONS

A detailed description of the frESCO Platform user flows for configuring and executing the core platform functionalities (i.e., data collection, data search and retrieval, data analytics), as well as different supplementary functionalities (i.e., User Profile Management, Organization Profile Management, Notification Alerts), was provided in Deliverable D4.4 ("frESCO Integrated Platform – Alpha, Mockups Release"). This deliverable defined the alpha version of the frESCO Platform in terms of the workflows for data collection, data search and retrieval and data analytics that the platform users might use. In more detail, this deliverable described extensively:

- The data collection user flow, which describes the workflow that must be set up for the creation, configuration, and execution of data collection jobs.
- The data analytics user flow, which aims to provide valuable insights to the frESCO Platform users, on the data assets they own, with the deployment of appropriate pre-trained algorithms.
- The data search and retrieval user flow, which enables the users to browse within the frESCO platform for data assets of interest and retrieve them according to their preferences. Additionally, the users can get a brief overview of their assets and create appropriate data asset profiles.
- The supplementary platform functionalities that will be made available within the frESCO Platform (i.e., User Profile Management, Organization Profile Management, Notification Alerts).

8 REFERENCES

GA 2020 - frESCO Grant Agreement

CA 2020 - frESCO Consortium Agreement

Pub/Sub Messaging 2021, UNITED NATIONS, ACCESSED 25 OCTOBER 2021, <https://aws.amazon.com/pub-sub-messaging/>

frESCO Consortium. (2020). frESCO D2.5 “Report on the frESCO conceptual architecture”