



Policy Cloud
Cloud for Data-Driven Policy Management

CLOUD FOR DATA-DRIVEN POLICY MANAGEMENT

Project Number: 870675

Start Date of Project: 01/01/2020

Duration: 36 months

D7.11 DATA MARKETPLACE: DESIGN AND OPEN SPECIFICATION M32

Dissemination Level	PU
Due Date of Deliverable	(31/08/2022), Month 32
Actual Submission Date	28/07/2022
Work Package	WP7, Communication, Exploitation, Standardisation, Roadmapping & Business Development
Task	T7.2
Type	Report
Approval Status	
Version	v1.0
Number of Pages	p.1 – p.52

Abstract: The deliverable D7.11 Data Marketplace: Design and Open Specification describes the final design and specifications of the PolicyCLOUD Data Marketplace. The latter will be a unified web-based platform consisting of two (2) core services, its front-end and back-end services, offering to its users various ready-to-use solutions, by supporting different kinds of assets.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.



PolicyCloud has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870675.

Versioning and Contribution History

Version	Date	Reason	Author
0.1	06/06/2022	Revision of the document's first version and determination of the new ToC	Argyro Mavrogiorgou, Thanos Kiourtis, Vasilis Koukos, Eleftheria Kouremenou (UPRC)
0.2	14/06/2022	Updates in Sections 1, 2	Argyro Mavrogiorgou, Thanos Kiourtis, Vasilis Koukos (UPRC)
0.3	17/06/2022	Updates in Sections 3.1, 3.3, 3.4, 4, 5	Argyro Mavrogiorgou, Thanos Kiourtis, Vasilis Koukos (UPRC)
0.4	28/06/2022	Updates in Sections 2.2, 3.2	Argyro Mavrogiorgou, Eleftheria Kouremenou, Thanos Kiourtis (UPRC)
0.5	13/07/2022	Check and revision of all Sections, New vizualized material for Section 3.2	Argyro Mavrogiorgou, Thanos Kiourtis, Fotis Karagiannis (UPRC), Marieke Willems (TRUST-IT)
0.6	22/07/2022	Internal review	Panayiotis Michael (ICCS), Giannis Ledakis (UBI)
0.7	25/07/2022	Updates based on received feedback	Argyro Mavrogiorgou, Thanos Kiourtis (UPRC)
0.8	26/07/2022	Quality check	Argyro Mavrogiorgou (UPRC)
1.0	28/07/2022	Submitted version	Atos

Author List

Organisation	Name
UPRC	Argyro Mavrogiorgou
UPRC	Thanos Kiourtis
UPRC	Vasilis Koukos
UPRC	Eleftheria Kouremenou
UPRC	Fotis Karagiannis
TRUST-IT	Marieke Willems

Abbreviations and Acronyms

Abbreviation/Acronym	Definition
AI	Artificial Intelligence
API	Application Programming Interface
AWS	Amazon Web Services
CRUD	Create Retrieve Update Delete
DB	Database
DL	Deep Learning
EC	European Commission
EOSC	European Open Science Cloud
EU	European Union
FS	File System
GB	Gigabyte
GUI	Graphical User Interface
HTTP	Hypertext Transfer Protocol
ID / IDs	Identifier / Identifiers
IDE	Integrated Development Environment
IoT	Internet of Things
IT	Information Technology
JSON	JavaScript Object Notation
ML	Machine Learning
NoSQL	Not only SQL
OS	Operating System
PDT	Policy Development Toolkit
REST	Representational State Transfer
SQL	Structured Query Language
UI	User Interface
WSGI	Web Server Gateway Interface

Contents

Versioning and Contribution History.....	2
Author List.....	2
Abbreviations and Acronyms	3
Executive Summary	7
1 Introduction.....	8
1.1 Objective of the Deliverable.....	8
1.2 Structure of the Deliverable.....	8
1.3 Summary of Changes.....	9
2 Related Work.....	10
2.1 Categories of Data Marketplace	10
2.2 Relevant Work in Data Marketplaces.....	14
2.2.1 Acumos Marketplace	14
2.2.2 5GTANGO Catalogue.....	15
2.2.3 FINSEC Marketplace	17
2.2.4 Our World in Data	18
2.2.5 Amazon Web Services Marketplace.....	19
2.2.6 HERE Olp.....	20
2.2.7 Streamr	21
2.2.8 Data Intelligence Hub	22
2.2.9 Open Data Hub.....	23
2.2.10 IoT Catalogue	24
3 Scope, Users & Assets of Data Marketplace	26
3.1 Scope.....	26
3.2 User Journeys.....	26
3.2.1 Adoption of Organic Food in Hotel Menus	27
3.2.2 Association of Terrorism with Geolocation	28
3.2.3 Optimization of Organic Food in Hotel Menus.....	29
3.2.4 Future Trends of Radicalization.....	30
3.2.5 Launch of New Wine Products	31
3.2.6 Environmental Interventions for Better Quality of Life.....	32

3.3	Offered Assets	33
3.4	Users	34
4	Architecture of Data Marketplace	36
4.1	Back-end	38
4.1.1	Assets Storage Layer	38
4.1.2	Assets Management Layer	39
4.1.3	Interaction Layer	42
4.2	Front-end	43
4.3	Overall Conceptual View	45
4.3.1	Use Case: Upload Assets	46
4.3.2	Use Case: Update Assets	47
4.3.3	Use Case: Retrieve Assets	47
4.3.4	Use Case: Delete Assets	48
4.3.5	Use Case: Search Assets	49
5	Conclusion	50
	References	51

List of Tables

Table 1 - Taxonomy for data ecosystems.	10
Table 2 - Types of Marketplaces.....	11
Table 3 - Classification of Marketplaces.	12
Table 4 - User Journey for a marketing director.	27
Table 5 - User Journey for a psychology researcher.	28
Table 6 - User Journey for a development consultant.	29
Table 7 - User Journey for a policy maker.....	30
Table 8 - User Journey for a marketing director of a wine company.	31
Table 9 - User Journey for an environmental researcher.	32
Table 10 - Matching (S)CRUD operations with HTTP methods and MongoDB statements.	41

List of Figures

Figure 1 - Snapshot of the “Marketplace” page of the Acumos Marketplace.	14
Figure 2 - Snapshot of the “Model Details” page of the Acumos Marketplace.	15
Figure 3 - 5GTANGO overall architecture design and the role of the Catalogue.....	16
Figure 4 - Snapshot of the Finsecurity.eu platform, including a list of its solutions.	17
Figure 5 - Snapshot of Our World in Data Homepage.	18
Figure 6 - Snapshot of the AWS Marketplace’s Popular Categories.	19
Figure 7 - Snapshot of the HERE OLP’s Marketplace.	20
Figure 8 - Snapshot of the Streamr Homepage.	21
Figure 9 - Snapshot of Streamr’s categories.....	21
Figure 10 - Snapshot of Data Intelligence Hub’s Homepage.....	22
Figure 11 - Snapshot of the Open Data Hub Homepage.....	23
Figure 12 - Snapshot of the IoT-Catalogue.	24
Figure 13 - Example of a Product’s presentation in the IoT-Catalogue.	25
Figure 14 - Data Marketplace’s interacting users.....	35
Figure 15 - Data Marketplace Architecture.....	36
Figure 16 - Data Marketplace Layers and main functionalities.	37
Figure 17 - Assets Storage Layer.	39
Figure 18 - Assets Management Layer.	40
Figure 19 - Interaction Layer.....	42
Figure 20 - Presentation Layer.	43
Figure 21 - Site-map of the front-end.....	44
Figure 22 - Overall Conceptual View.....	45
Figure 23 - Upload use case.....	46
Figure 24 - Update use case.....	47
Figure 25 - Retrieve use case.....	48
Figure 26 - Delete use case.....	48
Figure 27 - Search use case.....	49

Executive Summary

This deliverable (entitled “D7.11 - Data Marketplace: Design and Open Specification M32”) is a follow up on the design and the specifications of the PolicyCLOUD Data Marketplace as defined in the previous version of the deliverable (i.e. the “D7.4 - Data Marketplace: Design and Open Specification”). Several updates have been introduced in the deliverable, including, among others, the final types of assets to be offered by the Data Marketplace, additional user journeys and state of the art marketplaces, whereas there are additional updates on the specifications of the Data Marketplace’s layers (especially for the Assets Storage Layer and the Assets Management Layer). Finally, some use cases of the Data Marketplace are introduced under the Overall Conceptual View Section.

In essence, the Data Marketplace is a unified web-based platform consisting of two (2) core services, the front-end and the back-end services, offering its users various ready-to-use solutions. More specifically, it provides to the wider research and innovation community various assets (i.e. objects/solutions) in different domains.

As a Marketplace, the final design provides information on how several users (and contributors) can interact with the PolicyCLOUD Data Marketplace in order to ingest and retrieve the respective assets, through potential user journeys that are analyzed. Moreover, the deliverable describes the different types of offered assets, since the Data Marketplace went beyond policies and tools and it stores and offers additional elements, including rich descriptions of the offered assets, algorithms, validation datasets, and in general, various outcomes of the project.

To support the aforementioned functionalities, the PolicyCLOUD Data Marketplace specifications exemplify the respective layers of the Data Marketplace, the functionalities supported by each layer, the interaction points with different stakeholders, as well as the technical details that have been used for the implementation of the Data Marketplace. Finally, the document reviews the current state of the art marketplaces to identify the baseline technologies and approaches for the realization of similar marketplaces, which are compared to the corresponding specifications of the Data Marketplace.

1 Introduction

PolicyCLOUD aims to be a cloud-based data-driven policy management platform, enabling its stakeholders to model, analyze, evaluate and optimize their policies using a variety of tools and services. Among the components that have been developed and offered from the consortium is the Data Marketplace.

The main idea of the Data Marketplace is to create a community of users who will be able, through the developed platform, to provide and share various assets related to the scope of the PolicyCLOUD. It also aims at being one of the project's main ambassadors to different external communities (third parties), by disseminating and demonstrating the main outcomes of the project. To this end, the platform has been developed to provide the means for storing and retrieving (after searching) several types of assets. All the diverse types of assets that are provided by the Data Marketplace are thoroughly analyzed in [Section 3.3](#) (Offered Assets).

In summary, the Data Marketplace is consisted of a public web-based environment with many different APIs and functionalities, in order to cover all the different requirements of the project's stakeholders. In the following Sections of this document, the design and the specifications of the PolicyCLOUD Data Marketplace are analyzed in detail.

1.1 Objective of the Deliverable

The main objective of this deliverable is to provide the ground for the realization of the PolicyCLOUD Data Marketplace. It will analyze the architecture and the final specifications of the Data Marketplace. In addition, the deliverable outlines the potential users of the Data Marketplace (i.e. providers and consumers) and clarifies the reason for its implementation and existence, describing (extra) diverse user journeys. Finally, it lists the types of assets (offered assets) that are populating the Data Marketplace and are provided through it to the end-users.

1.2 Structure of the Deliverable

This document is structured as follows: [Section 2](#) introduces relevant work and approaches that serve as market platforms to identify potential links and baseline technologies for the Data Marketplace. [Section 3](#) describes the scope of the Data Marketplace and the users that will interact with it, along with some user journeys. It also lists the offered assets towards these users of the Data Marketplace. [Section 4](#) is the core Section of the deliverable, presenting in deep detail the architecture and the specifications of the Data Marketplace, analyzing its functionalities/capabilities that will be supported. The deliverable concludes with a reference to the presented material in [Section 5](#).

1.3 Summary of Changes

This Section highlights the updates made to the previous version of this deliverable D7.4 (Data Marketplace: Design and Open Specification):

- Included Sub-section 1.3 (Summary of Changes).
- Updated Section 2 (Related Work) with respect to the SotA Marketplaces and their most widely used types. More specifically, Section 2 has been split to two (2) Sub-sections, Sub-section 2.1 and Sub-section 2.2. Sub-section 2.1 describes a variety of Marketplace typologies/categories that have been developed in recent research along with the important works that have been recognized and used as a basis for the final version of the architecture and the specifications of the PolicyCLOUD Data Marketplace. Sub-section 2.2 contains the already described (from the first version) related Marketplaces and has been expanded with additional Marketplaces that were exploited for the final design of the PolicyCLOUD Data Marketplace.
- Extended Sub-section 3.2 (User Journeys) in order to analyze more user journeys, representing different kinds of end-users of the platform. More specifically the user journeys that are analyzed are the following: Adoption of Organic Food in Hotel Menus (S3.2.1), Association of Terrorism with Geolocation (S3.2.2), Optimization of Organic Food in Hotel Menus (S3.2.3), Future Trends of Radicalization (S3.2.4), Launch of New Wine Products (S3.2.5), Environmental Interventions for Better Quality of Life (S3.2.6).
- Updated Sub-section 3.3 (Offered assets) in order to describe the final types of the Data Marketplace's offered assets. More specifically, the final types of offered assets are the following: Tools, Policies, Datasets, Documents, Webinars, Tutorials and Other (more information can be found in the corresponding Section).
- Updated Sub-section 4.1 (Back-end Architecture) with respect to its final specifications.
- Extended Sub-section 4.2 (Front-end Architecture) in order to depict the visual sitemap of the front-end of the Data Marketplace along with a relevant description for the supported functionalities.
- Added new Sub-sections in Sub-section 4.3, describing the core use cases of the Data Marketplace (i.e. use cases for the functionalities on how to: Upload assets (S4.3.1), Update assets (S4.3.2), Retrieve assets (S4.3.3), Delete assets (S4.3.4), and Search assets (S4.3.5)).

2 Related Work

This Section describes a number of different typologies/categories of Data Marketplaces that have been suggested over the recent years. What is more, it outlines the relevant works that have been identified and have been exploited as a baseline to realize the final version of the PolicyCLOUD Data Marketplace.

2.1 Categories of Data Marketplace

Among the plethora of the different typologies/categories of Data Marketplaces, Gelhaar et al. (2021) [1] have proposed the most remarkable ones. More specifically, the authors have suggested a typology framework that utilizes three (3) Meta-dimensions, seven (7) Dimensions and eighteen (18) Characteristics, along these dimensions, as depicted in Table 1.

Meta Dimensions	Dimension	Characteristics			
Technical	Infrastructure	Centralized		Distributed	
	Openness	Open		Closed	
Governance	Interdependence	Tightly Coupled		Loosely Coupled	
	Control	Central		Decentral	
Economic	Domain	Scientific		Government	Industry
	Purpose	Innovation		Interaction	Transaction
	Organization	Keystone-centric	Platform-centric	Marketplace-Based	Decentralized

TABLE 1 - TAXONOMY FOR DATA ECOSYSTEMS.

As depicted in Table 1, the authors developed a metanalysis iterative model, for identifying Data Ecosystems classification variables (Meta-characteristics) and categories (Meta-dimensions), based on both relevant scientific literature and empirical case studies. After the identification of an initial set of 357 scientific papers, they focused on fifty (50) papers that “formed a good sample of extant knowledge about data ecosystems” and 7 empirical case studies. In deeper detail, their research captured the following Meta-dimensions, with the corresponding Dimensions and Characteristics:

- The **Economic Meta-Dimension** considers dimensions from a business-model and competitive dynamics perspective of data ecosystems. The *Domain Dimension* refers to the environment setting, where a data ecosystem emerges that determines the focus and the relevant characteristics of the data of the ecosystem (i.e. *Scientific*, *Governmental* and *Industry*). The *Purpose Dimension*, aims to create *Innovation* and *Interaction* among partners for achieving better communication and transfer and sharing of experience and knowledge, also accomplishing successful *Transactions* among organizations, through interfaces. Finally, the *Organization Dimension* refers to the different kinds of relationships, interactions and organization of the actors of the ecosystem. These are categorized as: (i) *Keystone-Centric*, in which ecosystem actors are organized around a keystone actor who plays a central role in data

provision to the ecosystem, (ii) *Platform-Centric*, in which infrastructure and services for supporting share and usage and resolving of interoperability and usability issues are provided by the platform, (iii) *Marketplace-based* that provides additional components and functions, such as business models and data sharing provisions, and (iv) *Decentralized-Structured*, featuring the absence of the central actor who is substituted by a common goal and a co-creation of value.

Apart from this research, Smith [2] proposed a typology framework that is User Oriented, along with three (3) Types of Marketplaces and seven (7) Key Features, as illustrated in Table 2.

Key Features	Personal	Business	Sensor
Value proposition	Allow consumers to monetize their data	Allows organizations to exchange data	Allows sensor owners to monetize their devices
Transaction Type	Business to consumer	Business to business	Machine to machine
Data Type	Personal and Sensitive	Public and fact – level	IoT sensor stream
Interface	App(sellers) and API (buyers)	API	API
TX confirmation	Wait for seller	Immediate	Immediate
Quality assurance	Trusted Sellers	Crowdsourced reputation	Trusted marketplace operator
Pricing	Pay per user	Pay-per-datapoint	Pay per hour

TABLE 2 - TYPES OF MARKETPLACES.

In deeper detail, the three (3) types of Marketplaces are as follows:

- **Personal Data Marketplaces** that allow consumers to monetize and sell personal data, such as email addresses, social media streams and their location, either directly or upon their own terms. Data buyers are integrating such data into (e.g. internal products, using them for marketing to target markets, etc.).
- **Business Data Marketplaces** that refer to efficient business-to-business data exchange of mainly structured data and large data providers and consumers. Hence, both industry specific data and scientific experiments results can be exploited.
- **Sensor Data Marketplaces** that trade real time data feeds from remote devices, such as data for pollution, power grid and vehicle telematics data feeds.

For all these types of Marketplaces the same Key Features can be applicable with the corresponding values (Table 2), referring to: *Value Proposition*, *Transaction Type*, *Data Type*, *Interface*, *Tx Confirmation*, *Quality Assurance* and *Pricing*.

Finally, Spiekerman [3], for the categorization of sixteen (16) Data Marketplaces, utilizes nine (9) characteristics for their description, also outlining the year of introduction and the year of being withdrawn, where applicable (Table 3Table 3).

Data marketplace	Value proposition	Market positioning	Market access	Integration	Data transformation	Platform architecture	Price model	Revenue model	Status	Founded	Closed
Dawex	Transaction	Neutral	Hybrid	Unspecific	Raw Data	Centralised	Fixed- price	Freemium	Active	2015	
IOTA	Transaction	Neutral	Hybrid	Specific	Raw Data	Decentralised	Progres- sive	Transaction Fee	Beta	2017	
Databroker DAO	Transaction	Neutral	Hybrid	Specific	Raw Data	Decentralised	Progres- sive	n/a	Beta	2017	
Streamr	Transaction	Neutral	Hybrid	Unspecific	Aggregation	Decentralised	Progres- sive	n/a	Active	2017	
Data Intel- ligence Hub	Data	Neutral	Hybrid	Unspecific	Raw Data	Centralised	Multiple	Transaction Fee	Active	2018	
Advaneo	Data	Neutral	Hybrid	Unspecific	Raw Data	Centralised	Fixed- price	Transaction Fee	Active	2018	
Otonomo	Data	Neutral	Hybrid	Specific	Aggregation	Centralised	Fixed- price	Transaction Fee	Active	2015	
Datafairplay	Transaction	Neutral	Hybrid	Specific	Normalisa- tion	Centralised	Progres- sive	Transaction Fee	With- drew	2014	2018
InfoChimps	Transaction	Neutral	Hybrid	Unspecific	Raw Data	Centralised	Fixed- price	Transaction Fee	With- drew	2009	2013
Qlik	Data	Provider	Hybrid	Unspecific	Raw Data	Centralised	Package	Freemium	Active	2017	
xDayta	Transaction	Neutral	Open	Unspecific	Raw Data	Centralised	Fixed- price	n/a	With- drew	2013	2015
Kasabi	Transaction	Neutral	Open	Unspecific	Normalisa- tion	Centralised	Fixed- price	Freemium	With- drew	2010	2012
Here OLP	Data	Provider	Hybrid	Specific	Aggregation	Centralised	Multiple	Freemium	Active	2018	
Azure Data Marketplace	Transaction	Neutral	Hybrid	Unspecific	Raw Data	Centralised	Fixed- price	Transaction Fee	With- drew	2010	2017
International Data Spaces	Data	Neutral	Hybrid	Unspecific	Raw Data	Decentralised	Multiple	Transaction Fee	PoC	2016	
Caruso Data- place	Data	Neutral	Hybrid	Specific	Aggregation	Centralised	Multiple	Membership Fee	Active	2017	

TABLE 3 - CLASSIFICATION OF MARKETPLACES.

More specifically, the categorization types and their characteristics refer to the following ones:

- **Value Proposition** refers to the specific value that is added by the platform to its users, where it can be either *Transaction-Centered* or *Data-Centered*. Transaction-Centered focus on bringing two (2) parties together, by either providing the necessary infrastructure or by direct switching. Such examples are IOTA, Streamr, and DAWEX as depicted in the table. In addition, the Data-Centered ones provide tools for data analysis, visualization and support for gaining new insights from the data. Examples are Hub, Advaneo, and Caruso.
- **Market Positioning** refers to whether Marketplace operators are also Data Traders of Data Providers (Provider), such as large companies, or whether they are neither Data Suppliers nor Data Buyers (Neutral), which usually concern the exchange of data by smaller companies.
- **Market Access** refers to the degree of access regulation to the Marketplace. Limitation of access to selected partners is specified as *Closed* Marketplaces. Alternatively, *Open* Marketplaces target a broad and not- specific group of potential users and a broader target group, with limited control over data quality and data use. *Hybrid* forms, although addressed to selected partners, they also give access to qualified newcomers or, at least to a part of the platform.
- **Integration** refers to the domain spectrum that the Marketplace is addressed or targeted to. Thus, it refers to the degree of broadness or specialization (e.g. industries or sectors), either for cross domain (e.g. DAWEX) or for specific domain (e.g. Caruso).

- **Data Transformation** refers to whether data are made available as raw data packages or unprocessed data streams (*Raw Data*) or, alternatively, in normalized, standardized, uniform, aggregated or report-based format (*Normalisation*), or in logical packages combined for further analysis, with guarantees of high and constant data quality (*Aggregation*).
- **Platform Architecture** refers to whether data are provided by different suppliers, through a central location (e.g. a cloud) for immediate processing (*Centralized* Platform), or whether the content data remain in the ownership-sovereignty of the provider, in spite of more difficult data storage and processing (*Decentralized* Platform), or alternatively, the combination of both with emphasis on provision of supplementary technical infrastructure.
- **Price Model** refers to six (6) different price or charging policies for data acquisition and Marketplace access and use. These include: *Free of Charge* from public or non-profit organizations and authorities. *Fixed Price* or *Subscription Fee* for a fixed period of time. *Package Price* for data or service combined Packages that cost more, in total, but offer lower cost per package, by comparison to individual purchase. *Pay per Use* of specific units accessed or purchased. *Progressive Price*, in which the charged price increases, following an increase in demand for data or licenses and is favored in cases where the wider dissemination of data is not considered appropriate.

To conclude, the classification by Spikerman is more business, market, value, and pricing oriented, than the one of Gelhaar et al, which is more process oriented, while the one of Smith places the main emphasis on monetized offerings, as well as on the application of Blockchain technologies. To this end, it should be noted that although all the abovementioned typologies share common typology elements, there are also, quite a number of differentiated ones. More specifically, Common typology elements are mainly related to a Centralized-Decentralized Infrastructure, Organization and Control, to Open or Closed Access, to Paid or Non-Paid Access and type of Provisioners and Customers (e.g. Public, Private Customers or Provisioners), etc. The differentiated ones include, among other elements, Quality Assurance (Trusted Sellers, Crowdsourcing Reputation, Reputable Operator), Market Positioning and Purpose (Innovation, Interaction, Transaction), and Interdependence (Tightly Coupled, Loosely Coupled).

2.2 Relevant Work in Data Marketplaces

2.2.1 Acumos Marketplace

Acumos [4] is a platform for open-source code that concerns AI, ML, and DL models. It can be exploited by data scientists and developers for the purpose of directing them to an AI-centric process for software production. Among others, it includes features and provisions for toolkits, chain models, marketplace, and visual IDE for AI applications, which could be exported as Docker images to run in private environments or in cloud.

The Acumos Marketplace is a web-based platform designed to make it easy to discover, explore, and use AI models. It has both public and private catalogues, where the former are the only models displayed to non logged-in users. On the “Marketplace” page (Figure 1) the users can find the offered models, being able to search them based on keywords, filter them based on various tags or their categories, and also sort them by their popularity, their number of downloads, their rating, and other metadata.

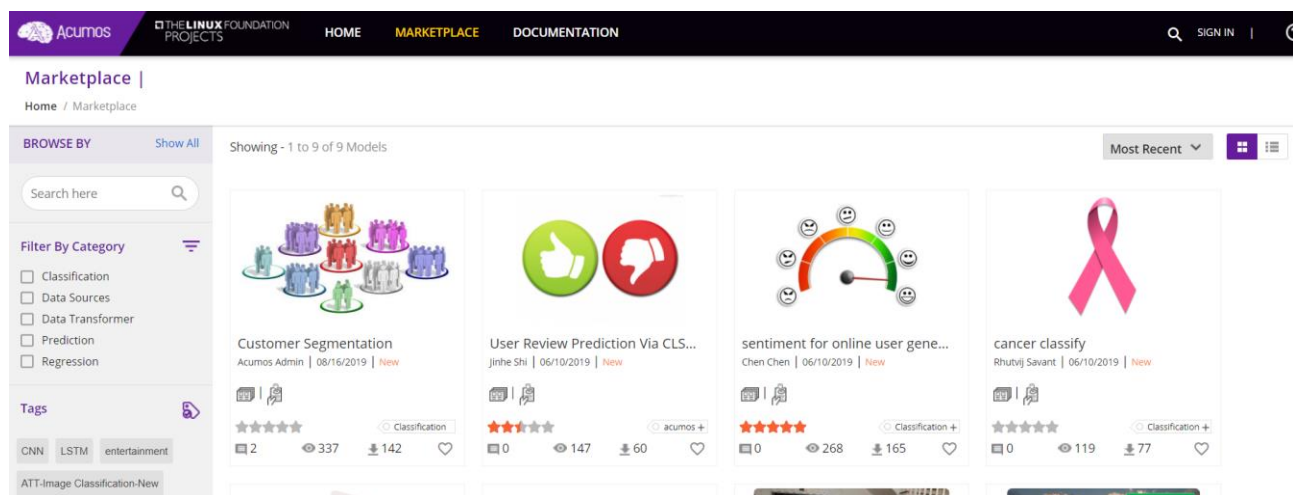


FIGURE 1 - SNAPSHOT OF THE “MARKETPLACE” PAGE OF THE ACUMOS MARKETPLACE.

Moreover, there is a “Model Details” page (Figure 2) from which the users can find more information about a selected model. More specifically, they can find information for the model such as its description, signature, related documents and model artifacts, and other details of the author/provider, including even some metadata such as the number of views, comments, and downloads.

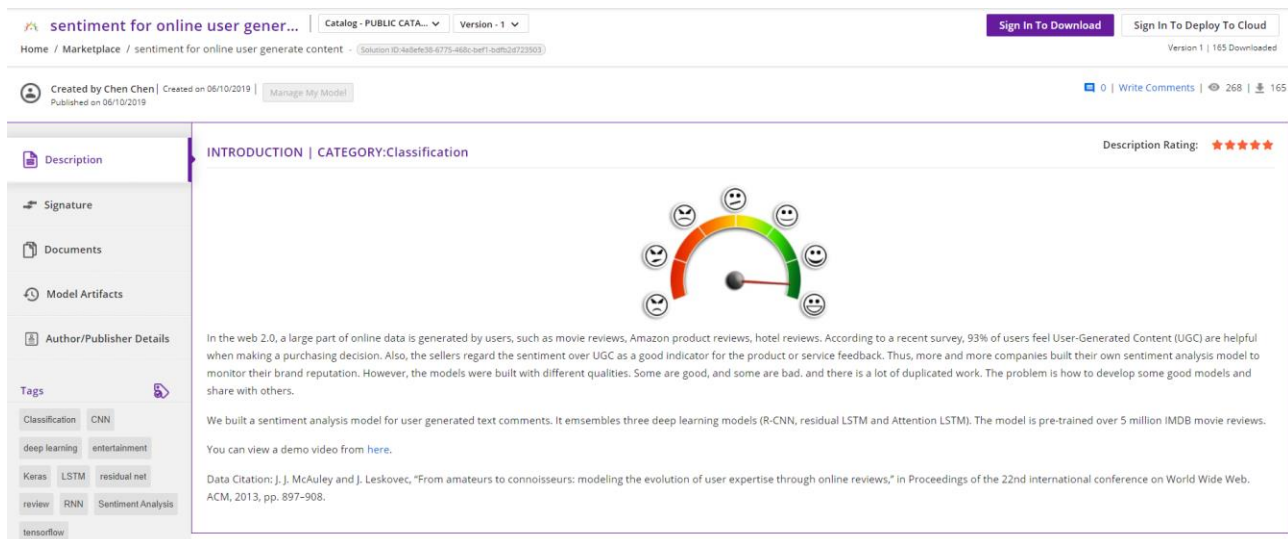


FIGURE 2 - SNAPSHOT OF THE "MODEL DETAILS" PAGE OF THE ACUMOS MARKETPLACE.

Another important feature of the Acumos Marketplace is that it enables its users (who have been assigned the Publisher role) to publish their own models upon request. Before the publication of the models, the publish requests for new solutions go through content control, and at the end, they are approved or declined.

The PolicyCLOUD Data Marketplace has been developed and implemented supporting similar functionalities in order to enable users both to search and to retrieve the offered assets of the Data Marketplace, also being able to upload their own assets (with content control). As a result, a "Model Details" page (i.e., Asset's page) has been built, where the core details for each asset are presented and its corresponding retrieval can occur.

2.2.2 5GTANGO Catalogue

5GTANGO [5] was an EU project whose purpose was to enable the flexible programmability of 5G networks and devise and realize a radical shift in the development of software for 5G-ready applications. In the context of the project, a market platform named as "Catalogue" was developed in order to host and offer different network-related services.

The Catalogue is an instrumental component of the 5GTANGO environment, presenting different parts of it, as it is depicted in Figure 3. Primarily, it hosts the different descriptors of the 5GTANGO packages. Since 5GTANGO aimed offering a multi-platform environment, it enables the developers to orchestrate and deploy their services using different Service Platforms. In this context, the 5GTANGO Catalogue has been adapted to support the storing and retrieval of new packages.

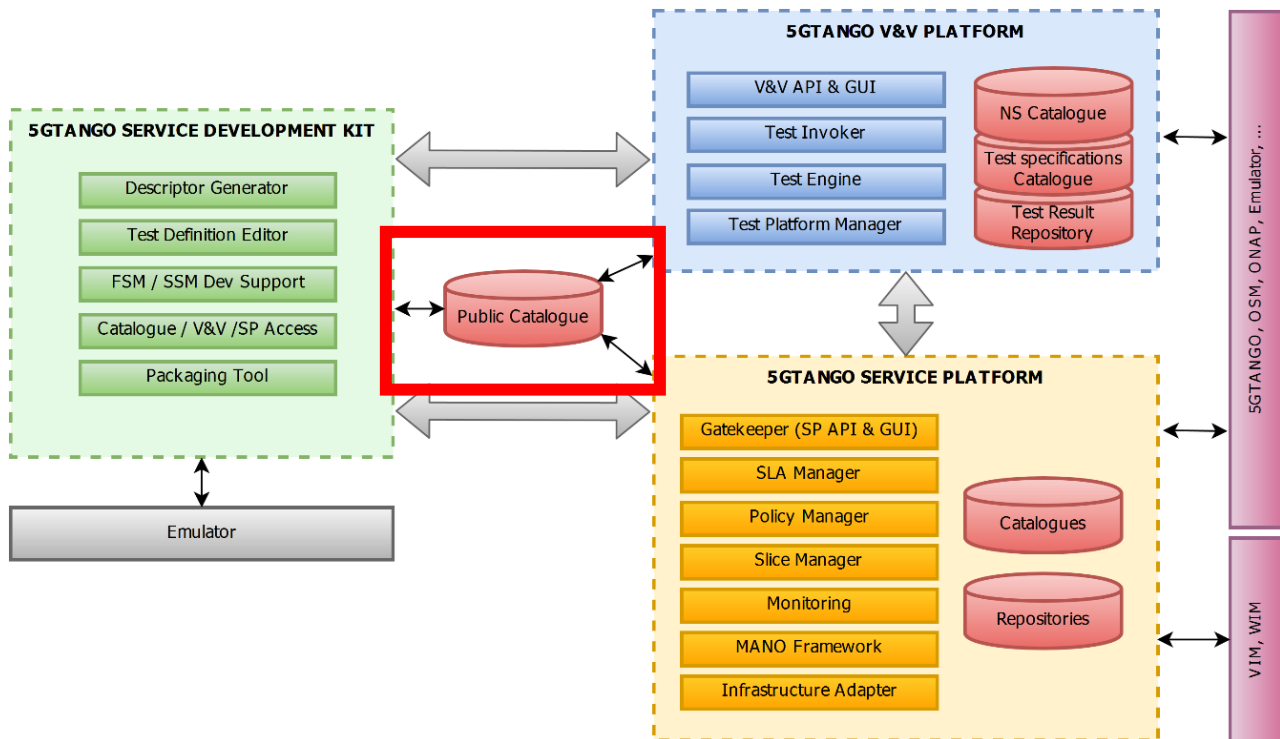


FIGURE 3 - 5GTANGO OVERALL ARCHITECTURE DESIGN AND THE ROLE OF THE CATALOGUE.

Moreover, the 5GTANGO Catalogue is aligned with the principle of persistent storage by extending the hosted descriptors with valuable fields for successful data integration, accuracy in the format of the document, confirmed time of creation, etc. In this way, it enables the development of enhanced operations for Creating, Retrieving, Updating and Deleting (CRUD) descriptors inside it, while reassuring the correct data format of the stored documents.

Going beyond conventional data storage, the 5GTANGO Catalogue provided intelligent functionalities in this 5G environment. Since the types of information vary, one of the requirements satisfied by the Catalogue, was the full-text search capability in structure-agnostic documents. Since the schema of the diverse documents (i.e., descriptors of different types of assets) is variable, the Catalogue provided searching capabilities without the necessity of indexes. Thus, it provides seamless retrieval abilities in deep-hierarchical machine-readable document structures.

Furthermore, besides the plain NoSQL document store for the diverse descriptors, the Catalogue provided a scalable file system for hosting the artifact files, required for the instantiation lifecycle of the services. The Catalogue also provided a set of endpoints where the CRUD methods were supported for the different descriptors of the project. Finally, as previously addressed, the Catalogue was responsible to store different objects with respect to the management of the services' lifecycle and more.

The PolicyCLOUD Data Marketplace follows a similar approach in its back-end side, by utilizing the main functionalities that have been described above, such as:

- Support of different types of assets on the Storage layer ([Section 4.1.1](#)).
- Support of descriptions of the stored assets ([Section 4.1.2](#)).
- Support of advanced search and retrieval of the stored assets ([Section 4.1.2](#)).
- Support of various interaction methods ([Section 4.1.3](#)).

2.2.3 FINSEC Marketplace

The FINSEC project has recently established and launched the Finsecurity.eu [6] market platform. FINSEC develops integrated (cyber/physical) security solutions for the critical infrastructures of the financial sector. Therefore, Finsecurity promotes and offers solutions and services for the security of the critical infrastructures of the financial sector. It serves as a promotional channel for the solutions and services that have been developed in the FINSEC project (Figure 4), while providing access to relevant knowledge assets such as whitepapers and training presentations about security and digital finance services.

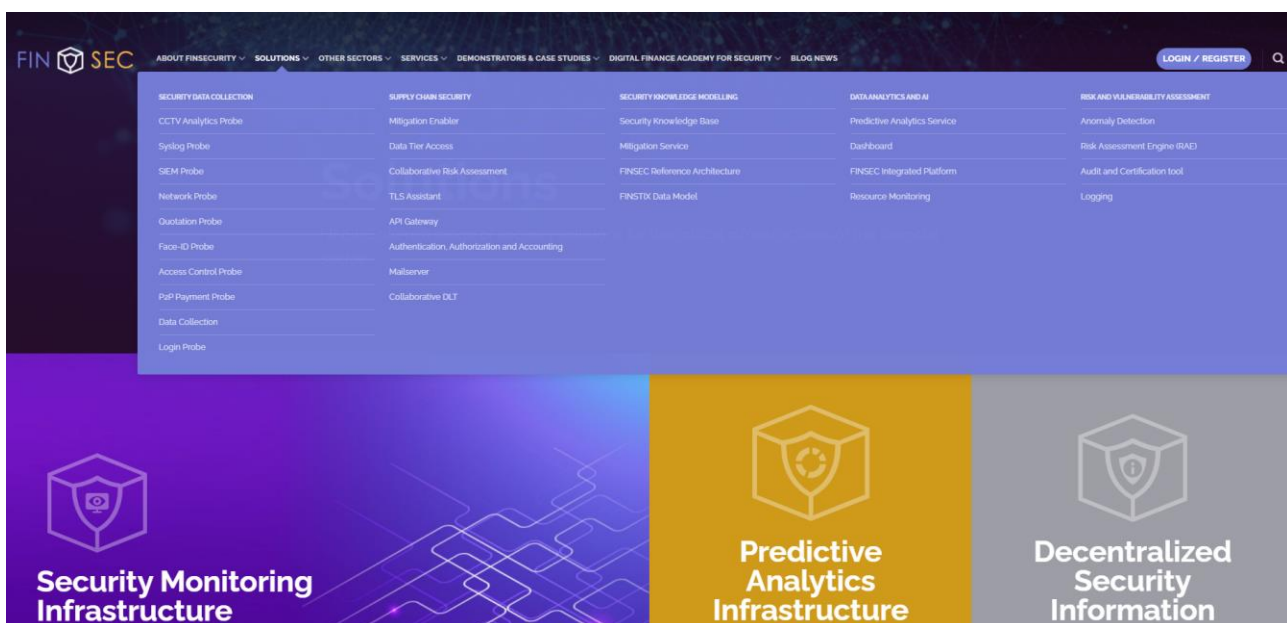


FIGURE 4 - SNAPSHOT OF THE FINSECURITY.EU PLATFORM, INCLUDING A LIST OF ITS SOLUTIONS.

Finsecurity offers selected contents to registered members only. It therefore provides a registration mechanism, along with user management. Registered users have access to premium content and updates. The Finsecurity community is gradually expanded based on members of other projects in critical infrastructures security, whereas these members are also expected to contribute additional solutions.

From what has been described above, the PolicyCLOUD Data Marketplace and the Finsecurity market platform have common objectives, supporting a common strategy for managing and promoting the offered solutions and possibly common communities.

2.2.4 Our World in Data

The Oxford University based Marketplace, namely Our World in Data [7], is a well-known non-profit and open access data marketplace. It provides data for 10 major topics and 297 sub-topics, ranging from health, food and agriculture topics to demographic change, living conditions, community and wellbeing topics, offering the ability to its users to be navigated among all the supported categories (Figure 5).

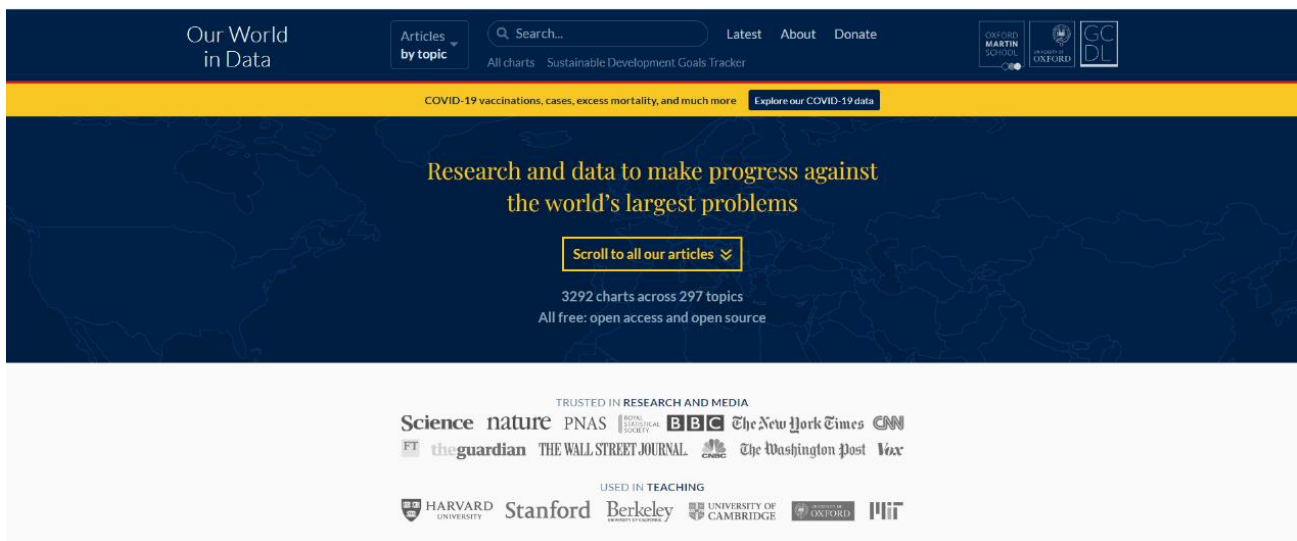


FIGURE 5 - SNAPSHOT OF OUR WORLD IN DATA HOMEPAGE.

Among its offered data and assets, it offers more than 3285 visualized charts, for most of the countries in the world, including interactive maps, diagrams and tables of various metrics, bivariate and multivariate plots of data and metrics, also offering the possibility for data and metrics benchmarking.

The PolicyCLOUD Data Marketplace has been inspired by the Our World in Data marketplace regarding the presentation of the available content. More specifically, the home pages of both marketplaces present the latest content (“assets” for the PolicyCLOUD Data Marketplace and “datasets”/“charts” for the Open World in Data), whereas the offered assets can be filtered based on their specific type/topic. Finally, both marketplaces are offered and maintained by open organizations.

2.2.5 Amazon Web Services Marketplace

Amazon Web Services (AWS) Marketplace [8] is a leading Data and Analytics Marketplace, offering solutions classified in eight (8) distinct major categories, ranging from Infrastructure Software solutions (i.e. Operating Systems, Data Analytics, Security, Network Infrastructure and others) and DevOps solutions (i.e. Application Development, Application Servers, Application Stacks, Monitoring and others), to Data Products solutions (i.e. Retail, Location & Marketing Data solutions, Financial Services Data, Public Sector Data and others) and ML solutions (i.e. ML for NLP, Images, Videos, Computer Vision, Speech Recognition and others) (Figure 6).

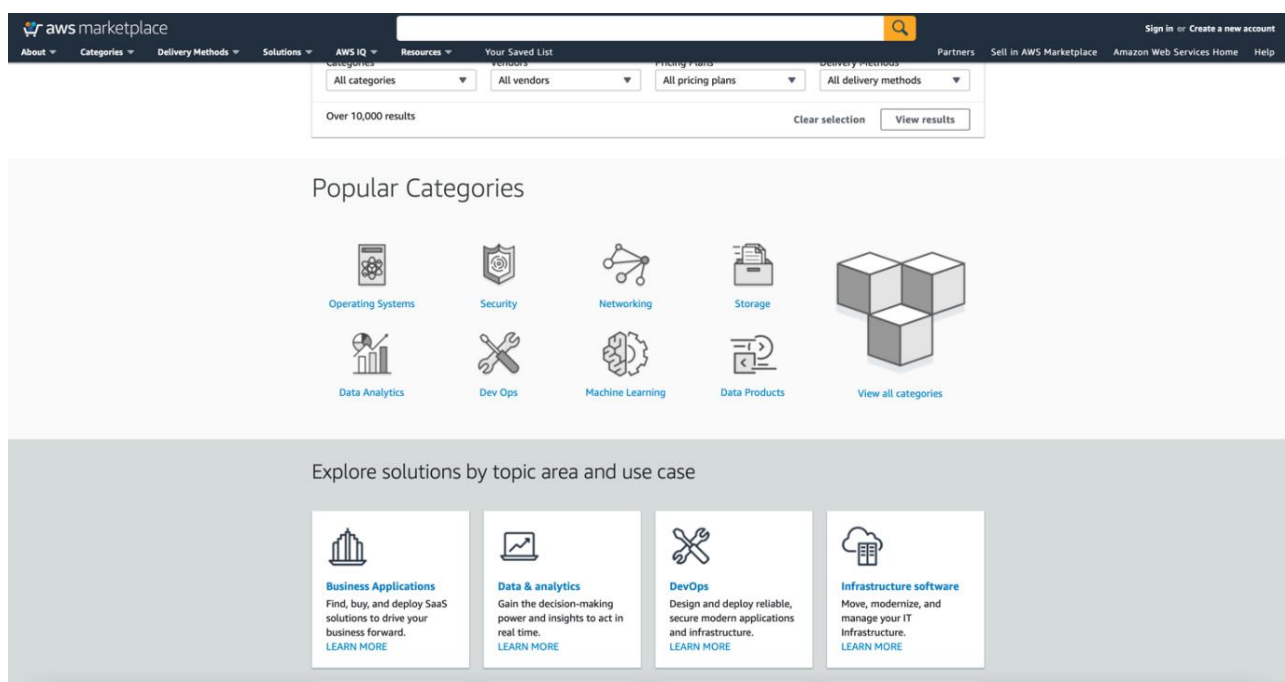


FIGURE 6 - SNAPSHOT OF THE AWS MARKETPLACE'S POPULAR CATEGORIES.

AWS Marketplace is a curated digital library that facilitates the discovery, acquisition, entitlement, provisioning, and administration of third-party software. Like in PolicyCLOUD Data Marketplace, the visitors may discover hundreds of software listings in popular categories as well as across specialized sectors. Additionally, users may investigate and purchase expert services to setup, implement, and maintain their third-party software. With AWS Marketplace, users can reduce procurement delays, apply the controls their business needs to operate with trust, and promote innovation inside their organization.

Like all the marketplaces analyzed so far, AWS Marketplace and PolicyCLOUD Data Marketplace have many approaches in common, from the categorization of the offered assets and solutions and their corresponding presentation on the appropriate pages, to the filtering for the search for assets and solutions according to the needs of the users, as well as and the metrics for the numbers that have been achieved by the marketplaces (statistics).

2.2.6 HERE Olp

HERE Olp [9] is a data marketplace and cloud computing service with a cloud-based environment for location-centric data interchange, visualization, and solution creation, as well as a variety of dynamic and static data. The HERE Olp Marketplace provides services and data that are related to geolocation and driving solutions (e.g. automated driving, fleet management, urban mobility and others). Most of their services can be visualized in modern maps in order to make location data “come to life” and find complex trends and patterns (Figure 7).

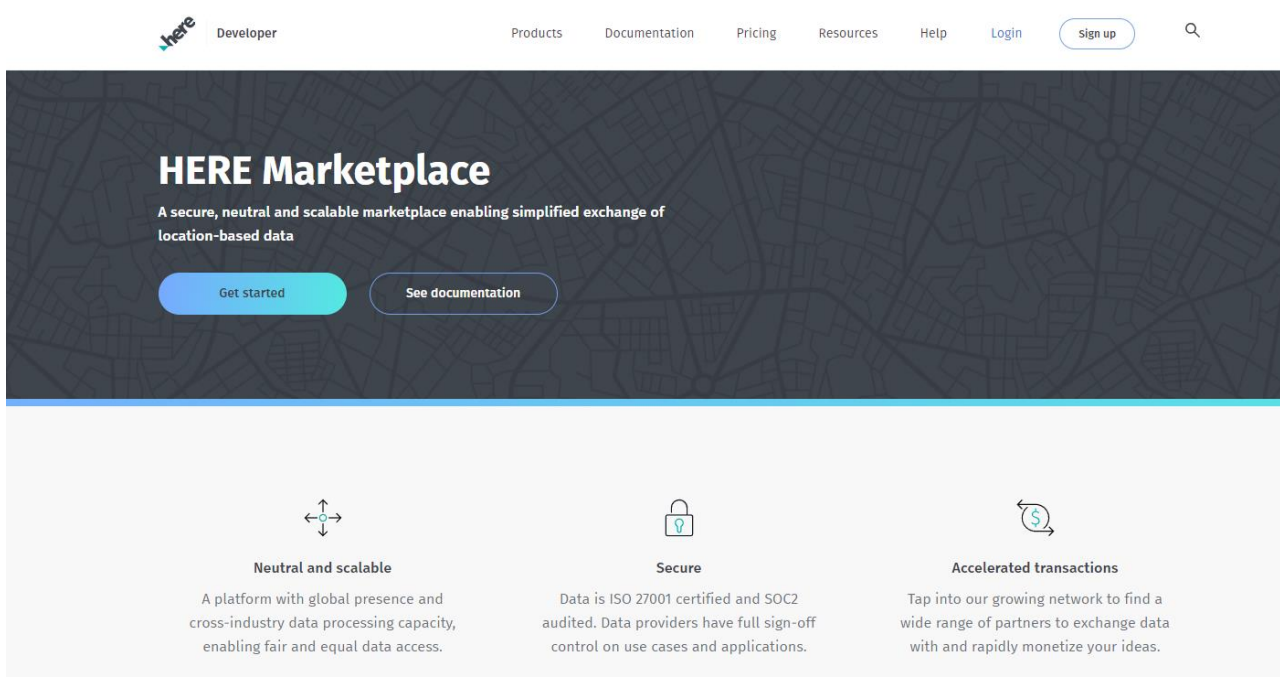


FIGURE 7 - SNAPSHOT OF THE HERE OLP'S MARKETPLACE.

The PolicyCLOUD Data Marketplace, like HERE Olp Marketplace, contains a plethora of data, algorithms, and visualization tools, whereas it provides a straightforward method of accessing location data and algorithms sourced from a network of providers. An important topic that so far has not been mentioned in the comparison with other marketplaces is that the HERE Olp Marketplace gives to its providers the ability to share their data in a controlled, secured, and privacy-compliant fashion. It also adheres to applicable security and privacy regulations and is compliant with the General Data Protection Regulation (GDPR) of the European Union, which is also considered into the legal and ethical concepts that are applied in the context of the PolicyCLOUD Data Marketplace.

2.2.7 Streamr

The Streamr Network [10] is a distributed publish-and-subscribe system that is centered on topics. Each stream and pub-subtopic has its own peer-to-peer overlay network, which is constructed and kept up to date by a set of trackers that are similar to those used by BitTorrent. It is a decentralized real-time data network in which users may create, distribute, and consume data streams using an open, scalable version of the Web3 protocol. Furthermore, it offers robust monetization opportunities, having contributors who live in a wide range of locales and time zones, from Boston to Wroclaw, Helsinki, London, and Melbourne.

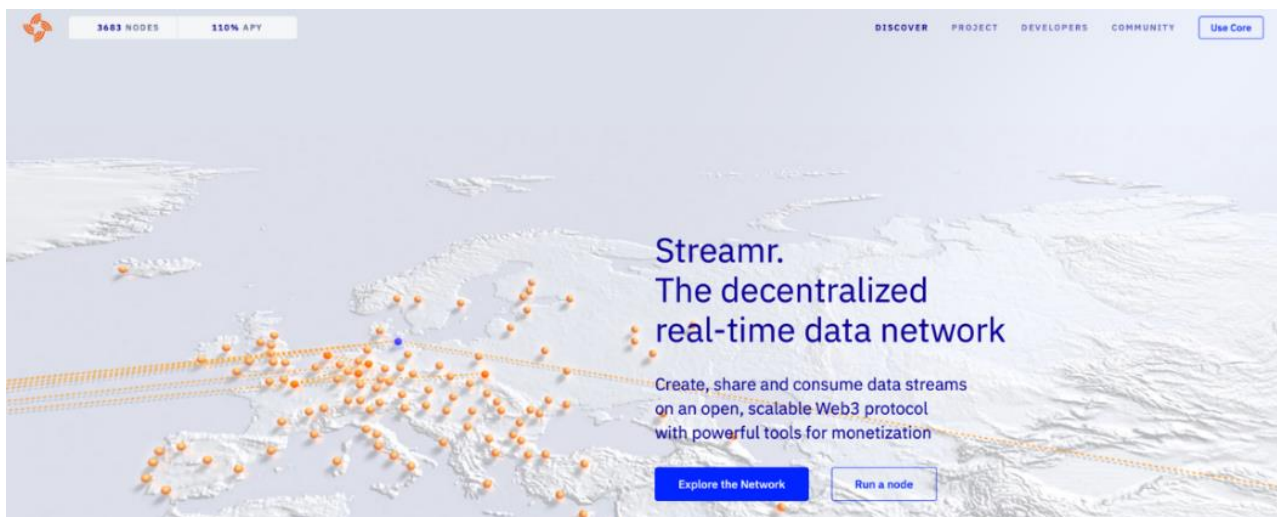


FIGURE 8 - SNAPSHOT OF THE STREAMR HOMEPAGE.

Similarly to the PolicyCLOUD Data Marketplace, Streamr includes users who are able to provide their products or subscribe to the already provided ones. At this point it is worth mentioning that most Streamr's products are available for a fee or may be offered with limited usage per hour (e.g. 100 data per hour). The PolicyCLOUD Data Marketplace offers its assets completely free of charge unless the providers upload services/assets with a fee or a subscription in the form of a link, in order to be purchased outside the Data Marketplace (External assets).

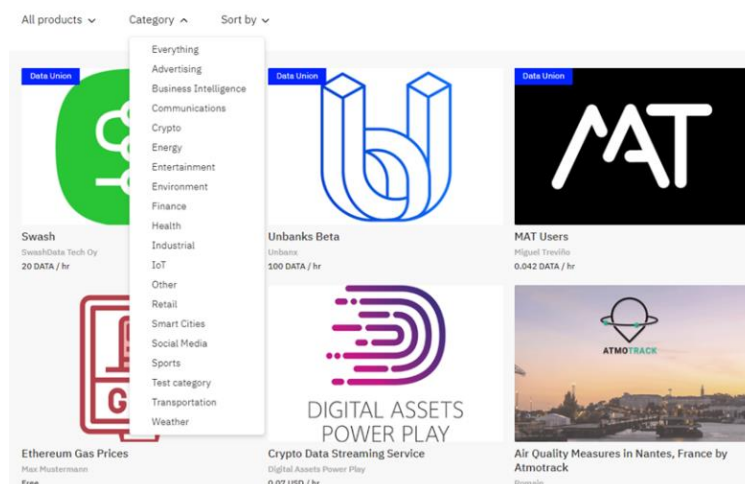


FIGURE 9 - SNAPSHOT OF STREAMR'S CATEGORIES.

2.2.8 Data Intelligence Hub

The Data Intelligence Hub [11] is a secure marketplace for data, data analysis tools, and algorithms across multiple companies and industries. It acts as an interface through which providers and users of data and solutions come into contact and interact effectively. The Data Intelligence Hub, as Figure 10 presents, covers a variety of subject areas and those with the largest amount of data (at least until the release of this deliverable) are the ones of: “Public Administration & Non-Profit”, “Finance, Insurance & Banking”, “Internet, Software & Technology” and “Economy, Politics & Environment”.

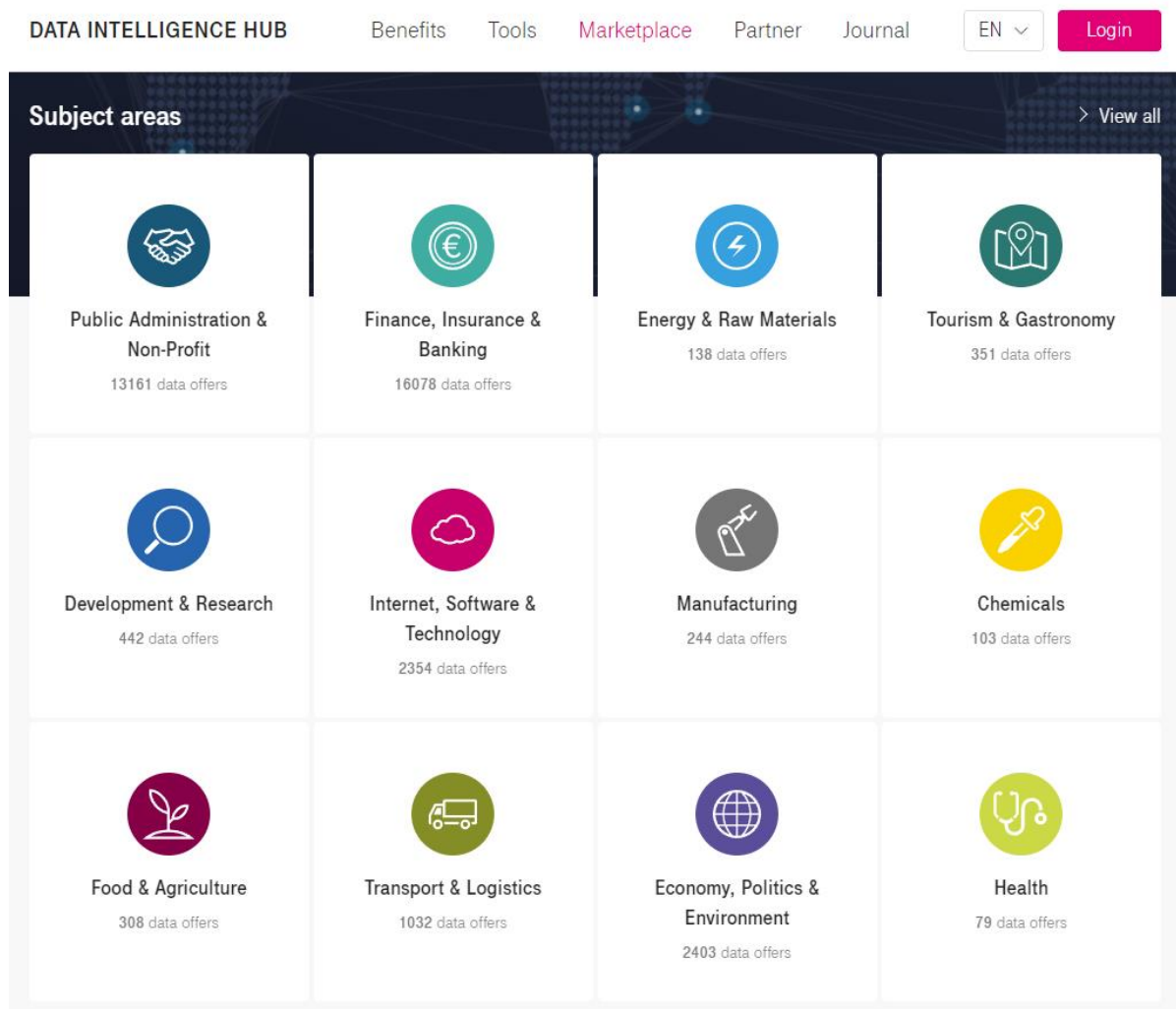
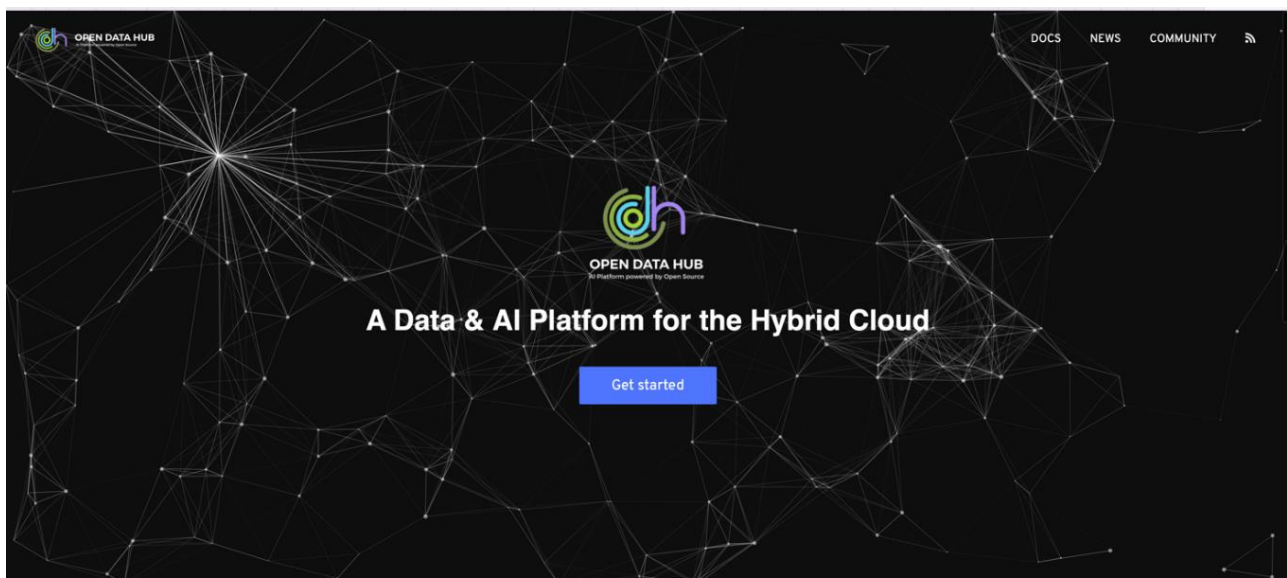


FIGURE 10 - SNAPSHOT OF DATA INTELLIGENCE HUB'S HOMEPAGE.

Moreover, the Data Intelligence Hub supports a very user-friendly interface with detailed information about the provided data and an advanced filtering system. Thus, as in the PolicyCLOUD Data Marketplace, there is no need for its users to be IT experts in order to use their platform. Also, the Data Intelligence Hub contains a significant number of ready-to-use tools and frameworks for data processing and gaining insights. Finally, it provides coding environments and tools, such as R Studio, Jupyter Notebook, and others, to let developers swiftly update their content.

2.2.9 Open Data Hub

Open Data Hub [12] is a design for constructing an AI as a service platform using Red Hat's OpenShift® Container Platform and Ceph Object Storage. Red Hat's internal data science and AI infrastructure is built on top of upstream technologies like Kafka/Strimzi and KubeFlow. Data scientists may use Jupyter notebooks with popular tools like TensorFlow™, Scikit-learn, Apache Spark™, and more to create models. To better meet critical business needs, teams may devote less effort to building and maintaining infrastructure thanks to the Open Data Hub.



What is Open Data Hub?

Open Data Hub is a blueprint for building an AI as a service platform on Red Hat's Kubernetes-based OpenShift® Container Platform and Ceph Object

FIGURE 11 - SNAPSHOT OF THE OPEN DATA HUB HOMEPAGE.

Open Data Hub is a meta-project that, like PolicyCLOUD Data Marketplace, includes open-source software solutions. It also encourages collaboration between open-source software communities, user-enterprise businesses, software providers, and academic institutions, all of which adhere to open-source best practices. Without having a complete knowledge of modern machine learning and artificial intelligence software stacks, the open-source community has the ability to experiment and develop intelligent applications.

2.2.10 IoT Catalogue

The IoT Catalogue [13] is a web-based catalogue and decision-support tool for solutions of the Internet-of-Things (IoT). The IoT Catalogue targets specifically developers/integrators of IoT systems addressing questions such as: What IoT solutions exist for a given problem? What components compose a given IoT solution? What is their cost? Where to buy them from?. All the components used in a solution are represented with detailed information such as manufacturer, product page and its vendors and allow the user to choose where to buy based on the store location, price, etc. The components are categorized in different types being type-specific information added to each component. In this tool, several solutions can be considered when taking into account different environments and their specific requirements.

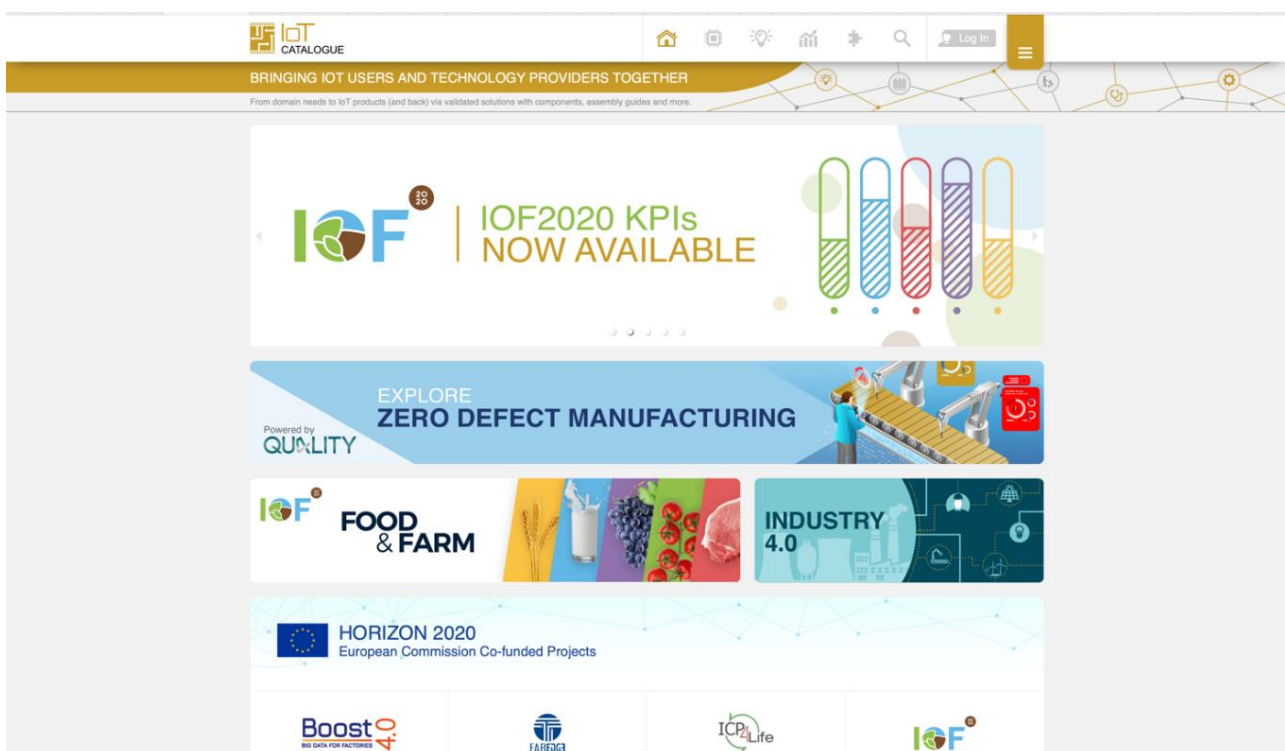
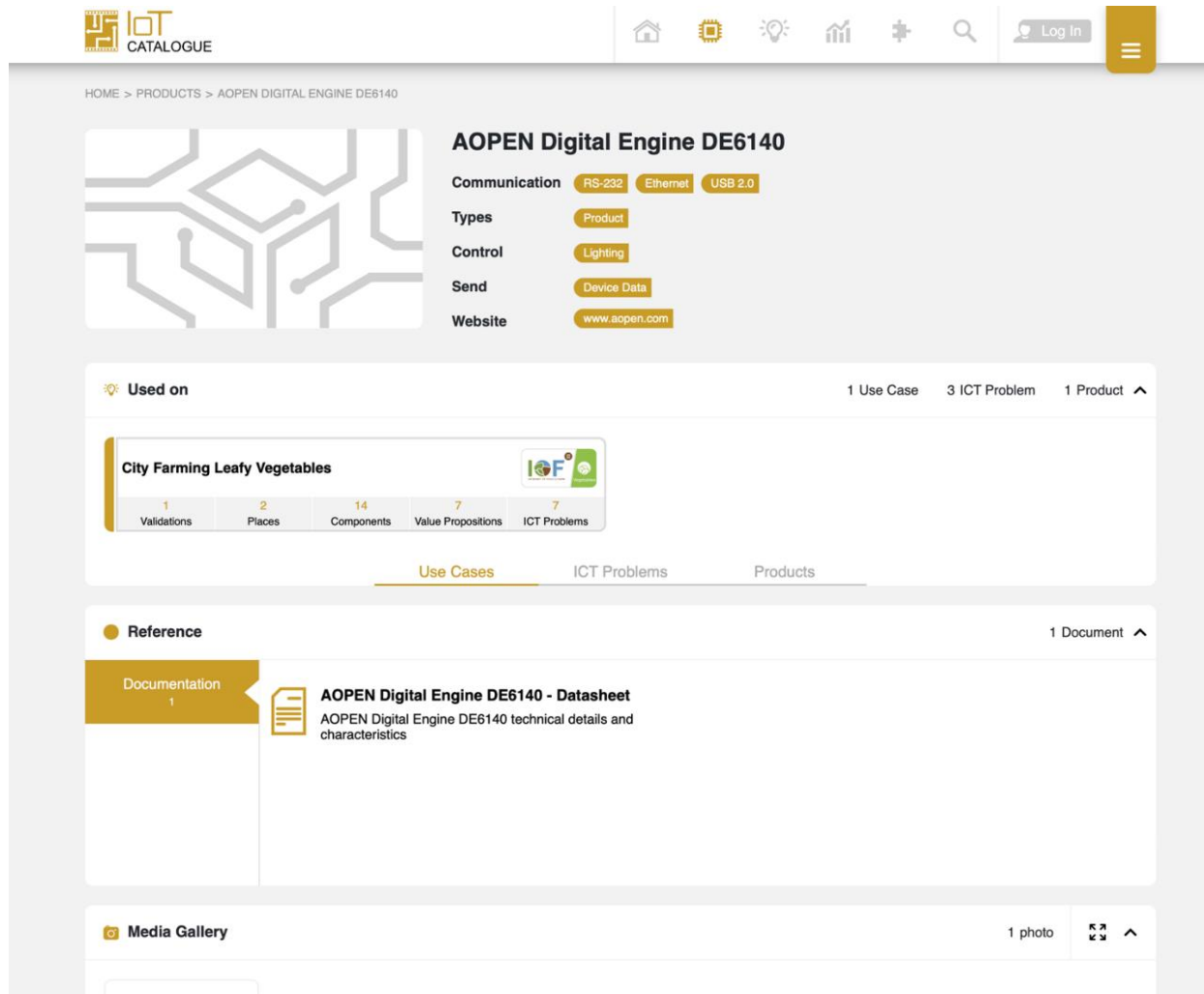


FIGURE 12 - SNAPSHOT OF THE IOT-CATALOGUE.

Similar to the “Discover” page of the PolicyCLOUD Data Marketplace, the IoT Catalogue offers a search page where users can filter and obtain data, that are mostly deriving from IoT devices, based on various search parameters (e.g. the domain, the type, the wireless protocol and others). Moreover, the page that presents a product (Product page, Figure 13) has similar design to the respective “Asset” page of the PolicyCLOUD Data Marketplace, providing a summary and several metadata. In general, the IoT Catalogue is the base for many marketplaces that have been developed and it hosts products from more than ten (10) different projects (e.g. the Internet of Food and Farm 2020 project, the KYKLOS 4.0 project and others).



The screenshot displays the 'AOPEN Digital Engine DE6140' product page within the 'IOT CATALOGUE'. The page layout includes a top navigation bar with a home icon, a search icon, and a 'Log In' button. The breadcrumb trail reads 'HOME > PRODUCTS > AOPEN DIGITAL ENGINE DE6140'. The product title 'AOPEN Digital Engine DE6140' is prominently displayed. Below the title, there are several categories with associated tags: 'Communication' (RS-232, Ethernet, USB 2.0), 'Types' (Product), 'Control' (Lighting), 'Send' (Device Data), and 'Website' (www.aopen.com). A 'Used on' section shows a list of use cases: 'City Farming Leafy Vegetables', with counts for 'Validations' (1), 'Places' (2), 'Components' (14), 'Value Propositions' (7), and 'ICT Problems' (7). Below this, there are tabs for 'Use Cases', 'ICT Problems', and 'Products'. The 'Reference' section shows a 'Documentation' tab with a count of 1, and a document titled 'AOPEN Digital Engine DE6140 - Datasheet' with a description: 'AOPEN Digital Engine DE6140 technical details and characteristics'. The 'Media Gallery' section at the bottom shows a count of 1 photo.

FIGURE 13 - EXAMPLE OF A PRODUCT'S PRESENTATION IN THE IOT-CATALOGUE.

3 Scope, Users & Assets of Data Marketplace

This Section describes the scope of the PolicyCLOUD Data Marketplace and summarizes the main users that may interact with it along with the respective user journeys. Moreover, it lists the different types of assets that are available through the platform and the types of its users.

3.1 Scope

The scope of the Data Marketplace in relation to the other components of PolicyCLOUD (like PDT, etc.) is different and unique. More specifically, the Data Marketplace offers through its platform ready-to-use sets of solutions related to various subjects. At the same time, it enables its users to upload different types of assets (according to the list in [Section 3.3](#)) that can be provided to all the Marketplace's users in order to be able to use them to solve/handle any of their needs. At this point it should be emphasized that all the uploaded assets are available to all registered users of the Data Marketplace without any restrictions/limits. Thus, the assets' providers are informed when registering in the Data Marketplace that their offerings are available to everyone.

As analyzed in the next Sub-section through the user journeys that are presented, some reasons that will motivate the users to visit the Data Marketplace may be to search and retrieve assets that will help them either to resolve some of their business issues, or to achieve some of their educational/research goals, or even use them just for their personal interests.

Therefore, the Data Marketplace is considered as a user-based repository of assets being beneficiary for different kinds of end-users. Its users are able to provide to the community of the platform ready-to-use solutions/tools to various subjects related to the areas of interest of the project.

3.2 User Journeys

The following Sub-sections present several user journeys of potential Data Marketplace users, their interaction with the platform and the reasons why they may be interested in visiting it and taking advantage of its offered functionalities. It should be mentioned that these user journeys cover multiple disciplines and domains in order to be clearly identified that the Data Marketplace is not intended to cover solely the requirements of policy makers (i.e., main focus of PolicyCLOUD), but it is specified in such a way towards addressing the needs from a wider perspective of sectors and disciplines. The described examples include users looking to retrieve (download) assets for their benefit and users being interested in disseminating their assets by uploading them to the Data Marketplace. The journeys also describe what are the expectations of the users, explain possible difficulties that they may have faced in the context of their work, and finally how they resolve them by retrieving relative solutions from the Data Marketplace.

3.2.1 Adoption of Organic Food in Hotel Menus



GOALS	ACTIONS	RESULTS
Create an account.	He wants access to all the Marketplace's content that is unavailable to unregistered users. Thus, in the 'Register' page, he fills in his personal data to create an account.	He confirms his email, he creates the account, he logs in, and he navigates across the various assets of the Data Marketplace.
Recognize attitudes for the organic food in hotels.	He finds a dataset from the 'Discover' page that includes hotel restaurant reviews from web contents, booking platforms and social media for various Greek hotels.	He extracts useful information about the product's penetration based on the hotel type, location, satisfaction with organic food.
Identify influence of organic food products to the hotels' ratings, based on the offered products' variety.	He finds and downloads a ML algorithm from the 'Discover' page, in order to apply Regression analysis in the hotel ratings based on the presence and satisfaction with organic food.	He finds out that there exists a tight relation between organic food and hotel ratings, since almost 45% of the overall ratings depend on the high-quality food.
Identify variety of offered organic product in current hotel menus.	He exploits a content analysis tool from the 'Discover' page, so as to find online information for all the existing online hotel-restaurants menus retrieved by hotels' sites and promotional material.	He discovers that currently the hotels in Greece, offer a variety and mix of organic food with some ingredients being repeated in almost every recipe.
Detect specific organic product combinations.	He finds and uses a cluster analysis tool from the 'Discover' page, in order to extract possible combinations from the hotel-restaurants menus.	He extracts various combinations of organic food products from specific locations in Greece that he can promote.
Identify market segments of adoption of organic food products.	He exploits a statistical tool from the 'Discover' page, so as to apply statistical associations between hotels' clusters, considering their characteristics.	He realizes the segmentation of the hotels based on the users' profiles, for different offerings and hotel characteristics.
Find the profiles of non-adopting hotels as potential targets for conversion to adopters.	He exploits an analytical tool from the 'Discover' page, for analyzing the patterns of the characteristics of the adopting and non-adopting hotels, on the basis of cluster analysis.	He identifies non-adopters for targeted promotional actions, and conversion ways towards becoming adopters, including hotels from specific regions with various ecological beliefs.

TABLE 4 - USER JOURNEY FOR A MARKETING DIRECTOR.

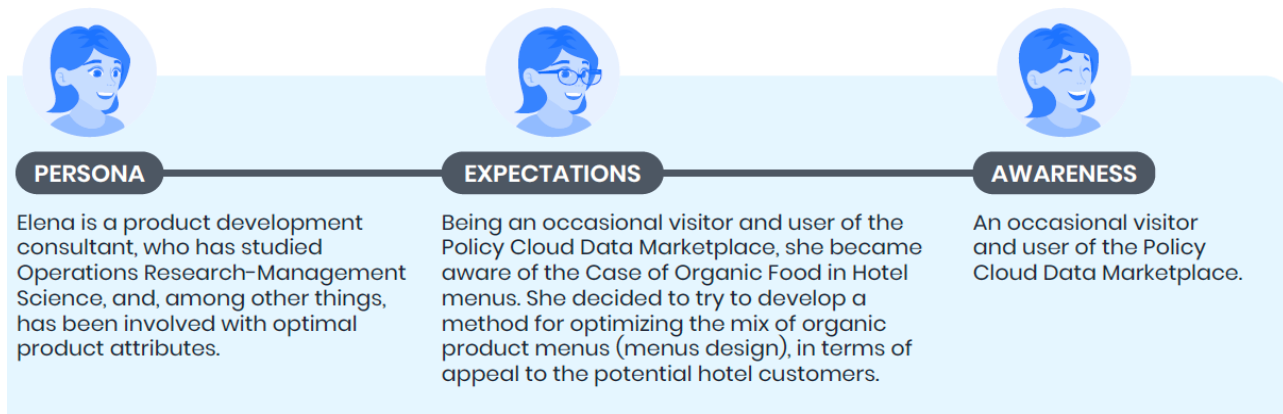
3.2.2 Association of Terrorism with Geolocation



GOALS	ACTIONS	RESULTS
Extract knowledge about how terrorism is being affected by the geolocation.	She searches and finds various datasets that contain relevant information, but all of them are not available to the unregistered users of the Marketplace.	She begins the procedure of creating an account in the Data Marketplace.
Create an account.	She wants access to all the Marketplace's content that is unavailable to unregistered users. Thus, in the 'Register' page, she fills in her personal data to create an account.	She confirms her email, she creates the account, she logs in, and she navigates across the various offered assets of the Data Marketplace.
Identify the correlation between terrorism and geolocation characteristics.	She accesses a sample dataset from the 'Discover' page, which contains a list of terrorist attacks that took place in diverse geolocations across Europe. Since this dataset (i.e., asset) is of high interest, she then downloads it.	She extracts useful information for her research from the dataset, by manually categorizing terrorist attacks based on the different geolocation characteristics.
Find additional tools to help her with her research.	She fills in the search bar the keyword "terrorism", and she continues to explore the offered assets, reading their content.	In the list retrieved by the search functionality, she finds all the relevant assets for her research.
Identify the common characteristics among the diverse geolocations that have been attacked.	She chooses to read a description of an algorithm that performs association rule mining in a research dataset including information for various geolocation attacks in Spain. To verify her selection, she reads the asset's reviews, and then downloads it.	She executes the asset by herself to outline the results in her research, performing various experiments upon it. She then applies the asset on top of the previously downloaded dataset, to compare the derived results.
Find statistical results about terrorism and geolocation.	She finds and downloads a statistical tool with data visualization capabilities about different types of terrorism attacks and the corresponding geolocation characteristics of these attacks.	She discovers the impact of geolocation into the evolution of terrorism, by visualizing the extracted results on top of the previously downloaded datasets.
Express satisfaction of Data Marketplace functionalities.	She reviews and writes separate comments in all the assets that she has downloaded from the Marketplace.	She helps other users to decide whether these assets will be useful for them or not.

TABLE 5 - USER JOURNEY FOR A PSYCHOLOGY RESEARCHER.

3.2.3 Optimization of Organic Food in Hotel Menus



GOALS	ACTIONS	RESULTS
Develop and upload an organic product menus' recommendation tool to the Marketplace.	She visits the 'Discover' page of the Data Marketplace in order to find similar tools to the one she wants to upload.	She does not find any relevant tool.
Understand the upload procedure, in order to upload her own asset (i.e., tool) in the Data Marketplace.	She clicks the 'About' page of the Data Marketplace, and she reads the provided Frequently Asked Questions (FAQ) section, where she finds the needed information for the upload process.	She understands the provided material, and she is able to upload her own asset. However, a mandatory step is to create an account in the Data Marketplace.
Create an account.	In the 'Register' page, she fills in her personal data to create an account, in order to be able to upload her own asset (i.e., recommendation tool).	She confirms her email, she creates the account, she logs in, and from now on she can access the 'Create' page of the platform, where she can initiate the upload process of her asset.
Upload her recommendation tool in the Data Marketplace.	She navigates to the 'Create' page, providing the needed information for her asset.	She writes a description for her asset, she adds the source file of her recommendation tool, and submits it. This description is then approved by the Marketplace admins, being accessible by registered users.
Review the external databases of the topic of interest to find correlations among the hotel ratings and the organic food menus.	In the meantime, she searches relevant databases from web, social media, and external portals, to discover relevant hotel reviews, based on the presence and consideration of organic food.	She downloads the data that is relevant to her goal. Then, she applies her implemented tool to extract the optimal degree of presence and consideration of organic food on hotel ratings.
Upload the results that she extracted before.	She navigates to the 'Discover' page, to find her description and upload further assets (i.e., the extracted results).	She uploads the new files and submits her updates.

TABLE 6 - USER JOURNEY FOR A DEVELOPMENT CONSULTANT.

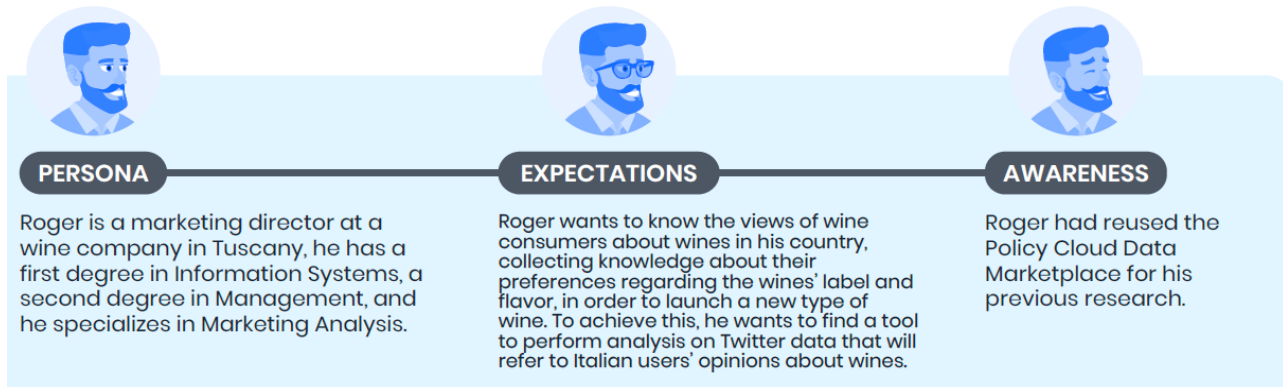
3.2.4 Future Trends of Radicalization



GOALS	ACTIONS	RESULTS
Understand current and future trends of radicalization efforts in France, through keywords detection from social media.	He visits the 'Home' page of the Data Marketplace. He reads the 'Top rated descriptions' along with the 'Suggestions', in order to see whether any relevant descriptions are illustrated or not.	Among the provided descriptions, he finds an interesting dataset with relevant data extracted from Twitter. However, he does not have access to the dataset's details, since he is not a registered user.
Create an account.	He clicks in the 'Home' page the 'Register Now' button, and he fills in his personal data to create an account.	After confirming his email, he creates the account, logs in, and navigates again to the 'Home' page to find the required asset, accessing its full description.
Find additional information and analysis tools about social media information upon radicalization in France.	He fills in the search bar the keyword "radicalization", for finding relevant results in the Marketplace.	He redirects to the 'Discover' page that shows all the results based on the chosen keyword. He explores the results, putting additionally the filter "per date", to see the more recent assets.
View more information about the content that fits his needs, and study and download its relevant material.	He chooses a ML-based geographic analytic tool that captures and analyzes radicalism data from Twitter, per region in France, and he downloads it.	He executes the downloaded tool, and he saves the captured results.
Study and compare results on radicalization in different regions in France.	He uses the visualization heatmap that is provided as a supplementary asset in the ML geographic analytic tool, to visualize the captured Twitter data per region.	He observes the heatmap, analyzing and comparing the percentages of radicalism per region in France, concluding into the current trends.
Make a new policy regarding the current and future radicalization trends in France.	He compares the results that he extracted from the Marketplace tool with external relevant data/policies that he already had in his position.	He creates a new policy, by incorporating in his current research results the new captured elements.

TABLE 7 - USER JOURNEY FOR A POLICY MAKER.

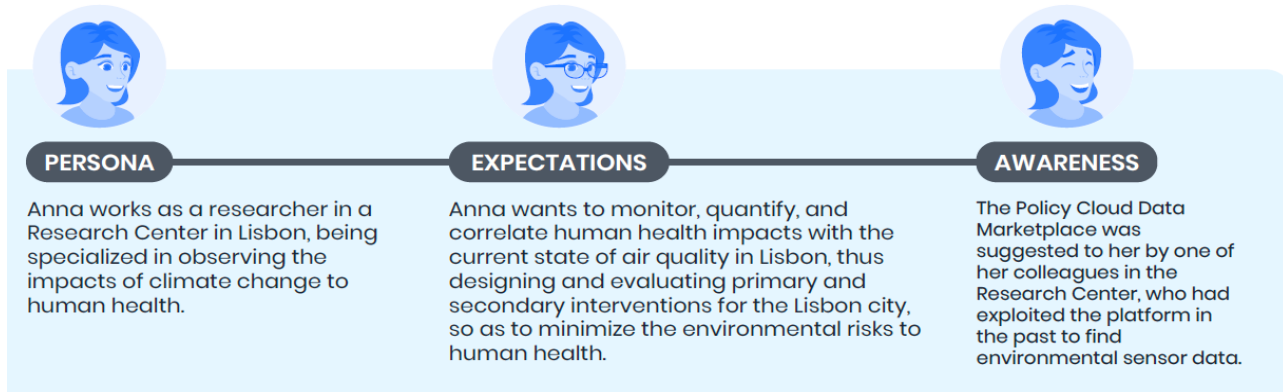
3.2.5 Launch of New Wine Products



GOALS	ACTIONS	RESULTS
Collect knowledge about the preference of Italians for wines' labels and flavors, so as to launch a new type of wine.	He visits the 'Discover' page of the Data Marketplace in order to find datasets and algorithms to help him with the analysis. He is already registered in the Marketplace, so he just has to login to the platform through the 'Login' page, to access its assets.	He successfully logs in, he looks at the assets' types in the 'Filter' section, and he chooses the option "datasets", in order to access all the datasets that are provided in the Marketplace.
Find all the available datasets deriving from Twitter.	He sees all the shown datasets, but since there is a plethora of them, he needs to narrow down the search results. Thus, in the 'Filter' section, in the search field he fills in the keyword "twitter", to access only the datasets that have derived from Twitter.	He observes the datasets, to find if any of them contains information about wines.
Find data about the preferences of Italians upon wine varieties.	He observes all the datasets, and he finds a dataset that includes Twitter text, categorized by hashtag referred to Twitter users' preferences upon wine varieties across Europe.	He carefully studies the datasets, to verify that it is useful for his research. He then downloads the dataset.
Get knowledge about the preferences of Italian Twitter users for different wine varieties.	He visits again the 'Discover' page to find the relevant algorithms to help him with the analysis of the data. Thus, at the assets' types in the 'Filter' section, he chooses "tools" option, in order to access all the relevant provided tools.	He finds out that there exists an NLP analysis tool, dedicated to the analysis of Twitter data, based on specific hashtags. He then downloads the NLP tool.
Understand the Italian users' opinions and preferences about wines.	Having downloaded both the dataset and the NLP analysis tool, he reads the tool's instructions of the provider, to understand how to use it.	He executes the tool by selecting the dataset and the hashtags that he wants to use. He gathers the extracted results, and he concludes in the wine preferences of the users.
Express satisfaction of the downloaded assets.	He writes a comment and he adds a full-star rating in the assets that he has downloaded, expressing his overall satisfaction.	He helps other users to decide whether these assets will be useful for them or not.

TABLE 8 - USER JOURNEY FOR A MARKETING DIRECTOR OF A WINE COMPANY.

3.2.6 Environmental Interventions for Better Quality of Life



GOALS	ACTIONS	RESULTS
Monitor the air quality in Lisbon to create decision algorithm for preventive actions for better human health, and quality of life.	She searches and finds various datasets that contain relevant information, but all of them are not available to the unregistered users of the Marketplace.	She begins the procedure of creating an account in the Data Marketplace.
Create an account.	Since she wants to have access to all the Marketplace's content, in the 'Register' page she chooses to register through her Google Account.	She automatically creates her account by logging into her Google Account, and navigates across the various assets of the Marketplace.
Exploit all the useful datasets for her research.	She navigates in the 'Home' page of the Data Marketplace, selecting the "Datasets" option by the "Discover our Assets" part of the page.	She observes the datasets and realizes that there exists a specific dataset that fits exactly her research purposes.
Find a dataset that provides information for the air quality in Lisbon.	She chooses the found air quality dataset, and she downloads it.	She applies her risk prediction tool upon the found dataset, storing the captured results for further exploitation in her research.
Disseminate past research results upon the impact of air quality in human health in the city of Coimbra.	In the meantime, she realizes that her research work so far could be exploited in the context of the Data Marketplace, and she decides to also upload her assets in the platform. Thus, she uploads in the 'Creation' page her own risk prediction tool and the external data that she has used, accompanied by a demonstration of how it can be used by other users.	She writes a description for her assets, she adds all the source files (tool files, data files, demo video), and submits it. However, this description is rejected by the Marketplace admins, since the uploaded video file was corrupted. Upon rejection, a notification email is sent to her, informing her to fix the corrupted file.
Fix the rejection issue to successfully disseminate her past research results.	She visits again the 'Creation' page, following the same procedure to upload all of her assets, after having fixed the corrupted file.	She writes again her assets' description and submits it. The new information is fulfilling the Marketplace quality criteria, and her description is approved.

TABLE 9 - USER JOURNEY FOR AN ENVIRONMENTAL RESEARCHER.

3.3 Offered Assets

In the context of PolicyCLOUD, the Data Marketplace offers several types of assets (objects) that may result from the separate procedures and mechanisms implemented in the PolicyCLOUD platform. The main assets' categories that are currently exploited and added into the Data Marketplace are listed and described below:

- **Tools**: This asset type refers both to algorithms and tools in general, implementing specific solutions. By the term algorithm, we refer to the type of asset that includes code in any programming language (e.g. Java, Python, Python Notebooks, etc.), which once retrieved can be executed to produce a result (e.g. to validate/create a new policy) or even integrate it into an existing project of the end-user (depending on the provided asset and its author/provider). In terms of tools, there may be analytics tools, management tools, etc. The idea behind the tools is that instead of offering parts of code, there will be executable programs (i.e. tools) that integrate different solutions for a subject/a couple of subjects.
- **Policies**: This asset type refers to ready-to-use policies, which are the project's core outcome and may have been derived by the PDT, whereas they can also be examples or templates of policies for any subject/area covered by a use case of the PolicyCLOUD project, or for any use case that can be adopted and applied into the project.
- **Datasets**: This asset type refers to datasets from different either internal or external sources and pools of data (e.g. CSV, JSON files, etc.). These datasets either may "feed" various algorithms and tools, that as mentioned above, are a primary asset type of the Data Marketplace (e.g. feed a machine learning algorithm as training or test datasets), or may be used as examples/test datasets in order to produce new policies.
- **Webinars**: This asset type refers to webinars and events related to PolicyCLOUD's topics, whether they are organized and accomplished by the PolicyCLOUD's partners or not, but are related to the project's research and development topics.
- **Tutorials**: This asset type refers to instructional videos/tutorials about various solutions whether they are about PolicyCLOUD's outcomes or they are just related to its research and development topics.
- **Documents**: This asset type refers to the project's general documents such as how-to guides, manuals, tutorials in text format, academic research papers, project's presentations, and other detailed documentations regarding various project's subjects.
- **Other**: This asset type refers to assets that do not fall into the above types of assets.

In addition to these types of assets, there is an extra asset type, considered as a horizontal asset type / sub-type for all the asset types, which is called “Externals”. The assets that fall into this type are assets that are not stored in the Data Marketplace, and they are referred to with links (URLs) to the other (external) platforms, services and webpages. In order to support this horizontal type of assets, an appropriate field has been added to the schema of the information collected for all assets (descriptions), in which the providers can add such information/assets. This asset sub-type is useful:

- In the case that there are assets that cannot be added/stored to the Data Marketplace, but are still relevant to the project, and therefore should be included in order to help the end-users or even to “promote” them.
- In order to offer assets that are already stored in public repositories (e.g. in Github).

Finally, it should be mentioned that all these types of assets emerge from the needs identified by all the PolicyCLOUD partners and have been extended with additional suggested assets as a fruitful way to exploit and disseminate the project’s results to the community. More specifically, the list above has been formed based on the requirements analysis that had been accomplished over various inputs collected during the design and development period of the first version of this deliverable (i.e., D7.4) and in the scope of Task 7.2, and has been further updated after the first period of Data Marketplace’s operation.

3.4 Users

As described in the Introduction Section of the deliverable, the Data Marketplace aims to create a community of users who will be able to provide and share various assets through the platform. Thus, in the Data Marketplace, it is envisioned that there will be many and different users - different by the meaning that they will have diverse backgrounds, specialties, interests, and objectives.

From the first version of this document, it has been described that the users can be distinguished into human users and machine users. In deeper detail, in addition to the normal “human” users, it was considered as a valuable extension to make the Data Marketplace more interoperable, being able to offer its assets and services to other services, platforms, and systems, which, with the appropriate interfaces, will be able to provide or retrieve the offered assets in a more direct way. Consequently, the users of the Data Marketplace are able to interact with it either through the website that has been developed (Front end - [Section 4.2](#)) or directly with the Marketplace’s back-end API (Back-end - [Section 4.1](#)). Human users are able to use both ways, but the machines/services are able to retrieve or provide new assets in the Data Marketplace, with a more direct and automated process, interacting directly with the back-end.

In addition, users are divided into providers and end-users. Providers can upload and at the same time publish to the Data Marketplace their assets. If the providers are the authors of the provided assets, they can also be called producers/owners. The end-users of the Data Marketplace, also named as consumers, are the users who will benefit from the platform, being able to retrieve the offered assets and/or services for which they are interested in, and use them by deploying and executing them, depending on their needs.

In general, the users that will form the community that will be created around the Data Marketplace, as depicted in Figure 14, are expected to be:

- **Policy makers**, who can contribute and enhance the Data Marketplace with policies.
- **Developers**, who are able to share their algorithms/tools and datasets to the platform in order to provide the end-users with their complete solutions regarding the relevant field of use.
- **Scientists**, who can contribute assets that would like to share with other scientists, as in the case of the developers.
- **Services**, which during or after their execution generate and store new assets on the Data Marketplace (e.g. policies/datasets). An example of a service – provider could be the PDT.
- **Any other interested third parties**, who want to achieve some personal, business, or educational objectives and will benefit from the Data Marketplace. Thus, the platform is intended to serve the needs of a wider community, allowing everyone to join (through registration) the Data Marketplace and obtain the respective assets.

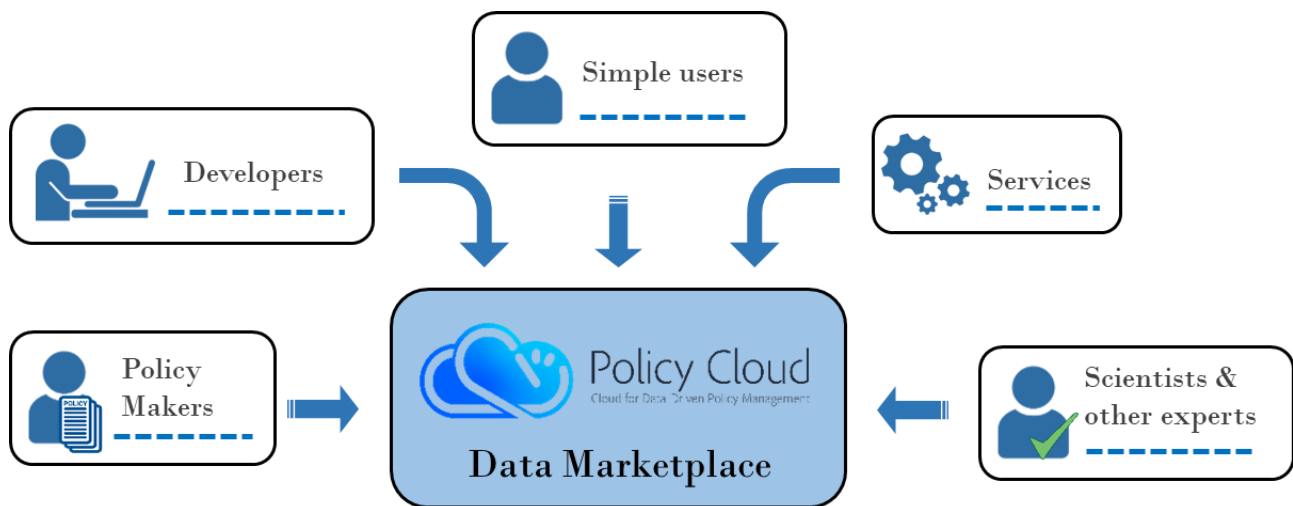


FIGURE 14 - DATA MARKETPLACE'S INTERACTING USERS.

Additional examples to the abovementioned types of users are the PolicyCLOUD consortium partners (e.g. technical members, analysts, policy makers, etc.) that fall in the classes described above, and other authorized users (e.g. members from other collaborating projects, etc.).

4 Architecture of Data Marketplace

As outlined in the introduction of this deliverable, the Data Marketplace is a public web-based environment with various different APIs. It is also considered as a unified and standalone platform, able to store several types of assets that may derive/result from the separate procedures and mechanisms that are implemented in the scope of the project, and also solutions authored or developed by other experts.

From the architecture perspective, the Data Marketplace is structured around two (2) core services, the back-end and the front-end. This is basically the reason why the Data Marketplace is considered as a unified platform (i.e. unified platform of these two (2) services). This separation contributes towards the platform's enhancement in terms of functionality (e.g. reduce maintenance costs, facilitate its management, etc.), also providing additional information and capabilities (e.g. enables direct access to the stored assets through the back-end).

Generally, the Data Marketplace provides several functionalities that are mapped to different layers. The back-end includes three (3) layers (i.e. Assets Storage Layer, Assets Management Layer, and Interaction Layer), while the front-end includes the fourth layer of the Data Marketplace (i.e. Presentation Layer). The Data Marketplace in full is consisted of all of these four (4) different layers (as depicted in Figure 15 and Figure 16) that realize the expected capabilities as presented below, and further discussed in greater detail in the next Sub-sections.

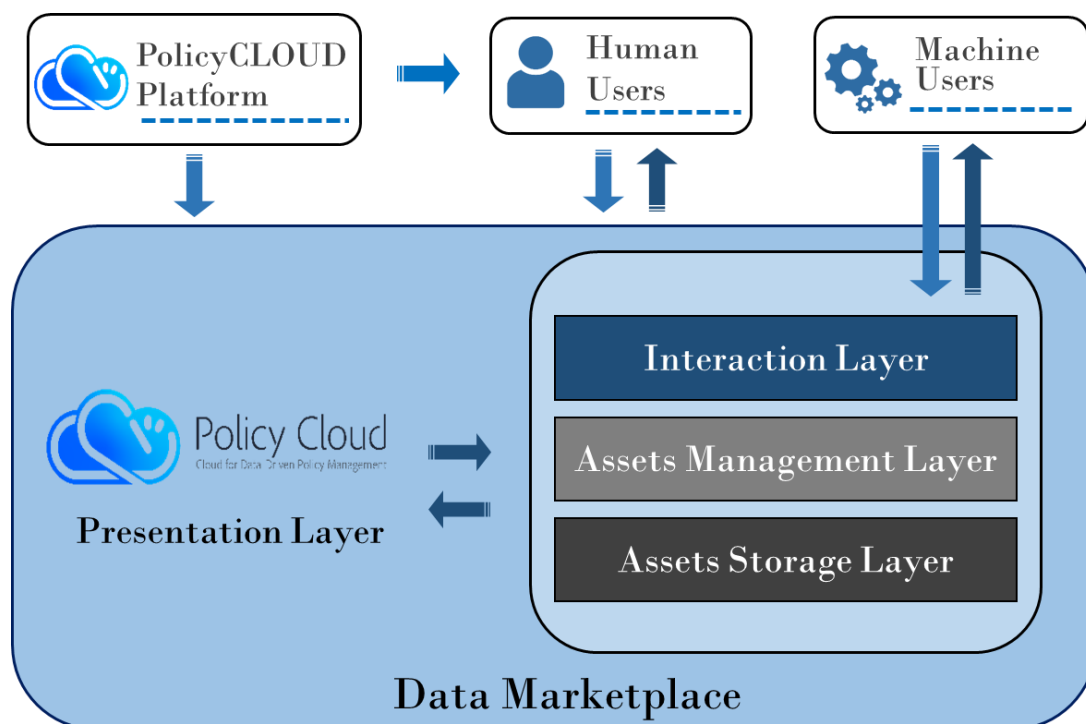


FIGURE 15 - DATA MARKETPLACE ARCHITECTURE.

- The **Assets Storage Layer** (part of the back-end) is the layer in which the platform's offered assets are stored.
- The **Assets Management Layer** (part of the back-end) delivers all the needed principles and techniques for the management of the Marketplace's assets.
- The **Interaction Layer** (part of the back-end) supports the communication between the Data Marketplace and its users (i.e. human users, and machine users), by providing discrete APIs for exploiting each different type of asset.
- The **Presentation Layer** (part of the front-end) provides the User Interface (UI) towards the different types of users that are willing to use the platform.

As described in [Section 3.4](#), the Data Marketplace supports access to its offerings to both human users (i.e. policy makers, developers, scientists, interested third parties) and machine users (through the respective interfaces). In this context, the human users are able to interact with the Data Marketplace through the front-end (through the Presentation layer) that reflects a user-friendly platform (providing the UI), while other additional services (e.g. project's services, 3rd parties) can interact directly with the back-end (through the Interaction layer). This separation is also represented in Figure 15.

In terms of storage, as described in [Section 3.3](#), the Data Marketplace stores various types of assets (objects) in any format, whereas relevant functionalities have been developed for the assets' management. In addition to these, the platform considers issues related to the security of the assets it stores and offers. It also considers, complies, and informs the end-users on privacy issues (i.e. GDPR). Hence, user registration is a required functionality to ensure that all the regulations are enforced. Finally, as already described in Section 3.1, all uploaded assets are available to all registered users of the Data Marketplace without any restrictions/limits, and the assets' providers are informed when registering in the Data Marketplace that their offerings are available to everyone.

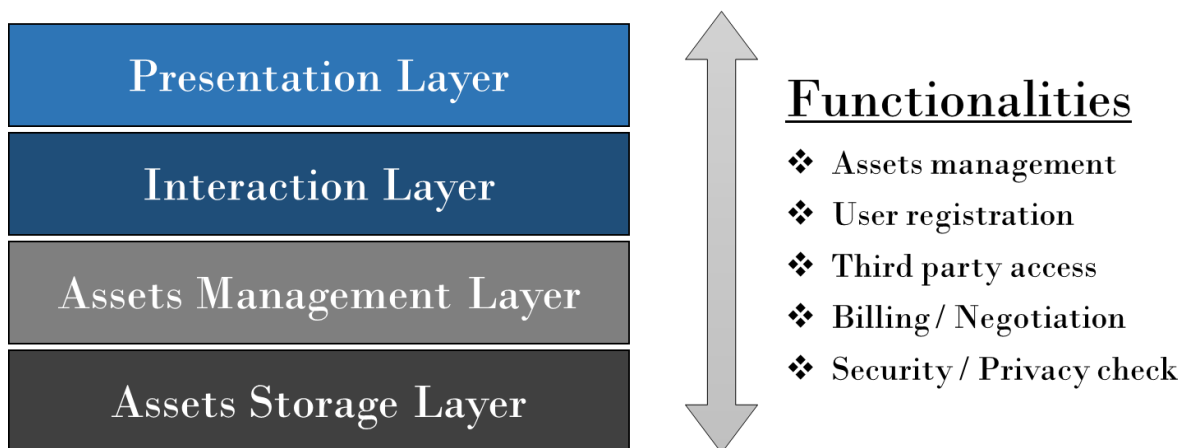


FIGURE 16 - DATA MARKETPLACE LAYERS AND MAIN FUNCTIONALITIES.

4.1 Back-end

The back-end is the core base of the Data Marketplace and has been developed using a variety of components/tools. All of its sub-components are containerized in Docker images [14] that, among others, offer more efficient management and maintenance, enabling continuous updates and integration. Python [15] is used as the programming language that along with the Flask [16] framework, which is a Web Server Gateway Interface (WSGI) developed in Python, implements RESTful APIs to handle the respective HTTP requests.

The descriptions of the offered assets and their metadata are stored in a MongoDB [17] database that according to the initial specifications would be used in combination with GridFS [18] specification for storing and retrieving large files/objects, of any format. From the various measurements made during the development of the back-end, it was found that the retrieval of the assets using the GridFS was slower (because it loaded a large amount of data, i.e. GBs), and thus it was preferred not to use the GridFS and instead to use the operating system (OS) that hosts the back-end which is more efficient, especially when the assets are videos that (except for the common retrieval) need a streaming functionality. Moreover, Gunicorn [19], a Python WSGI HTTP Server for UNIX, is utilized with NGINX [20], an open-source high-performance HTTP web server and reverse proxy, since Flask is not optimum for production mode, and thus, both tools extend the Flask framework in order to enable access to multiple users at the same time.

The following subsections analyze and describe the functionalities of the back-end's three (3) layers (i.e. Assets Storage Layer, Assets Management Layer, and Interaction Layer).

4.1.1 Assets Storage Layer

The Assets Storage Layer is responsible for storing the assets/objects that are offered by the Data Marketplace (Figure 17). An essential component of it, is a database that can store files in any format as well as additional information about the files provided. For this purpose, the appropriate database type is the NoSQL technology. NoSQL databases are non-tabular and store data in a different way than relational databases do. NoSQL databases come in a variety of types based on their data model, where the main types are document, key-value, wide-column, and graph databases. They provide flexible schemas and scale easily with large amounts of data and high user loads.

Based on the above, a NoSQL database is used for the Data Marketplace and in particular, a document-based database so that it can efficiently store files. One of the most performant and efficient document databases is MongoDB, which stores data in flexible, JSON-like documents, meaning fields that can vary from document to document and data structure that can be changed over time. In addition, in order to store files that exceed the maximum size limit that MongoDB sets, the operating system (OS) that hosts the back-end, is used to store large files and assets with their binary data.

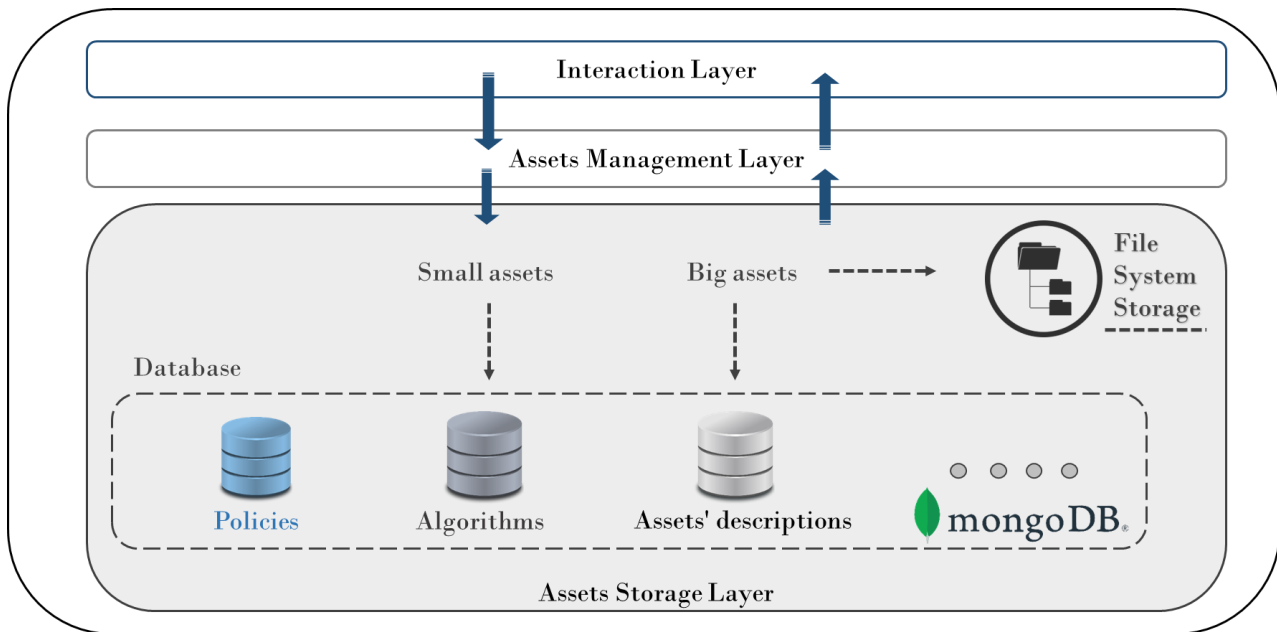


FIGURE 17 - ASSETS STORAGE LAYER.

Thus, for this layer of the Data Marketplace, the combination of MongoDB and the OS is used to store both small and large files. Appropriate collections have been created in MongoDB per destination branch/field of application of the files (i.e. a collection for tools, another collection for policies, etc.).

The information provided for the offered assets by their providers, called descriptions (described in [Section 4.1.2](#)), are also stored in MongoDB in JSON format, which is fully supported by MongoDB, as mentioned above. Moreover, JSON files are easy to be created, retrieved, and read even by a simple user, although in the Presentation Layer (front-end), the containing information is presented to the end-users in an even more user-friendly way. JSON files also extend the interoperability of systems, especially in this case, which includes the systems of the back-end and the front-end.

4.1.2 Assets Management Layer

The Assets Management Layer is responsible for the entire lifecycle of the data and assets within the platform, and offers all the principles and techniques for their management (Figure 18). More specifically, the Assets Management Layer is the layer that handles the assets from the moment they are entered to the platform through the APIs and then stored in the database (in the Assets Storage Layer), until their final deletion from the platform.

In general, through this layer, the Data Marketplace supports the CRUD operations. CRUD stands for "Create", "Read" ("Retrieve" in this case), "Update" and "Delete" that are the four (4) basic functions of persistent storage. On top of these operations, the platform supports also the searching functionality, as described below:

- Create functionality is the functionality where new assets/objects are ingested in the Data Marketplace. This operation is triggered following the upload of a new asset by a provider and it results to the creation of appropriate entries in the database for the new asset (its metadata, the binary data which are stored to the OS, etc.). An important information for each object refers to a unique alphanumeric identifier through which the indexing can be done as well as its subsequent retrieval.
- Search functionality is the functionality that enables the users to search for assets, based on various parameters from metadata given by the providers or possibly other information generated by the system during its creation (e.g. unique identifier, number of downloads, etc.).
- Retrieve functionality is the functionality that is executed when a user wants to download an asset from the Data Marketplace, after first seeing/finding it through the search functionality. The retrieval results to obtaining the corresponding asset from the Storage Layer and delivering it to the appropriate API.
- Update functionality is the functionality that is triggered when a user replaces an asset with a newer version of it or modifies some of its metadata.
- Delete functionality is the functionality that handles the deletion of the assets from the platform when the owners, or other users with the appropriate rights (i.e. admins), decide to delete them.

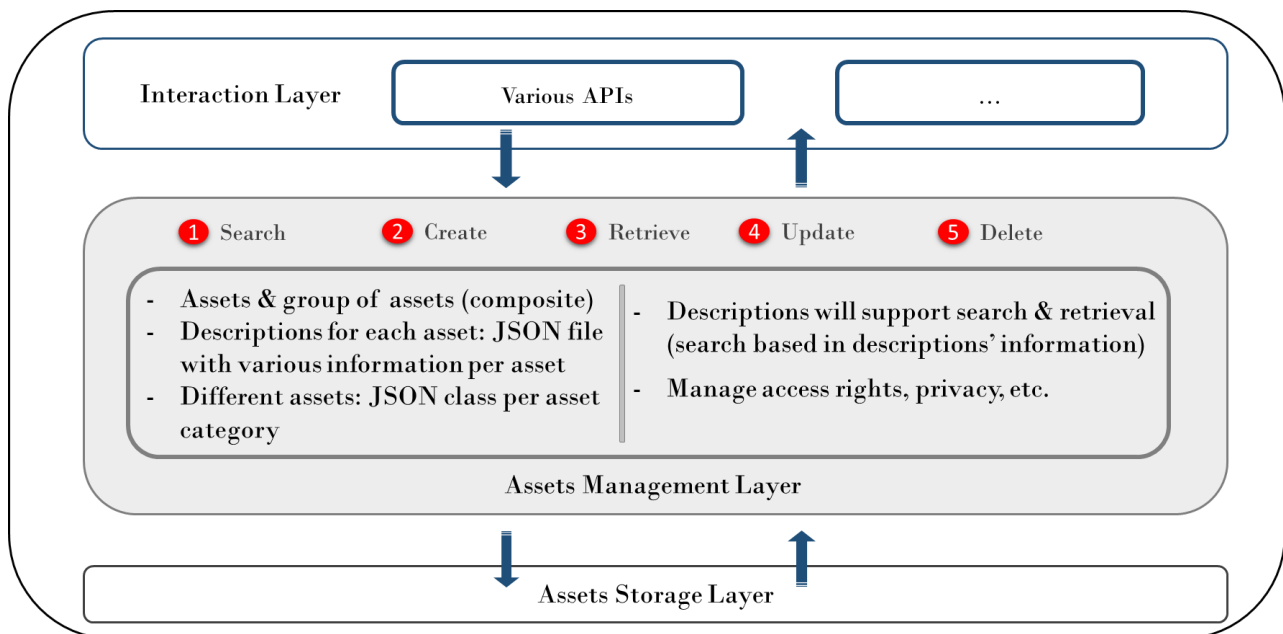


FIGURE 18 - ASSETS MANAGEMENT LAYER.

The main part of the Data Marketplace, i.e. all the layers except the Presentation Layer (because of its separation from the back-end), is managed by RESTful APIs. A RESTful API is an Application Program Interface that uses HTTP requests to GET, PUT, POST and DELETE data/assets. The latter indicate some widely used HTTP methods that are related to the abovementioned Search and CRUD operations/functionalities. Table 10 matches these HTTP methods with the above supported functionalities and the corresponding statements in MongoDB (the database that is exploited in the Data Marketplace).

Operation	HTTP Method	MongoDB
Create	POST / PUT	Insert
Retrieve	GET	Find
Update	PUT / POST	Update
Delete	DELETE	Remove
Search	GET	Find

TABLE 10 - MATCHING (S)CRUD OPERATIONS WITH HTTP METHODS AND MONGODB STATEMENTS.

An important feature of the Data Marketplace is the search functionality. When a provider uploads a new asset to the Data Marketplace, is also prompted to provide additional information about the asset (i.e. metadata). This information is related to the usability of the file, the type of the file/asset, its input parameters, as well as the exported results if it is an algorithm, who is the legal owner of the asset, other useful comments, etc.

This set of metadata related to an asset is the content of a JSON file called “description file” of the asset. Each asset has a description file so that it can be searched and retrieved by the end-users, who are able to search for files according to this information (i.e. metadata). Therefore, the submission of these description files per asset is necessary, since their existence on the platform depends on them.

This process is performed automatically from the front-end as, in order for the users to upload an asset to the Data Marketplace, they must first create a description by filling in a specific form. Thus, the front-end, after the submission of this form, converts the data to JSON data that is sent as a JSON file to the back-end in order to store it in the database. In cases where the users interact directly with the APIs of the Interaction Layer (i.e. the machine users/other services), they need to create on their own an appropriate description file. Based on the above, as the offered assets vary and since different information may be needed for each type of offered asset, a standard template with the minimum required fields per asset type has been created, which can be extended based on the providers’ needs. A representative example of a description file (as configured with the specified minimum required fields) in JSON format is cited below, for uploading as an asset the K-means algorithm:

```
{
  "name": "Kmeans example", "description": "This is a simple K-Means example in Python.",
  "type": "algorithms", "subtype": "machine learning", "owner": "The legal owner of the asset.",
  "comments": "Private section - only displayed to logged in users.",
  "fieldOfUse": ["data mining", "..."], "links": [{"title_of_the_link:url", "...:..."}]
}
```

At this point it should be emphasized that all layers of the back-end, except for the management of assets and their descriptions, manage also the users of the Data Marketplace. For this purpose, many APIs have been developed that manage the registration and the log-in of the users to the Data Marketplace, the configuration of their data, the deletion of their accounts from the Data Marketplace and more. Great reference to these APIs - with full description - has already been made in the deliverable "D7.5 - Data Marketplace Software Prototype" that describes the technical part of the Data Marketplace (prototype), whilst any change in these APIs will be delivered in the updated version of D7.5 (i.e. D7.12).

4.1.3 Interaction Layer

The Interaction Layer has to do with the different aspects and interfaces of the RESTful API that handle the back-end's operations. More specifically, it offers the interfaces that connect the outside world with the platform. The Data Marketplace has different APIs that allow all the users of the system (i.e. providers, end-users, human and machine users) to interact with the platform.

In short, appropriate APIs have been developed that receive HTTP requests and through them the platform's CRUD operations are triggered. There exist different APIs for handling the descriptions of the assets and different APIs that handle the lifecycle of the assets. For example, there are APIs that enable users to search the description files and/or metadata stored in the database to find information that fits their needs and there are APIs that enable the upload of assets. These APIs have been developed so that they are common for all types of users of the Data Marketplace, either the user is a human-user or a service-user (e.g. a platform or a system). In general, the back-end is able to respond to HTTP requests with exchanged information in JSON format and only (binary data for assets).

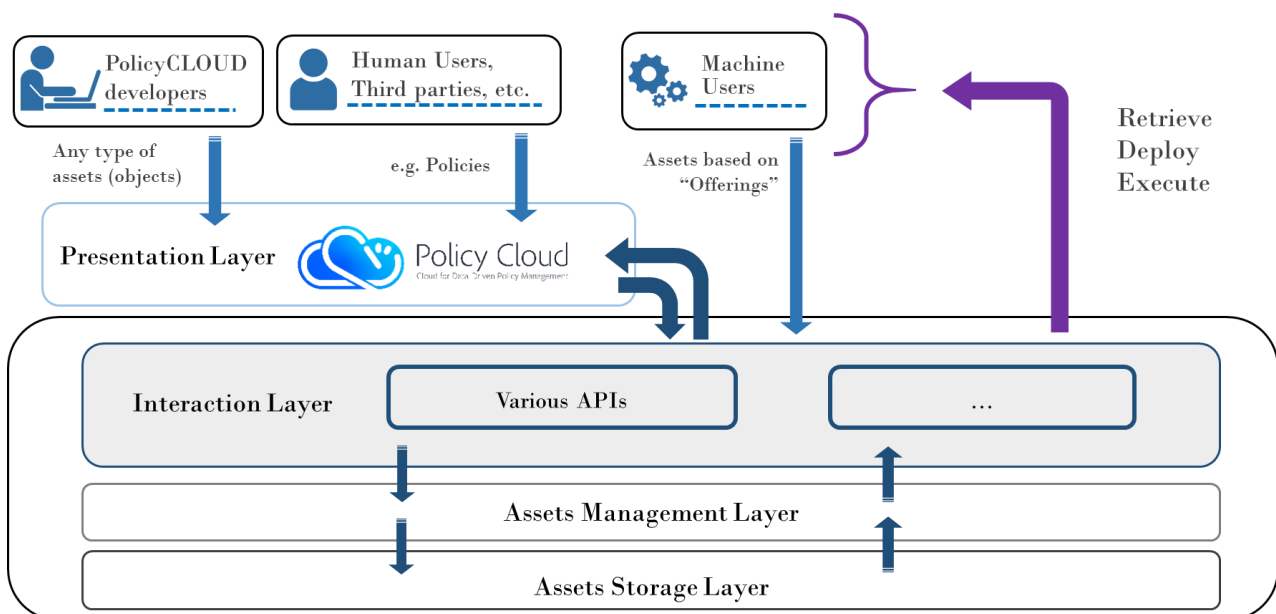


FIGURE 19 - INTERACTION LAYER.

Moreover, there are different APIs per asset type, in the same way as with the collections of the database (as described in [Section 4.1.1](#)), and through appropriate APIs, the users of the Data Marketplace are able to authorize themselves to the platform. Finally, as described above, the back-end provides direct access to its interfaces for all types of users (especially for machine users/services). Based on that and the specifications of the back-end and the front-end, the Data Marketplace has two (2) domain names to be accessed, one for the back-end (<https://mdb.policycloud.eu>) and one for the front-end (<https://marketplace.policycloud.eu>).

4.2 Front-end

The front-end is the fourth layer of the Data Marketplace and it is a web-based server that presents the offered assets to the users, with a friendly UI. It is implemented using various web technologies (HTML, CSS, etc.) and it is functional using PHP and JavaScript. It also exploits the WordPress [21] and various plugins of it, in order to properly manage the content that is presented.

In general, the front-end converts all the interfaces of the back-end (RESTful API) into user friendly interfaces and provides automated forms and processes that make it easier for users to interact with the back-end and benefit from its stored assets. Therefore, it acts as an intermediate among the users and the back-end that sends the respective HTTP requests to the latter, and presents its responses.

Figure 20 depicts the interactions between the front-end and the back-end and the additional components of the overall Data Marketplace.

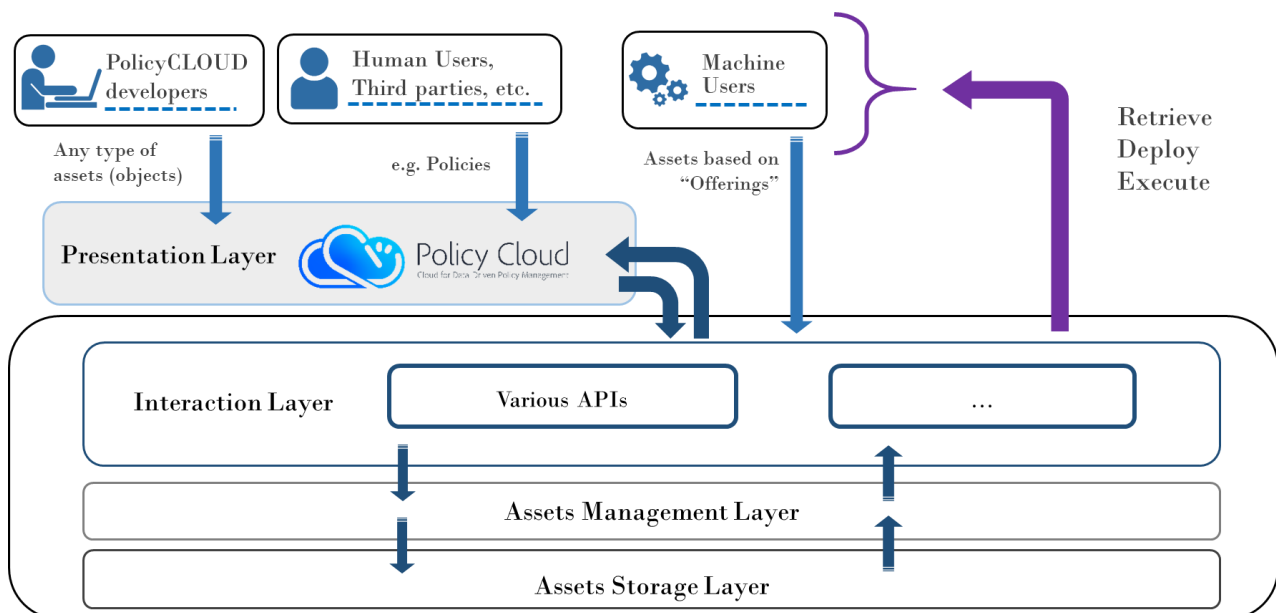


FIGURE 20 - PRESENTATION LAYER.

In short, the front-end allows users to register and log-in to the Data Marketplace (user-based platform), upload their offered assets by filling in an appropriate form whose fields are the contents of the description files of the assets (as mentioned in [Section4.1.2](#)); search for assets according to various fields (title, asset's type, fields of use, provider, other metadata, etc.) that can be further filtered or even sorted by the number of views or the date they were uploaded to the Data Marketplace, etc. Also, there is a page that presents in detail the information of the assets, and through this page, the users are able to retrieve the assets. The initial GUI interfaces of the front-end have already been presented and described in the deliverable "D7.5 - Data Marketplace Software Prototype" and their final version will be delivered in the updated version of D7.5 (i.e. the D7.12).

Figure 21 illustrates the visual sitemap of the front-end part of the Data Marketplace, for providing the individual pages that the front-end part consists of, also indicating any content-related extra information. In deeper detail, on the initial level, there exists the *Homepage*, which serves as the primary page of the Data Marketplace. The pages on the second level of the sitemap are accessible via the navigation bar that relies on the top of the *Home* page, whilst on the third level of the sitemap, the users can browse to "child pages". If a page is limited but users can still access it, the sitemap displays a square with the words "restricted view" after the page that links to it and a square with the words "full view", indicating that users have full access to the pages' functionalities. Hence, in the Data Marketplace, there are specific pages that require authentication in order to be accessible by the users. These pages refer to the *Create* and *Account* pages, which are displayed just after the "login" square.

More specifically, as depicted in Figure 21, the *Home* page is the primary page that can be accessed by all the site visitors. Certain pages are open to all users, while others are restricted, as shown by the sitemap. The *Discover* page is the presentation page that is visible to the general public, where all the underlying assets can be filtered based upon their specific categories, through the filter that is provided. Moreover, all the members (i.e. registered users) can visit the *Asset* page by clicking on any Asset from the *Discover* page, whilst the *About* page is a publicly accessible page offering information, instructions, and "Frequently Asked Questions (FAQa)" about the Data Marketplace. Also, the *Contact* page contains a contact form that users can use in order to submit their personal information and communicate with the Data Marketplace's administrators. It should be noted that both the *About* and the *Contact* pages are open to the general public. Apart from these, there exists the *Login* page, where the users can enter their username and their password to gain access to the entire platform. After logging in, users can access the *Create* page, where all the different kinds of end-users (i.e. developers, policymakers, etc.) can add their own descriptions/assets. After logging in, users can view their submitted assets, comments, ratings, etc. on the *Account* page. The logged in users can also see the entire Discover and Asset pages and discover the detailed information about each asset (i.e. have full view access).

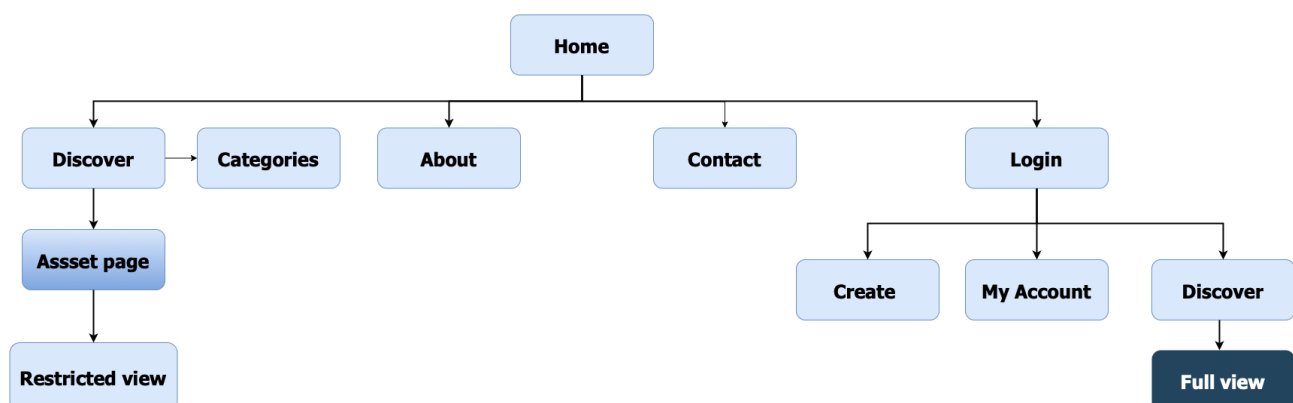


FIGURE 21 - SITE-MAP OF THE FRONT-END.

4.3 Overall Conceptual View

Following the description of the individual elements, the overall conceptual view of the Data Marketplace is presented in Figure 22, which depicts all the platform's layers along with their key offered functionalities, the providers and the end-users.

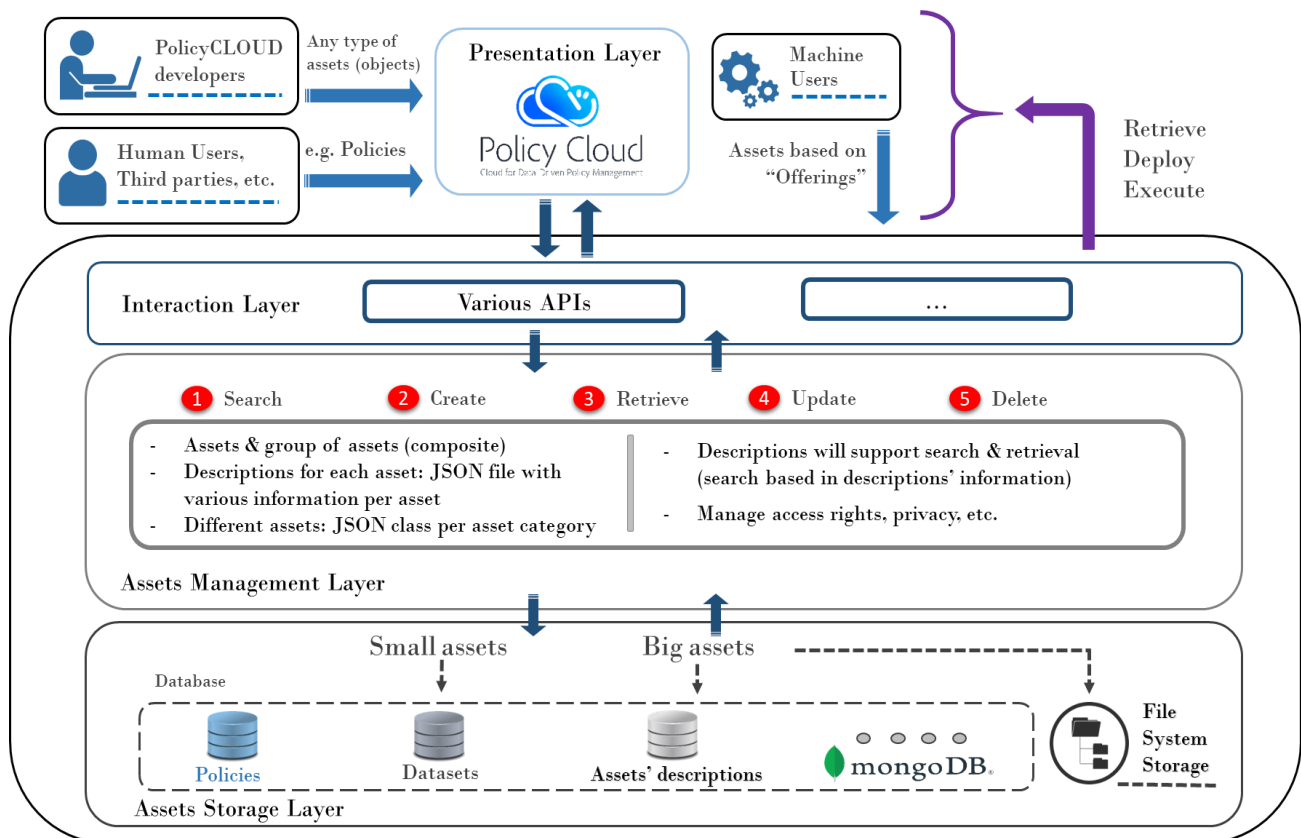


FIGURE 22 - OVERALL CONCEPTUAL VIEW.

The overall information flows are depicted in the figure through the respective arrows that represent the main interactions:

- The users of the Data Marketplace interact with it through the front-end (Presentation Layer), from which HTTP requests are sent to the back-end platform (requests depending on the case: search, upload/store new assets, retrieve, update, or delete an existing asset).
- These HTTP requests are received by the corresponding APIs of the Interaction Layer. Users (especially machine users) are able to interact directly with the APIs, sending the HTTP requests by themselves (using an appropriate tool).
- After receiving the requests, the Assets Management Layer undertakes the processing of the requests, using the developed functionalities. Specifically, it interacts with the Assets Storage Layer to retrieve useful information that, after processing, is sent to the users via APIs, in response to their HTTP requests.

- When a provider intends to upload a new asset, the provider should also submit the asset's description file (via front-end it is generated automatically by filling in the appropriate fields), which should contain metadata for the asset. Both the asset and the description are stored in the Assets Storage Layer, the descriptions are stored in the MongoDB and the assets are stored in the OS.
- In order to retrieve the assets, the Assets Management Layer, through the retrieve functionality, finds the requested asset from the OS (using the metadata stored in the database) and delivers it to the end-user via the platform's APIs. The retrieval of the assets can also be done in a similar way by other systems/services that are linked to the Data Marketplace.

In the following Sub-sections, additional core use cases are introduced to highlight the interaction of the end-users with the Data Marketplace.

4.3.1 Use Case: Upload Assets

Figure 23 presents the use case of uploading a new asset in the Data Marketplace. Note that in the figure, the area with the grey background cites the case where users upload the assets (and their descriptions) through the front-end, while the area with light-blue background reflects the information with respect to the back-end interactions. The IDs (either for assets or descriptions) refer to unique IDs that are associated to each asset/description.

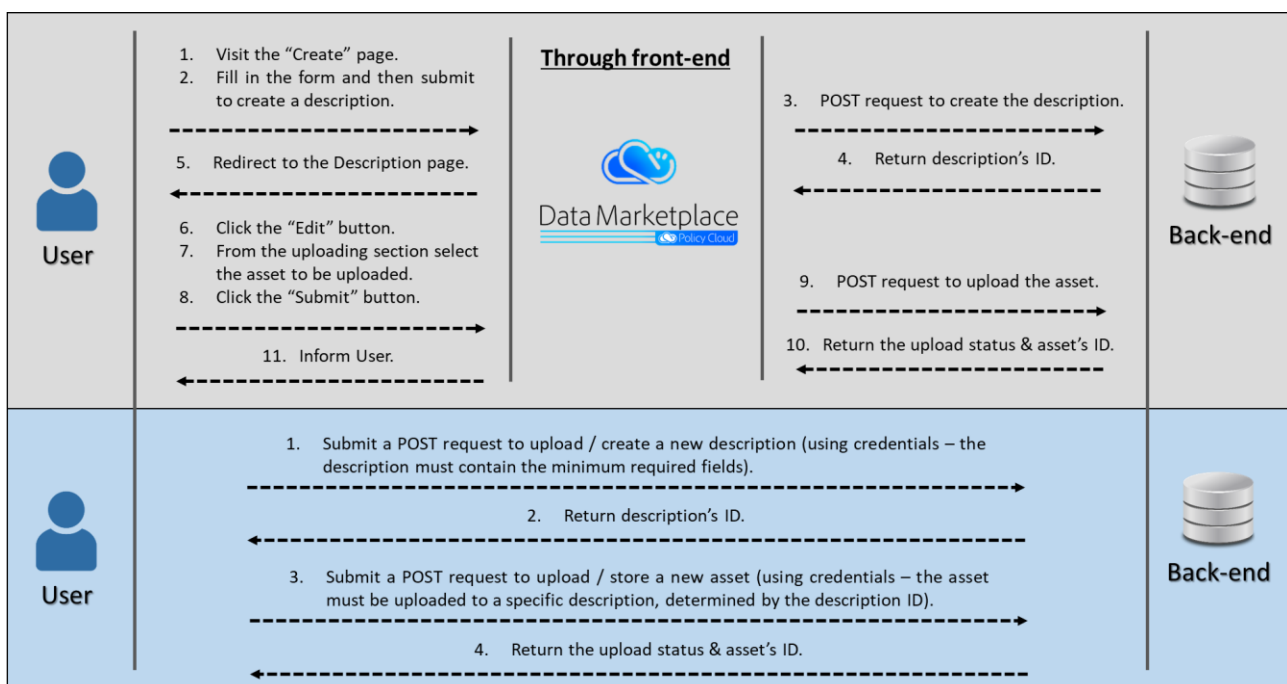


FIGURE 23 - UPLOAD USE CASE.

4.3.2 Use Case: Update Assets

The update use case has a similar process with the upload use case (Section 4.3.1), with the difference that in this case, the assets should be already stored in the Data Marketplace. Corresponding to the previous use case, Figure 24 depicts the interactions in terms of front-end and the back-end of the Data Marketplace.

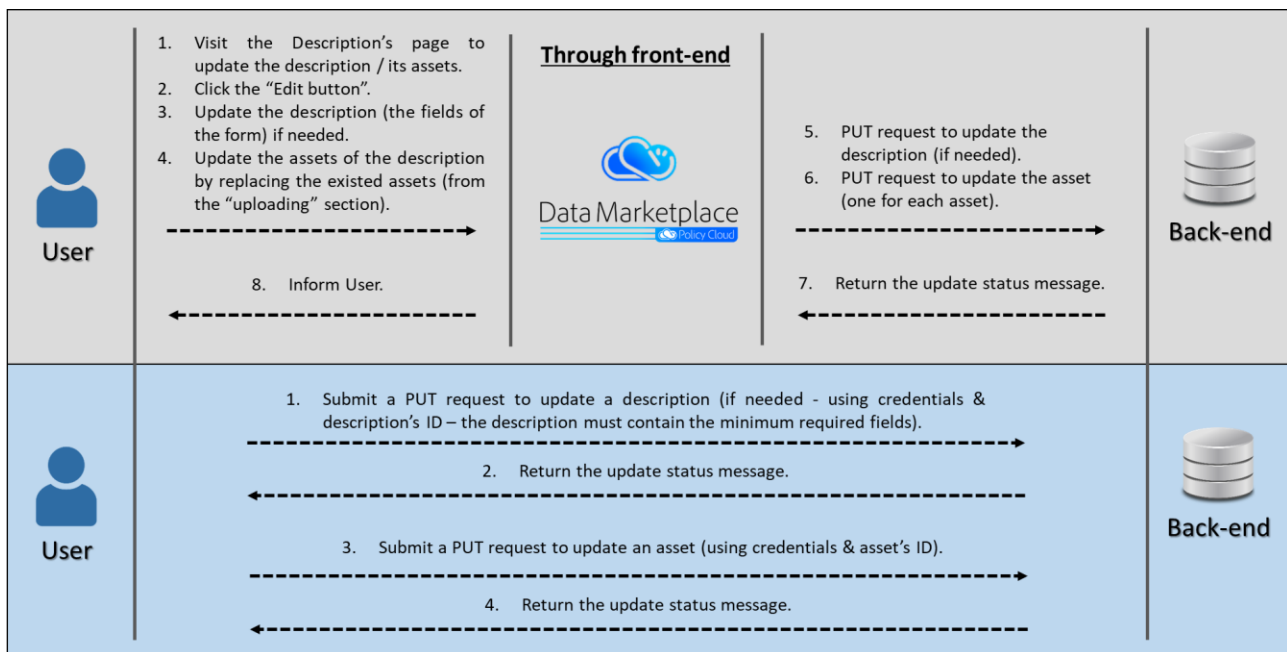


FIGURE 24 - UPDATE USE CASE.

4.3.3 Use Case: Retrieve Assets

Figure 25 depicts the process of retrieving an asset from the Data Marketplace, either through the front-end or directly through the back-end. A point that needs to be further extended for the following figure (but also for the previous figures) are the credentials. As described in Section 4.1, the Data Marketplace takes into account issues related to user's security of the assets, and thus the access of the assets is done through credentials, after the user's registration.

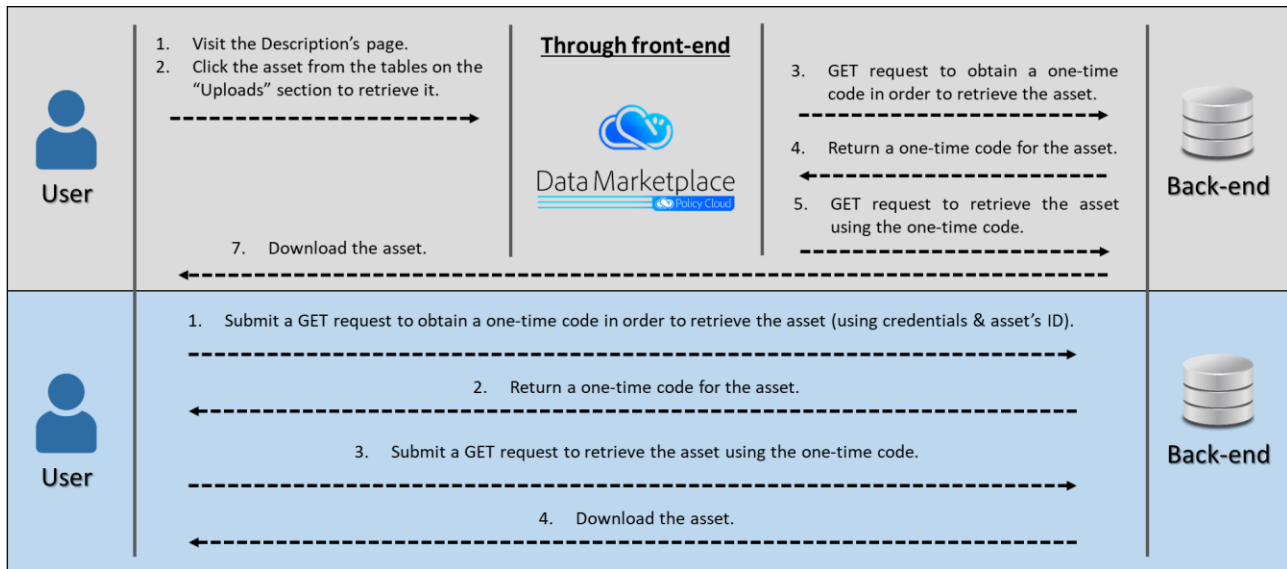


FIGURE 25 - RETRIEVE USE CASE.

4.3.4 Use Case: Delete Assets

The process for deleting assets is presented in Figure 26 for both the front-end and the back-end services of the Data Marketplace. This process is available only for the owners of the assets and for the administrators of the Data Marketplace.

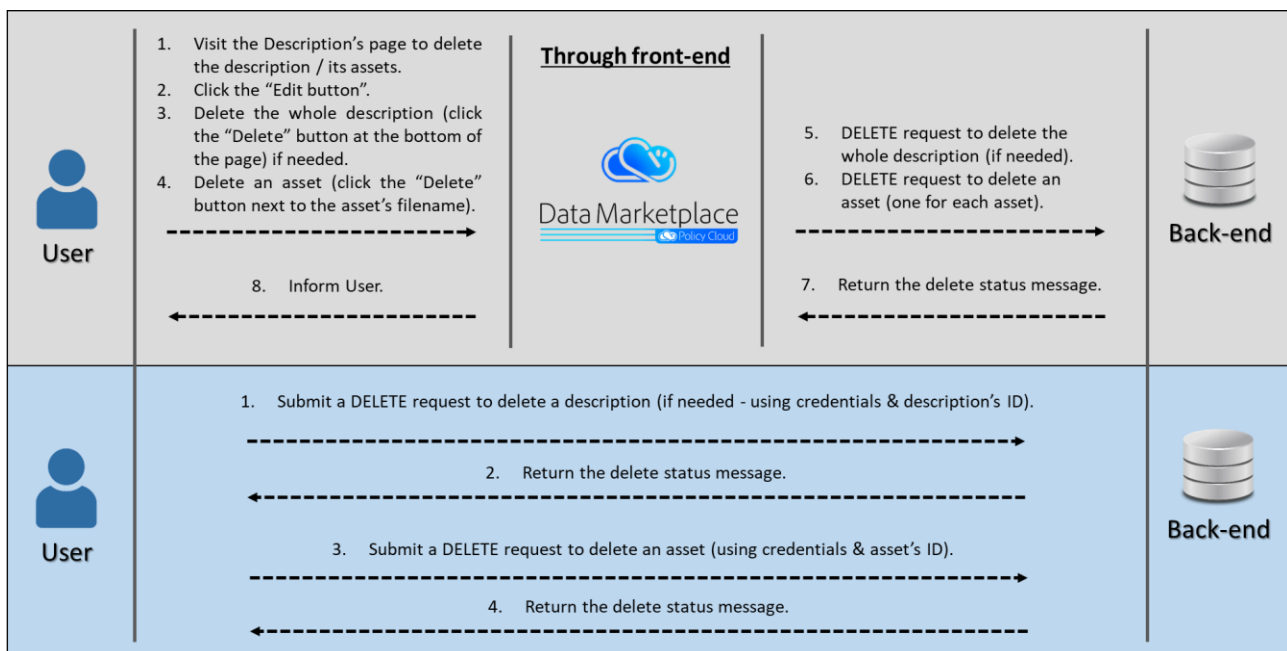


FIGURE 26 - DELETE USE CASE.

4.3.5 Use Case: Search Assets

One of the main capabilities of the Data Marketplace refers to the search functionality. Figure 27 cites the corresponding information flows for the back-end and the front-end. The search functionality is applied to various metadata that are the content of the description files (as described in Section 4.1.2). Thus, the search functionality is applied directly to the descriptions. As depicted below, search parameters/filters are set in the URL as query parameters (GET request).

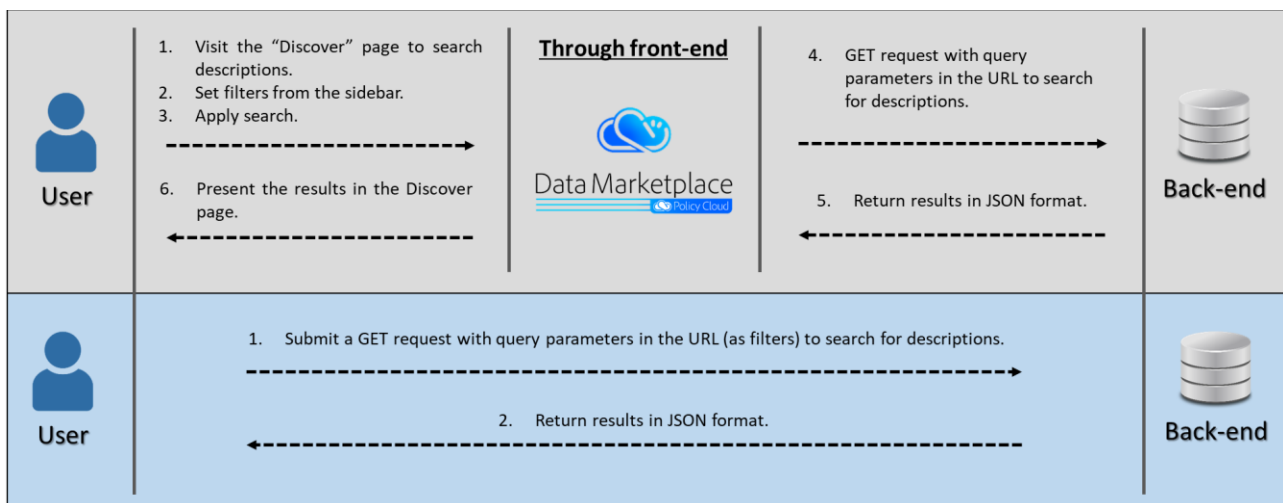


FIGURE 27 - SEARCH USE CASE.

5 Conclusion

This document reports the work that has been currently done in the scope of task T7.2 - Data Marketplace, whose main objective is to provide the final design and specifications of the project's Data Marketplace and develop it.

The specifications that have been introduced in this deliverable and its previous version (i.e. the D7.4 - Data Marketplace: Design and Open Specification), were the base for the realization and implementation of the Data Marketplace, encompassing the described main functionalities regarding the storage of assets, their advanced search and retrieval, and their combination into unique turn-key offerings. The latter has been demonstrated through the already submitted and realized deliverable, entitled D7.5 - Data Marketplace: Software Prototype.

Overall, the Data Marketplace integrates ready-to-use solutions and assets of the project, including policies, tools, algorithms, and datasets to feed them, as well as various project's outcomes (e.g. documents, tutorials, etc.). In general, the Data Marketplace is open to everyone who wants to join its community and benefit from its offerings. The community developed around the Data Marketplace consists of experts (e.g. data scientists, developers, policymakers, etc.) who can offer such solutions and every interested third party who wants to achieve any type of objectives in the respective fields of use.

In conclusion, now that the design of the Data Marketplace has been concluded, the task will focus on the development of its final prototype, as well as the population of the Data Marketplace with assets, and its dissemination to external communities. In M34 (October 2022) it is planned to be released the updated version of the D7.11 - Data Marketplace: Software Prototype (i.e. the D7.12 - Data Marketplace: Software Prototype) that is the last deliverable for the Data Marketplace.

References

- [1] Gelhaar J., Groß T., & Otto B. (2021). A Taxonomy for Data Ecosystems, Proceedings of the 54th Hawaii International Conference on System Sciences
- [2] Smith J. (2018). Data Marketplaces: The Holy Grail of our Information Age, Towards Data Science, August 7
- [3] Spiekermann M. (2019). Data Marketplaces: Trends and Monetization of Data Goods, Intereconomics, version 4
- [4] Acumos, Homepage, <https://marketplace.acumos.org>, retrieved 2022-07-25
- [5] 5GTANGO, Homepage, <https://www.5gtango.eu>, retrieved 2022-07-25
- [6] FINSEC, Solutions, <https://finsecurity.eu/solutions>, retrieved 2022-07-25
- [7] Our World in Data, Homepage, <https://ourworldindata.org>, retrieved, 2022-07-25
- [8] AWS Marketplace, Homepage, <https://aws.amazon.com/marketplace>, retrieved 2022-07-25
- [9] HERE Marketplace, Platform, <https://developer.here.com/products/platform/marketplace>, retrieved 2022-07-25
- [10] Streamr Marketplace, Homepage, <https://streamr.network/marketplace>, retrieved 2022-07-25, <https://corea.streamr.network/marketplace>, retrieved 2022-07-25
- [11] Data Intelligence Hub Marketplace, <https://portal.dih.telekom.net/marketplace>, retrieved 2022-07-25
- [12] Open Data Hub, Marketplace, <https://opendatahub.io>, retrieved 2022-07-25
- [13] IoT Catalogue, <https://www.iot-catalogue.com>, retrieved 2022-07-25
- [14] Docker, Homepage, <https://www.docker.com>, retrieved 2022-07-25
- [15] Python, Homepage, <https://www.python.org>, retrieved 2022-07-25
- [16] The Pallets Projects, Flask, <https://palletsprojects.com/p/flask>, retrieved 2022-07-25
- [17] MongoDB, Homepage, <https://www.mongodb.com>, retrieved 2022-07-25
- [18] MongoDB, GridFS, <https://docs.mongodb.com/manual/core/gridfs>, retrieved 2022-07-25
- [19] Unicorn, Homepage, <https://unicorn.org>, retrieved 2022-07-25

- [20] NGINX, Homepage, <https://www.nginx.com>, retrieved 2022-07-25
- [21] WordPress, Homepage, <https://wordpress.com>, retrieved 2022-07-25