Deliverable D3.3 –Semantic Broker Service

Responsible Partner:	SINTEF	18.07.2023
Contributor(s):	Treesa Rose Joseph (SINTEF), Bjørn Tore Løvfall (SINTEF)	18.07.2023
Reviewer(s):	Anders Eklund and Kristine Wiik (SINTEF)	21.07.2023
Coordinator:	CMCL Innovations	21.07.2023
Dissemination Level:	Public	
Due Date:	M32 (July 2023)	
Submission Date:	31.July.2023	

Project Profile

Programme	Horizon 2020
Call	H2020-NMBP-TO-IND-2020-twostage
Topic	DT-NMBP-40-2020
	Creating an open marketplace for industrial data (RIA)
Project number	953163
Acronym	DOME 4.0
Title	Digital Open Marketplace Ecosystem 4.0
Start Date	December 1st, 2020
Duration	48 months

Document History

Version	Date	Author	Remarks
V0.1	27.06.2023	Treesa Rose Joseph	Technical specifications and content
V0.2	30.06.2023	Bjørn Tore Løvfall	Added some comments and a few lessons learnt
V0.3	11.07.2023	Anders Eklund and Kristine Wiik	Review and add reference
V0.4	21.07.2023	Amit Bhave	Review and comment
V0.5	24.07.2023	Adham Hashibon	Add lessons learnt
V1	25.07.2023	Willem van Dorp	Finalize document

Executive Summary

This document is supplementing the Ontology-driven Semantic Broker that comprises of:

- Discovery Service
- Registration Service
- Knowledge Service

We here share the details of how the broker integrates with the rest of the DOME 4.0 platform and its key features and internal workings.

Table of Contents

List of Figures

1. Introduction

This document reports on the activities and outcome of Task 3.3 of the DOME 4.0 project, named "Ontology-driven Semantic Broker". The primary result of this task is the development of the semantic broker component, which enables data discovery by establishing smooth connections to various data providers. It facilitates the easy search of data by querying various data providers. It supports the registration of both data providers and consumers on the DOME 4.0 platform. Furthermore, and most importantly, it performs a semantic, ontology compliant search to identify data source and data service candidates and facilitating the negotiation of terms of data exchange. The development of this system builds upon the overall architecture of DOME 4.0 which is shown in Figure 1. It essentially facilitates the development and provision of a matching algorithms that maps the requirements of end users with service and data providers in a meaningful (i.e., semantic) manner. It utilises the development and support for both semantic and syntactic keywords in in D3.3 (the data set ontology), it allows augmenting the semantic matching with a simple ad hoc keyword matching, catering hence for legacy non semantic data sources as well as modern semantic ones.
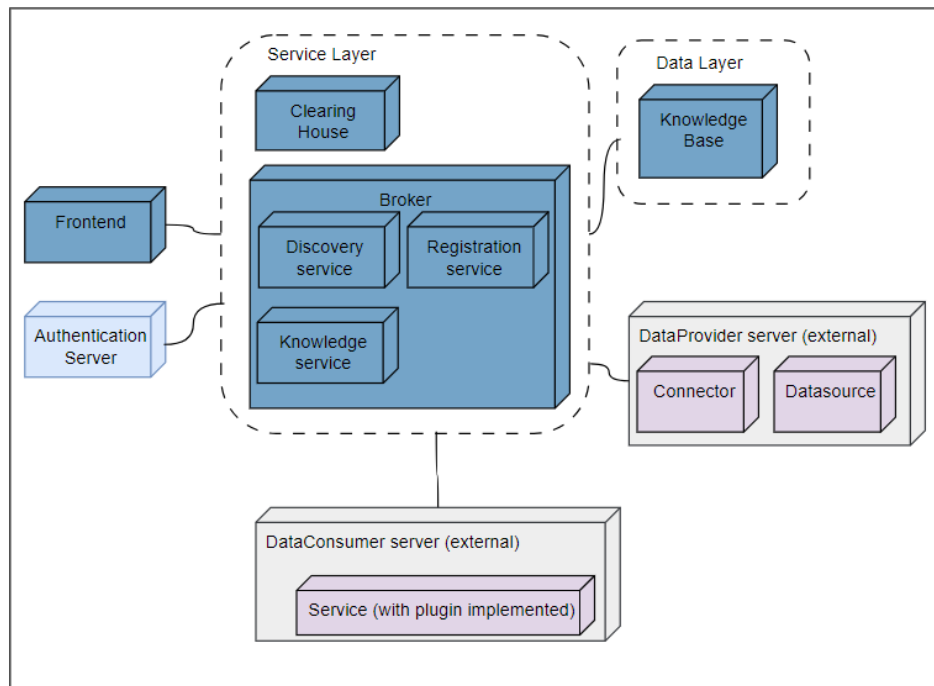
Figure 1: Components that interact with the Semantic Broker

1.1 Objective

As the name suggests, the main aim of the semantic broker is brokering of data. The semantic broker performs a semantic and syntactic keyword search to identify data sources/providers and data consumer candidates and facilitates the terms of data exchange by making use of the clearing house. It also helps in the onboarding of data providers and consumers.

When a user performs a search on the DOME 4.0 platform, the broker queries the knowledge base, finds and matchers (i.e., brokers) the relevant data providers, and performs a search on them. It also displays all registered compatible data consumers (i.e., WP2 data tools and services) that can be used to work with data retrieved from the knowledge base.

Another objective is that it should be easy for third party data providers and consumers to add new databases and tools/services to the platform and make it semantically discoverable by the broker. The broker provides registration APIs that help providers and consumers register/document their data, plugins, and connectors with the DOME 4.0 platform.

While most of this functionality is already covered, achieving a fully ontologically semantic broker depends on the availability of specific functionality from the backend knowledge services, clearing house, and connector. Currently, intensive work is undergoing to complete the semantic services relying on the SimPhoNy Open Semantic Platform (OSP) which will provide deep integration with ontologies, dynamics semantic brokerage and full provenance across all services on DOME 4.0. This work will be further developed as part od Task 3.6 and will be readily integrated with the broker enhancing the services overall. A full detailed report will be augmented with D3.6.

1.2 Key Functionality

The Semantic broker acts as an intermediary between users/data-consumer-tools and data-providers. It makes use of other components of the platform like the clearing house, knowledge base, etc. to help establish that link and make data flow seamless. Figure 2 shows an overview of user scenarios and the components of the Broker that are responsible.

The main functionalities are:

- Users can search multiple databases registered on the DOME 4.0 platform and filter results according to their need (i.e., semantic, and syntactic matching scores).
- Users can find compatible tools and services based on the type of data or application that they are accessing depending on semantic and syntactic keywords.
- Users can register/catalogue their data on the platform, so they are available for other services.
- Users can register their database as a data provider with the platform so that they are discoverable and accessible by other users.
- Users can register tools and services that can consume data from the platform to visualize data or run various simulations.
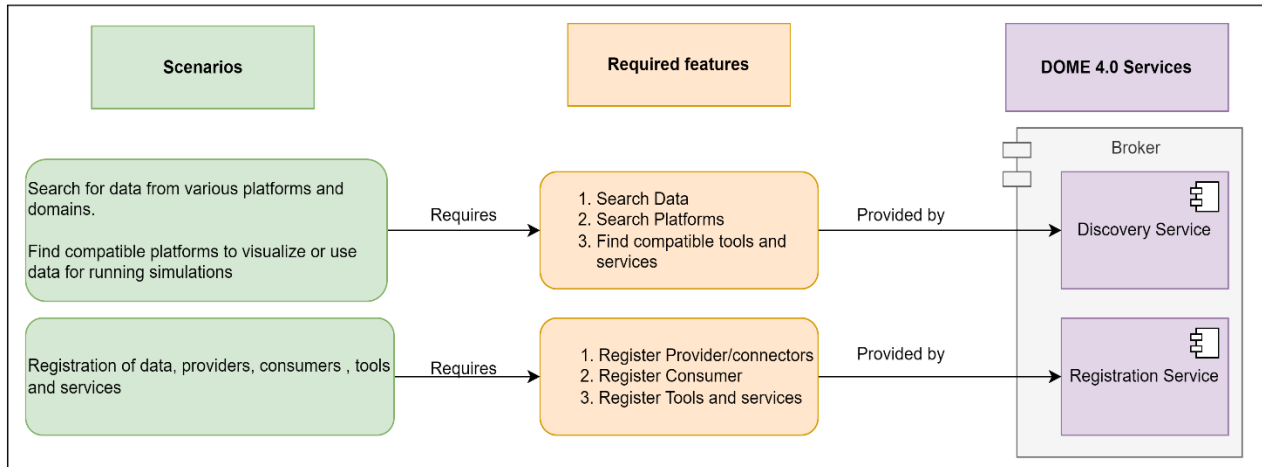
Figure 2: Broker user-case scenarios and the services that are responsible to provide the required features

2. Implementation

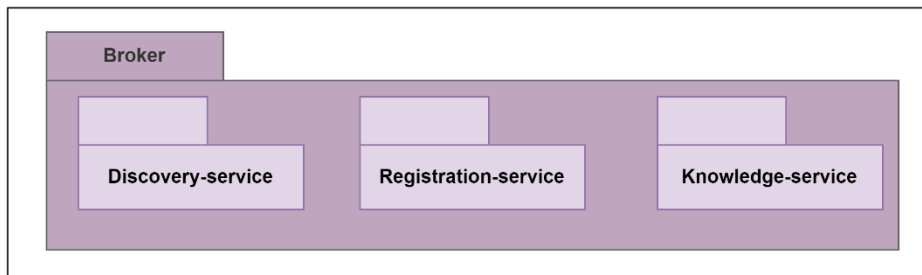## 2.1 Components of Semantic Broker

Figure 3: Components of the Semantic Broker

The Semantic Broker is part of the service layer of the DOME 4.0 platform. The broker consists of three services that work towards efficient data brokerage: the Discovery Service, the Registration Service, and the Knowledge Service as shown in Figure 3.

2.1.1 Discovery Service

The Discovery Service is a set of REST APIs that enables the search for data providers and their information based on semantic mapping of the keywords entered by the user/agent and the backend data services. It

allows users to search for data providers and their offerings based on specific criteria. The service facilitates the retrieval of data and metadata from data providers, allowing users to fetch information based on applied filters. This functionality relies on the connectors implemented by the data providers, which need to support and provide these APIs to fetch such data. The Discovery Service helps identify compatible platforms to visualize and use data for running simulations and other workflows. This service should have both internal and secured exposed search APIs. The internal APIs will be used by the frontend service, while the exposed APIs will enable users to directly query all providers or a particular platform. The exposed APIs can be used by a machine to extract data from DOME 4.0 platform.

The service is also able to discover relevant data catalogues and provide references to them, but since these are not registered as data providers (with integrating Connectors) their data must be fetched by the user in separate processes outside the DOME 4.0 platform.

2.1.1.1 Discovery Process

The Discovery process begins when a user initiates a search within the DOME 4.0 platform. When the search keyword reaches the discovery service, it collaborates with the knowledge service to retrieve relevant data providers and relevant data catalogs. The discovery service then verifies with the clearing house whether the user is authorized to access the returned providers.

If the user has authorization to access the data providers, the discovery service proceeds to query the connectors associated with those providers. The connectors also perform authorization checks for the user. Once the authorization flows of the connectors are completed, the data is fetched and returned to the discovery service, which subsequently presents it to the user.

In cases where the user lacks authorization to access the data, only metadata will be visible. Depending on the platform type, the user may purchase the data on a paid platform or agree to the terms and conditions on an open platform to get data. The data discovery process flow is represented in Figure 4.
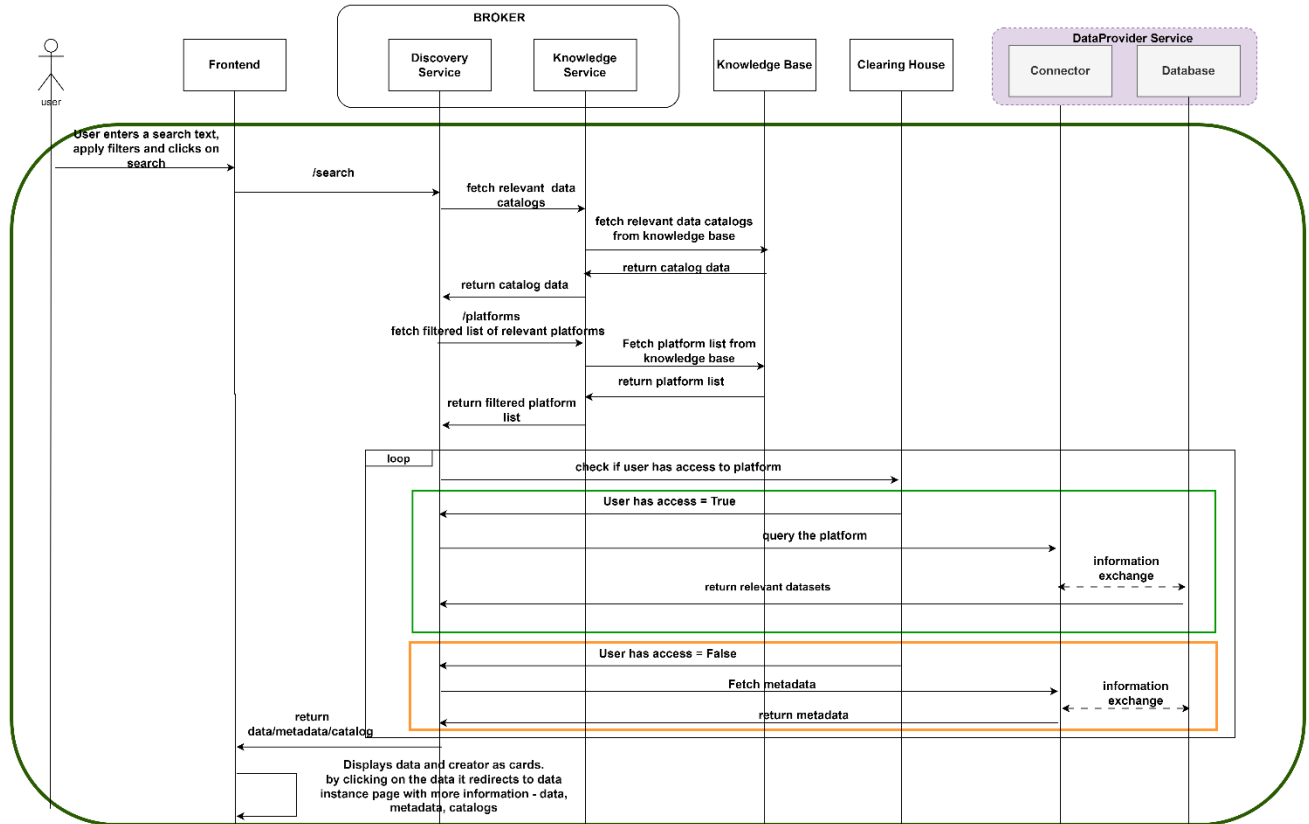

Figure 4: Data Discovery process

2.1.2 Registration Service

The Registration Service is a set of REST APIs that enables the registration of data providers and data consumers (data compatible tools and services). The service allows data providers to register their data catalogs or register connectors for their databases, enabling the DOME 4.0 platform to access their data. The service also enables the registration of data consumers, which can consume data from the DOME 4.0 platform, as well as other tools and services that are compatible with the platform. The Registration Service is essential for the efficient management of the DOME 4.0 platform. It enables the platform to onboard new data providers, data consumers, and other tools and services. This service is used by the frontend service internally. Based on the information given in the registration process, the semantic broker will be able to match relevant data sources with relevant tools or services, to automatically provide the user with easy access to relevant tools and services.

2.1.2.1 Provider Registration

During the registration process, the user or data provider utilizes the connector template provided by DOME 4.0. They make necessary modifications to the template, enabling DOME 4.0 to uniformly query data and authenticate databases. Once the connector is developed, the data provider proceeds to the provider registration page and completes the form on the DOME website. After submitting the form, the information of the data provider is stored in the ecosystem ontology within the knowledge base of the DOME 4.0 platform.

Simultaneously, a corresponding provider is created within the clearing house, and it is registered with OAuth2. The OAuth2 credentials and URLs are then provided to the data provider. At this stage, the data provider must update the developed connector to establish mutual authentication between DOME 4.0 and the connector. The provider registration flow is represented in Figure 5. For this authentication flow, we make use of an open-source identity and access management solution called Keycloak [1].
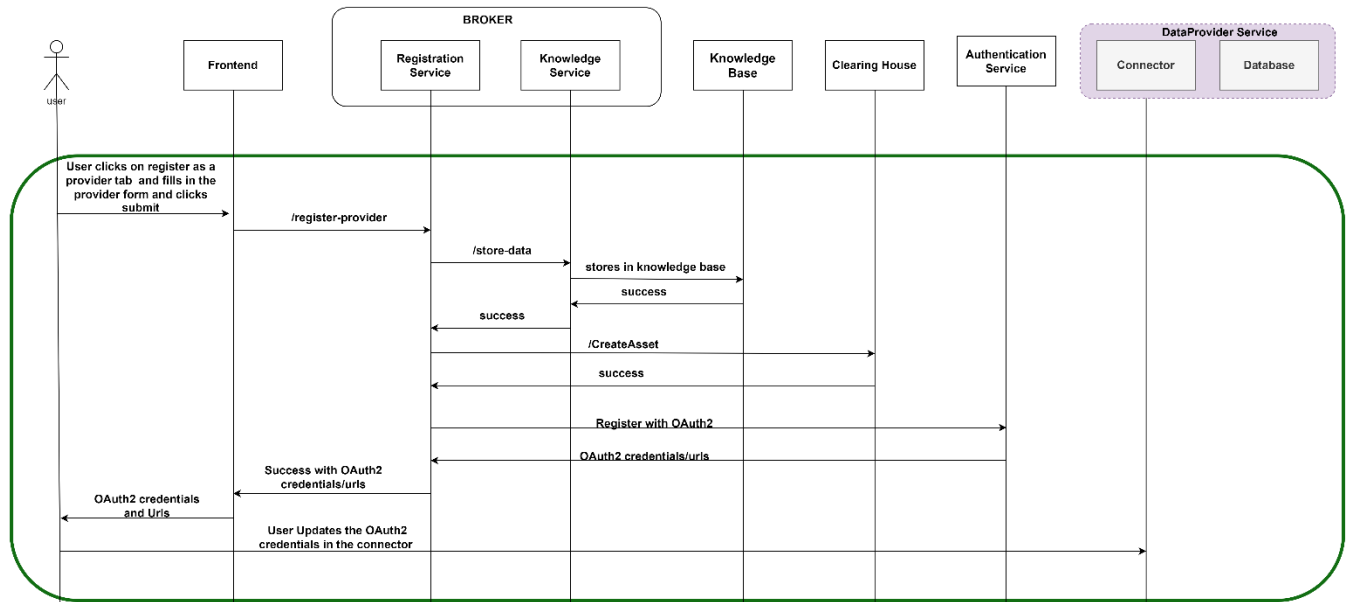

Figure 5: Data Provider registration process

2.1.2.2 Consumer Registration

Here consumer registration refers to the tools and services (developed in T3.7) that can be registered with the DOME platform. Other data consumers are users, and they are registered with DOME when they login to the website (I.e., create an account with DOME). During the registration process, the data consumer utilizes the plugin template provided by DOME 4.0 (developed as part of D3.7). They modify the plugin template to establish connections with tools and services. Third-party providers of these tools and services who wish to register on DOME are required to visit the consumer registration page on the DOME 4.0 platform. The consumer registration page provides guidance to the tools/service providers throughout the registration process. This involves completing the form and documenting essential features of their service. These features encompass details such as the accepted types and formats of data inputs and the outputs provided by the service.

Once the registration form is submitted, the information pertaining to the data consumer (tools/services) is stored in the ecosystem ontology (D3.2) within the knowledge base of the DOME 4.0 platform. Concurrently, a corresponding consumer entry is created within the clearing house and registered with OAuth2. The data consumer is then provided with OAuth2 credentials and URLs. At this stage, the data consumer must update the developed plugin to facilitate mutual authentication between DOME 4.0 and the plugin of the tool/service. The consumer registration flow is illustrated in Figure 6.
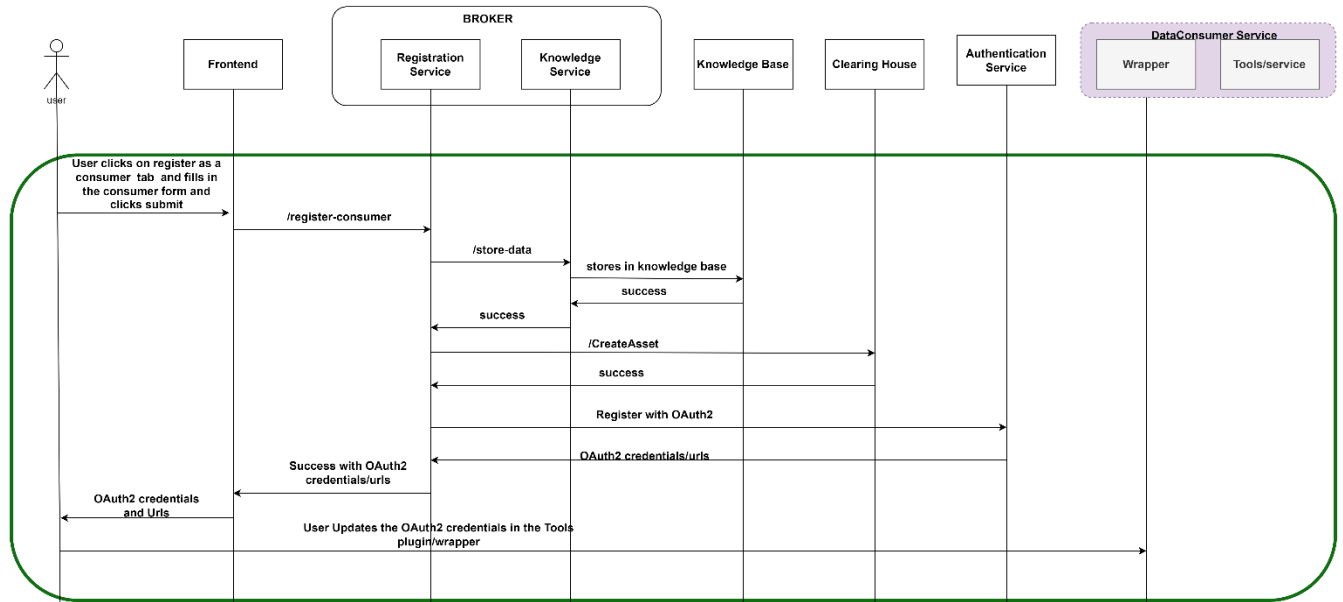
Figure 6 Consumer registration process

2.1.2.3 Data Catalog Registration

To register a data catalog, users can access the "Upload Catalog Data" tab in the DOME 4.0 platform and complete the provided form. Once the form is submitted, the registration service registers this catalog information in the knowledge base. The data set ontology developed under T3.1 is used in the knowledge base and then populated with the catalog information. Figure 7 shows data catalog registration flow.
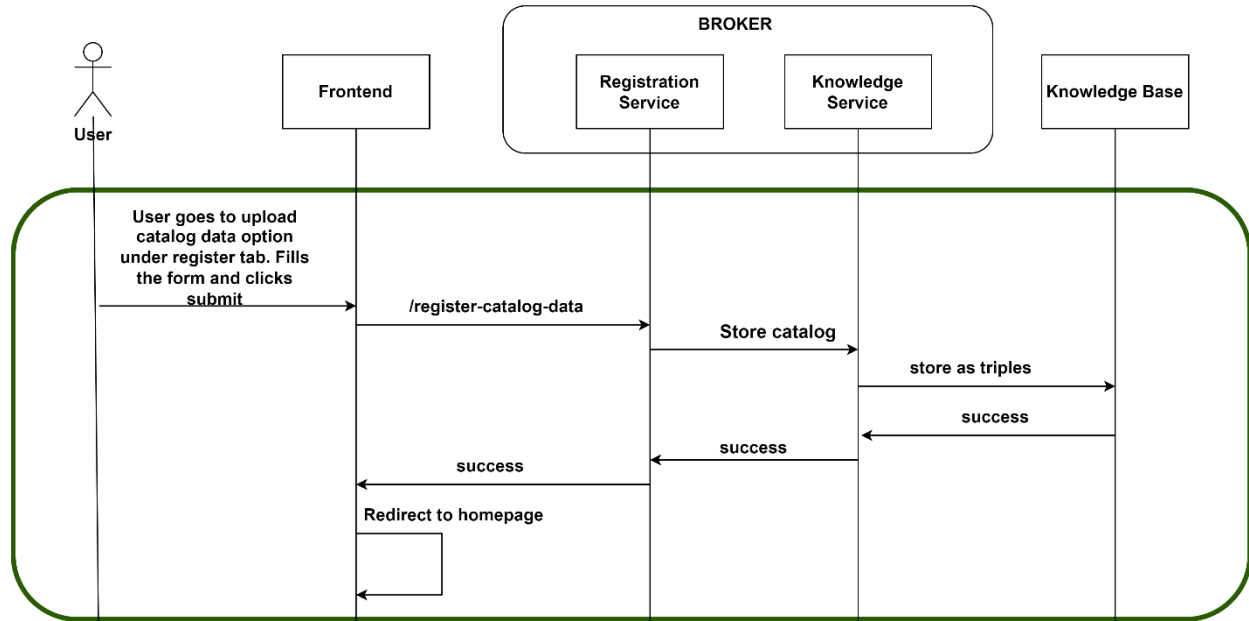
Figure 7:Data Catalog registration process

2.1.3 Knowledge Service

The Knowledge Service plays a crucial role in the ontology-driven semantic broker. It serves as a connection point to the knowledge base, which houses the relevant ontologies (e.g.: the ecosystem ontology (T3.2) and dataset ontology (T3.1)) describing the relations utilized by the broker to offer its diverse functionalities. The Knowledge Service ensures that the broker has access to the necessary semantic resources.

The Knowledge Service consists of a set of REST APIs that interact with the Knowledge base/triple store. It separates the APIs connecting to the triple store, making it easier to replace the current knowledge base with another in the future and to maintain the platform overall. This service is used by the discovery and registration services internally on the DOME 4.0 platform and for the purpose of securing metadata and ontologies not accessible to users.

This service contains APIs that search the knowledge base to find relevant data providers and compatible data consumers w.r.t to the information stored during the registration process.

3. Conclusions / Next steps

As part of this task, we have developed the broker which consists of the discovery service, the registration service, and the knowledge service. We can register providers, consumers and data catalogs, search for data and providers and find compatible data consumers for data.

Next Steps:

- As part of T1.4, we improve the user interface. i.e., improve the registration pages once the connector implementation is completed as part of T3.4.
- As paer od T3.6, further deep semantic integration with ontology is ongoing. It will enable more efficient semantic matching (brokerage) of the user search keywords and the backend stored information and metadata regarding third party resources and platforms. It will enable furthermore a native data store on DOME 4.0 itself.

4. Lessons learnt

The semantic broker is an integrator for a lot of the components on the DOME 4.0 platform. It is challenging to design and implement this in parallel with the other components. At the same time, the broker is required for testing that all the individual components fit together. The lesson is that the agile approach with several small demonstrators (Minimum Viable Products) was crucial to avoid divergence in the implementation of the different components that the broker was depending on.

One of the aspects that could be optimised relatively easily and is in fact ongoing work, is to replace the backend proprietary closed source triple store with one that is a) open source, b) supports ontology natively, and c) allows still the use of plethora of backends for persistent storage. Work is ongoing through Task 3.6 to finalise the integration of a semantic data backend for DOME 4.0 which will enable management of the entire ontology stack (including the eco system ontology, and use case ontologies, clearing house, provenance).

5. Deviations from Annex 1

No deviations.

6. References

[1] Keycloak, "Keycloak," [Online]. Available: https://www.keycloak.org/docs/latest/. [Accessed 11 07 2023].

7. Acknowledgement

The author(s) would like to thank the partners in the project for their valuable comments on previous drafts and for performing the review.

Project partners:

#	Type	Partner	Partner full name
1	SME	CMCL	Computational Modelling Cambridge Limited
2	Research	FHG	Fraunhofer Gesellschaft zur Förderung der Angewandten Forschung E.V.
3	Research	INTRA	Intrasoft International SA
4	University	UNIBO	Alma Mater Studiorum – Universita di Bologna
5	University	EPFL	Ecole Polytechnique Federale de Lausanne
6	Research	UKRI	United Kingdom Research and Innovation
7	Large Industry	SISW	Siemens Industry Software NV
8	Large Industry	BOSCH	Robert Bosch GmbH
9	SME	UNR	Uniresearch B.V.
10	Research	SINTEF	SINTEF AS
11	SME	CNT	Cambridge Nanomaterials Technology LTD
12	University	UCL	University College London

8. Table of Abbreviations

Abbreviation	Explanation
API	Application programming interface
URL	Uniform resource locator
UI	User interface