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 D4.3
Data Collection, Security, Storage & Management Services Bundles

Deliverable number	D4.3
Deliverable name	Data Collection, Security, Storage & Management Services Bundles
Lead beneficiary	Suite5
Description	The deliverable focuses on the rapid prototyping and beta release of the frESCO Platform data collection, security, storage, governance and management services
WP	WP4
Related task(s)	T4.2, T4.3, T4.5
Type	Other
Dissemination level	Public
Delivery date	23.12.2021
Main author	Alexandros Tsitsanis (Suite5), Georgios Zafeiris (Suite5)
Contributors	CIRCE, UBITECH, KONČAR, KRK, COMSA, VOLT, VERD

Document history

Version	Date	Changes	Author
V1 – first draft	09.12.2021		S5
V1 – reviews	14.12.2021	Minor corrections	Dionysios Pramangioulis – UBITECH
V1 - reviews	15.12.2021	Minor Corrections	Andreas Muñoz Zuara - CIRCE
Final version	21.12.2021	Addressed the corrections suggested by the partners who performed the review of the document	S5
Final deliverable submission	23.12.2021	Minor grammar and format changes	CIRCE

ABBREVIATIONS

Abbreviation	Name
API	Application Programming Interface
CIM	Common Information Model
CSV	Comma-separated values
D	Deliverable
DoA	Description of Action
DER	Distributed Energy Resources
DMP	Data Management Platform
GDPR	General Data Protection Regulation
JSON	JavaScript Object Notation
M	Month
UI	User Interface
WP	Work Package
XML	Extensible Mark-up Language

TABLE OF CONTENT

EXECUTIVE SUMMARY

The current deliverable D4.3 "Data Collection, Security, Storage, and Management Services Bundles" summarizes the efforts made under Task 4.3 " Core Big Data Ingestion, Curation and Management Services" and Task 4.5 " Data Assets Security, Encryption and Privacy Mechanisms" of WP4 " Big data collection and management, automation, privacy and interoperability". The goal of this deliverable is to deliver the WP4 Services Bundles beta release. As explained in the deliverable D2.5 "Report on the FRESCO conceptual architecture", where an initial representation of the Big Data Management Platform was provided, the frESCO Platform is recognized as the overall component of the architecture, while the various services bundles to be implemented are renamed to modules. Each module introduces its respective features.

In this deliverable, the following documentation is provided:

- The implementation status of the various features of each Module, together with potential features to be considered in future releases of the frESCO Big Data Management Platform.
- The frameworks and libraries for the implementation of each module, together with their respective version.
- The Swagger interface used for the technical documentation for each module's APIs.

The current deliverable is the beta version of WP4's Modules for M18. The delivered modules were integrated at the mock-up level (in the alpha release documented in frESCO Deliverable D4.4 ("frESCO Integrated Platform – Alpha, Mock-ups Release") and have paved the way for the frESCO platform beta release, which will be documented in frESCO Deliverable D4.6 ("frESCO Integrated Platform – Beta Release") in M20.

1 INTRODUCTION

1.1 Purpose of the deliverable

The Big Data Management Platform is recognized as the main component of the frESCO architecture. The platform will help with granular communication and data interchange between a variety of sources, including buildings, DER management systems, weather data sources, and wholesale energy pricing, as well as the appropriate mass importing and storage of large volumes of associated data assets. The present deliverable provides the beta release of the frESCO Platform data collection, security, storage, governance and management services bundles, accompanied with their back-end implementation details.

As documented in deliverable D2.5 "Report on the FRESCO conceptual architecture" and in deliverable D4.4 "frESCO Integrated Platform – Alpha, mock-ups release", the various services bundles implemented under WP4, are renamed to modules. The frESCO Platform consists of the following modules, Data Collection, Data Security & Storage, Data Analytics, Data Search & Retrieval and Platform Governance.

In the present deliverable, the Data Collection, Data Security & Storage, Data Search & Retrieval and Platform Governance modules are thoroughly described, along with the different features introduced per module. The Data Analytics Module will be presented in detail, in Deliverable D4.5 "frESCO Baseline Data Analytics – Draft Release".

D4.6 "frESCO Integrated Platform – Beta Release" will encompass the User Interface implementation of the different modules developed under WP4.

1.2 Scope of the deliverable

The primary goal of D4.3. is to provide a complete overview of the beta release of WP4 modules. D4.3 goes over the modules of the frESCO Platform in greater depth, as well as their features. Each of the modules, Data Collection, Data Security & Storage, Data Search & Retrieval, and Platform Governance, is thoroughly detailed by:

- Outlining each module's implemented features, the features that are under development and the features that will be considered for future releases of the frESCO Platform, should they be recognized as necessary.

- Identifying the libraries and frameworks that each module is developed with.
- Documenting the creation of relevant APIs, that will facilitate the communication and information exchange among the modules.

1.3 Structure of the deliverable

The following is the structure of the document:

Section 2 provides a description of the platform's frESCO modules (Data Collection, Data Security and Storage, Data Search, and Platform Governance), as well as the various features of each module.

The implementation status of the features of each module, the frameworks and libraries utilized for the development of each module, along with the different APIs that have been created and are documented in Swagger, are described in sections 3-6.

Section 7 provides an overview on the availability of the frESCO modules during the development process, any limitations imposed and the relevant licensing and means for access.

Section 8 concludes this deliverable D4.3 “Data Collection, Security, Storage & Management Services Bundles”, providing a brief summary of what has been presented in the different sections of the deliverable.

2 MODULES SYNOPSIS

The frESCO Big Data Platform consists of the Data Collection, Data Security & Storage, Data Search & Retrieval, Platform Governance and Data Analytics modules. In the present deliverable, only the first four (4) modules and their respective features will be described and further analysed in the sections to follow. The Data Analytics module will be further elaborated within deliverable D4.5 “Baseline Data Analytics – Draft Release”.

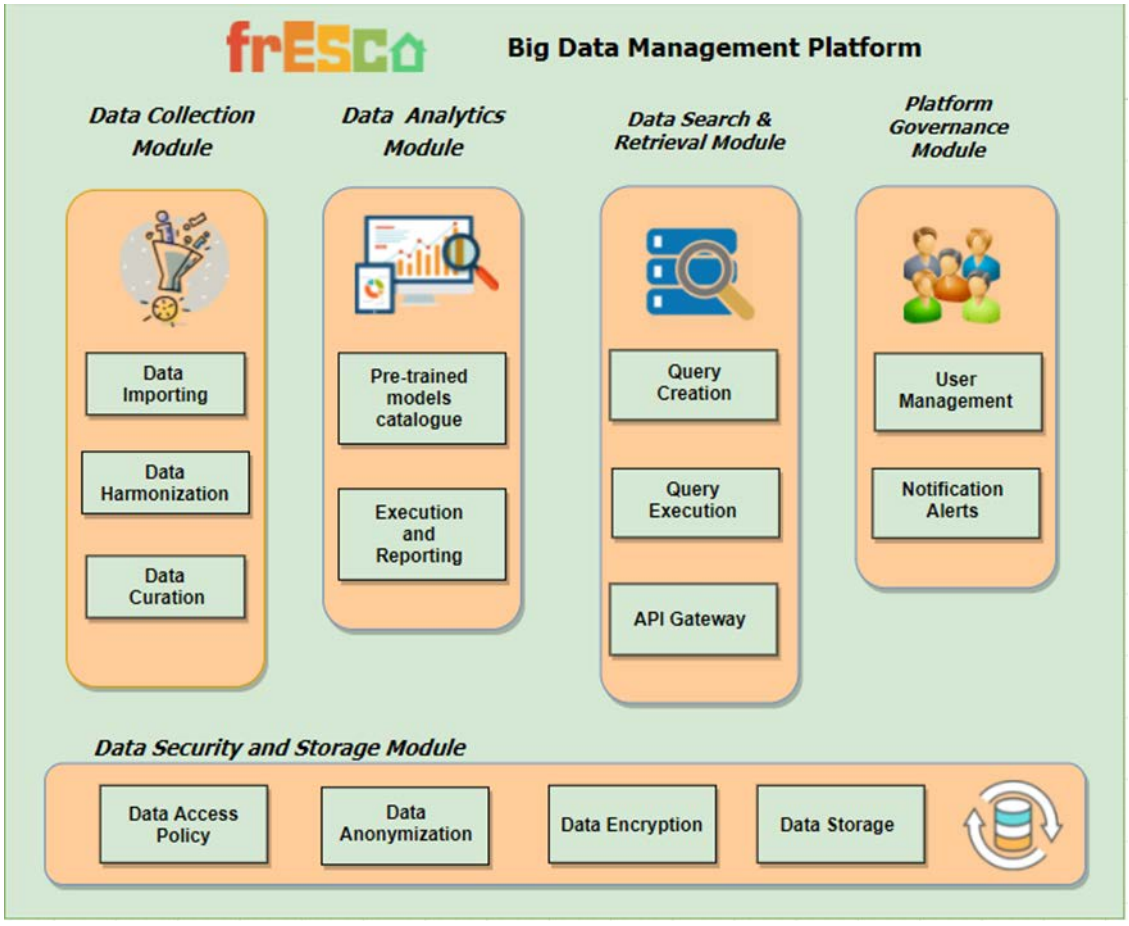

Figure 1 frESCO Big Data Management Platform and Modules Introduced

2.1 Data Collection Module

The frESCO Data Collection Module offers a range of features that are involved in the data collection process, namely: i) Data Importing, introducing appropriate mechanisms and methods for importing data into the frESCO Platform, (ii) Data Harmonization, allowing for

matching imported data assets with the CIM and (iii) Data Curation, for Data Quality Assurance Enhancement.

- The data collection configuration is enabled by the effective handling of data management processes, allowing data providers to set the data import method and schedule, after which the data will be uploaded to the frESCO Integrated Platform. Data from various sources is imported into the frESCO Big Data Management Platform in a straightforward manner. The data asset providers will use the platform to specify how, when, and what data they own and are ready to share, as well as to configure the parameters of the data import process, define their needs and preferences for how data will be inserted in the platform (data provider, own API, file uploading), and perform proper harmonization actions on all attributes of the data, based on the frESCO Common Information Model.
- Also, the way that the attributes of a data asset are harmonized with the entities of the frESCO Common Information Model (CIM) is determined by the Data Harmonization feature. As a result of the matching between the CIM and the data asset, common domain-specific knowledge of the data is established, which makes it easier to utilize for data-related services in the frESCO Platform. The use of appropriate harmonization and transformation rules on data assets allows for the renaming of attribute names and the conversion of measurement units, to ensure compliance with the CIM.
- Regarding the Data Curation feature, the user determines the constraints and limitations that the data ingested into the module may have, as well as the precise actions that must be performed if any of these constraints are exceeded, during the curation configuration process. This configuration refers to a set of curation rules that are implemented by performing the necessary data curation procedures. A number of data validation methods are available depending on the data type of each attribute and are complemented by two types of corrective measures: removing entries and modifying values.

2.2 Data Security & Storage Module

The Data Security and Storage Module has a dual function. Firstly, it takes on the task of addressing the Data asset providers' data security and privacy concerns over the data that will be imported and processed in the frESCO Data Management Platform. Secondly, by providing resilient mechanisms and multiple indexing approaches, this module addresses the need for reliable data storage and indexing.

The frESCO Data Security & Storage Module provides data security-related features to maintain and protect data assets' privacy, remove the risk of unwanted data access or data leakage, and provide data asset providers complete control over their access. These core functionalities are contained in the Data Security & Storage module's three fundamental features: (i) anonymisation, (ii) access policies and (iii) encryption.

- The Anonymisation feature is responsible for keeping data safe from unintentional disclosure of personal information. Anonymization activities that reflect into defined parts of the data (i.e., particular attributes) are undertaken prior to making the data available on the frESCO platform. Users of the frESCO platform are provided with appropriate tools in order to anonymize their data, as well as any attributes in their data that they believe identifiable information is included.
- The Access Policies Feature allows data asset providers to specify the visibility levels for their data assets, followed with respective access policies. By establishing the visibility levels, the data asset provider can 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 and the inclusion of particular exceptions. This feature is responsible for assessing the applicable access policies throughout a user's search for a data asset within the frESCO platform, as well as during data retrieval through the API Gateway, in order to specify the corresponding results.
- The Encryption feature of the module adds data encryption capabilities to data asset providers' protection systems, reducing the probability of unauthorised data access and/or data leakage of any type.

With regard to Storage, the module is introducing a variety of storage and indexing developments that are in charge of keeping track of configurations, data assets, sensitive data, and logs, among other things. This feature is responsible for the span of storage solutions utilized in the frESCO Platform, which are briefly detailed below.

- Storage of data: Keeping track of multiple types of data (i.e., data assets, analytics results and reports), as well as their metadata in order to make it available to all of the frESCO platform's modules and services.
- Storage of log data: Keeping track of log-related information for the frESCO Data Management platform's operation and usage, as well as data from users and organizations, as well as any administrative data required for the frESCO Platform's seamless operation.
- 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 Search & Retrieval Module

The Data Search & Retrieval Module is a critical part of frESCO's Big Data Management Platform, allowing users to search for and discover data that can be proven valuable, determine and define which of these available data are important, and last, have a clear and comprehensive picture of the results.

- The Query Creation feature, introduced by the Data Search & Retrieval module, provides data asset providers with all necessary mechanisms for data discovery and exploration. Taking advantage of this feature, they can look for data that is relevant to their needs, study the results, and go further into the data to find candidates for retrieval.

- The customizable Query Execution feature, also introduced in the Data Search & Retrieval Module, extends the Query Creation feature's data discovery and exploration capabilities by offering a built-in way to execute the query created, as well as a quick summary of the results that will be generated. As a result, the user can promptly assess whether the findings meet his or her expectations and make any revisions or corrections that are required.
- The module's configuration of APIs for data retrieval is critical since it serves as a single point of entry for apps (both frESCO and third-party) to retrieve data from the frESCO Platform that they're authorized to access, as well as analytics findings. The API gateway accepts all API calls before retrieving the data required to respond to the call and returning the appropriate result.

2.4 Platform Governance Module

It establishes the means and processes for users to register securely and reliably on the platform. Users are provided access to the data they are eligible to use, as well as the opportunity to get analytics results, through proper authentication and security mechanisms. Users will also be notified via platform notifications when data import or data analysis occurs on the platform, based on their preferences. The two fundamental features implemented within the module are: (i) Secure and Reliable User Management and (ii) Dynamic Personalized Notifications.

- The User Management feature is interdependent with security-related activities within the frESCO Platform, allowing users to register, authorize, and authenticate themselves.
- The Notifications Alerts feature is in charge of detecting events linked to the state and progress of data collection jobs and analytics results, in order to provide suitable content to the frESCO Platform’s users.

3 DATA COLLECTION MODULE

3.1 Implementation Status

The release of the Data Collection Module, documented in the present deliverable, implements a set of features that are described as follows, also outlining their implementation status.

At this point, the implemented features of the Data Collection Module are:

- Configuration of the data ingestion process settings: comprises in detail the needs and preferences regarding the way that the data will be imported into the platform (API, file uploading).
- Importing of data from files: enables retrieval of data from files in formats that can be processed (e.g., csv).
- Reliable and secure data upload: Reliable and secure procedures are utilized in order for any organization's data to be uploaded to the frESCO Big Data Management Platform infrastructure.

Regarding the features with ongoing implementation, they can be listed as:

- Provision of harmonization guidance to the frESCO platform users: The frESCO UI to provide appropriate guidance to the users of the platform, to harmonize all of the imported data's attributes based on the frESCO Common Information Model (CIM).
- Definition of the steps for data curation: The data curation feature will ensure the quality and value of the data that will be imported.
- Data Importing from APIs: This feature will enable data to be retrieved from demonstrator systems APIs as well as Open Data APIs (e.g., weather data, other local sources).
- Matching of the imported data to the frESCO CIM: This feature will make the necessary adjustments to the data attributes to correspond to the CIM entity names.
- Harmonization actions on the imported data: This feature will ensure that by performing the necessary harmonization actions on the imported data, the data

values are matched to the CIM measurement units, whenever relevant, and cast to the relevant data types.

- Execution of data curation rules: The user determines the constraints and limitations that the data ingested into the module may have, as well as the precise actions that must be performed if any of these constraints are breached, during the curation configuration process. A curation rule is created by combining a validation option with a corrective action for a given attribute, and all curation rules are applied to the transformed data (rising from the process of semantic harmonization to the Common Information Model), resulting in curated data.

3.2 Frameworks and Libraries

The back - end of the Data Collection Module is developed with Python 3.10, also using a range of open-source technologies, namely Flask 1.1.4, Pandas 1.3.0 and Docker 20.10.11.

The front- end of the Data Collection Module is developed with Vue.js.

3.3 Interfaces

The Data Collection Module communicates with the other modules and features in the frESCO Big Data Management Platform through selected APIs that have been created and are documented in Swagger as depicted in the following figure.

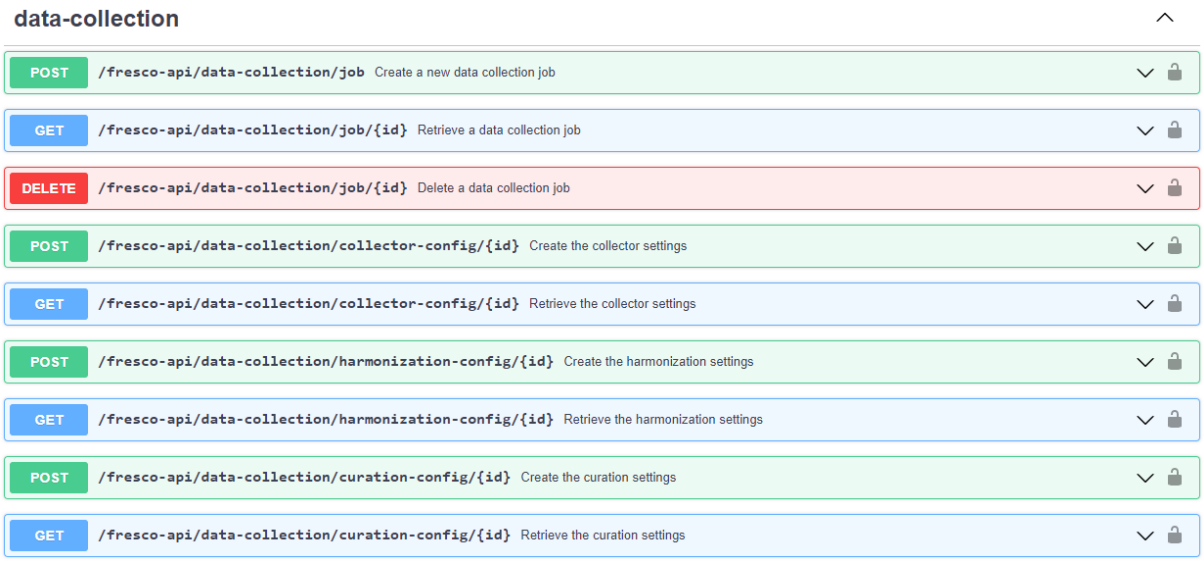

Figure 2 Swagger API Documentation for Data Collection Jobs

4 DATA SECURITY & STORAGE MODULE

4.1 Implementation Status

The release of the Data Security & Storage Module, documented in the present deliverable, implements a set of features that are described as follows, also outlining their implementation status.

At this point, the implemented features of the Data Security & Storage Module can be listed as:

- Management and handling of access policies in the backend: Policies can be used to determine when access is allowed.
- Enforcement of data access policies in the backend: When a request is made to access data that has been imported into the frESCO DMP, the access policies that have been established are applied. This feature affects both the results and the data received through the Data Search & Retrieval module, allowing for faster data access.
- Data storage: Storing various data (e.g., data assets, analytics models, and analytics results) as well as its metadata in order to make it available to all of the frESCO platform's modules and features.

Regarding the features with ongoing implementation, they can be listed as:

- Creation, configuration, and change of access policies: Using a graphical user interface, users can create advanced access policy rules for their assets. Policies can be used to determine when access is allowed. The rules are saved, and the data asset provider can easily modify them using the given interface.
- Log Data Storage: Storing log-related information for the frESCO Data Management platform's operation and usage, spanning from users and organizations, as well as any administrative data required for ensuring the seamless operation of the frESCO Platform.
- CIM Storage: Storing the different versions of the frESCO CIM, together with all relevant entities and attributes.

Some additional features for which implementation has not started but will be considered in future releases (should these become necessary) of the frESCO Data Management Platform, are presented as follows:

- Anonymization of data attributes: GDPR requires data asset providers to identify any data attributes that contain sensitive or identifying personal information.
- Data encryption rules execution: The Data Encryption feature shall provide appropriate encryption techniques and facilitate the data asset provider, in order to prevent unauthorized access to the data.
- Data decryption: The Data Encryption feature shall provide all the necessary services for the decryption of the data, so that it may be accessed by users who have been allowed access to the frESCO platform.

4.2 Frameworks and Libraries

The back - end of the Data Security & Storage Module is developed with Python 3.10, also using a range of open-source technologies, namely Flask 1.1.4, and Pandas 1.3.0.

The access policy management library of the Data Security & Storage Module is CASL.

4.3 Interfaces

The Data Security & Storage module communicates with the other modules and features in the frESCO Big Data Management Platform through selected APIs that have been created and are documented in Swagger as depicted in the following figure.

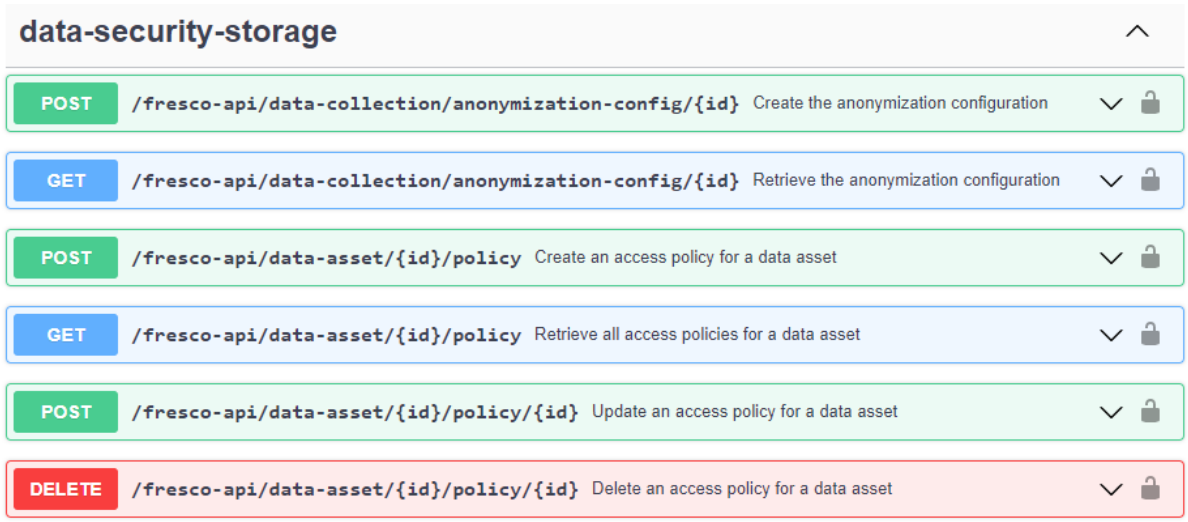

Figure 3 Swagger API Documentation for access policies

5 DATA SEARCH & RETRIEVAL MODULE

5.1 Implementation Status

The release of the Data Search & Retrieval Module, documented in the present deliverable, implements a set of features that are described as follows, also outlining their implementation status.

At this point, the features for which implementation is ongoing, can be listed as:

- User-friendly data search, introducing keyword and metadata searches: The Query Creation feature shall allow users to create queries using both a flexible free-text search and filtering on the data assets' information, giving them additional options in how they find and search for data. With this feature, the platform users will be enabled to search for data assets based on their metadata and content.
- Data asset profile creation: With this feature, any result of a data collection job is saved as a data asset in the frESCO Big Data Management Platform. The users of the platform are provided with the possibility to assign a title and respective metadata to a data asset and explain what is contained within it.
- Query Execution and Query results: The Query Execution feature shall be able to translate the data creator's query configuration into a query that can be executed

efficiently in the platform, and accordingly, the results that match the query are returned, processed, and presented to the user.

- Data and analytics results retrieval: Authorized apps shall be able to use the frESCO Open APIs to configure data retrieval from a single dataset or from analytics findings. Filters, which will be represented by API request parameters, and the selection of certain data attributes shall be utilized for fine-tuning the returned results. The API Gateway will create a unique identification, provide endpoint usage instructions, and contain a test API capability for the fast evaluation of the results.

Some additional features for which implementation has not started but will be considered in future releases (should these become necessary) of the frESCO Data Management Platform, are presented as follows:

- Search for Analytics Reports: A simple data search strategy that covers many asset categories, including analytics results.

5.2 Frameworks and Libraries

The Data Search & Retrieval module is developed with PostgreSQL 14, MongoDB 5.0 and Elasticsearch 7.14.

5.3 Interfaces

The Data Search & Retrieval module communicates with the other modules and features in the frESCO Big Data Management Platform through selected APIs that have been created and are documented in Swagger as depicted in the following figure.

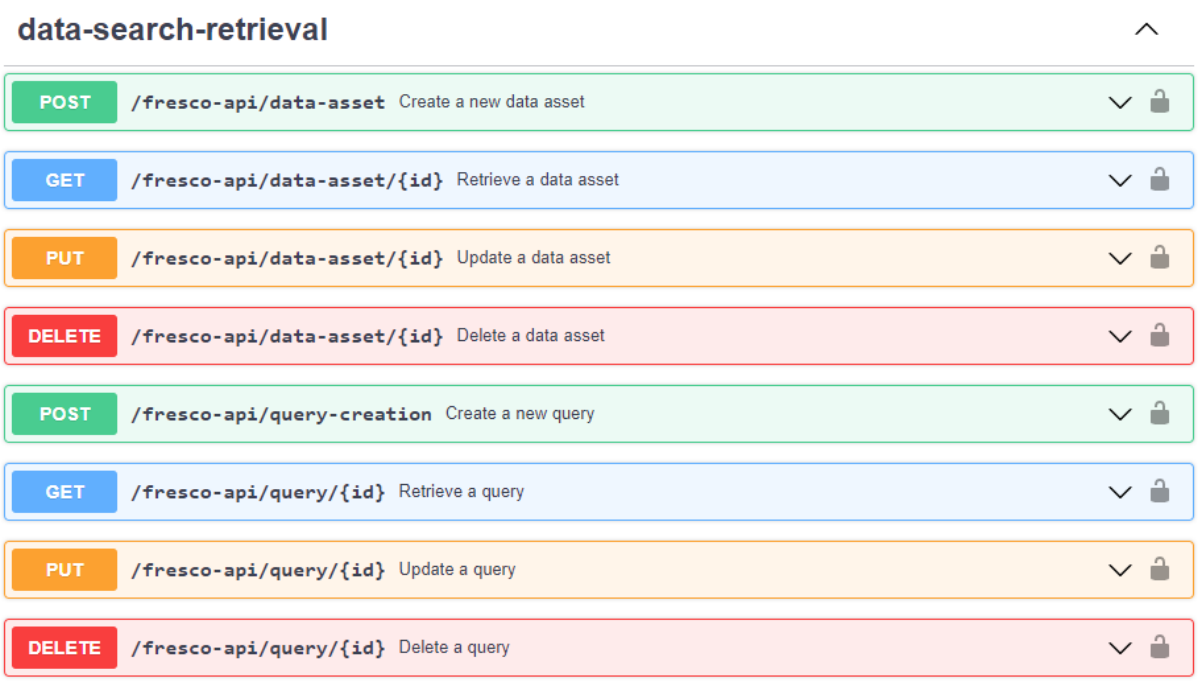

Figure 4 Swagger API documentation for Queries

6 PLATFORM GOVERNANCE MODULE

6.1 Implementation Status

The release of the Platform Governance documented in the present deliverable, implements a set of features that are described as follows, also outlining their implementation status.
At this point, the implemented features of the Platform Governance Module can be listed as:

- User Management: This feature represents the typical functionalities of an identity provider, creating and managing identification information for those who are authorized to access the frESCO platform.

Regarding the features for which implementation is ongoing, they can be listed as:

- Authentication and authorisation services for features and functionalities: The User Management feature shall offer authentication and authorisation services to limit access to authorized users.
- Notifications on different events in the frESCO Platform: With this feature, the platform users will be provided with appropriate information regarding the progress of a data import job's execution (successful or failed) and the progress of a data analysis execution (successful or failed).

Some additional features for which implementation has not started but will be considered in future releases (should these become necessary) of the frESCO Data Management Platform, are presented as follows:

- Notification management: Users shall be provided with the possibility to explore and act on various notifications, as well as delete them as needed. Also, the platform users shall be able to personalize which notifications they want to receive based on their particular preferences.

6.2 Frameworks and Libraries

The back-end of the Platform Governance Module is developed with Python 3.10, using also a range of open-source technologies, namely Django 3.2.9.

Websockets 10.1 and Kubernetes 1.22.2 are utilized for server-side notifications implementation and for the system deployment respectively.

6.3 Interfaces

The Platform Governance module communicates with the other modules and features in the frESCO Big Data Management Platform through selected APIs that have been created and are documented in Swagger as depicted in the following figure.

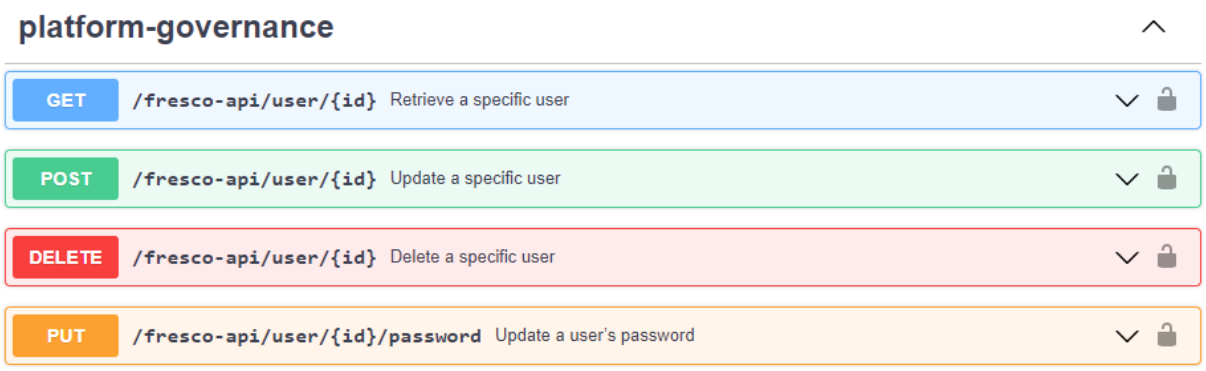

Figure 5 Swagger API documentation for Platform Users

7 AVAILABILITY AND LIMITATIONS OF THE FRESCO MODULES

7.1 Installations

The Data Collection, Data Security & Storage, Data Search & Retrieval and Platform Governance Modules are available in a private repository. Since all subcomponents are already packaged as Docker containers, detailed instructions for the module's deployment are provided in the corresponding private code repository.

7.2 Licensing and Access Granting

The Data Collection, Data Security & Storage, Data Search & Retrieval, and Platform Governance modules contain only closed-source code. Through the integrated frESCO platform, the versions of the various modules that are to be deployed, shall be accessed (as documented in D4.4 for the alpha release and in the upcoming D4.6 for the beta release).

8 CONCLUSIONS

The objective of deliverable D4.3, "Data Collection, Security, Storage, and Management Services Bundles," was to deliver the initial version of the modules that were implemented in WP4. It also included a brief report with the implementation status of each feature of the respective module, the frameworks and libraries utilized and the corresponding interfaces.
To that aim, the deliverable extensively detailed the initial version of the four (4) of the five modules implemented in WP4, namely the Data Collection, Data Security & Storage, Data Search & Retrieval, and Platform Governance modules. The Data Analytics Module will be described in detail in Deliverable D4.5. More specifically, this deliverable represented for every one of the described modules:

- According to frESCO Conceptual Architecture deliverable D2.5, the actual implementation status of each feature developed that make up each module.
- The details for each module's Frameworks and Libraries, including the version used.
- The Swagger interface that is used to create the technical documentation for the APIs supplied by each module.

Based on the foundations that were established in the early mock-up release of the frESCO Platform (described in D4.4), the current deliverable is the first version of the modules released in M18 in accordance with the frESCO DoA, offering the back-end implementation details and sets the ground for the upcoming beta version of the frESCO Integrated Platform that will be provided M20 (D4.6), which will include the matching User Interface implementation. Nonetheless, the module delivery in WP4 is a live operation that will continue until M32. Any changes and improvements performed, based on the project's progress and any new requirements that may emerge will be included in the updated versions of the frESCO Big Data Management Platform.

REFERENCES

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

frESCO Consortium. (2020). frESCO D4.4 “frESCO Integrated Platform – Alpha, mock-ups release”

frESCO Consortium. (2020). frESCO D4.5 “frESCO Baseline Data Analytics – Draft Release”

frESCO Consortium. (2020). frESCO D4.6 “frESCO Integrated Platform – Beta Release”