D2.2 - Technical Specifications and Platform Architecture I

Title	D2.2 - Technical Specifications and Platform Architecture I
Revision Number	1.0
Task reference	T2.2
Lead Beneficiary	UPRC
Responsible	Argyro Mavrogiorgou, Thanos Kiourtis
Partners	ATOS, BPFI, DAEM, ECO, ENG, GFT, IBM, IDSA IQB, JOT, JRC, KM, LXS, MC, MOH, NOVO, SFS, TRB
Deliverable Type	Report
Dissemination Level	PU
Due Date	2023-09-30 [Month 9]
Delivered Date	2023-09-28
Internal Reviewers	INSO FTS
Quality Assurance	UPRC
Acceptance	Coordinator Accepted
Project Title	FAME - Federated decentralized trusted dAta Marketplace for Embedded finance
Grant Agreement No.	101092639
EC Project Officer	Stefano Bertolo
Programme	HORIZON-CL4-2022-DATA-01-04

This project has received funding from the European Union’s Horizon research and innovation programme under Grant Agreement no 101092639

Revision History

Version	Date	Partners	Description
0.1	2023-07-23	UPRC ATOS INNOV UBI GFT JRC ENG NUIG	TOC and first contributions
0.2	2023-08-09	UPRC ATOS INNOV UBI GFT JRC ENG NUIG EUBA JOT	Contributions
0.3	2023-09-22	UPRC ATOS INNOV UBI GFT JRC ENG NUIG EUBA JOT	Contributions for 0.3 version
0.4	2023-09-22	UPRC ATOS INNOV UBI GFT JRC ENG NUIG EUBA JOT UNP	Contributions for 0.4 version by UNP
0.6	2023-09-22	UPRC ATOS INNOV UBI GFT JRC ENG NUIG EUBA JOT UNP	Version for peer review
1.0	2023-09-28	UPRC ATOS INNOV UBI GFT JRC ENG NUIG EUBA JOT UNP	Final Version for submission

Definitions

Acronyms	Definition
ACM	Association for Computing Machinery
AI	Artificial Intelligence
API	Application Programming Interface
APP	Application, usually referred to the Project WEB application
BDVA	Big Data Value Association
BI	Business Intelligence
BPMN	Business Process Model and Notation
CDN	Content Delivery Network
CEP	Center (for) Energy Policy
CO	Confidential only for members of the Consortium (including the Commission Services)
CPU	Central Processing Unit
CRUD	Create Retrieve Update Delete - Basic Operations in DBMS
DB	Data Base
DCAT	Data Catalog Vocabulary
DL	Deep Learning
DLT	Distributed ledger technology
ERC	Ethereum Request for Comments
ESG	Environmental, social, and corporate governance
EU	European Union
FAIR	Findable Accessible Interoperable Reusable
FDAC	Federated Data Assets Catalogue
FIBO	Financial Industry Business Ontology
FIGI	Financial Instrument Global Identifier
FML	Federated Machine Learning
GDPR	General Data Protection Regulation
HTAP	Hybrid transaction/analytical processing
IBM	International Business Machines
ICT	Information Communication Technologies
IDS	Intrusion Detection System
IDSA	International Data Spaces Association
IEEE	Institute (of) Electrical (and) Electronic Engineers
IP	Internet Protocol
IT	Information Technology
JDBC	Java Database Connectivity

JSON	JavaScript Object Notation
JWT	JSON Web Token
KPI	Key Performance Indicator
LLM	Large language model
ML	Machine Learning
MSP	Membership Service Provider
NIS	Network and Information Systems
NLP	Natural language processing
ODRL	Open Digital Rights Language
OSS	Open Source Software
OWL	Web Ontology Language (W3C)
PSD2	Second Payment Service Directive
PSDII	Second Payment Service Directive
RA	Reference Architecture
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
REST	Representational State Transfer
SA	Supervisory Authority
SAX	Situation Aware Explainability
SDK	Software Development Kit
SLA	Service Level Agreement
SOA	Service Oriented Architecture
SQL	Structured Query Language
SSL	Secure Sockets Layer
TLS	Transport Layer Security
TR	Technical Requirement
UI	User Interface
UML	Unified Modelling Language|Universal Markup Language
VDIH	Virtual Digital Innovation Hub
WP	Workpackage
XACML	EXtensible Access Control Markup Language
XAI	eXplainable Artificial Intelligence

Executive Summary

The document presents the first part of the work performed within the FAME project to design and build a stable Solution Architecture (SA) that has as a goal to develop, integrate, validate, and offer as a publicly accessible service Europe's first federated, decentralized, trusted and energy efficient data assets' marketplace for Embedded Finance (EmFi). To this context, the main focus of the deliverable is to set the technical background of the FAME SA, and clarify that the aim of the project is to realize a practical and mature Federated Data Space to facilitate finance and non-finance organizations, accessing Data Marketplaces and other Data Spaces in accordance with the European Union (EU) Data Strategy and Data Regulations for securely accessing, sharing, trading, and analysing data to support the creation of added value services (e.g., EmFi applications), ultimately contributing to the data economy of the European Single Market.

The work on the SA has been conducted as a teamwork that involved all the Consortium beneficiaries, being a continuous and incremental process, whose stable snapshot is represented in the current deliverable. An updated version of this deliverable will be produced at the 18^{th} month (M18) of the project, considering all the improvements and due verifications over the period.

The current SA fulfils the project's milestone M2 fully specifying the first SA of FAME. The FAME partners have selected a concrete methodology to work on the SA, identifying it in the "C4 model", which is presented in the document. Based on abstractions that represent how software architects and developers think about and create software, the methodology is an "abstraction-first" approach to diagramming a software architecture. The current document demonstrates that all the functionalities of FAME are properly covered by this model, being fully analyzed in the remaining of this deliverable.

The survey performed prior to concluding to this SA is also provided, consisting of a detailed literature review on FAME related state-of-the-art definitions (e.g., data lakes, data warehouses, data marketplaces, data spaces), including their challenges and provided opportunities. The business value as well as the technical value of FAME is also presented, all of which have led to the specification of the current FAME SA. What is more, this deliverable's survey underlines that some already existing Reference Architectures and Solution Architectures - which are thoroughly explained - provide substantial input to FAME, to support the functionalities of the different components, pilots, and use cases of the project. Thus, the resulting FAME SA provides a schema for building solid workflows and ensures full communication and interaction among all the FAME building blocks and its interactive end-users.

Table of Contents

List of Figures

List of Tables

1 Introduction

For Europe to operate cooperatively and in accordance with European principles, such as self-determination, privacy, openness, security, and fair competition, the digital market is crucial. The expanding data economy needs a legal framework to define data protection, fundamental rights, safety, and cybersecurity. The harmonization of digital markets is one of the fundamental policy achievements of the European Union (EU), with the European Strategy for Data serving as its primary tangible product. The Data Governance Act comes next, with the goal of promoting usable data by boosting EU-wide data sharing procedures and enhancing trust in data intermediaries. The Data Act, a legislative proposal that intends to establish a framework that will promote business-to-government data sharing, is also another vision of the EU data economy.

Data Space is a notion that is still evolving, but it has been specifically described by the European Strategy for Data, which directs European activity towards the data economy. The approach is so comprehensive that it identifies nine (9) common European Data Spaces: Health, Industrial, Agriculture, Finance, Mobility, Green Deal, Energy, Public Administration, and Skills. While naming data spaces provides guidance, it does not reveal their nature or essence as ecosystems, which may have unique characteristics and place more focus on layers that are thought to be common to all Data Spaces. For many years, the Big Data Value Association (BDVA) community of specialists has focused on Data Spaces to comprehend and consider the complex nature of the idea. A separate space that crosses organizational, sectoral, and geographic barriers is what BDVA specifically refers to when it refers to the "European data sharing space", which is a place made up of connecting a variety of distinct spaces. Data Spaces can also be thought of as the phrase for an ecosystem that includes data-sharing technology, an appropriate regulatory environment, and novel new business elements.

Towards this vision, in this deliverable the FAME project's Solution Architecture (SA), overall added-value, and performed background analysis is presented, introducing FAME's vision into acting as a single-entry point Data Space of low complexity, offering the toolsets to gain valuable insights for Embedded Finance (EmFi) applications, by exploiting cutting-edge technologies (e.g., Artificial Intelligence (AI)/eXplainable AI (XAI), Machine Learning (ML)/Deep Learning (DL), Blockchain). FAME's main goal is to provide EmFi with Europe's first standard-based, secure, interoperable, and federated data and analytics marketplace. The SA of the FAME platform (Section 7) will use already existing reference models for constructing its Data Space, connecting various external data sources like existing Data Spaces, Data Marketplaces, and Datastores among others, also providing interfaces for trading, pricing, and managing data assets. FAME will also offer trusted and energy-efficient analytics, tools for compliance with key regulations of the industry, federated access to data assets from various providers, advanced semantic interoperability features for EmFi, novel decentralized techniques for programmable, value-based trading and monetization of data assets, and more.

1.1 Objective of the Deliverable

The purpose of this deliverable is to introduce the FAME SA. It emphasizes on the presentation of the structuring principles of the SA, building on experience, recommendations, guidelines, and design principles from other Reference Architectures (RAs) of existing Data Marketplaces (e.g., i3-MARKET, INFINITECH, DataVaults) and Data Spaces (e.g., IDS, Gaia-X).

Prior to that, the deliverable aims to present the roadmap towards the presented FAME SA, including its value proposition, its stakeholders, the addressed challenges, and the provided opportunities. The FAME SA's involved components are thoroughly described and technically specified, providing their added-value to the overall SA. Hence, the deliverable presents the initial version of the FAME SA that will drive the technological developments of the project, as well as the design and initial integration of the underlying use cases. As part of these development and integration processes, feedback on the relevance and appropriateness of the SA will be solicited. This feedback will be considered in the development of the final version of the FAME architecture as part of the deliverable D2.6 - Technical Specifications & Platform Architecture II that will be released at M18 of the project.

1.2 Insights from other Tasks and Deliverables

The FAME SA signals an important milestone for the FAME work in WP2. Specifically, it provides the structuring principles that will drive the design and development of an EmFi Federated Data Space, considering current Data Spaces' challenges and opportunities, as well as the wider requirements of this challenging sector. As such, FAME SA is:

- Driven by an earlier WP2 deliverable, namely D2.1 - Requirements Analysis, Specifications and Co-Creation I [1], where the current deliverable considers the extracted both business and technical requirements of the underlying use cases, along with the EmFi application domain's needs, towards providing a first stable version of the SA.

Other than that, the current deliverable provides insights and technical specifications for the detailed design of the technical components in WP2, WP3, WP4 and WP5 of the project that are thoroughly described. It must be noted that while the design and development of these components in-principle can be independent of their use in a specific architecture, the FAME SA will have an impact on the way these components interact among them, having an influence on their design and development. Overall, the present deliverable has an instrumental role in the project, as it drives many tasks and FAME-related deliverables, being linked with one of the milestones of the project's workplan.

1.3 Structure

The deliverable is structured as follows:

- *Section 1* (current Section) includes an introduction to the related motivation behind the FAME project's vision including the objectives of the current deliverable, as well as the relation of this deliverable with the existing ones and the project's technical WPs.
- *Section 2* introduces some key terminologies that are going to be used across the document and the overall FAME project, facilitating its readability and understanding.
- *Section 3* includes the basic concepts and overall idea behind the FAME project, analyzing some of the most critical data associated technologies/approaches behind its realization. Also, the surrounding challenges of the concepts with which FAME is interacting are depicted.
- *Section 4* includes the overall value of FAME, depicting the status of its related activities concluding to a high-level overview of the project. Both the business value and the technical value of FAME are provided, introducing its potential stakeholders.
- *Section 5* provides all the necessary information regarding existing RAs, principles, and functionalities that will be exploited in the context of the FAME SA. For these, it depicts an overview of their internal architecture, as well as their added-value to FAME and their contributions towards achieving FAME's set objectives.
- *Section 6* contains a short introduction to the FAME's mechanisms and capabilities, explaining their value, the challenges they are tackling, and the interconnections among them.
- *Section 7* includes the overall FAME SA, introducing its C4 architecture model followed by a detailed explanation of each of the different views of the C4 model and the interacting technical components whose technical specification is also provided.
- *Section 8* contains the conclusions that have been derived from the FAME SA and its overall components' technical specifications, concluding to the achieved KPIs.

2 Terminology

On top of the list of acronyms cited above, this Section introduces some key terms that are used across the document and the FAME project in general. These key terms refer to the following:

- **Data Space**: A decentralized infrastructure for transparent and trustworthy data sharing and exchange in data ecosystems within a certain application domain, based on commonly agreed principles and capabilities, consisting of data platform(s), data marketplace(s), and data sovereignty.
- **Data Platform**: An environment that facilitates the exchange of value between two or more parties, with the multiple parties interacting through the platform.
- **Data Marketplace**: A multi-sided place (i.e., platform) where data providers and data consumers can find each other to stimulate data exchange or access.
- **Data Sovereignty**: The capability of a person or organization to make all data-related decisions on their own.
- **Embedded Finance (EmFi)**: Integration of financial services like payments and trading, into nonfinancial businesses' infrastructures without the need to redirect to traditional financial or insurance institutions.
- **FAME**: A Federated Data Space for the EmFi application domain, providing a single-entry point to various data assets, supporting related data indexing, searching, monetization, and trading, additionally offering a list of energy efficient analytics.
- **FAME Federation**: A group of participants (e.g., consumers, producers) connected together with agreed government, access, and security rules through FAME.
- **(Data) Asset**: A data resource or artefact (e.g., system, application output file, document, database, web page, dataset, service, algorithm, AI/ML model, data insights, data visualizations, software, publications) that carries data that is relevant for the value chain of an organization or institution, or which has strategic or operative value to generate revenue through exchanging it.
- **(Asset) Offering**: A description of the commercial and licensing terms under which a data asset can be obtained, where the latter may contain one or multiple offerings simultaneously.
- **Asset Metadata**: Information about a data asset that helps to describe, structure, or administer that data asset, providing a structured reference that helps to sort and identify attributes of the information it describes.
- **(FAME) Asset**: Any data asset or technology component related with FAME.
- **(FAME) Services**: The basic components that trigger any functionality within the FAME Federated Data Space.
- **(FAME) Dashboard**: A user interface for FAME stakeholders to interact with the overall FAME Federated Data Space and its functionalities.

3 FAME Baseline Concepts

Current Section includes the basic concepts and overall idea behind the FAME project. In essence, some of the most critical data associated technologies are analyzed, whereas some related state-of-the-art methodologies and data approaches are introduced, all of which have led to the realization of the FAME vision and concept. Furthermore, the surrounding challenges of the concepts with which FAME is interacting are depicted, being categorized into additional layers with respect to the limitations they may offer towards the vision of a common EU data economy.

3.1 Concepts

3.1.1 Data Associated Technologies

Data and information are the most important assets for most organizations. While organizations can use their data to improve their businesses, their data can also have significant value beyond their business. At the same time, nowadays organizations are aware of the competitive edge that they can gain by also incorporating external data into their data strategy and business models, trying to optimize their overall information processes and insights. It is an undeniable fact that in such a data-driven business climate, data is playing a key role in capturing market intelligence and actionable insights to augment business operations. Hence, in today's data-driven world, organisations are constantly seeking innovative approaches to manage and harness the power of data. Towards this direction, data management platforms, tools, and associated technologies (i.e., Databases, Data Warehouses, Data Lakes, Data Meshes, Data Hubs, Data Catalogues, Data Platforms, Data Ecosystems, Data Marketplaces, Data Spaces) are increasingly getting a global focus, having created a plethora of diverse options that lead to controversies and hot debates. Such fact is also verified by a Gartner study that has shown that more than 25% of customers thought that a Data Hub was a Data Lake solution [2]. Gartner's research illustrated how much confusion there is about what the different concepts entail, which is also intensively spotted in the research community [3][4]. While all the existing concepts aim to improve data capabilities within organisations, they differ in their fundamental principles, architectural design, and organisational impact.

The following paragraphs provide more clarity on the meaning of these terms and showcase how their concepts differ from each other, making clear their general scope and usage.

- **Database**: A database stores information from a *single data source* for one particular function of an organization, being able to process many simple queries upon the data quickly.
- **Data Warehouse**: A data warehouse stores large amounts of *structured data from multiple sources in a centralized place*. The goal of using a data warehouse is to combine disparate data sources (inside or outside of an organization) in order to analyse the data, look for insights, and create Business Intelligence (BI) in the form of reports and dashboards.
- **Data Lake**: A data lake stores data from *disparate sources that is stored in its original, raw format* (either structured, or semi-structured, or unstructured data format). The basic scope is to store raw data from all sources without the need to process or transform it at that time, allowing organizations to ingest and manage large volumes of data in an aggregated storage solution when they might not be entirely sure how they will use that data in the future (whether for business intelligence or data products).
- **Data Fabric**: A data fabric is a *centralized storage approach* to data management architecture, using automated/intelligent systems to connect *data stored in multiple places and in multiple formats*. In essence, a data fabric expands on the architecture of a data warehouse, including building blocks such as data access, discovery, transformation, integration, security, governance, lineage, and orchestration. The goal is to advocate for

setting up a unified data layer to provide a single source of truth for data, ensuring data quality, consistency, and security, while allowing different end-users to access and manage data easily.

- **Data Mesh**: A data mesh is a *decentralised and domain-oriented storage approach* that enables collection, integration, and analysis of data from disconnected systems concurrently, so there is no need to pull in data from disparate systems into a single location and preprocess them for analysis. In essence a data mesh decentralizes data from a single source so that it can be readily available to multiple users. It is ideally targeted for a business environment where data needs to be integrated from many disintegrated systems or processes for fast analysis.
- **Data Hub**: A data hub does not store data itself but takes care of the *flow of data between source/target systems and users*. The goal is to indicate exactly what actions need to be performed with the underlying data, taking the form of a hub-and-spoke architecture where systems can distribute data through the data hub, rather than through point-to-point integration where every system is connected to any other system with which data needs to be shared.
- **Data Catalogue**: A data catalogue primarily serves as a *unified inventory or directory of an organization* (much like a library catalogue for books), *maintaining metadata* about the data including their location, format, schema, usage, relationships, ownership, and quality among others. The goal is to act as an inventory or a metadata management system, enabling users to quickly find and access the data they need for their tasks.
- **Data Platform**: A data platform serves as the *ultimate storehouse for data coming from diverse sources*, transforming it into a single source of truth, and *allowing the scaling of complex processing and analytics operations* that turn data into valuable insights. In essence, a data platform is a *software framework or environment* that provides a foundation for developing and running software applications. It can be thought of as a set of tools, libraries, and services that are used to create, deploy, and manage the built applications. Thus, it can be viewed as an integrated solution that encompasses the features of data lakes, data warehouses, data hubs, etc. towards the realization of the needed applications. Without a data platform each component is typically handled by different tools to make data flow from source to end-user, creating complex environments. Hence, the basic goal is to provide an environment for end-users to build and run their applications, where the hardware and the software components of the platform work together to support the application throughout its life.
- **Data Ecosystem**: A data ecosystem refers to the *combination of enterprise infrastructure and applications* that is utilized to aggregate and analyse information. In essence it refers to all the programming languages, algorithms, applications, and general infrastructure that is used for *collecting, analysing, and storing data*. The goal is to enable organizations to better understand their data and take the proper decisions.
- **Data Marketplace**: A data marketplace is a platform where *data* (e.g., raw data, processed data, analytics-ready data products) can be *bought, sold, or exchanged* between organizations and/or individuals. In essence it is a platform that connects data buyers (consumers) with data sellers (providers), acting as an easy-to-navigate hub that connects data providers with data consumers, offering a range of datasets from various industries and domains. Thus, the goal is to serve as a hub where data is curated, organized, and made available for users to able to easily explore, purchase, or exchange it.
- **Data Space**: A data space is a *decentralised infrastructure for trustworthy data sharing and exchange* in data ecosystems, based on *commonly agreed principles*. Into this context, data is not stored centrally but rather at the source, and thus only transferred as necessary, supporting a decentralised nature that allows actors to keep the sovereignty on the data. In essence, a data space brings together relevant data infrastructures and governance frameworks to facilitate

data pooling and sharing. The goal is for the data to stay with the providers and made available via secure peer-to-peer communication with common semantics and data sovereignty.

As relevance of data grows every year, the European Commission invested for the development of Data Spaces, envisioned as of strategic importance for the growth of the European data economy [5]. The aim is to enable and stimulate the development of Data Value Chains, keeping sovereignty and trustworthiness under European premises and values. Figure 1 depicts the relationship and the relevance among all the abovementioned concepts, highlighting the positioning of Data Spaces that is further investigated and adopted in the context of FAME.

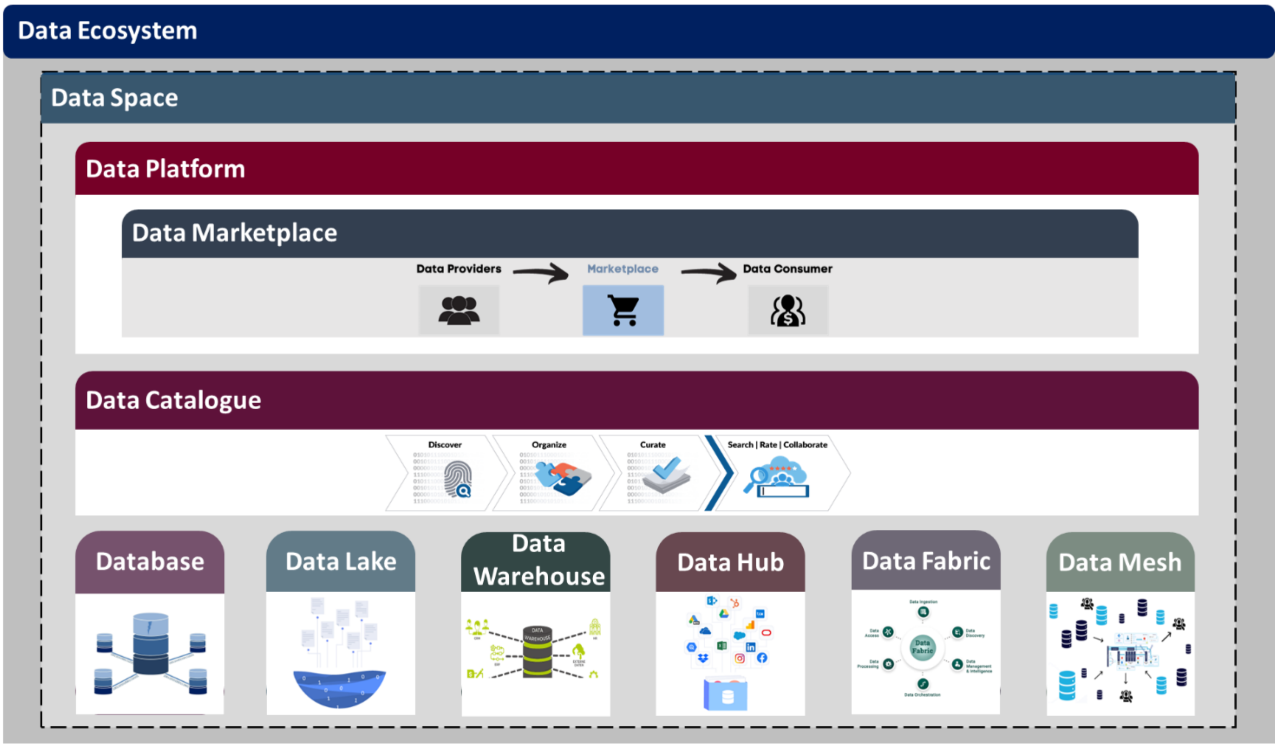


Figure 1 – Relevance of existing data management associated technologies

3.1.2 Centralized, Decentralized, & Federated Data Approaches

In today's data-driven world, organizations face critical decisions regarding the storage and management of their valuable data. One fundamental choice for performing such tasks is to select to apply a data architecture between a centralized, a decentralized, or a federated manner. Each approach has its merits, and understanding the advantages and disadvantages can help organizations make informed decisions that align with their unique needs and goals.

The following paragraphs provide more insights on the meaning of those three (3) diverse data architecture types, outlining their overall idea and concepts in the context of being applied into any kind of Data Platform, Data Space, etc. [6][7][8].

¶ **Centralized Data Approach**: A centralized data approach refers to the practice of *storing data in a single, centralized location*, typically within a data centre or a cloud environment. Thus, all the participating source systems copy their data to a single, centrally located data repository where they are organized, integrated, and stored using a common data standard/formatting.

¶ **Decentralized Data Approach**: A decentralized data approach involves *distributing data across multiple locations or systems*, where *each system performs its own actions* upon the data, finally contributing its results/data into a central place. However, since each system is not interacting with each other, it is not aware of the processes/standards/formatting applied upon the data. In essence, in a decentralized data approach every system makes its own decisions, and the result of the underlying Data Platform is the aggregate of the decisions of the individual Systems.

- **Federated Data Approach**: A federated data approach involves *individual systems that maintain control over their own data* but agree to *share some or all this information to other participating systems upon request*. Users of the system submit questions through a common intermediary interface, which searches the various source systems. In essence, by following a federated data approach it allows an organization to leave its data where it is, using a common (single source) platform to provide a unified view of the data. Through this way data consumers can retrieve information from multiple, disparate systems with a single query, in real time.

Choosing between following a centralized, decentralized or a federated data approach is not a one-size-fits-all decision. However, nowadays federated data management has emerged as an effective solution for managing raw data and empowering organizations and data consumers to put valuable data to use [9]. All of the involved parties are able to (i) maintain ownership of whatever data they produce, (ii) aggregate and standardize their data and then make it available on behalf of the data owners, and (iii) perform their data actions by themselves or allow data consumers to edit their data with their approval. Figure 2 depicts the different applied ideas among the abovementioned data approaches, highlighting the concept of the federated data approaches that are deeply investigated and adopted in the context of FAME towards realizing the creation of a Federated Data Space.

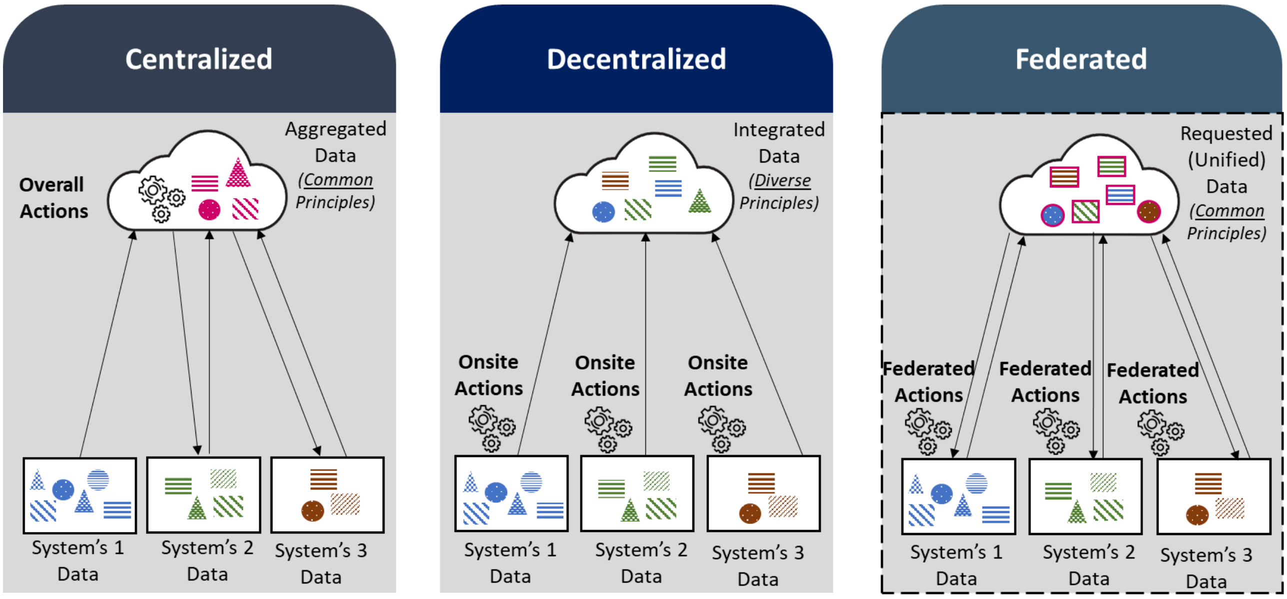


Figure 2 – Difference of centralized, decentralized, and federated data approaches

3.2 Challenges

A successful Data Space takes a lot of work to be developed and is not completely risk immune. Organizations face a wide range of difficulties when attempting to create and keep up a Data Space that caters to the interests of all stakeholders. The difficulties can be divided into two (2) primary categories: (i) intra-organizational (problems faced by data producers and consumers, as data sharing players), and (ii) inter-organizational (lack of adequate data sharing ecosystems).

The difficulty in evaluating data worth due to a lack of data valuation standards and assessment tools is the first major intra-organizational challenge. The very arbitrary and party-dependent nature of data value as well as the general absence of producers' data sharing foresight make this issue even worse. The second issue is the challenge that data producers confront in balancing the perceived worth of their data (after sharing) against the risks it exposes (upon sharing), even when they follow the rules. In a business environment that is already fiercely competitive, some examples include the perceived loss of control over data (due to the fluid nature of data ownership, which is still difficult or impossible to be defined legally), the loss of trade secrets (due to accidental disclosure or malicious reverse engineering), and the risk of evading legal constraints given potential data policy breaches (including General Data Protection Regulation (GDPR) and the disclosure of private identities).

The lack of reliable and meaningful data sharing ecosystems that compel quick widespread engagement continues to be the top inter-organizational concern. The absence of strong governance models, legal and ethical frameworks, and trustworthy intermediaries that ensure data quality, dependability, and fair usage are the main factors. This is made worse by the fact that new best practices and standards (including interoperability, provenance, and quality assurance standards) are not being widely followed and whose maturation rate is likewise falling short of expectations. The rapid transition towards decentralized mixed-mode data sharing and processing architectures also creates considerable scaling issues. From a technical perspective, data sharing solutions need to satisfy European concerns like ethics-by-design for democratic AI.

All these challenges can be grouped into four (4) broad categories, referring to: (i) technical, (ii) business/organizational, (iii) legal compliance, and (iv) national/regional (furtherly analyzed in the below sub-Sections). It is possible to develop a viable Data Space that satisfies the demands of all the stakeholders and promotes the expansion of the European data ecosystem and economy by overcoming these obstacles and addressing each of these problems.

3.2.1 Technical Challenges

The need to create a cross-border, cross-sector sharing Data Space and provide platforms the ability to handle "mixed" private, individual, and open public data creates new technical difficulties and exacerbates those that already exist. Following the emergence of opportunities for data sharing that extend beyond conventional raw data and its transformations along the processing chain to metadata, models, and processing algorithms, it is necessary to revisit the impact of known challenges (e.g., the Vs of Big Data: volume, velocity, variety, veracity) along the data lifecycle. The primary difficulties refer to:

- **Sharing-by-Design**: At the time of data generation, the majority of data producers do not yet think that data sharing is a possibility. Existing data lifecycle management models need to accomplish a better job of incorporating all the pertinent processes, such as discovering the right data and preparing it for dissemination. Both the availability of the data itself and the maturity of the data services (such as cleaning and aggregation) in data sharing ecosystems are essential for the development of the data economy. Additionally, by separating the various types of data that can be shared into the categories mentioned above, the "variety" challenge becomes more complex, and interoperability solutions must take this change into account.

- **Digital Sovereignty**: A mixed data sharing space will only become a reality if data producers are assured to keep their ownership rights, allowing them to decide who can use their data, how it can be used, and under what conditions. To ensure digital sovereignty, additional research into appropriate data rights management frameworks or alternative ownership models is necessary.
- **Decentralization**: Decentralized data storage designs are favored over centralized data storage arrangements to ensure that data creators retain control over their data. As a result, when discussing data volumes and data velocity (data streams), it is increasingly important to consider both the scalability of real-time operations over dispersed data at rest in arbitrary geographic distributions, and the distributed processing of data in motion, which does not require intermediate storage. There is also a rising need for standard data exchange protocols in decentralized architectures.
- **Veracity**: For data sharing ecosystems to continue to function, data integrity is still essential. Data at different processing stages will need to carry traceable information about their sources and processes (i.e., metadata about their initial state, algorithms, and processes they went through). To increase confidence, support for enhanced provenance is necessary.
- **Security**: Closed (proprietary, personal) data must be unlocked for interchange and sharing inside a trustworthy network, which necessitates a proper solution for problems like confidentiality and digital rights management. Furthermore, even in a decentralized peer-to-peer network, secure access control must be ensured. As a result, all the nodes and participants in the data sharing space must adopt standardized security solutions and exchange protocols.
- **Protection of Privacy**: Although there are technological solutions for secure and reliable data sharing (such as privacy-enhancing and privacy-preserving technologies, including digital identity management), these solutions must continue to mature in order to increase their acceptability and adoption.

3.2.2 Business & Organizational Challenges

The socioeconomic viability of a pan-EU Industrial Data Platform (IDP) integrating various Data Spaces and providing Data Marketplaces is anticipated to provide the following business issues:

- **Values of EU**: IDPs created by the EU must uphold ideals like democracy, fair competition, and equality of treatment. These qualities can set businesses apart in the international market and get rid of dubious "shortcuts" that would benefit rivals from around the world. Additionally, new business models must show how they adhere to EU principles and how they are superior to current commercial solutions in this environment.
- **Overall Competition**: A major competitive advantage for the EU in the international market is the union of the digital and service industries. Therefore, it is necessary to find value-added data-driven services that could make "Made in EU" products competitive on a worldwide scale. Furthermore, co-opetition models require more research, as do SMEs (which make up 99% of the EU industrial fabric) and the function of PPP intermediaries like the Digital Innovation Hub (DIH).
- **Changing Ecosystems**: Shared data ecosystems must ensure that data producers have total control over who has access to and uses their data in the industrial domain. However, it might be challenging to determine ownership legally. Additionally, there are no agreed-upon standards or clear rules for implementing data sovereignty in adaptable and dynamic businesses. Additionally, it is uncertain how peer-to-peer networks of the future will maintain sovereignty and confidence without centralized management.

- **Changing Skills**: Divergent opinions exist regarding the precise effects of new data-driven technology and automation on employment and jobs. Personnel re- and up-skilling are examples of short-term initiatives. In the long run, nevertheless, a full redefinition of workflows, procedures, and patterns of human-machine interaction is necessary. Furthermore, the present-day educational system is still not designed to consistently accommodate for new and undiscovered occupations.
- **Digital Transformation**: Data-driven changes are required for people (and roles), platforms (and spaces, marketplaces), products (and services), processes (and organizations), partnerships (and participatory innovation models), and performance (and data-driven KPIs). It is necessary to develop strategies and resources to aid EU industry in this transformation.
- **Trust**: Understanding the commercial value of data produced by industry at all levels is essential for data markets. An issue is the lack of trust in the quality of data that is shared. Widespread, automatic data exchanges will not happen without quality requirements. Algorithms, such as algorithm bias, should also be subject to efforts to improve data accuracy. Additionally, costs associated with data preparation (such as cleaning and quality assurance) as well as risks (such as possible access to trade secrets and intellectual property sharing) must be taken into account. Additionally, careful adherence to GDPR requirements is required when sharing personal data in Business-to-Business (B2B) applications. Ad hoc and on-the-fly B2B data sharing mechanisms and contracts, given under clearly defined data sovereignty rules, must be taken into consideration in order to establish trusted data networks.
- **Standards for Valuation**: Data Marketplaces open up new possibilities and business models with the monetization or valuation of data assets at their core. The pricing of data poses new issues, such as determining whether this is done by the producer, by market demand, or by a broker or other third party. Another issue is determining whether the value of a particular data asset is fixed or dependent on the buyer-seller relationship. To help firms assess the value of involvement, guidelines and price models must be developed.

3.2.3 Legal Compliance Challenges

A complicated data policy environment has been created by all the various rules that have been implemented over the past ten years within the context of the digital single market. Therefore, a deeper knowledge of how data regulation interacts with and links to within data platforms is required. The following are the issues that need to be resolved immediately:

- **Protection of Data**: Data rights and consent might change, thus new business models should not presume that a large enough percentage of private users have the knowledge, skills, and motivation to properly understand data consequences. To ensure that users always have access to and control over their private data, the practice of using their private data should be prohibited. Regulators and data platform developers need to provide more direction on this.
- **Data Flowing Free**: Data is far from moving freely, despite the fact that we refer to it as the fifth European freedom. Even more so in an AI setting, legal issues relating to data ownership, access, portability, and retention continue to be urgent subjects of discussion. The creation of new business models and the usage of data in AI are hampered by outmoded legislation (such as database rights). A Data Marketplace environment makes it challenging to handle the issue of data ownership because it is difficult to be legally identifies. The idea of data sovereignty can be a solution to confidentiality and security requirements in the absence of a “GDPR for non-personal data”, but it also presents operational difficulties.
- **Preservation of Privacy**: Open innovation is being fueled in various ways by public blockchains and open data projects. It is important to carefully consider privacy protection in light of national and European legal compliance in this era of openness.

- **Compliance to Regulation**: How to be compliant, when, where, and which regulation comes into effect, as well as how to gather knowledge on implementing the regulation are still issues that data-driven SMEs and businesses that seek to create data platforms must address.

3.2.4 National & Regional Challenges

Since the European Commission can make changes to its policies and laws, industry and academia adopt disruptive technologies far more quickly than member states. These issues need to be at the top of the political agenda in the context of an emerging data economy made possible by the convergence of digital technologies:

- **Skills related with Workforce**: Public organizations are having a hard time keeping up with the rapid development of digital technologies. At the same time, it might be challenging to predict the additional knowledge and abilities that public organizations will require. Budget concerns, which may not be high on the public agenda, are also related to organizational and individual skill development.
- **Resistance to Change**: Processes will be transformed by digitization, and data and AI will help society learn more about itself. The work profiles of the workforce change when the organization undergoes transformation. Roles will shift, requiring re- and up-skilling and disrupting employment.
- **Evaluation of Investment**: Public organizations serve the public, including both industry and citizens. Governments are driven to identify new services based on data by the ongoing need to increase efficiency and effect. Investment decisions in development, however, are challenging to be made and are viewed as dangerous. Consequently, the issue is to assess investment in data-centric businesses and guarantee that the economic outcomes have an influence on society.
- **Policies being EU-Wide**: It is difficult to move from regional innovation initiatives to EU-level comparisons. Data can be used to assess the effects of innovation programs, but due to regional differences in needs, comparisons between regions are challenging.
- **Compliance to Policies**: Additional infrastructure expenditures and data certification from devices like Internet of Things (IoT) appliances and edge nodes are needed to make real-time, data-driven policy compliance verification possible.

4 FAME Value Proposition

Current Section includes the overall value of the FAME project. In more detail, it initiates by depicting the status of related activities concluding to the vision and the capabilities offered through FAME. A high-level overview of the project is introduced, explaining how the overall FAME contributions are going to target the introduced challenges of Section 3. Based on this input, both the business value and the technical value of FAME is provided, introducing its potential stakeholders as well as the offered opportunities (i.e., added-value) for them.

4.1 FAME in a Nutshell

Considering the current status of Data Spaces and their cross-cutting challenges (analyzed in Section 3), Data Spaces' stakeholders are not able to easily access, manage, share, trade, or exchange data assets of their interest. Most of the times, they must follow complicated, unstructured, and energy-hungry roadmaps towards extracting useful information, usually leading to dead ends, increased levels of complexity, and "spaghetti junctions", in which even simple functionalities such as data rights management mechanisms, must vary for each platform and interacting party. What can be also seen as a blocking issue are the bottlenecks of identifying fairly priced, interlinked, and even correlated data assets, often related to different tools, models, services, and applications that have been constructed and implemented from diverse user groups with different levels of technical knowledge and design requirements. Even in the case that such tasks are successfully accomplished, most of the times this leads to complex, energy-harvesting, and time-consuming processes requiring multi-aspect data engineering efforts and skillset.

FAME's vision is to act as a single-entry point of low complexity, offering the toolsets to gain valuable insights in the EmFi application domain, by exploiting cutting-edge technologies (e.g., AI/XAI, ML/DL, blockchain). In essence, FAME will represent a practical and mature Federated Data Space that will be accessible through a single-entry point to facilitate finance and non-finance organizations, Data Marketplaces/Data Spaces, and single users (being thoroughly described in Section 4) towards securely accessing, sharing, trading, and analysing their data in a federated manner. Inside this ecosystem, both data privacy and security, as well as environmental and carbon footprint aspects will be continuously monitored, and related actions will be triggered on time. Such concept is synopsized in Figure 3, outlining the FAME supported technological aspects, the beneficiaries (i.e., demand side) of its overall ecosystem, and the data assets that can be exploited (i.e., supply side).

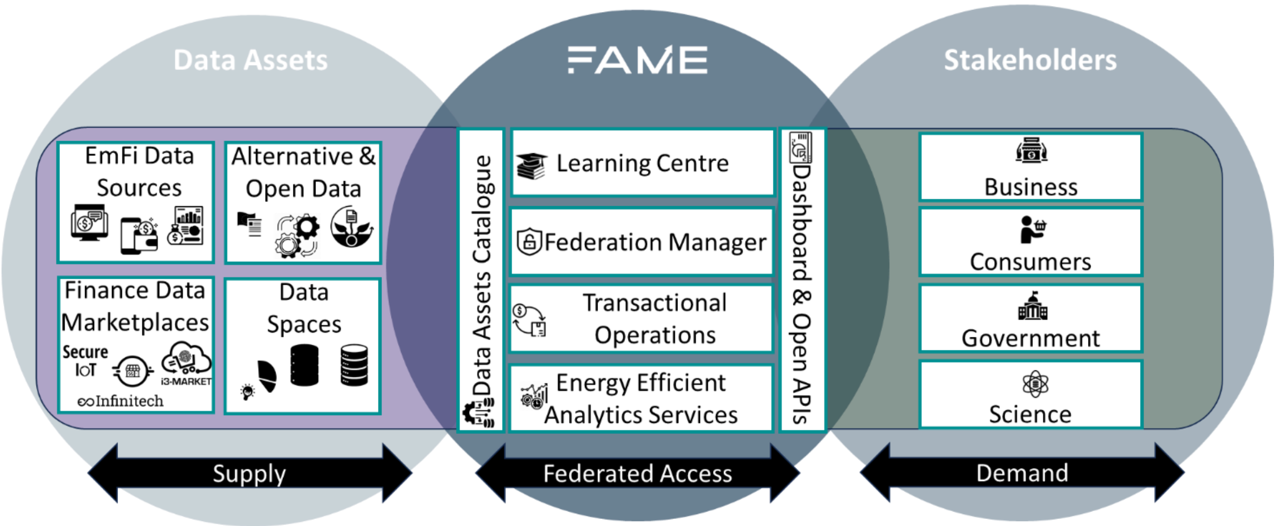


Figure 3 – FAME in a Nutshell

More specifically, as introduced above, FAME is a joint effort of high expertise in data management, data technologies, data analytics, data economy and digital finance to develop, deploy and launch to the global market a unique, trustworthy, energy efficient, and secure Federated Data Space for EmFi application domain. The core of the FAME project is a Federated Data Assets Catalogue that alleviates the proclaimed challenges and limitations of Section 3, towards demonstrating the full potential of the data economy. In this direction, FAME enhances state-of-the-art Data Spaces' infrastructures by constructing a Federated Data Space that supports decentralized and programmable data assets' trading and monetization, offering at the same time the capability to apply trustworthy and energy efficient analytics upon those data assets. Towards this direction, FAME provides novel functionalities in three (3) core directions: (i) Secure, interoperable and regulatory compliant data exchange across multiple federated cloud-based data providers in-line with emerging EU cloud initiatives, (ii) Decentralized and programmable data assets' trading and pricing leveraging blockchain tokenization techniques, and (iii) Integration of trusted and energy efficient analytics based on innovative technologies such as XAI to gain higher trust in data analytics, Situation Aware Explainability (SAX) to offer sound and contextually enriched explainability of data analytics, incremental energy efficient analytics, as well as power-efficient edge analytics.

Through the latter core directions, all the challenges (Section 3) are considered and addressed through the implementation of the following FAME components (furtherly analyzed in Section 7), which represent the fundamental building blocks of FAME's overall functionality:

- **Authorization & Authentication**: It will address the challenges of "Security" through providing the possibility of securely granting access to an end-user, acting as a single-entry point to the FAME Federated Data Space.
- **Federation of External Sources**: It will address the challenges of "Sharing by Design", "Global Competition", "Dynamic Ecosystems", and "Digital Transformation" through giving the ability to integrate with external data sources (i.e., Data Spaces, Data Marketplaces, Single Databases, etc.) and sharing their content. At the same time, it will address the challenges of "Free-Flowing Data", "EU Values" and "Global Competition" through offering the possibility to properly index FAME's assets with specific identifiers, enabling the FAIRness of the data.
- **Assets Policy Management**: It will address the challenges of "EU Values", "Data Protection", "Regulatory Compliance", "EU-Wide Policies", and "Policy Compliance" through ensuring the secure management of the FAME indexed data assets, regulating who can view and have access to them.
- **Assets Provenance & Tracing**: It will address the challenges of "Data Sovereignty", "Veracity", "Privacy Protection", "Trust", and "Privacy Preservation" through providing metadata that identify the nature, the meaning, and the provenance of the FAME data assets, ensuring their authenticity and integrity.
- **Assets Pricing**: It will address the challenges of "Valuation Standards" through properly calculating FAME data assets pricing and/or trading costs, considering a plethora of dynamic and non-dynamic pricing schemas and variables.
- **Assets Trading & Monetization**: It will address the challenges of "Decentralization", and "Valuation Standards" through giving the possibility to the FAME end-users to navigate and trade the preferred FAME data assets.
- **Assets Searching**: It will address the challenges of "EU Values", "Global Competition", and "Dynamic Ecosystems" through giving the opportunity to the FAME end-users to perform data queries upon the existing data assets, following specific semantic models.

- **ML/AI Analytics**: It will address the challenges of "Free-Flowing Data", "EU Values" and "Global Competition" providing the ability to the FAME end-users to deploy and deploy and execute specific ML/AI models for their preferred scenarios.
- **SAX Analytics**: It will address the challenges of "Dynamic Ecosystems", and "Trust" offering the ability to the FAME end-users to perform trustworthy and explainable analytics to the involved business processes, the AI/ML models and the analytics' results.
- **Analytics CO_2 Monitoring**: It will address the challenges of "EU Values", and "EU-Wide Policies" through providing the estimation of energy usage and consumption of the produced analytics' models.
- **FML Deployment**: It will address the challenges of "Dynamic Ecosystems" through giving to the FAME end-users the opportunity for data sharing in federated learning scenarios.
- **EmFi Training**: It will address the challenges of "Dynamic Skills", "Workforce Skills", "Resistance to Change", and "Investment Evaluation" through providing the proper users' skillsets training upon both the FAME Federated Data Space and its overall functionalities and the guidelines and principles of Data Spaces (focusing on EmFi applications), by offering a plethora of dedicated training programs and a related learning centre.
- **Dashboard**: It will address the challenges of "Free-Flowing Data", "EU Values", and "Global Competition" through offering an end-to-end user-friendly integrated User Interface (UI) as the means to access and use all the offered FAME services and functionalities.

4.2 FAME Business Value

The fundamental idea behind the business value of FAME is to realize a **Federated Data Space of different type of data assets** (i.e., data, algorithms, tutorials, etc.), which provides specialized functionalities to enable the discovery and utilization of data assets and technology components to different entities (i.e., data providers, data consumers, application owners, etc.) to support the development and rollout of **data-driven EmFi applications**.

Hence, FAME is an enabler for a **Data Space** in the **financial sector** helping business organizations, system integrators, data scientists, etc. in the Financial, Insurance and Industry sectors to leverage the development and rollouts of EmFi Applications providing full data governance management (i.e., sovereignty, privacy, security, trustworthiness, etc.) according to European values and regulations.

At the same time, FAME works as a **multisided Data Marketplace** that bridges consumers and providers of assets providing tools for exchanging assets (datasets, technology components, tutorial materials and more) with embedded (by design) methods for trading, security, and trust.

Entities federated in FAME are **identifiable, sovereign, secure, trustworthy operating within a unique ecosystem** that minimizes the risks to exchange data with uncertain provenance, reliability, security, in violation with EU regulations. Therefore, interactions among FAME federated users are by design to insure sovereignty, privacy, security, trustworthiness over the data.

To achieve the abovementioned, FAME introduces state-of-the-art **secure, intelligent, environmentally friendly, computationally efficient tools** with the scope to minimize the costs of development and rollout of EmFi applications. It also provides tools and services as well for **machine to machine (M2M)** interactions to enable fully automated value chains.

Via its federated mechanism, FAME provides **core functionalities for financial services** and also runtime applications to support **development and production** environments.

FAME is also used by **private citizens, research institutions and public organisations**, as well as by **corporate and small businesses (FinTech-InsuranceTech)** to discover and trade assets with high specialization in different sectors.

Utilization **examples are innumerable** and of great value for the sector: from banks-insurance-score ranking companies that want to share customers' risks profiles sharing only the relevant information without violating the privacy of the customer, to personalize insurance products from high-value environmental data.

Thus, the **FAME Federation concept** is an emerging area in finance, that will transform current FinTech and InsuranceTech sectors and other organizations that want to provide Embedded Finance applications in an efficient, cost-effective and EU values and regulations compliant way to contribute to the European Data Economy.

Finally, FAME is inspired and follows the **European data strategy**, which aims to make EU a leader in a data-driven economy and society. FAME shares the EU vision of creating a **single data market** and aims to contribute to the development of specific Data Spaces adopting technology concepts from other EU projects (IDSA and Gaia-X among others). The key principle is to shift from a monolithic and centralized way of accessing/sharing data to a federated and decentralized way, favouring data governance in-line with EU principles of trust and sovereignty.

Summarizing, FAME will contribute to the European data strategy since:

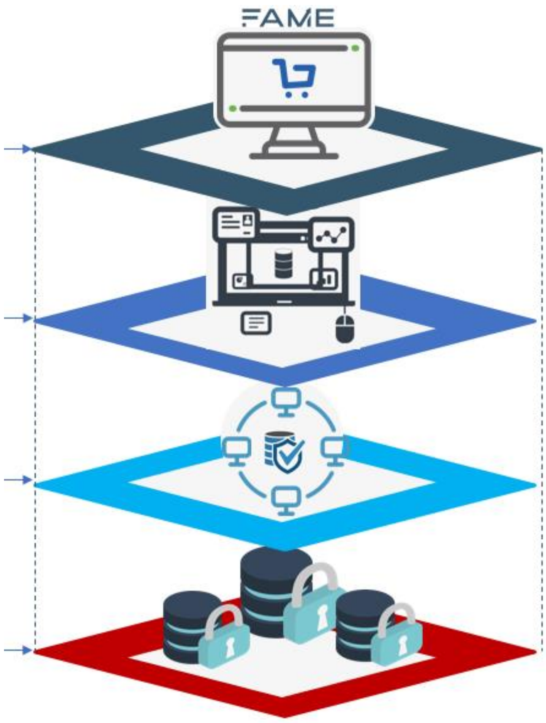
FAME **Federated Data Space** will be a **functional platform to federate data and technology assets** for EmFi data-driven applications in compliance with the EU data strategy and regulations.

FAME will define a new **technology and services stack** based on IT solutions, state-of-the-art methodologies, technologies, and components to leverage the costs of application **design, development, and execution of EmFi.**

FAME will provide **tools** to **onboard, discover and manage assets** for both static and dynamic usage to **diverse sectors of stakeholders** (business, consumers, science, government).

Towards this direction, the **FAME Federation Stack** will provide:

- **Cost saving ecosystem to develop EmFi Applications** for showcasing the pricing, trading, and monetization of datasets, ML/AI models, training courses, etc.
- **Integration of novel technologies** such as SAX, Incremental Energy Efficient Analytics, Assets' Tokenization, etc.
- **Secure, interoperable, and regulatory compliant data assets exchange** across multiple federated asset providers in-line with emerging European initiatives.
- **Federation of Assets Producers/Consumers** for integrating Data Spaces and Data Marketplaces, enabling core functionalities for datasets and technologies, with distributed, federated and interoperability models.



4.3 FAME Stakeholders

As indicated in the previous sub-Section, major industrial players, national and European legislative bodies, and other important stakeholders have recently shown a lot of interest in:

- Aligning and integrating proven data sharing technologies and solutions, preventing the need to reinvent the wheel and promoting scale.
- Eliminating data silos that can be broken down using architectures, standards, protocols, and governance models that prioritize protecting personal data, enabling distributed and decentralized solutions, and facilitating fair and secure data sharing and exchange.
- Adopting business models that allow participating stakeholders, including local, national, and European authorities and institutions, research entities, and even private individuals, to exploit the value of data assets bilaterally or multilaterally.
- Promoting and accelerating the use of data technology and the data economy in industries with non-data-driven business models.
- Making use of already-existing pan-European programs and networks to enable data analytics across a European data sharing ecosystem, which includes research institutions, businesses, government, and international organizations.

When speaking about Data Spaces, such stakeholders may be categorized into the roles of: (i) *data providers and consumers*, (ii) *technology providers*, (iii) *intermediaries*, (iv) *data space operators*. The participants who provide and interact with data are known as *data providers and consumers*. These participants include the data producer who creates the data, the data owner who has the rights to access and use the data, the data acquirer or provider who collects the data and makes it available via the Data Space catalog, the data consumer who accesses the data from the catalog, and the application provider who creates the applications. The *intermediaries* are independent companies that provide the publishing, resource-finding, and transaction-registration services, with vocabularies, ontologies, and metadata broker services being a few examples of the services provided by them. For a Data Space to function properly and provide a secure and reliable environment, the *technology providers* offer components like connectors, user management systems, or monitoring systems. To achieve this, the *data space operators* concentrate on managing the space, by handling activities like management of requests, software maintenance, etc. They oversee the administration of the Data Space, certify players, and establish the functional plan, among other things. However, it should be noted that all these roles are not exclusive, and the same user can adopt several roles.

To this context, FAME supports the interaction and integration with all the aforementioned roles, and more particularly with external data EmFi applications, Data Marketplaces/Data Spaces, and single users, all of whom may act as data providers/consumers, technology providers, intermediaries, and data space operators, depending on each different scenario. It should be mentioned that *application owners*, *data assets providers/data marketplaces/data spaces*, as well as *data/content consumers* can be equally considered as key players of each of the abovementioned roles. In this context, all the parties can contribute their own data assets or access the FAME's existing ones, also gaining access to all the platform's functionalities via open Application Programming Interface (APIs).

All of these parties can be further grouped into four (4) main categories: ***business*** (industry), ***consumers*** (citizens and non-tech users), ***science*** (research and academia), and ***government*** (local, national, and European governments, and public bodies). Nevertheless, *data providers/consumers*, *technology providers*, *intermediaries*, and *data space operators* can all equally belong to each of the aforementioned categories, for whom a plethora of opportunities can be identified to have a broader socioeconomic value. The overall involved stakeholders are depicted in Figure 4.

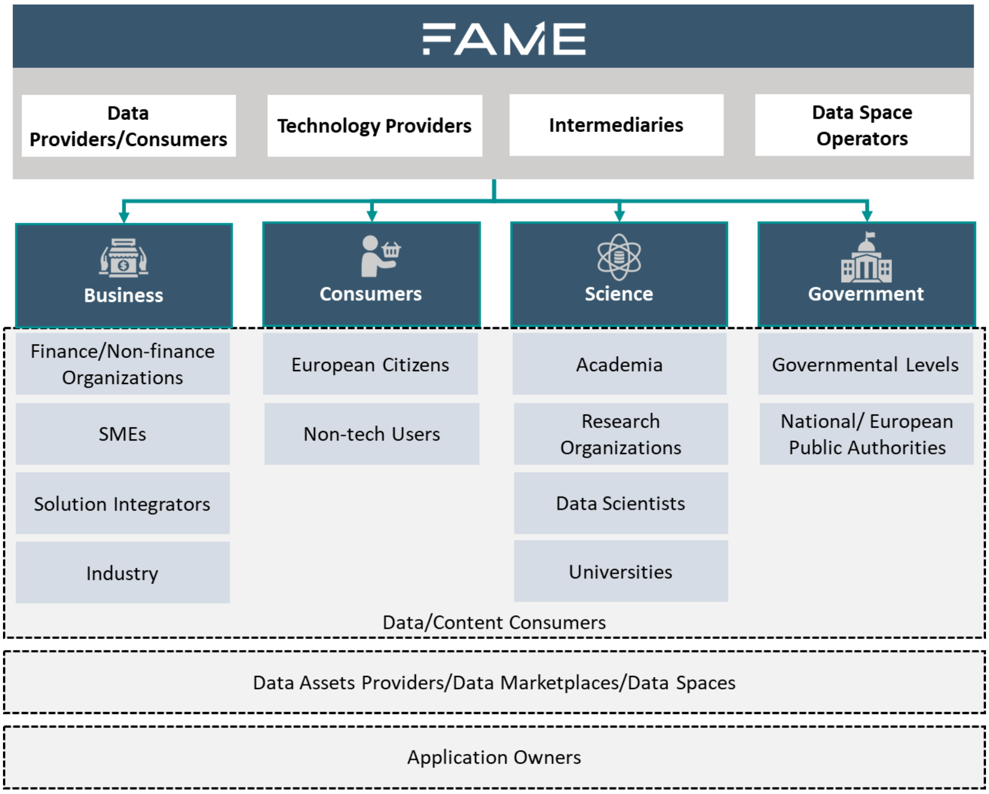


Figure 4 – FAME stakeholders

Concerning the opportunities that have been identified for the roles of Figure 4, these can be listed as follows:

Opportunities for *Business* Stakeholders: Organizations in the finance and non-finance sectors, as well as SMEs and major businesses, may be considered business stakeholders, as well as solution integrators, industrial players, application owners, data providers, and training content consumers, who may benefit from the following growing prospects:

- **Open data marketplaces towards industrial data trading**: With assurances that the producers maintain data sovereignty and are compensated, industrial data can be shared both inside and outside of a value network.
- **Greater accessibility of massive, diverse data ecosystems for AI**: The potential of AI added-value must be fully realized, especially in crucial sectors including business services, manufacturing, wholesale, retail, and infrastructure providers (such as 5G operators).
- **Cutting-edge business concepts powered by data and new value ecosystems**: Moving from "data for business intelligence" to "data for AI" also entails shifting from internal processes to cross-domain ecosystems that are more collaborative and participatory.
- **Personal data being safe**: Following explicit consent and true anonymization, personal data will be more frequently considered for cross-sectoral applications. Cross-sectoral services will keep developing, driven by the usage of digital services by private clients.

Opportunities for *Consumers*: European citizens, referring primarily to non-technical users but also to application owners, data providers, and training content consumers can benefit from easy and secure data sharing in various ways:

- **Personal data controlling**: People will be able to control what information is shared, where it is stored, and who has access to or uses it while maintaining the freedom to change their minds and fully understanding the repercussions of their choices.

- **Cross-sectoral and personalized Business to Consumer (B2C) services**: By bringing production closer to consumers, digitization expands choice and personalisation regardless of geography.
- **Monetization of personal data**: With the introduction of Consumer to Business (C2B) business models, which allow individuals to maintain ownership over their data while earning fair financial or economic benefits directly, new European legislation encourages users to disclose their data.
- **New professions**: Additional innovation will provide new career paths and more jobs, and national and local authorities will continue to fund education, retraining, and upskilling, since they see the long-term importance of these endeavours.

Opportunities for *Science*: Academia, referring to research organizations, data scientists, and universities, as well as application owners, data providers, and training content consumers are expected to benefit from the following opportunities:

- **Broader and more significant social impacts of research data**: It will become simpler to find, integrate, or otherwise collaboratively process and analyze data in different scientific disciplines as standards for data and metadata representation, sharing models, licensing, and exchange protocols converge.
- **Enhancing open innovation and science through data being available**: Researchers with access to external data will find it simpler to participate in open innovation initiatives, increasing innovation in businesses. Academics may create and execute more difficult challenges due to the increased data accessibility, which enhances scientific crowdsourcing to advance science, whereas it finds answers that will benefit industry.
- **Revenue prospects provided by new data-driven business models**: By allowing the controlled interchange and monetization of research data from stakeholders, it will open up new business potential for academic institutions for integrating and analyzing various data.

Opportunities for *Government*: All governmental levels, as well as national and European public authorities, also considering application owners, data providers, and training content consumers will profit from the following prospects:

- **Using data commons to improve governmental services**: Opening public domain datasets and systems presents potential for developing new services or improving the existing ones, hence enhancing accessibility, and streamlining e-Services.
- **Digital services enhancement by AI**: It will assist in foreseeing and analyzing national and European data in a moral and privacy-preserving way.
- **EU Statistics being real-time**: Real-time monitoring across important industries can be provided at the national and EU levels by an integrated European data sharing space.
- **Access to government services promotes a lean corporate environment**: For more efficient business planning, public governmental services can relate to industrial data.
- **Policymaking being evidence-based**: It will enable decision-makers and governmental organizations to work with private actors to improve and hasten policy cycles and investigate new areas of policymaking in a data-driven manner.
- **Data as proof of policy coherence**: The accessible data can be used as proof to confirm whether certain policy-related requirements implied by European-wide regulations and policies have been satisfied. This may open new roadmaps for the delivery of organizational procedure certifications.

5 FAME Technical Background

Current Section provides all the necessary information regarding existing RAs, principles, and functionalities that will be exploited in the context of the FAME SA. In more detail, RAs related with Big Data, Data Marketplaces and Data Spaces related projects, as well as activities and design principles are introduced, depicting an overview of their internal architecture, and their capabilities towards facilitating the construction of FAME. The relevance of these RAs and their added-value to FAME SA is furtherly described, concluding to the FAME related technical advancements.

5.1 Significance of Reference Architectures

The architectural design of many modern software systems, such as Data Applications, Data Platforms, and Data Spaces, has grown increasingly difficult due to the size and the complexity of these systems. In this situation, RAs have been demonstrated to be highly pertinent to support the architectural design of systems in numerous essential application domains, including but not limited to, health, avionics, transportation, agriculture, and finance.

Towards this direction, the primary goal of a RA is to provide a common vocabulary, reusable designs, and best practices that are used as a guidance for more concrete software (e.g., systems, platforms) architectures in a specific domain (e.g., healthcare, smart cities, agriculture, finance). Typically, a RA includes *common architecture principles, patterns, building blocks, and standards*, outlining the *components needed to compose a system, the externally visible properties of those components, and the relationships among them*. In essence a RA is not a solution architecture (i.e., they are not implemented directly), but primarily intended to provide a methodology and/or set of practices and templates that are based on the generalization of a set of successful solutions for a particular category of solutions. Hence, a RA provides guidance on how to apply specific patterns and/or practices to solve particular problems. It acts as a “reference” for the particular architectures that businesses will use to address their own problems in this way. A RA is never meant to be implemented as-is; rather, it should be utilized as a benchmark or a place to start for the architectural efforts of different organizations. Thus, a <u>RA should be technology and domain-independent, abstract, and flexible</u>, being defined at various levels of detail, from high level principles to detailed implementation guides [10][11].

Summarizing, a RA can serve as a t*emplate that can be used to create a specific architecture (i.e., solution architecture) for a software*. What is worth mentioning is that it can be used to **develop either standard or custom applications or even scalable system solutions** (i.e., solutions scaling up or down as needed the RA’s provided concepts).

By adopting a RA within an organization, the latter is provided with a structured approach to design and deploy software systems, helping to ensure that the final product meets the desired requirements. By leveraging RAs, organizations accelerate delivery through the re-use of an effective solution, thus reducing time to market, improving quality, and increasing efficiency. On top, organizations can be benefitted since a RA provides them with: (i) provision of *a frame of reference* that helps them to get an overview of a particular domain, being provided by a starting point, (ii) systematic *reuse of common functionalities and configurations* throughout the development of their systems, (iii) *risk reduction* through the use of proven and partly qualified architectural elements included in the RA, (iv) enhanced quality by facilitating the achievement of software quality aspects already addressed by the RA, (v) *interoperability* among different systems and their software components establishing common means for information exchange, and (vi) *regulatory compliance* accounting principles, practices, and processes that are already in place.

However, to obtain such benefits, these architectures should be suitably described (i.e., represented/modelled) aiming at reliably communicating the knowledge that they contained [12][13].

Nowadays, there exists a variety of different types of RAs, deriving from diverse application domains [14], all of them however sharing a common goal: to provide a starting point for organizations that need to solve a particular problem. Such architectures have been described in many different approaches, such as using textual description, informal models, modelling languages (e.g., Unified Modelling Language (UML)), and in widely adopted system architecture models (e.g., 4+1 model, C4 model), as in the case of the FAME SA (described in Section 7).

The following sub-Sections provide a list of state-of-the-art RAs that have been considered into the realization of the FAME SA, taking into consideration the technical experience and knowhow of the consortium partners in different domains.

5.2 Overview of Reference Architectures for Data Marketplaces

5.2.1 i3-Market

The i3-MARKET project [15] addresses the growing demand for a single European Data Market, by innovating marketplace platforms, demonstrating with industrial implementations that the data economy growth is possible. i3-MARKET aims at providing technologies for trustworthy (secure and reliable), data-driven collaboration and federation of existing and new future marketplace platforms, with special attention on industrial data. The i3-MARKET architecture is designed to enable secure and privacy preserving data sharing across data spaces and marketplaces, through the deployment of a Backplane across operational data marketplaces. Hence, i3-MARKET is not trying to create another new Marketplace, but to implement the Backplane solutions that allow other Data Marketplaces and Data Spaces to expand their market, facilitate the registration and discovery of data assets, facilitate the trading and sharing of data assets among providers, consumers, and owners, whereas also providing tools to add functionalities they lack for better data sharing and trading processes. Towards this direction, the i3-MARKET project has built a blue-print open-source software architecture called "*i3-MARKET Backplane*" [16] that addresses the growing demand for connecting multiple Data Spaces and Data Marketplaces in a secure and federated manner.

5.2.1.1 Architecture Overview

The overall architecture defines all the required components and modules, their basic functionality and behaviour, as well as their interfaces and interaction patterns in accordance with the user stories and the requirements specified in the project. In particular, the high-level architecture covers:

- The i3-MARKET Backplane solutions with its core functionalities.
- The interaction of existing Data Spaces and Data Marketplaces with the i3-MARKET Backplane and each other (for secure data access) based on open interfaces.
- The engagement of data providers, consumers, owners via smart wallets and applications, and the interactions with the i3-MARKET Backplane for the sake of privacy preservation and access control to their personal or industrial data assets.

In deeper detail, the i3-MARKET architecture has been designed following the 4+1 architectural view model [17], which is a standard model, commonly used for documenting software architectures. One of the major views of this model is the logical view that shows the functionality that the system provides to the end-users, having a twofold objective. On the one hand, this view shows the i3-MARKET system (green box in Figure 5) and the link between the stakeholders and the Data Marketplaces, while on the other hand, the logical view pursues to show the internal decomposition of the i3-MARKET system into the logical subsystems and components that implement the i3-MARKET Backplane API and the Secure Data Access API (SDA API).

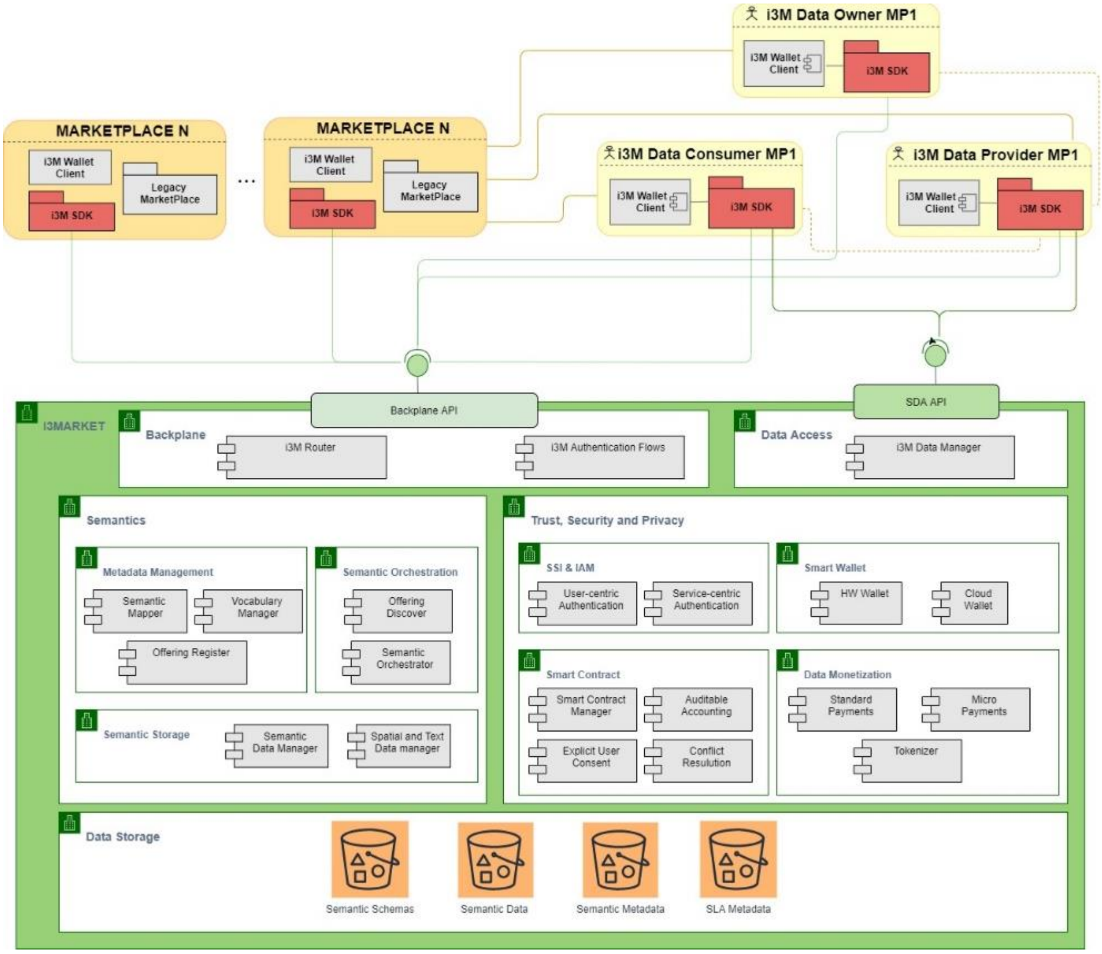


Figure 5 – Logical View with i3-MARKET

In general terms, i3-MARKET supports the actors with the i3-MARKET Backplane's functionality by means of the two (2) following main entry points:

¶ The *Backplane API* and the *SDA API* (depicted as green lines in Figure 5), or in other words, the direct access to the i3-MARKET Backplane. These two (2) APIs enable access to all the integrated building blocks. This is the use case of these actors that follow a more ad-hoc integration with i3-MARKET.

¶ The *i3-MARKET SDK (i3M SDK)* (depicted as orange boxes in Figure 5), supports the end-users' developers with the integration of the Backplane API and the SDA API. This product is intended for the actors that pursue a more "assisted" support.

Regarding the link with the stakeholders and marketplaces, in the case of the Data Marketplaces' actors, i3-MARKET assists them with a full version of the Backplane API and the i3M SDK (Backplane module), which gives support for interacting with the Backplane API. In the case of the data owners, data providers and data consumers, the normal operating mode is the access to i3-MARKET Backplane through their own Data Marketplace. However, for some particular Data Marketplaces' cases, data owners, data providers and data consumers, they may be able to directly interface with the i3-MARKET system through the available SDKs and APIs.

To guarantee the authentication mechanisms proposed by i3-MARKET, a Wallet Client should be installed into the end-user side to store the user private keys.

Apart from the i3-MARKET Backplane's offered entry points, the Backplane mostly includes a set of semi-independent subsystems with self-contained functionalities that will be furtherly discussed below. Most of these subsystems have broken down their functionalities into atomic and loosely coupled sub-components exposing their functionality through a REpresentational State Transfer (REST) API, which yields a microservices-based nature [18] to the i3-MARKET system.

Figure 6 shows a detailed landscape of the current set of microservices (cubes), APIs (little yellow rectangles), components (blue rectangles), and storages (white rectangles) on i3-MARKET. Each arrow in the figure denotes a dependency between the subject and the object involved in the arrow. Finally, remark that the Remote Procedure Call (RPC) Distributed Ledger is one and single instance, but it has been put as several instances for picture legibility.

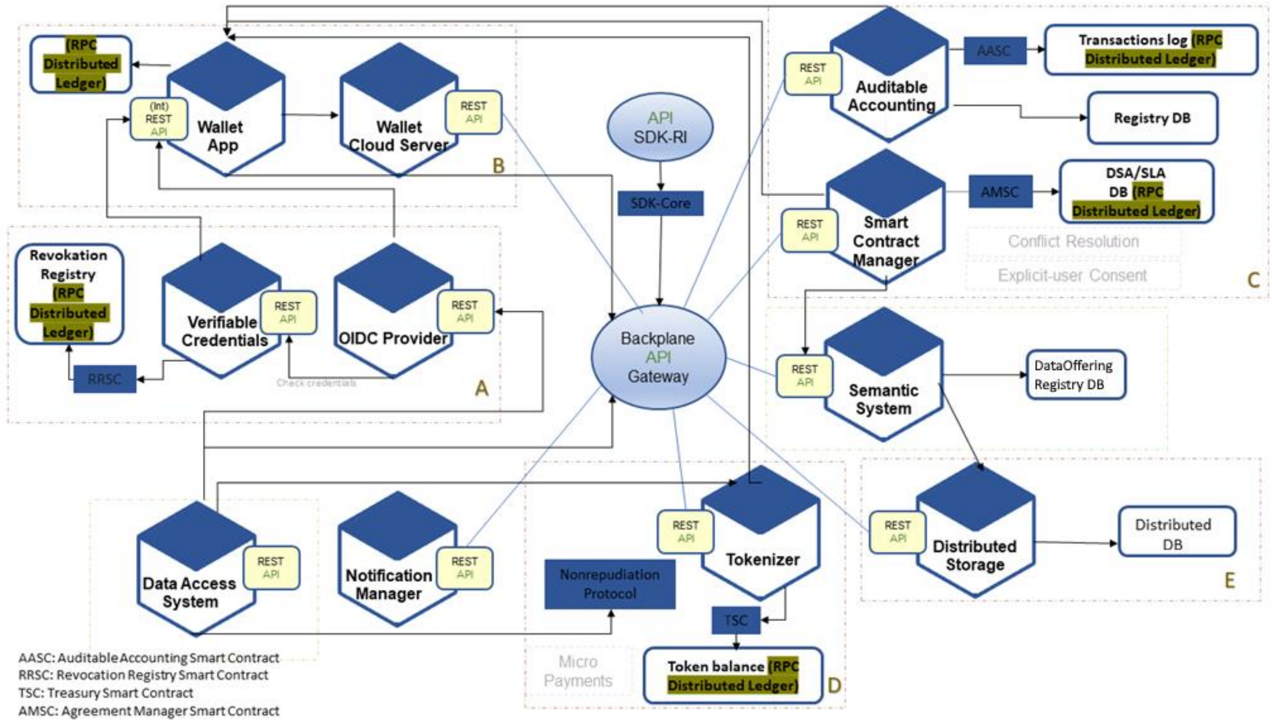


Figure 6 – i3-MARKET microservices layout

As Figure 6 shows, there exist the following dependencies among the i3-MARKET components:

¶ *SDK System*

o *SDK-core* libraries for making easier the development of applications that make use of the Backplane API. It interfaces with the Backplane Gateway.

o *SDK-RI* common pilots-driven complex workflows based on the Backplane services. It interfaces with the SDK-core library.

¶ *Trust, Security and Privacy System*

o *SSI and IAM Subsystem (label A)*

Á *User-centric Authentication*: A component for providing the management of Self-Sovereign Identity based on DID and VC and the compatibility with OIDC standard, being supported by the "Verifiable Credential" microservice. It also interfaces with the Wallet because the Verifiable Credential assumes that the user created and controls with his/her crypto wallet their identities and with the RPC Ledger Storage for updating the revocation of credential.

Á *Service-centric Authentication*: A component for providing the authentication and authorization of the end-users with the standard OIDC/OAuth flows, integrating the User-centric Authentication component. This functionality is supported by the "OIDC Provider" microservice that implements the OIDC compatibility (based on verifiable credential). It interfaces with the Verifiable Credential for allowing the token creation based on the verifiable credentials and with the Wallet for sending the credentials.

- *Smart Wallet Subsystem (label B)*
  - *Wallet APP*: An application for storing user's private keys, which interfaces on the RPC Ledger Storage.
- *Smart Contract Subsystem (label C)*
  - *Smart Contract Manager*: A component/microservice for providing a gateway to access the Smart Contracts, being conceived mainly for managing the SLA and DSA Agreements Smart Contract (business smart contracts). It interfaces with the RPC Ledger Storage for storing the Data Sharing Agreement object and the Semantic Engine for crating data purchase.
  - *Auditable Accounting*: A component/microservice for capturing logging and auditing interactions between the components, recording the registries in the blockchain. It also interfaces with the RPC Ledger Storage for registering auditable data.
- *Data Monetization Subsystem (label D)*
  - *Non-Repudiation Protocol*: A library that interfaces with the Backplane API for interacting with auditable accounting.

- *Semantic System*
  - *Semantic System*: A service for managing the assets' discovery and semantic data model in the i3-MARKET. It interfaces with Contract Manager managing contractual parameters, it depends on the registry DB store and the Distributed DB service's API. It might interact with ledger for Verifiable Credentials and DID IDs, and it interacts with the Notification manager Service for reporting new data assets.
- *Data Access System*
  - *Data Access*: A service for providing the means for allowing the transfer of data between the data provider and the data consumer. It interacts with the Non-Repudiation Protocol library and Backplane API for enforcing smart contract.
- *Storage System*
  - *Distributed Storage Subsystem (label E)*
    - *Distributed Storage*: A component/microservice for storing i3-MARKET assets' index.
- *Backplane Gateway*: A component responsible for providing a gateway for all the internal services conforming the Backplane. This gateway is the single-entry point for all the clients.

5.2.1.2 Relevance to FAME & Added-Value

The i3-MARKET project addresses the challenge of being integrative following design methods used in industry and OSS implementation best practices, interoperable by using semantic models that define a common conceptual framework and information model that enables cross-domain data exchange and sharing. It is intelligent from the perspective of smart contracts generated automatically and associating those financial operations into a set of software tools that facilitate that data assets can be commercialised via intra-domain or cross-domain almost transparently in a secure and protected digital market environment. As stated above, the i3-MARKET has provided its software in the form of a backplane with a set of software support tools and as public available open-source solutions addressing the challenge of enabling the coexistence of data spaces with data marketplaces for enlarging the European Digital Market Ecosystem. Towards this direction, the i3-MARKET contributions to FAME's objectives are envisioned to be:

User-Centric Authentication & Authorization Infrastructure [**Obj 2**]
The FAME Authentication & Authorization Infrastructure will leverage on IDM Authentication and Authorization based on self-sovereign identities, OIDC and OpenID Connect standards, W3C Decentralized Identifiers (DIDs) and Verifiable Credentials and access control to data, with support of decentralized system Besu, OAuth 2.0 and JSON Web Token (JWT) open standards. Into this context, the FAME Authentication & Authorization Infrastructure can leverage the relevant i3-MARKET technologies already put in place, referring to its User-Centric Authentication & Authorization solutions, which will provide baseline support for self-sovereign identities and DIDs. The latter essentially are a new type of identifiers that enable verifiable, decentralized digital identity, OpenID Connect (OIDC) integrations, Verifiable Credentials (VC) that can represent all of the same information that a physical credential represents, and in addition of technologies, such as digital signatures, makes verifiable credentials more tamper-evident and more trustworthy than their physical counterparts.

Authentication & Authorization Infrastructure Interfacing Multiple Data Providers [**Obj 2**]
The possible interaction and improvements on the i3-MARKET smart wallets' solutions implement functionalities for enabling the Self-Sovereign Identity (SSI) services of FAME.

Decentralized Programmable Value-Based Data Trading and Monetization [**Obj 4**]
The exploitation of the specifications and know/how on i3-MARKET's Smart Contracts management for data sharing agreements contracts and tokenization can aid FAME in the management and compilation of its supported data sharing/trading agreements smart contracts, being aligned with conditions, operations, and token flow processes.

Federated Data Marketplace Platform [**Obj 1**]
Additional i3-MARKET services can be adopted in the context of the FAME implemented marketplace regarding the functionalities of a Notification and Subscription Service, a Pricing Manager Service, as well as a Rating Service. These tools could facilitate the notification of information in the FAME platform and help its stakeholders with suggested prices or providers' ratings based on their provided data assets' characteristics, on their data conditions considering operations and processes of the data flow and in general their provided data assets.

5.2.2 INFINITECH

The INFINITECH project [19] is a joint effort of Europe's leaders in Information and Communications Technology (ICT) and finance/insurance sectors towards providing the technological capabilities, the experimentation facilities (testbeds and sandboxes) and the business models needed to enable European financial organizations, insurance enterprises and FinTech/InsuranceTech innovators to fully leverage the benefits of Big Data, IoT and AI technologies. The latter benefits include a shift towards autonomous (i.e., automated and intelligent) processes that are dynamically adaptable and personalized to end-users' needs, while being compliant to the sector's complex regulatory environment.

5.2.2.1 Architecture Overview

The overall high-level view of the INFINITECH RA is illustrated in Figure 7, providing a generic perspective in order to leave great level of flexibility depending on the implementation.

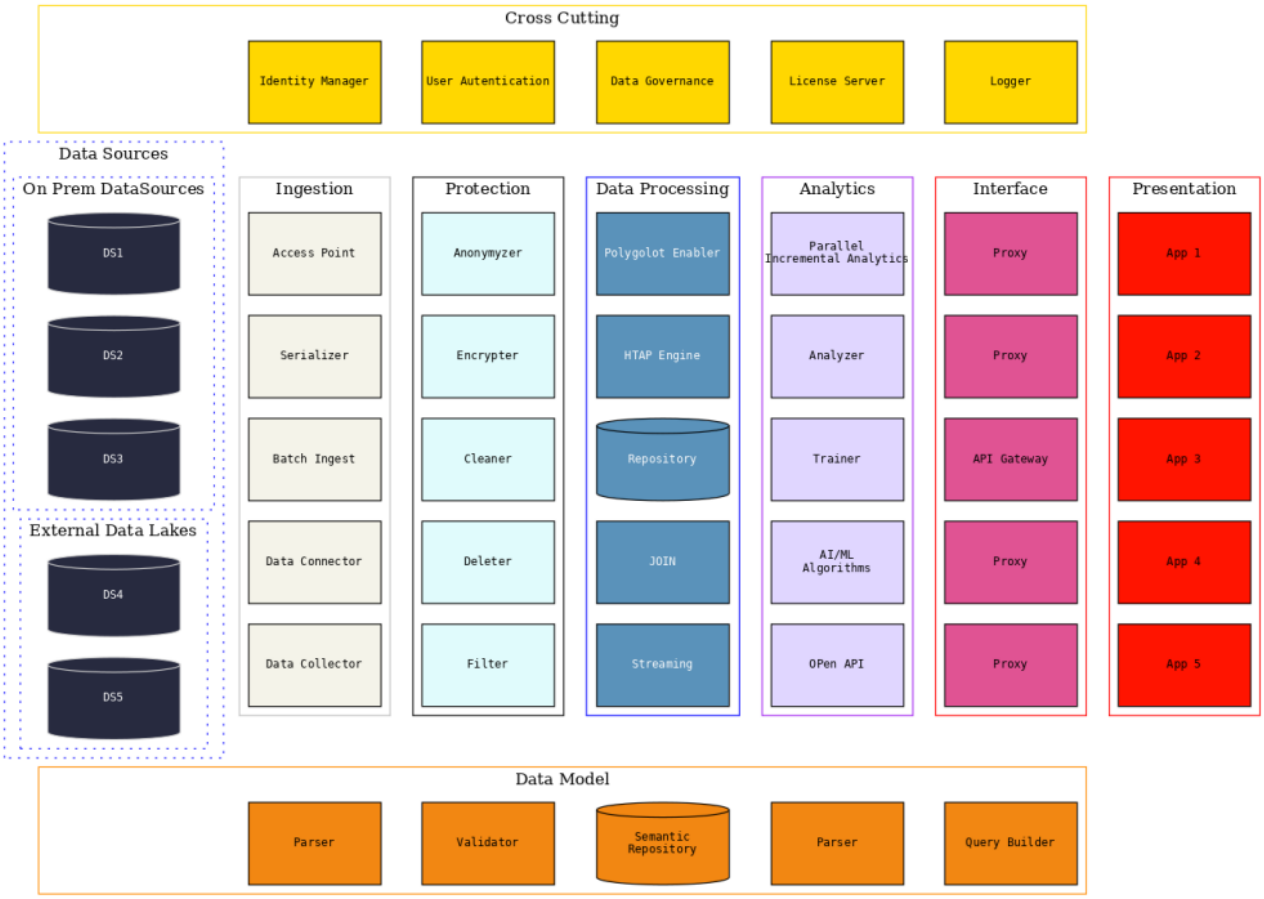


Figure 7 – High level view of Infinitech architecture

In greater detail, the overall architecture supports the following layer/components:

- *Data Sources*: At the infrastructure level there are the sources of data (i.e., database management systems, data lakes holding non-structural data, etc).
- *Ingestion*: A layer of data management usually associated with data import, semantic annotation, and filtering from the data sources.
- *Security*: A layer for management of the clearance of data for security, anonymization, and cleaning of data before any further storing or elaboration.
- *Management*: A layer responsible for the data management aspects, including the persistent storage in the central repository and the data processing enabling advanced functionalities such as Hybrid Transactional and Analytical Processing (HTAP), polyglot capabilities, etc.
- *Analytics*: A layer for the AI/ML/DL components.
- *Interface*: A layer for the definition data to be produced for user interfaces.
- *Cross Cutting*: A layer with service components that provides functionalities orthogonal to the data flows (e.g., Authentication, Authorization, etc.).
- *Data Model*: A cross cutting layer for modelling and semantically annotating the data in the data flow.
- *Presentation/Visualization*: A layer usually associated with the presentation applications (i.e., desktop, mobile apps, dashboards, and the like).

It should be noted that the INFINITECH RA does not impose any pipelined, or sequential composition of nodes. However, it is recommended to consider each different layer and the relative components to solve specific problems of the use case.

5.2.2.2 Relevance to FAME & Added-Value

INFINITECH is a large-scale project on Big Data and AI in digital finance, which has produced dozens of technologies, proven in over fifteen (15) pilots covering the entire spectrum of digital finance. FAME will extend various components of INFINITECH, including the project's blockchain infrastructure for data provenance, the INFINITECH semantic data models and ontologies, as well as specific assets for the marketplace. In particular, the INFINITECH's contribution to FAME's objectives will be:

Federated Data Assets Catalogue (FDAC) based on Machine-Readable Data Models and Ontologies for EmFi [**Obj 3**]

FAME will specify several data models and ontologies for EmFi applications. INFINITECH has demonstrated the benefits of semantics for digital finance based on ontological models like the Financial Industry Business Ontology (FIBO) [20], the Financial Instrument Global Identifier (FIGI) [21], and the Legal Knowledge Interchange Format (LKIF) [22] that will be investigated and considered in the context of FAME.

Decentralized, Configurable, Dynamic, Value-based Data Assets Trading and Monetization Schemes [**Obj 5**]

FAME can learn from the experience and leverage a readily available high performance and energy efficient blockchain infrastructure that supports tokenization, which has been developed in INFINITECH project. FAME will reuse and extend the readily available permissioned blockchain infrastructure of the INFINITECH project, which provides support for ERC-20, ERC-721 and ERC-1155 tokens [23]. Specifically, the INFINITECH blockchain has extended the Hypeledger Fabric blockchain [24] with support for tokenization, including ERC-1155 support.

FAME Integrated Platform [**Obj 6**]

The federated Data Space infrastructure, the permissioned blockchain infrastructure and the trusted and power efficient analytics infrastructure of INFINITECH will be integrated in a single platform. The integration will be driven by FAME's architecture. FAME will also leverage the DevOps and continuous integration approaches of INFINITECH, while enhancing them with DataOps and MLOps techniques (i.e., Kubeflow, MLFlow, Data Version Control) to increase automation and improve the quality of the production models.

5.2.3 FINSEC

The FINSEC project developed the Finsecurity.eu platform [25], which is a single access point to knowledge assets about Critical Infrastructure Protection (CIP), with an emphasis on information about the protection of critical infrastructures of the finance sector. The ambition of the platform is to support the CIP community through easing access to knowledge and innovative solutions. It is used by the European Cluster for Securing Critical Infrastructures (ECSCI) cluster of projects [26] and linked to the knowledge hub of the Horizon Europe Coordination and Support Action (EU-CIP CSA) [27]. The platform comprises several data assets and content items, including whitepapers, description of CIP and cybersecurity solutions, webinars, courses, blog articles and more. It should be noted that access to the various assets and content items is generally public, yet some assets are only available to registered users of the platform that are authenticated and authorized via a username/password mechanism.

5.2.3.1 Architecture Overview

The technical architecture of the platform is illustrated in Figure 8. It is a classical three-tier/multi-tier architecture that comprises the following tiers/layers:

- The *Presentation Tier*, which comprises the front-end of the platform. This tier ensures the graphical representation and presentation of content items to the end-users.
- The *Application Tier*, which comprises the backend application logic of the platform. It is in charge of processing end-users' requests and generating user-specific content in-line with the authorizations of the end-user.
- The *Database Tier*, which manages the data assets and content items of the platform.

The platform also interfaces to other external platforms and services (e.g., to the FINSEC CIP platform), which enable the integration of demonstrator applications. These demonstrator applications are managed outside the Finsecurity.eu platform, yet visualized and presented inside the platform.

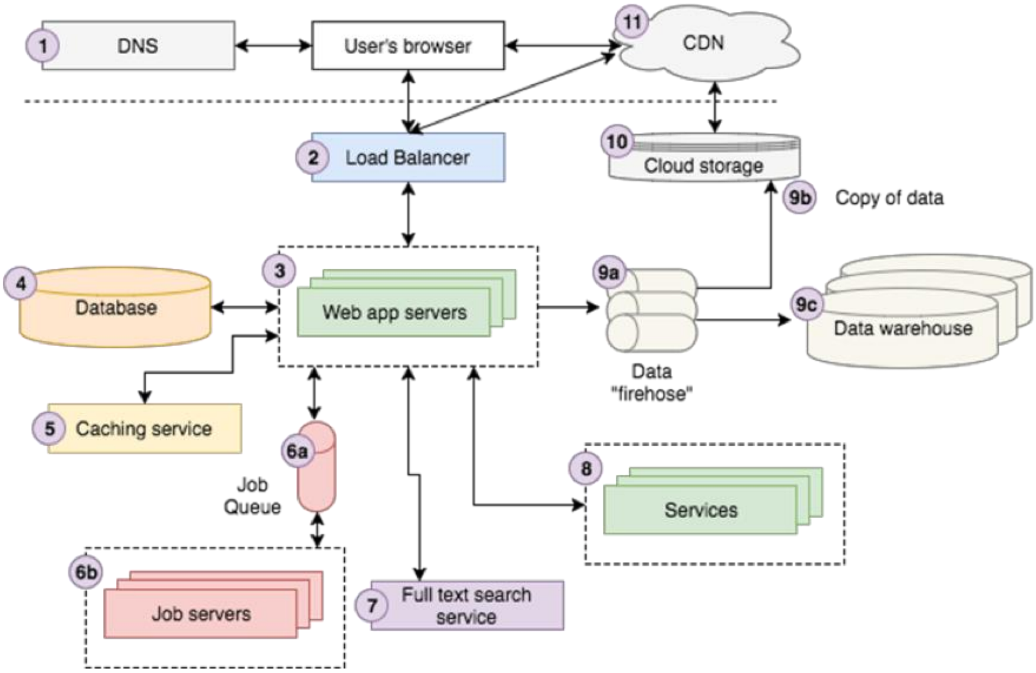


Figure 8 – Technical architecture of Finsecurity.eu platform

To support its robust and scalable deployment, the platform comprises the following additional technical components:

- *Public DNS Server*: Provides a network discovery service for the client applications i.e., the end-users.
- *Load Balancer*: Implements the balancing mechanisms that boost the scalability and high performance of the marketplace.
- *Backend Application*: Manages the processing and implementation of the "business logic" that is executed by the platform.
- *Database*: Is the main component of the Database Tier. As already outlined, it is in charge of storing and managing the dynamic content of the platform.
- *Caching*: Offers a caching mechanism for access times minimization. The mechanism keeps an index of pre-calculated data, which boosts content access performance.
- *Jobs Scheduling*: Consists of: (i) A Job Queue that enables asynchronous processing of tasks/jobs to reduce idle times and improve resource utilization; (ii) Job Servers, which ensure

the independence of the running jobs, while supporting the resource limits (e.g., performance throttling) that can help the vertical scaling of the platform's performance.

- *Search Service*: Supports the implementation of content search algorithms.
- *External Services*: Enable access to third-party resources and services via proper connectors.
- *Data Services*: Include data streaming, data replication, and data warehousing services. These services ensure scalable and efficient data management for the content items of the platform.
- *Cloud Storage*: Manages data and content distribution in ways that makes data items available to end-users (through the CDN network).
- *Content Delivery Network (CDN)*: Boosts content access efficiency and performance based on access and management of high-available storage networks across different regions.

The technical architecture supports microservices, leveraging proper containers and Service Oriented Architecture (SOA) principles [28], whilst the microservices can be scaled both vertically and horizontally. The implementation of the architecture leverages mainstream commercial public cloud services that provide high scalability, robustness, and availability.

5.2.3.2 Relevance to FAME & Added-Value

The integration of the Finsecurity.eu platform to the FAME SA is destined to offer the following added-value to FAME's general set objectives:

Federated Data Assets Catalogue (FDAC) integrating FAIR data assets from different data providers and marketplaces [**Obj 3**]

The Finsecurity.eu platform will provide access to content items and data assets about CIP and cybersecurity. This will foster the engagement of the CIP community in the FAME Federated Data Space, and subsequently increase both the exploitation of the overall FAME functionalities and the demand for the FAME assets.

Federated Data Assets Catalogue (FDAC) integrating FAIR data assets from different data providers and marketplaces [**Obj 3**]

The Finsecurity.eu platform includes different types of knowledge assets and different access modalities to them, which will be also adopted within the context of FAME. Nevertheless, it is also acknowledged that other existing Data Spaces and Data Marketplaces may offer access to similar types of data assets and content items.

Marketplace Platform agnostic Data Provider Interface (DPI) [**Obj 2**]

As an enterprise scale content platform based on full transparency over its implementation details, the Finsecurity.eu platform will provide FAME with opportunities to demonstrate how the FAME Federated Data Space can interface and communicate with external platforms. The FAME authentication and authorization infrastructure interface can demonstrate how FAME stakeholders can easily access the contents of Finsecurity.eu without needing to deal with the above-listed low-level architecture details.

Abstract data policy management functionalities [**Obj 2**]

The Finsecurity.eu platform will provide FAME with the opportunity to differentiate the handling of existing assets (Finsecurity.eu assets) with different access policies (i.e., assets available to all vs. assets available to registered users only). Such action will be accomplished through the FAME assets policies management component that will be able to shape such policies in a FAME-compliant policies format. It is however acknowledged that the public access to the FAME data assets will not offer opportunities for implementing complex security schemes (e.g., location-based security, content-based security) over the Finsecurity.eu integration.

5.2.4 SecureIoT

The main goal of the SecureIoT project [29] is to introduce, validate and promote a novel approach to the security of Internet of Things (IoT) applications, which emphasizes a timely, predictive, and intelligent approach to the identification and mitigation of security threats and incidents. The IoT Security Solutions Market Platform is built to offer the capabilities of the SecureIoT project to interested stakeholders. It constitutes a Multi-Sided Market Platform (MSP) that offers Security services on IoT based environments and extends the results of SecureIoT for the purposes of better exploitation through the building of an appropriate community that will serve the sustainability strategy of the project.

5.2.4.1 Architecture Overview

Figure 9 illustrates the conceptual architecture of the SecureIoT Marketplace, outlining its supported functionalities in accordance with the external users that are able to exploit them.

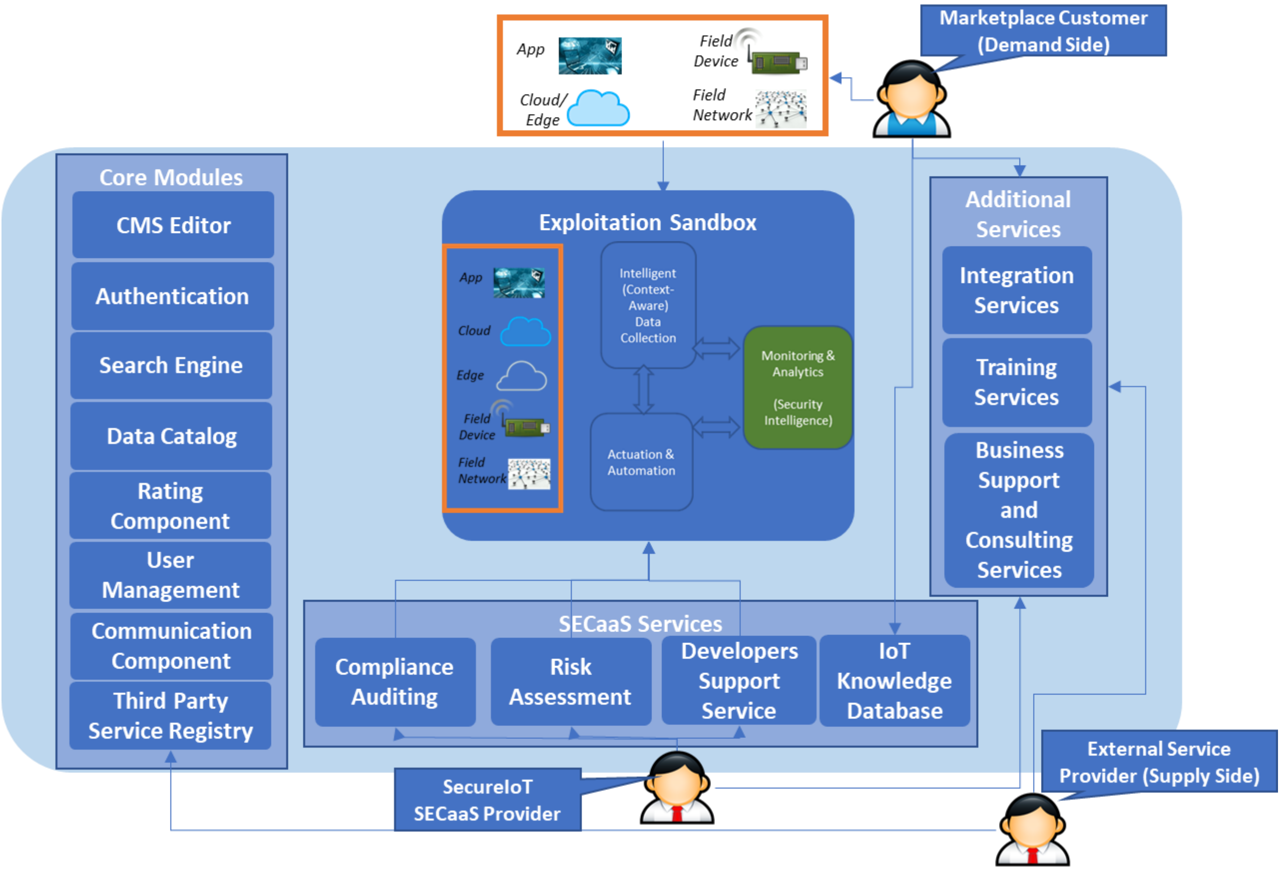


Figure 9 – SecureIoT marketplace conceptual architecture

The main layers of the marketplace are the following:

- *Core Modules*: The core modules layer that includes all the components of the SecureIoT Marketplace that implement the core marketplace functionalities. It includes modules for data catalogue, search, rating, authentication, user management, communication between the users (messaging and forum), and a Content Management Component (CMS editor). It also includes an add-on module that is responsible for the connection to any third-party service. In short, the supported functionalities of this layer refer to:
  - *Participants and Business Entities Registration*: Registers all the interested participants to the ecosystem.
  - *Authentication and Authorization*: Ensures authenticated and authorized access to the various services and sections of the platform.

- *Search and Discovery of Service Offerings*: Refers to the search engine for discovering available services based on appropriate metadata for the descriptions of the services.
- *Catalogue Publishing of Services*: Refers to publication and presentation of the ecosystem services, solutions, tools, and other entities.
- *Provision of Recommendations*: Refers to the context-aware proposition of relative services.
- *Collaboration Services*: Refers to the collaboration services (e.g., Forum/Messaging /Repository), including the relevant community support.
- *Review and Rating of Service Offerings*: Refers to the tools for rating services from the end-users’ viewpoints.
- *Manage and Tracking Registered Services*: Refers to the access to the status of subscriptions and services.
- *Solution Presentation*: Refers to the solution’s presentation through examples.
- *Services Presentation*: Refers to a comprehensive list of all services supported.
- *Knowledge Base*: Refers to the information services including articles, presentations, news, blogs, etc. Online training and education services in the form of self-contained presentations.
- *Marketplace Layout*: Refers to the aggregation of services and solutions in categories/subcategories with searchable metadata, thumbnails, descriptions, ratings. Management features for addition/deletion/categorisation, etc.
- *User and Organization Management*: Refers to the ability to manage organizations and users participating in an organization.
- *Basic Content Management Support*: Refers to the provision of content (publications, news, blogs etc) for visitors. Content should be editable by the moderator.

¶ *Exploitation Sandbox*: It is a part of the marketplace dedicated to offer to the interested end-users of the market an easy way to connect and use SecureIoT for testing in a sandboxed environment. The sandboxed environment provides the possibility to use it as is for understanding the platform, or to connect the end-users’ devices and datasets for more concrete testing. It also provides the possibility to use SECaaS that is already deployed within the sandbox environment.

¶ *SECaaS Layer*: A layer that includes the security services of the Marketplace, which are available in two (2) different forms: (i) through the exploitation sandbox, and (ii) by providing description, instructions and needed artefacts for their standalone usage.

¶ *Additional Services*: Additional services that represent integration services, training services and business support or consulting services. These services do not have any direct connection to SecureIoT or the developed SECaaS but can be beneficial to the end-users and important for the adoption of SecureIoT. In short, the additional services refer to:

- *Registration through 3rd Party Authentication Services*: Refers to the support for authentication with 3rd party services such as Google or LinkedIn since some users prefer this for faster registration.
- *Developers’ Support Services*: Refers to the developers joining the project’s platform will be offered with access to APIs and annotations and a dedicated IoT Developers Support as a Service function.
- *Training, Consulting and Technical Support Services*: Refers to the related services offered in the form of complementary (augmented) added-value services through partner value chains (expert human interface needed).

On top of the abovementioned, the SecureIoT Multi-Sided Market Platform supports the following functionalities:

- *Libraries*: Refers to the middleware libraries for SEC, SECaaS and general IoT, as well as open APIs for accessing the libraries including accompanying documentation.
- *Access to Services*: Refers to the ability for stakeholders to use, evaluate and consider the use of the: (i) IoT Security Risk Assessment and Mitigation as a Service, (ii) IoT Compliance Auditing as a Service, (iii) IoT Programming Support Services, (iv) IoT Knowledge Base, and (v) Relevant regulations and directives knowledgebase (e.g., GDPR, NIS, ePrivacy).
- *Access to Tools*: Refers to a coherent presentation and access to the code produced by the project on the topics of: (i) Interfacing, Data Collection and Collaboration, (ii) Multi-Level Security Measures and Security Analytics, and (iii) SecureIoT Services Implementation and Integration (SECaaS).
- *Use Case Paradigm Presentation*: Refers to the end-to-end implemented solutions serving as an example of integration: (smart manufacturing - Industrie 4.0, connected cars and IoT-enabled socially assistive robots).
- *Localization*: Refers to the support for an international environment through appropriate localization of the services including currency and language support.
- *Support of Privacy and Security Regulations*: Refers to the critical user and service data that are stored encrypted through a secure connection SSL/TLS. Privacy policy compatible with GDPR shall be created, considering the legal rights of the registered users (e.g., right to be informed, right to be forgotten), rules for long term storage of the data, and the specification of contact points for users of the platform.
- *Future Monetisation Module (marketplace/e-commerce)*: Refers to the pricing scheme (per unit/service/data volume/usage units or freemium). Also, the welcome addition would be to provide e-commerce/secure transaction management through 3rd party integration.

5.2.4.2 Relevance to FAME & Added-Value

The integration of the SecureIoT solutions to the FAME SA is going to provide the below added-value to FAME's general set objectives:

Federated Data Assets Catalogue (FDAC) accelerating the process of discovering interoperable, FAIR and interrelated data assets [**Obj 3**]

The core modules of SecureIoT that aim to provide a data catalogue with the ability of searching, rating, authentication, user management, and communication between the users will provide its design and implementation principles towards specifying the characteristics and the supported capabilities of the FDAC of FAME.

Upskilling and reskilling stakeholders in directions that enable them to understand the data economy and leverage the FAME marketplace [**Obj 7**]

SecureIoT will offer the technical knowhow with respect to its additional services that target on integration and proper skillset training, to build the FAME Learning Centre that will aim on providing training material for FAME stakeholders upskilling and reskilling.

Secure & Regulatory compliant integration of data spaces and marketplaces [**Obj 2**]

The support of Privacy and Security Regulations of SecureIoT will provide proper guidelines and best practices for constructing privacy data assets' policies based on existing regulations (e.g., GDPR), towards the envisioned FAME security infrastructure.

Implementation of decentralized pricing and trading schemes leveraging smart contracts [**Obj 4**]

The Monetisation Module related with marketplaces and e-commerce of SecureIoT that refers to pricing schemes, will act as a backbone for building the decentralized pricing and trading schemes of the data assets within FAME.

5.2.5 PolicyCLOUD

The PolicyCLOUD project [30] aims to be a cloud-based data-driven policy management platform, enabling its stakeholders to model, analyse, evaluate, and optimize their policies using a variety of Big Data tools and services. Among the components that have been developed and offered from the project is its Data Marketplace. The main idea of the PolicyCLOUD Data Marketplace is to create a community of users who will be able to provide and share various assets related to the scope and the investigated domains of the overall PolicyCLOUD platform. In essence, this Data Marketplace is a unified and standalone platform that can 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 store solutions authored or developed by other experts. Hence, this Marketplace is intended to offer through its platform ready-to-use sets of solutions related to various subjects, enabling its users to use them to solve/handle any of their needs.

5.2.5.1 Architecture Overview

From the architecture perspective, the Data Marketplace is structured around two (2) core services, the backend 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 backend).

Generally, the Data Marketplace provides several functionalities that are mapped to different layers. The backend 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 these four (4) different layers (as depicted in Figure 10) that realize the below capabilities:

¶ The *Assets Storage Layer* (part of the backend) is the layer in which the platform's offered assets are stored.

¶ The *Assets Management Layer* (part of the backend) delivers all the needed principles and techniques for the management of the Marketplace's assets.

¶ The *Interaction Layer* (part of the backend) 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 UI towards the different types of users that are willing to use the platform.

The overall conceptual view of the PolicyCLOUD Data Marketplace is presented in Figure 10, which depicts all the platform's layers along with their key offered functionalities, the providers, and the end-users. It is worth mentioning that the PolicyCLOUD Data Marketplace's users can be distinguished into human users and machine users (depicted in the top-left part of the Figure). In deeper detail, in addition to the normal "human" users (e.g., developers, scientists, interested third parties), platforms and systems are considered as a valuable extension to make the PolicyCLOUD Data Marketplace more interoperable, being able to offer its assets and services to other services, Thus, with the appropriate interfaces, external platforms and systems can provide or retrieve the offered assets in a more direct way. Consequently, the users of the Data Marketplace can interact with it either through the developed UI or directly with the Marketplace's backend API. 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 backend, as depicted in the Figure below.

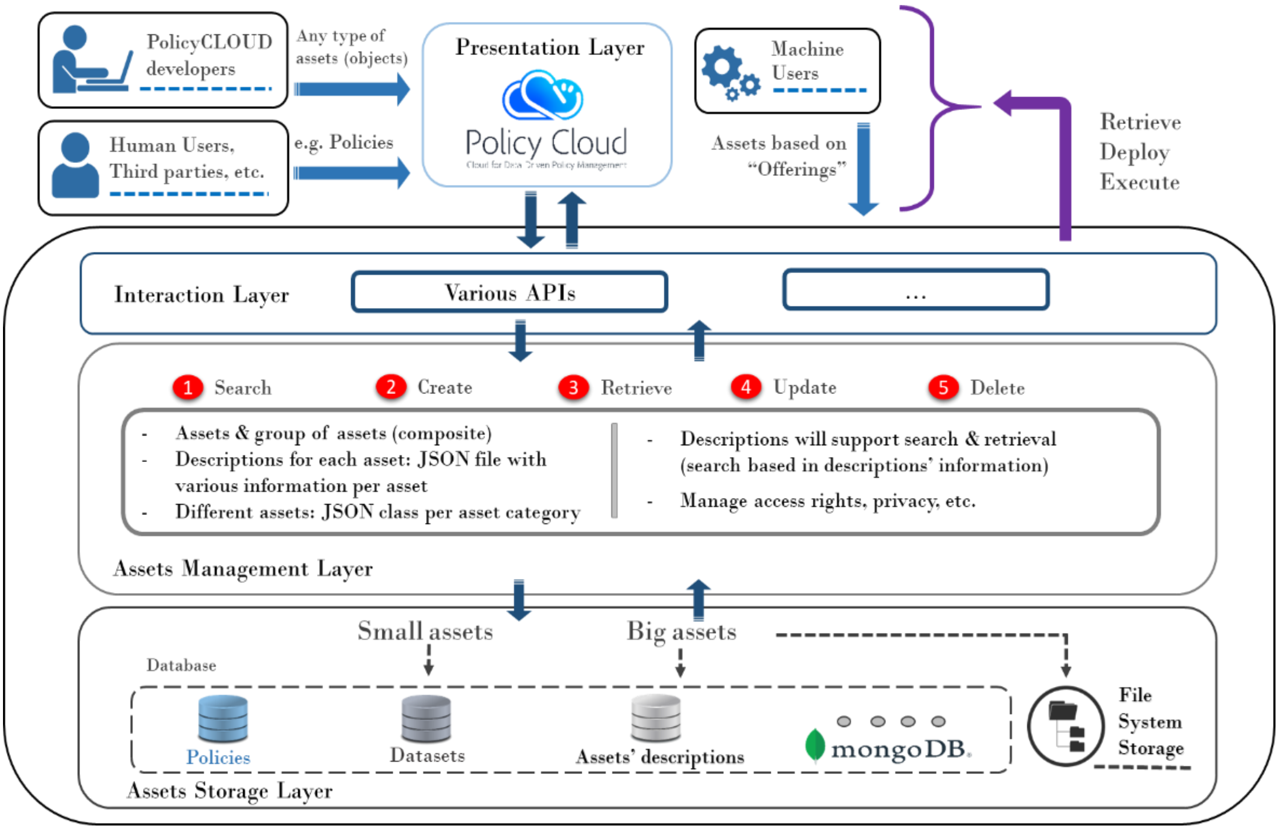


Figure 10 – PolicyCLOUD data marketplace conceptual architecture

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 backend 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) can 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 marketplace's data storage and the assets are stored in the operating system.

¶ To retrieve the assets, the *Assets Management Layer*, through the retrieve functionality, finds the requested asset from the Operating System (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.

5.2.5.2 Relevance to FAME & Added-Value

The integration of the PolicyCLOUD Data Marketplace to the FAME SA is expected to offer the following added-value to FAME's general set objectives:

Well-defined and documented Open APIs [**Obj 1**]
As in the case of the PolicyCLOUD Data Marketplace, both individual end-users and external marketplaces, systems, and platforms will be able to directly interact and exploit the FAME functionalities through its notion of offering the developed functionalities through open APIs.

Federated Data Marketplace platform supporting assets' discovery and exchange [**Obj 1**]
FAME will be developed and implemented supporting similar functionalities with the PolicyCLOUD Data Marketplace to enable its end-users both to search and to retrieve the offered assets of the platform, also being able to provide their own assets (with relevant content control). Such core functionalities will be basically supported by the FDAC of FAME, which will act as the backbone for managing the platform's data assets.

Federated Data Assets Catalogue (FDAC) integrating FAIR data assets from different data providers and marketplaces [**Obj 3**]
As in the case of the PolicyCLOUD Data Marketplace, FAME will support diverse types of data assets, including tools, datasets, webinars, tutorials, and documents among others, towards the facilitation of data exchange, repurposing, and reuse of a variety of data assets. On top, FAME will also adopt the idea of notating the data assets with additional metadata (as in the case of the PolicyCLOUD Data Marketplace), where for each of the data assets, FAME will provide full provenance and traceability, recording metadata about the assets' type, data volume, data quality features (e.g., completeness of values, readiness for machine learning use), user-friendliness scores, CO_2 and wastes, location specific characteristics (e.g., locality of data capture), timeliness, demand (e.g., number of searches and more), algorithmic metadata, and more.

5.2.6 DataVaults

The DataVaults Cloud Platform [31] is a cloud service offering a single-entry point for data seekers. It offers data seekers the functionality to search over available datasets, filter them using various criteria and buy the ones they are interested in. Once bought, those datasets are available in their personal "My Vault" page and can be downloaded from there. Points are redacted from the Data Seeker's wallet, once a purchase has been made and the data seeker can view the current balance of their wallet in the relevant page. The transaction is also stored in the blockchain and thus both the Private Wallet and the Personal App can be notified to update accordingly. Data seekers also have the option to request datasets without a pre-defined price and get notified about the outcome of those requests. Furthermore, data seekers can create and share questionnaires, as well as browse a record of all the transactions they have performed. They can also view some diagrams regarding their statistics and usage of the DataVaults platform.

5.2.6.1 Architecture Overview

The DataVaults infrastructure is divided in two (2) main parts: the *Personal DataVaults App* (residing westbound) and the *DataVaults Cloud Platform* (eastbound), as they are depicted in Figure 11.

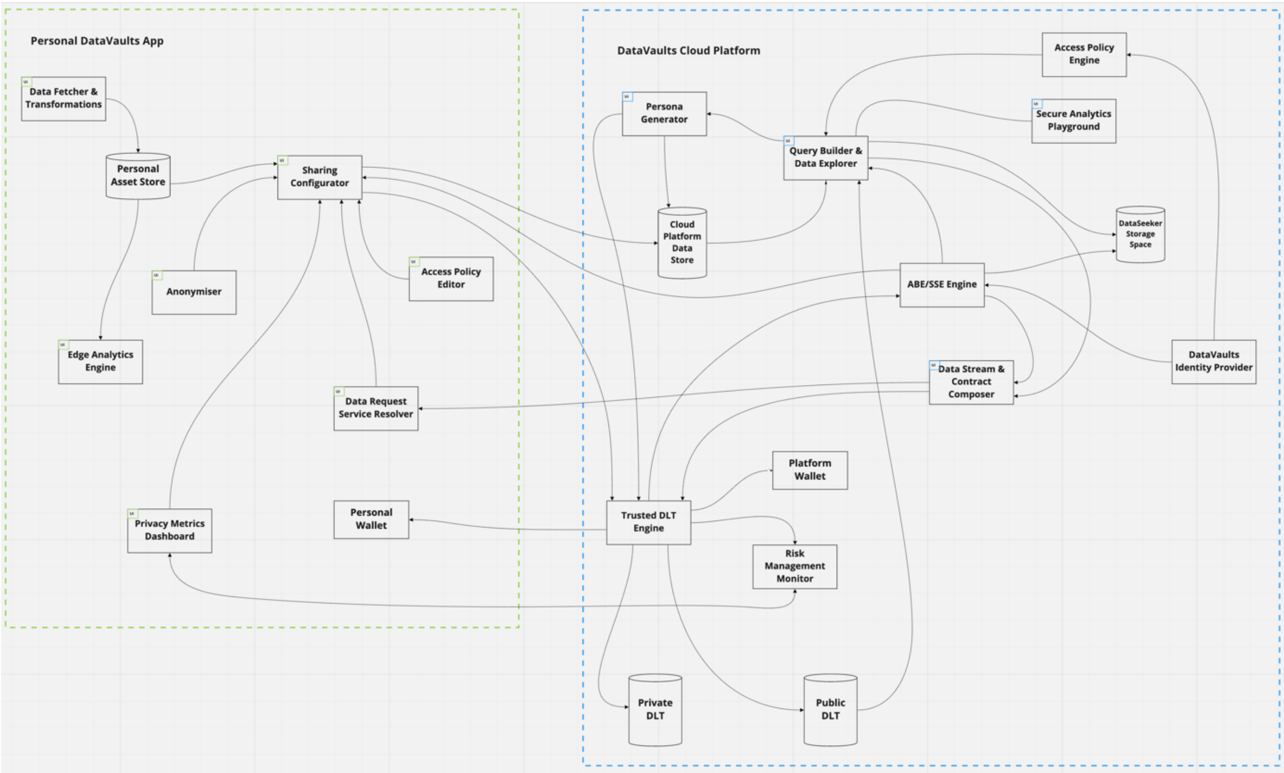


Figure 11 – DataVaults architecture overview

In greater detail the overall architectural flow is as follow:

- The *DataVaults Cloud Platform* is the main cloud-based infrastructure that supports the operations of data sharing and is the entry point for all the data seekers that are willing to onboard search for data, acquire it and conduct certain analysis on the platform. The DataVaults Cloud Platform is a wrapper around different sub-components.
- Once an asset is shared by the *DataVaults Personal App*, it gets stored in the *Cloud Platform Data Storage*, while at the same time information about this transaction is forwarded in the *Trusted DLT Engine*, which stores the transaction details both in the Private Distributed Ledge Technology (DLT) (indicating which user has shared what with the platform and under which terms/configuration) and in the Public DLT (indicating the location of the asset shared, alongside with sharing configurations but not revealing the data owners identity).
- Assets that reside in the *Cloud Platform Data Store* can be retrieved by using the *Query Builder & Data Explorer* that performs searches over this data. The *Query Builder* part of this component gets as input a query coming from a data seeker, performs a search in the *Cloud Platform Data Store* and retrieves the returned items' sharing configurations from the Public DLT. Part of the latter information (e.g., policies accompanying each asset) is used as input in the *Access Policy Engine* to decide whether access can be provided to the requiring data seeker or not. In the case that such items are encrypted, the necessary decryption keys are fetched from the *ABE/SEE Engine* (always depending on the data seeker's Attributes that are read from the DataVaults Identity provider) to query over the encrypted datasets (using SSE).
- In the case that a data seeker is willing to buy an identified asset, the *DataStream & Contract Composer* is triggered, which reads the information for this specific asset through the *Trusted DLT Engine* and forwards there a transaction. The Trusted DLT Engine then takes over, validating and storing the transaction in the Public DLT, performing the necessary shift of currency in the *Platform Wallet*, also validating a transaction in the Private DLT Engine and transferring the necessary amount in the *Personal Wallet* of the individual (i.e., data owner).

- Then, the asset is made available, and the necessary keys are used to decrypt the asset and store it in the *Data Seeker Storage Space*. In case that the transaction cannot be automatically executed (for example because the individual has chosen not to disclose a price), the data seeker can utilise the *DataStream & Contract Composer* to author a new contract proposal, which is forwarded to the Personal DataVaults App. This request is caught by the *Data Request Service Resolver*, that translates the request to a new sharing configuration that can be accepted or rejected by the Individual. In the former case, the sharing configuration is loaded in the *Sharing Configurator*, and the asset is shared under that specific configuration, while a transaction is then performed on both ledgers, as described above.
- Finally, the *Query Builder & Data Explorer* provides a file browser offering CRUD operations over the *Data Seeker Storage Space*, which is a repository space owned by a data seeker and is used to "download" any already "bought" dataset and upload any other data that the data seeker would like to include in an analysis. These assets are made available to the *Secure Analytics Playground* via the Data Explorer. Also, the Query Builder is used to identify (by DataVaults data scientists) data that are flagged as available to be used in a Persona. In such a case, data is automatically retrieved and used by the Persona Generator, and the developed personas are stored back in the *Cloud Platform Data Store*, while their sharing properties are forwarded to the Trusted DLT Engine to be made available for the data seekers.

5.2.6.2 Relevance to FAME & Added-Value

The integration of DataVaults to the FAME SA is expected to offer the following added-value to FAME's general set objectives:

Catalogue that will accelerate the process of discovering interoperable, FAIR and interrelated data assets [**Obj 3**]

As in the case of DataVaults, the technical knowhow of the predefined Query Builder & Data Explorer that aims on facilitating the performance of searches over the DataVaults related Data Marketplace will be used into FAME, to boost up the searching ability of the related FAME catalogue.

Readily available high-performance, energy efficient blockchain infrastructure that supports tokenizations [**Obj 4**]

Considering the way that DataVaults transactions are stored into the blockchain, DataVaults will offer the technicalities of storing transaction details both in the Private DLT and in the Public DLT, to provide sufficient feedback for the blockchain infrastructure of FAME for facilitating the envisioned tokenization technologies upon the data assets to be traded.

5.3 Overview of Reference Architectures for Data Spaces

5.3.1 IDS

The International Data Spaces Association (IDSA) [32] functions as a virtual data realm that utilizes established standards, technologies, and widely embraced governance models within the data economy. Its primary goal is to enable secure and standardized data exchange and data linkage in a trusted business ecosystem. This establishes a foundation for developing smart-service scenarios and promoting innovative cross-company business processes, all while ensuring data sovereignty for data owners.

Data sovereignty is a central aspect of the International Data Spaces (IDS). This entails the inherent ability of individuals or organizations to have complete self-governance over their data. The initiative of IDS introduces a Reference Architecture Model (RAM) that encompasses this very ability, along with associated factors such as the necessity for secure and trustworthy data interchange within business ecosystems.

5.3.1.1 Architecture Overview

IDS-RAM was originally defined as part of the research activities conducted in the Industrial Data Space project by Fraunhofer [33] and continues to evolve through the work of many research and industrial projects under the steering of the IDSA Architecture working group. Furthermore, IDSA promotes the IDS-RAM, IDS implementations and use cases, to establish an international standard for secure data exchange and data sharing facilitated by the IDS Connector, the central technical component of the International Data Spaces.

Focusing on the broad conceptualization of functionalities, capabilities, and the overarching processes engaged in establishing a *secure network of trusted data*, the IDS-RAM exists at a more elevated level of abstraction compared to typical architectural models of specific software implementations. Figure 12 illustrates the general structure of the IDS-RAM that uses five (5) layers and three (3) perspectives.

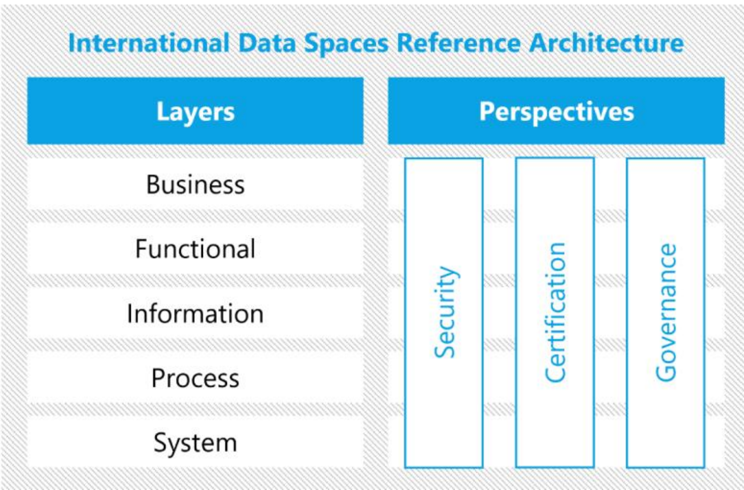


Figure 12 – General structure of IDS Reference Architecture Model

- *Business Layer* [34]: Specifies and categorizes the different roles which the participants of the International Data Spaces can assume, and it specifies the main activities and interactions connected with each of these roles. The Business Layer specifies the requirements such as establishing trust and technical frameworks for technically enforced agreements to be addressed in the Functional Layer and provides an abstract description that can be considered as a blueprint for the other, more technical layers.
- *Functional Layer* [35]: Defines the functional requirements of the IDS and the concrete features to be derived from the: (i) Trust that represents the fundamental features of data spaces, the roles, the identity management, and the user certification, (ii) Security and Data Sovereignty for performing authentication/authorization, usage policies usage enforcement, trustworthy communication security by design, and technical certification, (iii) Data Ecosystem that consists of the data source's description, metadata brokering, and vocabularies, (iv) Standardized Interoperability, (v) Value Adding Applications, and (vi) Data Markets that consider the monetary value concepts of data like clearing and billing, and governance.
- *Information Layer* [36]: Defines a conceptual model that makes use of linked-data principles for describing both the static and the dynamic aspects of the IDS constituents.
- *Process Layer* [37]: Specifies the interactions taking place between the different components of IDS; using the BPMN notation, it provides a dynamic view of the RAM. The following processes and their sub-processes are included: (i) Onboarding, (ii) Data Offering, (iii) Contract Negotiation, (iv) Exchanging Data, and (v) Publishing and using Data Apps.

¶ *System Layer* [38]: Is concerned with the decomposition of the logical software components, considering aspects such as integration, configuration, deployment, and extensibility of these components.

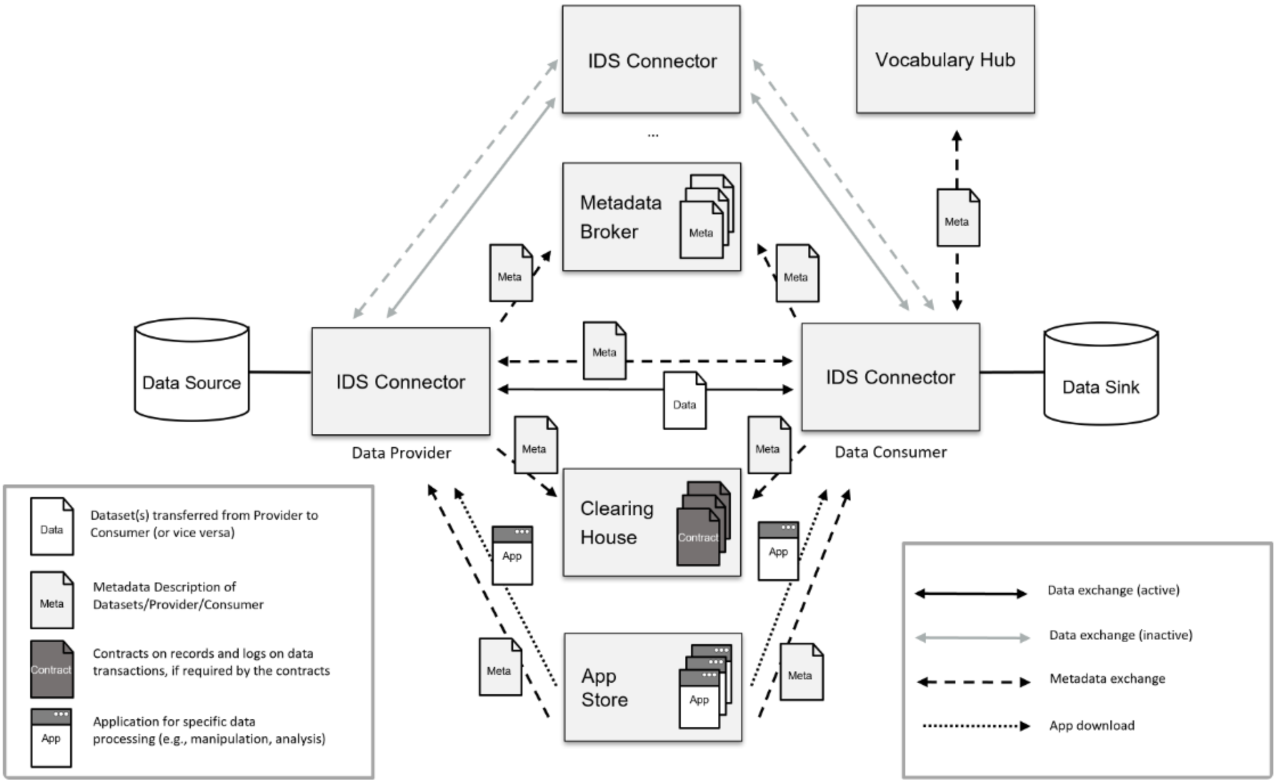


Figure 13 – Interaction of technical components

Figure 13 illustrates the interaction among the various existing technical components. It is crucial to note that the actual data is solely transmitted from the Data Provider to the Data Consumer. The remaining interactions with the other components rely on Metadata. Metadata delineates the appearance of the data, its format, user policies, data ownership, and more. The Metadata is submitted to a broker due to the absence of a centralized data lake. Connectors need to locate each other; hence they furnish the broker with their connection details and Metadata. A connector has the potential for augmentation with Data Apps. These Apps can be installed from an App Store within the connector to alter data before transmission or to perform analytical functions on the data within the connector.

In addition, as stated above, the IDS-RAM comprises three (3) perspectives that need to be implemented across all the five (5) layers:

¶ *Security* [39]: The IDS Security Architecture provides means to identify devices in the IDS, protect communication and data exchange transactions, and control the use of data after it has been exchanged.
¶ *Certification* [40]: Any organization or individual seeking permission to operate components in the International Data Spaces needs to pass the Operational Environment Certification that ensures secure processes and management of components.
¶ *Governance* [41]: The Governance Perspective defines the roles, functions, and processes of the International Data Spaces from a governance and compliance point of view.

5.3.1.2 Relevance to FAME & Added-Value

The IDS-RAM and its rulebook are currently investigated to be adopted within the FAME SA, bringing value to FAME in two (2) diverse ways, as further discussed below.

- **Architectural elements**: IDS-RAM and its rulebook describe several elements that could be relevant for the FAME SA.
- **Building blocks for technical implementation**: IDSA community provides several open specifications and open-source components for FAME to build on.

IDS Architectural elements for FAME: In this part, the elements from the IDS-RAM that are currently investigated that could have a role in different layers of the FAME SA are outlined.

- *FAME Dashboard*
  - *IDS Connector* [42]: The IDS network is formed by its IDS connectors, each enabling data exchange through exposed Data Endpoints. This eliminates the need for centralized data storage. IDS Connectors must be accessible to counterparts in different organizations, which might necessitate adjustments to firewall policies or the establishment of a demilitarized zone (DMZ). Accessibility via standard Internet Protocol (IP) and adaptability to various environments are essential features. Participants on data spaces (like FAME) that we want to federate use IDS connectors, which are deployable on-premises or in the cloud. Further information can be found in the respective data connectors report [43].
- *FAME Federation Manager*
  - *Policies Definition* [44]: IDS can contribute to specify the FAME policies: the two (2) main policy groups that are central to the functionality of a data space are access policies (which control access to contracts) and contract policies (which control the contract terms and the usage of data). On IDS rulebook there is more information about the policies.
  - *Usage Control* [45]: Access control restricts access to resources, whilst authorization is the process of granting permission to resources. There are several models of Access Control, such as Discretionary Access Control (DAC), Mandatory Access Control (MAC), Role-Based Access Control (RBAC), Attribute-Based Access Control (ABAC), etc. RBAC and ABAC are the most frequently used models. In IDS, data owners define ABAC policies for their endpoints, which can also be used in the context of FAME's access control process. Further information on how to design access control based on IDS can be found in the relevant report [46].
  - *Exchanging Data* [47]: The IDS-RAM defines how to exchange data between a Data Consumer and a Data Provider identifying two (2) phases: The Control Phase and the Transfer Phase. In the control Phase there is the data offering and contract negotiation that is specified in the data space protocol [48], while in the Transfer Phase, the Data Consumer and the Data Provider actually start to exchange data by invoking a Data operation via their IDS connectors. A similar scenario will be investigated to be adopted into FAME towards the exchange of data assets.
  - *Metadata Broker* [49]: The IDS metadata broker is an IDS Connector that contains an endpoint for the registration, the publication, the maintenance, and the query of Self-Descriptions. A Self-Description encapsulates information about an IDS Connector itself, its capabilities, as well as its characteristics. This Self-Description contains information about the offered interfaces, the owner of the component, and the metadata of the data offered by the component. A Self-Description is provided by the operator of the Connector. The Self-Description in total can be seen as metadata, a

concept that can be adopted into the supported FDAC of FAME, which is primarily based on managing the data assets' metadata.

  - *Vocabulary Hub* [50]: The IDS vocabulary hubs provide the developer of domain-specific vocabularies the tools and functions to create, improve, and publish their terms. While it is expected that these vocabularies follow the Resource Description Framework (RDF) pattern, further requirements like the Linked Data concepts or even formal ontologies are not enforced. The IDS relies in RDF to encode its attributes and data descriptions. Meanwhile, the IDS Information Model is the central vocabulary that all parties of any IDS share. A similar scenario will be investigated to be adopted into the FDAC of FAME.
  - *Information Model* [51]: The IDS information model provides fundamental concepts to describe data products based on the IDS core concepts and fundamental standards, Data Catalog Vocabulary (DCAT) for data assets, and Open Digital Rights Language (ODRL) for contract policies. As in the previous case, those concepts will be investigated to be adopted into the constructed information models of the data assets relying into the FDAC of FAME.
- *FAME Transactional Operations*
  - *IDS Clearing House* [52]: The IDS clearing house provides functions on logging services that record information relevant for clearing and billing as well as usage control. In the context of FAME, such concept can be exploited in the data assets' tracking, monetization, pricing, and trading procedures.

IDS Building blocks for FAME technical implementation: IDS-G [53] is a publicly accessible GitHub repository and provides the IDSA building blocks' specifications and further documentation from IDSA to the public. While the IDS-RAM is technology independent, the specifications in IDS-G describe the binding of the RAM to technological concepts and focus on documentation and specifications for IDS based solutions. Specifications for the following components are available:

- *IDS Data Connectors Implementations*:
  - *IDS Meta Data Broker* [54]: The specification of the IDS metadata broker as an index-service running in conjunction with an IDS Connector.
  - *IDS Clearing House* [55]: The specification of the IDS clearing house as an intermediary in IDS ecosystem. All the IDS Connectors may log information in this module to support any process that requires an auditable logging mechanism [56].
  - *IDS Information Model* [57]: The specification of an RDFS/OWL-ontology covering the fundamental concepts such as the types of digital contents that are exchanged by participants by means of the IDS infrastructure components.
- *IDS Deployment Scenarios* [58]: A repository that lists IDS deployments from various domains and cases. It serves as a source of inspiration and guidance for those who wish to experiment with data spaces or best deployment practices.
- *IDS Data Space Testbed* [59]: A testbed that can be used as a baseline for creating minimal viable data spaces for cases in the financial sector. This testbed allows to test functionalities, business concepts and build proof of concepts before investing in real-life solution.

By exploiting all the aforementioned guidelines and principles, the IDS-RAM aims at contributing into the achievement of the following FAME objective:

Federated multi-cloud, multi-stakeholder, standards-based data marketplace platform [**Obj 1**]

The FAME SA will be specified in-line with the IDS-RAM for providing federated data exchange in a multi-cloud environment, by exploiting various IDS architectural elements.

5.3.2 GAIA-X

Gaia-X [60] is a framework being created for a Federated and Secure Data Infrastructure, having as a primary goal the innovation through digital sovereignty. This is achieved by establishing a decentralized ecosystem in which data is made available, collated, and shared in a trustworthy environment, where users always retain sovereignty over their data. Towards this direction, the Gaia-X community consists of multiple stakeholders who are specifying and developing a set of functional and interoperable components consisting of: (i) Federation Services and other technical components, (ii) a Governance Framework, and (iii) a Trust Framework.

To make the Gaia-X concept operational, the Gaia-X Federation Services (GXFS) toolbox [61] has been developed, aiming to provide the minimum technical requirements/set of services needed to build and operate this cloud-based, self-managed data infrastructure ecosystem.

5.3.2.1 GXFS Elements

To be more specific, GXFS consists of several components (Figure 14) enabling federations in data ecosystems and providing interoperability across federations, being the OSS Toolbox for Building Federated Data Ecosystems. These components are categorized into the five (5) core groups of (i) Identity and Trust, (ii) Sovereign Data Exchange, (iii) Sovereign Federated Catalogue, (iv) Compliance, and (v) Portal, as further described below.

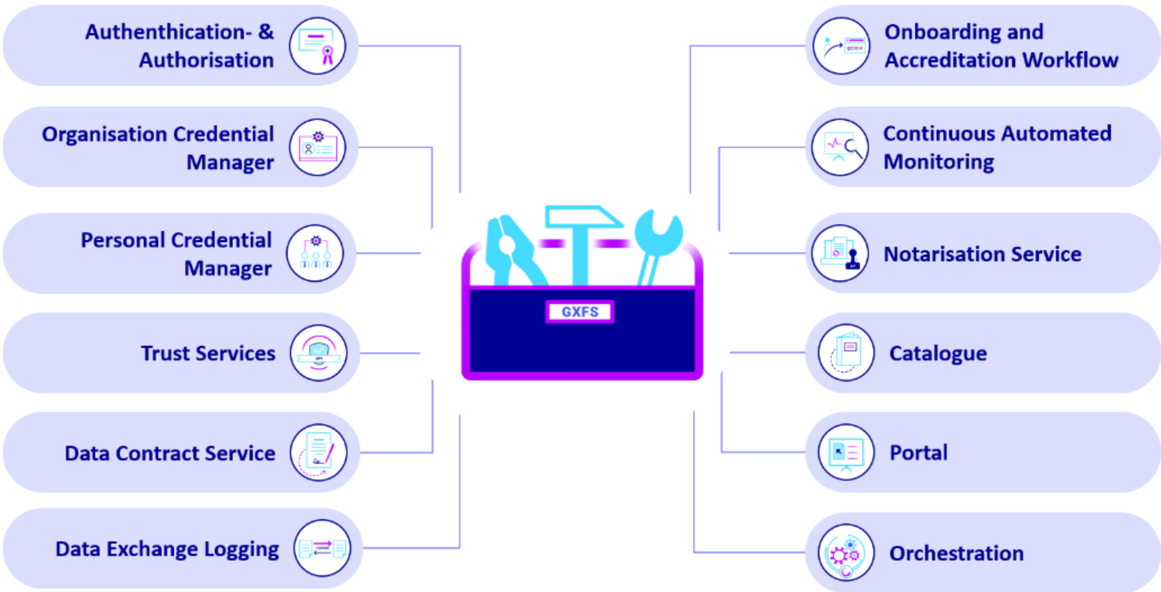


Figure 14 – GXFS supported elements

- *Identity and Trust Elements*
  - *Authentication & Authorization* [62]: This service enables Gaia-X participants to authenticate other users and systems in a trusted, decentralized, and self-sovereign manner without the need for a central source of authority.
    - Request verifiable, decentralized, and cryptographic credentials and identity attributes from other participants in a Federation.
    - Maintain control over what information is shared with others.
  - *Organization Credential Manager* [63]: This service establishes trust between the different participants within the decentralized Gaia-X ecosystem. It includes all trust-related functions required to manage and offer Gaia-X self-descriptions in the W3C Verifiable Credential Format, providing the following core functionalities:
    - Configure a self-determined and easy entry into a Federation for companies by, e.g., independently issuing digital participation credentials to employees.

    - Provide software services and data assets with a digitally verifiable seal.
    - Create cryptographic verifiable Self-Descriptions.
    - Manage credentials and certificates of employees, data, and services.
  - *Personal Credential Manager* [64]: This service enables Gaia-X users to manage their credentials themselves. To do this, the user needs secure storage (user wallet) and presentation capabilities in the authentication and authorization processes. Hence, the supported functionalities refer to:
    - Manage self-sovereign of one's own credentials, e.g., identity documents, certificates, or authorizations of individual participants (self-employed or employed).
    - Maintain control over which credentials are used for authentication and authorization purposes.
    - Authenticate using a mobile app or browser application.
    - Authenticate and authorize natural persons as well as machines and digital twins to enable trust-based machine-to-machine communication.
  - *Trust Services' API* [65]: This service ensures that a consistent level of trust can be established between all the components and the participants in a Gaia-X ecosystem. They are the central, technical implementation of cryptographic functions for enforcing policies in the SSI context for the use of the capabilities provided in a decentralized and self-governing manner. Consequently, the supported include:
    - Enforce usage policies.
    - Ensure chains of trust among multiple participants, organizations, and authorities.
    - Establish trust anchors with verification standards such as W3C Verifiable Credentials/Presentations.
    - Establish rule-based trust on an attribute basis.
- *Sovereign Data Exchange Elements*
  - *Data Contract Service* [66]: This service enables data exchange in a secure, trustworthy, and auditable way in a Gaia-X ecosystem. It provides interfaces for negotiating data contracts that define the agreed terms (Data Asset Usage Policy) for the planned data exchange. Thus, the supported functionalities include to:
    - Obtain legally binding consent between data provider and data user for data access, exchange, and use.
    - Sign cryptographically a data contract.
    - Subsequent provision of the signed contract.
  - *Data Exchange Logging Service* [67]: This service is used to run evidence whether data has been transmitted, received and rules and terms of use (data usage policies) have been respected or not within the Gaia-X ecosystem, offering functionalities to:
    - Track whether data has been transmitted and received or not.
    - Track whether data usage policies have been respected or violated, e.g., to clarify operational issues or detect fraudulent transactions.
    - Create an auditable transaction log that is only accessible to the contracting parties.
- *Sovereign Federated Catalogue Elements*
  - *Federated Catalogue* [68]: This service includes a catalog where Gaia-X resources, asset items, and participants can be found by potential consumers and end-users. Resources, asset items and participants are provided at Gaia-X using self-descriptions. Hence, the offered functionalities refer to:

        - Search and select providers and their service offerings in a Federation based on self-descriptions.
        - Monitor relevant changes in service provision.
        - Support of Self-Description Tools, including a Creation Wizard (for creating valid Self-Descriptions (claims) using interactive web forms), a Visualization Tool (for visualizing created Self-Descriptions, and a Validation Wizard (for validating the created Self-Descriptions (claims), e.g., check whether data types are correct and all mandatory information is present).
- *Compliance Elements*
    - *Authentication & Authorization*
        - Support the implementation of a validation process for participants, resources, and service provision prior to inclusion in a Federation's catalogue.
        - Document the validation process and create an audit trail to ensure compliance with generally accepted conformity assessment practices.
    - *Continuous Automated Monitoring* [69]: This service provides Gaia-X users with transparency about whether individual service offerings in a Gaia-X Federated Catalog are compliant with the rules or not. This compliance is based on certain requirements and rules that Gaia-X itself has set for its system. Thus, its core functionality is to:
        - Continuously automate rule compliance monitoring based on Self-Descriptions in a Federation's catalogue.
    - *Onboarding & Accreditation Workflows* [70]: This service ensures that all participants and offerings within the Gaia-X ecosystem undergo a validation process before being added to the Federated Catalog.
    - *Notarization API* [71]: This service authenticates given master data and transforms it into a W3C-compliant, digitally verifiable representation. These tamper-proof digital assertions about specific attributes are central to gaining the desired trust in provided self-descriptions of assets and participants. As for the service's supported functionalities these refer to:
        - Issue a verifiable credential following successful validation of a participant to confirm status as a registered participant in a Federation.
        - Process notarization requests and issue digital, legally binding, and trustworthy credentials.
- *Portal Elements*
    - *Portal* [72]: This service serves as a RA for interacting with core service functions via an intuitive user interface and corresponding backend implementation functions. The user interface provides mechanisms for interacting with core functions via API calls. As for the service's supported functionalities these refer to:
        - Operate a business web client for each Federation.
        - Integrate the individual Federation Services such as: (i) querying Federation databases and displaying search results for services and data within a Federation, (ii) profile management of member's account to create and edit self-descriptions, organizational data, login history, etc., (iii) credentials management, (iv) addition and authorization of new members of a Federation, and (v) dashboard with overview of all the active and inactive services, the status of the booked services, as well as the history of the used services.
    - *Orchestration* [73]: This service allows Gaia-X consumers to instantiate and manage infrastructure services, such as virtual machines, from the Federated Catalog search results via the Gaia-X portal.

- *IDM & Trust Architecture*:
  - Decentralized identity management.
  - Trust Layer with signature and validation mechanisms.
  - Service components/features supporting on-/offboarding processes.
  - Access management.

5.3.2.2 Relevance to FAME & Added-Value

All the GXFS services are currently being investigated to be used as base elements or references within the FAME SA concept, contributing into the achievement of the following FAME objective:

Federated multi-cloud, multi-stakeholder, standards-based data marketplace platform [**Obj 1**]

All the GXFS services that have core elements that work together and provide basic compliance and security rules, are based on standards and are open source. On top of business-specific rules, description schemes and specific business requirements, these elements provide maximum flexibility to create a specific federation, which is also one of the major targets of the FAME Federated Data Space. However, these elements/services represent only the minimum basic structure and functionalities of a federation. FAME will analyze and examine to what extent these elements can serve as a basis to facilitate the implementation of specific both business and technical requirements by the project partners, considering (i) whether previous partner developments can be built on this basis and thus optimized, and (ii) whether flexibility and controllability can be realized in this way, even for future rule changes, while at the same time achieving high economies of scale.

6 FAME Capabilities

The FAME Federated Data Space encompasses several innovative tools and mechanisms that provide specific capabilities, all in one realizing the vision of a unique, trustworthy, energy-efficient, and secure federated EmFi data marketplace that will offer novel decentralized programmable pricing and trading of data, among others. Hence, the FAME's capabilities will be related with the opportunities of typical Data Spaces and Data Marketplaces - as the ones already discussed in Section 3, additionally offering: (i) an one-stop shop federated data catalogue integrating and linking data assets from external data marketplaces and applications, (ii) a single-entry point authorization and authentication infrastructure, (iii) a decentralized, configurable, dynamic, value-based data assets' trading and monetization capability, and (iv) a set of EmFi applications related trusted and energy-efficient analytics in the form of additional data assets, instead of solely EmFi applications' related data. All the underlying capabilities are presented in this Section, while Section 7 provides details for the mechanisms/components that realize these capabilities.

Table 1 – FAME Overall Capabilities

#	Capability	Short Description
1	**Authorization & Authentication**	FAME will allow to authenticate end-users and authorize them to access the respective FAME functionalities and the existing content.
2	**Federation of External Sources**	FAME will support the discovery, indexing and semantic annotation of data assets coming from diverse sources lying outside of the FAME platform, referring to external data spaces, data marketplaces, datastores, etc.
3	**Assets Policy Management**	FAME will ensure the secure management of assets, regulating who can view and purchase the assets in the platform, also providing end-users with a comprehensive list of assets they own accompanied by their surrounding regulations.
4	**Assets Provenance & Tracing**	FAME will ensure the quality of the published assets' metadata, providing metadata that identify the nature, meaning and provenance of each asset, while ensuring the authenticity and integrity of such metadata.
5	**Assets Pricing**	FAME will provide a tool to calculate the cost of a published asset, considering both static (based on the nature of the assets) and dynamic (based on demand) variables.
6	**Assets Trading & Monetization**	FAME will grant to its end-users the ability to navigate the FDAC and identify the assets that are aligned with their interests, supporting the trading of access rights for a chosen asset.
7	**Assets Searching**	FAME will allow the end-users to perform data queries upon the existing assets, based on their provided semantic models.

8		**ML/AI Analytics**	FAME will allow the end-users to select, build, deploy and run an ML/AI model that is suitable for their chosen scenario/application, being also able to improve the overall energy efficiency of the produced model by applying incremental analytic capabilities.
9		**SAX Analytics**	FAME will support process-aware explainability services towards providing comprehensive insights upon analysed processes, being complemented by a capability for scoring the explainability of the different constructed models.
10		**Analytics CO_2 Monitoring**	FAME will estimate the CPU usage of the ML/AI models, which then will be translated into the energy consumed by each model so as to be optimized in terms of consumption.
11		**FML Deployment**	FAME will support energy efficient data sharing in federated learning scenarios where various end-users contribute assets.
12		**EmFi Training**	FAME will provide a pool of training assets' resources (focusing especially in EmFi applications), enabling the end-users to search, access, as well as comment and rate all the available resources, whereas integrating new training content.
13		**Dashboard**	FAME will provide an integrated user-friendly end-to-end UI facilitating the interaction of the FAME end-users with all the involved FAME services, components, and processes.

6.1 Authorization & Authentication

Authentication and authorization are two (2) fundamental security processes in the management of access to systems and resources. In the context of FAME, these processes are extremely crucial in protecting sensitive data and user privacy, and in maintaining the overall system's integrity.

More specifically, authentication will be the process of FAME for verifying the identity of an external user or system. Its primary function will be to ensure that the entity requesting access is who or what it professes to be. Such functionality will be of high importance for the end-users as well, since authentication will help them to ensure that only verified users have access to their accounts, thus protecting their personal and sensitive data from unauthorized access. At the same time, once FAME identifies the user through authentication, it will be able to deliver a personalized user experience, including preferences, settings, and other personalized data.

Once authentication has been achieved, the next step will be the authorization. In the context of FAME, authorization will grant or deny access to specific resources or permissions within the developed platform once a user's identity has been authenticated. Such functionality will be of high importance for the end-users as well, since authorization will allow the control of access to sensitive information on a need-to-know basis, restricting what users can see and do within the system based on their individual roles or privileges. At the same time, the supported authorization process will limit the access to information based on the user's roles, reducing the risk of accidental changes or unauthorized data exposure.

Hence, through authentication, users will be assured that their accounts and data are protected from unauthorized access, whilst through authorization, they will know that their actions within the system are appropriate to their role, being protected from accidental misuse or deliberate data breaches. To successfully support such concept, the FAME Authentication and Authorization Infrastructure will provide Self Sovereign Identity (SSI) capabilities, based on Distributed Identity and Verifiable Credentials, maintaining the most used authentication and authorization flows and standards in this moment, to facilitate the integration of stakeholders' applications and incentivize a wide adoption. To be more specific, the SSI concept has been chosen to be adopted for solving the following issues: (i) Identity and personal data are stored with the user, (ii) Claims and attestations can be issued and verified among users and trusted parties, (iii) Users selectively permission access to data, and (iv) Data only needs to be verified a single time. As for the Blockchain technology, providing decentralization, immutability, and cryptographic security, it will allow the creation of credentials that could be issued and verified without the need of a central certification authority and could be owned by the end-users and directly shared with third parties without involving the credential issuer.

6.2 Federation of External Sources

The ability to discover and index data assets outside of the FAME platform is essential to allow FAME to perform discovery, exchange, pricing, and trading of relevant assets published in external Data Spaces/Data Marketplaces. In FAME, the Federated Data Assets Catalogue (FDAC) will provide such functionality through a developed mechanism to describe and execute asset metadata importers - called Resolvers. The latter will be able to connect to the associated source of assets, use the available interfaces to explore the available assets, read the metadata of the assets described on the source's data model, extract the information, and add it to the FDAC. Among the information extracted by the Resolver, this will be the description of the asset, the location of the asset, any existing pricing information and knowledge about any policy related with the assets discoverability and purchase conditions. Since such information is relevant to multiple FAME components, whenever new information is added to the FDAC, automatic notifications will be generated to inform those components. What is more, it should be mentioned that a Resolver can be executed periodically to update existing information or be executed on-demand when an event requires it.

Since different sources of assets may expose different APIs and represent information using different data formats, multiple Resolvers may need to be implemented. However, the assets' sources may follow Reference Standards or Architectures that define the interfaces and the data models used for communication, contributing to reduce the amount of heterogeneity among diverse sources. Therefore, a Resolver implemented to follow one of these References will be able to be reused to fetch information from several sources that also comply with the Reference.

This characteristic also assumes relevance in the European ecosystem where RAs, like IDSA and Gaia-X, emerged and are getting popularity. To this context, a Resolver implemented to comply with the interfaces and the data models of one of those specifications will be able to be reused to connect to the multiple Data Spaces/Data Marketplaces that also implement it.

6.3 Assets Policy Management

The Assets Policy Management plays a crucial role in ensuring the secure management of assets within FAME. This component serves two (2) key functionalities, both of which are vital for the smooth operation of FAME.

Firstly, the Assets Policy Management will enable the complete lifecycle management of policies associated with the federated assets discoverable through FAME, determining who is eligible to view and potentially acquire specific assets within FAME. Leveraging the Attribute-Based Access Control (ABAC) model, the component will consider various user and organizational attributes to make informed policy decisions. By fulfilling its role as a Policy Decision Point (PDP), the Assets Policy Management will ensure that other components of the project always display the appropriate assets to authenticated and authorized individuals or organizations. It will act as a central authority, facilitating the enforcement of access controls throughout the system.

Secondly, the component will provide end-users with a comprehensive list of assets they own. This will encompass assets uploaded by the end-users themselves or any other member of their organization, as well as assets acquired by them and have active contracts. This functionality will offer to the end-users a clear and consolidated view of their assets, enhancing their ability to manage and track their assets portfolio effectively. This functionality will enhance transparency and accountability within the project, allowing users to have a clear view of the assets under their control.

Overall, the Assets Policy Management will play a critical role within FAME, since it will serve as the central authority for managing asset security policies and ensuring that the appropriate assets are displayed throughout the system. By acting as a PDP, it will enable effective access control enforcement by other components. Additionally, the component will provide to the end-users a comprehensive overview of their owned assets, enabling better asset management and control. With its pivotal functionalities, the Assets Policy Management will reinforce the project's security posture, will foster transparency, and will promote efficient assets' management practices.

6.4 Assets Provenance & Tracing

The trustworthiness of any data space depends on several conditions, a key one being the availability of the metadata that the end-users can rely on. This means two (2) things: firstly, providing metadata attributes that unambiguously identify the nature, meaning and provenance of the underlying data asset; and secondly, ensuring the authenticity and integrity of such attributes, so that their content cannot be tampered by malicious actors. The Assets Provenance & Tracing component provides these capabilities in the context of FAME. In essence this component is the implementation of a registry where the identities of verified sources and the digital fingerprint of catalogue entries are stored as permanent and immutable records, so that any inauthentic version of these key information items (e.g., a counterfeit entry from the catalogue of a federated data space) can be easily spotted. Being blockchain-based, the component's registry is shared by all the members of a federation.

From the perspective of a publisher (i.e., a user sharing a data asset), the integration of the Assets Provenance & Tracing component in a federated data space gives a high level of confidence that what has been provided as the description of the published item cannot be altered (e.g., by a malicious actor posing as the publisher). From the perspective of a would-be consumer, it adds trustworthiness to the FDAC, as the provenance of the asset and the integrity of its catalogue entry can be relied on.

6.5 Assets Pricing

The Assets Pricing component plays an important role in the development of the business side of FAME. This component consists of two (2) key functionalities, both of which are quite important for the transactional operation of FAME.

Firstly, it will suggest an objective price for different data assets and services leveraging the metadata of the data assets. To this context, it will specify and implement a set of different pricing schemes for different data assets and services, using metadata information, including the asset's completeness, volume, quality, timeliness, CO_2 wastes, user friendliness, trustworthiness, ML-readiness, and more. This solution will require access to the metadata of the various data assets, whereas to implement the above listed schemes, this component will leverage the data provenance and traceability API of the blockchain infrastructure.

In addition to metadata, weighting factors and values obtained through questions from sellers are also important for asset valuation. These will help to ensure the subjective aspect of valuation, as the primary goal of valuing individual assets is to achieve satisfaction, especially from the seller's perspective. The seller can choose the best price that would reflect all costs associated with acquiring the provided asset.

The result of this component will be a price recommendation for the end-user (seller), being targeted to help them better estimate the price of the asset they are offering. The sellers can either lean towards the recommended price or choose their own price, supporting a seller-oriented approach.

6.6 Assets Trading & Monetization

The Assets Trading & Monetization component is a key component of FAME, designed to facilitate the secure and efficient trading of data assets. It leverages the power of smart contracts to bring significant value to both providers and consumers of data assets.

For data asset providers, the Assets Trading & Monetization component will offer a robust and secure mechanism for monetizing their data assets. By utilizing smart contracts, it will ensure that providers are fairly compensated with ERC-20 tokens for the value they bring to the marketplace. This not only provides an immediate revenue stream for providers but also incentivizes the continual addition of high-quality data assets to the platform.

For data asset consumers, the component will simplify the acquisition process. Consumers will be able to easily browse the FDAC, identify data assets of interest, and acquire them through a transparent and secure trading process. The use of ERC-1155 tokens in this process will allow for the accrual of value in the case of certain data assets, enhancing the potential return on investment for consumers.

The Assets Trading & Monetization component will also play a crucial role in maintaining the integrity and transparency of the marketplace. All the trade details will be logged on a ledger, providing a clear audit trail that promotes trust among users and ensures compliance with regulatory requirements. In essence, it will streamline the trading process, will ensure fair compensation for providers, will simplify asset acquisition for consumers, and will foster a transparent and trustworthy marketplace.

6.7 Assets Searching

FAME's FDAC will offer to its end-users a platform to easily search and access data assets from various sources. By providing a unified search interface, real-time data availability, and enhanced discoverability, the end-users will be able to quickly find relevant data according to their needs, for analysis and decision-making. In addition to that, the searching engine will offer the requested data assets ranked not based on semantic criteria, but also considering additional criteria such as the assets' price, size, and format. Hence, the main goal will be to provide the data asset consumers the most relevant offer depending on their specific needs and requirements for data exploitation.

Thanks to the single-entry point for finding data assets' information through the developed searching mechanism upon the FDAC, FAME will not only ensure data security and governance to its end-users, but it will also promote collaboration and knowledge sharing while optimizing resource efficiency.

6.8 ML/AI Analytics

The ML/AI Analytics will bring to FAME the capability to understand EmFi applications related problems and get the necessary insights to automate decisions based on the input data that the clients provide. Realizing a catalogue of available ML techniques is the main goal of this component. Moreover, functionalities for training these algorithms, as well as the capability to infer output given new data samples, will be also implemented. State-of-the-art models such as Transformer, Recurrent Neural Network (RNN), Convolutional Neural Network (CNN) or classical ML models will be provided to address different EmFi applications' tasks, which comprise solving time series, sentiment analysis, ranking system, or recommendation system problems.

The ML/AI Analytics catalogue will provide the FAME end-users with tools for understanding either their own data, or the data provided by existing external ecosystems (e.g., data marketplaces, data spaces), in an intuitive manner without the need to have a deep understanding of ML techniques, allowing them to take decisions according to the model's outputs.

6.9 SAX Analytics

Explanations to outcomes or decisions in any business process is critical for the adoption of AI-augmented systems. Users need to trust and understand the systems for them to use them. However, state-of-the-art XAI techniques when applied to business processes fail to: (i) express the business *process model constraints*, (ii) include the richness of *contextual situations* that affect process outcomes, (iii) reflect the *true causal execution dependencies* among the activities in the business process, or (iv) make sense and be *interpretable* to human users. FAME will go beyond XAI state-of-the-art and develop SAX that tackle the short comes mentioned above.

The main functionalities of the FAME developed SAX capability will revolve around process-aware explainability services, aimed at providing comprehensive insights into analysed processes given their event logs as input. These capabilities will include the discovery of causal execution dependency views, allowing the end-users to understand cause-and-effect relationships among events and decisions. SAX will also enrich the event logs with contextual information, focusing on significant situations to gain a deeper understanding of the processes. The component will offer explanations, both local and global, for a set of instances, with features like deriving process outcome relevant attributes ranked by importance, streamlining attributes inferred through causal execution dependencies, and filtering/augmenting contextual-related attributes. Additionally, SAX will synthesize and streamline explanations, ensuring they are sound and interpretable, empowering end-users to make informed decisions and gain a clear understanding of the underlying processes.

SAX will also be complemented by the XAI Scoring framework, which will be an additional capability for scoring the explainability of the different models towards comparing alternative approaches. The provided explanations will be easy to be interpreted by human-users who will also have the possibility to provide their preferences over the explanations in terms of the considered attributes. In essence, the XAI Scoring framework will be a transformative tool that will evaluate the explainability of the produced models, providing to end-users the benefits of: (i) Trust gauge, since by understanding AI decision-making processes, end-users will be able to better trust and rely on the system's outputs, (ii) Human-focused user experience, since the framework will prioritize the user experience, ensuring that AI is not just technically sound but also user-friendly, and (iii) Benchmarking, since users will be able to compare different AI techniques, making informed decisions about which models best align with their objectives. The output of this framework can be also exploited to aid in refining the pricing of the AI data assets provided/produced into FAME, ensuring that valuations are both accurate and reflective of their true explainability and utility.

By using those components, the end-users will be able to understand what makes a model to perform the way it does or identify which data that they are feeding with the model is more valuable in the given estimations.

6.10 Analytics CO_2 Monitoring

Estimating the CPU usage of the models used within the FAME applications will be the main functionality of the Analytics CO_2 Monitoring component. This metric will later be translated into the energy consumed by the model. Using public information or public databases on the average kilograms of CO_2 emitted per kilowatt-hour per nation, this component will then optimise the use of the available models within FAME in terms of consumption.

As far as it concerns the end-users, this functionality does not have a direct impact on them, as it is aimed at monitoring and reducing CO_2 emissions. However, as an optional functionality, it is intended to show end-users in a graphical way the consumption of the models they use through appropriate visualizations and charts.

6.11 FML Deployment

The FML deployment of FAME will be destined to support privacy friendly and energy efficient data sharing in Federated Learning (FL) scenarios where many different clients contribute data assets. Without sending the raw data, FL trains the models on each user's device at the edge as an individual client. This means that each device only gathers the data required to carry out its role during the FL process. Through this way, FL secures private raw data, minimizes data leaking, and assures secure operations, since it retains raw data on the device and only communicates model changes to the central server.

6.12 EmFi Training

FAME will offer to the end-users of its platform with access to a pool of training assets' resources, notably resources related with EmFi applications. To this context, the EmFi Training functionalities of the FAME platform will enable its end-users to: (i) Search and access training resources, including courses, webinars, tutorials, how-to videos, and FAME-related demonstrators, (ii) Search and access knowledge resources, including research papers, whitepapers, blog posts and other knowledge-related content items, (iii) Search and access the contents of the training resources (e.g., courses) catalogue, (iv) Comment and rate training resources to foster collection and publishing of community feedback, and (v) Integrate new EmFi training resources. Some of the above-listed will be accessible only to properly authorized users, while others will be made available to all the users of FAME.

6.13 Dashboard

The main reason behind the Dashboard of FAME is to provide to all of its end-users (both technical and non-technical stakeholders), a user-friendly end-to-end UI that will facilitate their interaction with all the involved FAME services, components, and processes, communicating with them via their Open APIs. In particular, a dashboard will be designed, specified, and implemented in such a way that it will enhance user experience, driving user engagement, and ultimately determining the success of FAME. It will focus on providing an interface that will be aesthetically pleasing, simple to use, inviting the FAME end-users to further explore all the associated capabilities and components. As a result, user experience will be impacted, with an emphasis on streamlining complicated FAME-related activities (such as searching, monetization, and trading), guaranteeing fluid navigation, and ensuring that end-users can perform their tasks with ease.

Consequently, the Dashboard’s primary goal will be to maintain consumers’ interest throughout time. It will have visually appealing components, engaging tools, and considerate animations to produce a captivating and enjoyable user experience. As a result, the interacting end-users will be more likely to use FAME capabilities, discover its features, and spread their positive experiences, which will promote FAME acceptance and growth. To ensure that all the end-users can connect effectively and to make FAME accessible to a wider audience, it will also consider accessibility elements including appropriate colour contrast, font sizes, and sensible navigation options. This will improve its usefulness and reach.

Considering all these features, a strong brand identity will be built, helping towards transforming FAME to a widely known single-entry point of data assets and functionalities related with EmFi applications. FAME’s value will be reinforced, establishing a strong connection with all the interacting end-users, prioritizing both technical and non-technical stakeholders, thus creating a consistent experience across diverse related Data Spaces and Data Marketplaces.

7 FAME Solution Architecture

Current Section includes the overall FAME SA that has been designed considering all the related activities of Section 3, Section 4, Section 5, and Section 6 that have led the design of the current version of the SA. Since as it has been already introduced the C4 architecture model has been selected as the best candidate to represent FAME's SA, a further analysis of this model is provided. The FAME SA overview is then outlined, being followed by a detailed explanation of each of the different views of the C4 model and the interacting technical components. Afterwards, the technical specification of the different interacting technical components is provided, followed by a description of their goal, their related architectural figure based on the C4 model, as well as an analysis of their underlying techniques and baseline technologies.

7.1 C4 Architecture Model

Modelling and validating systems is often performed with low-fidelity software models and disjointed architectural specifications by various engineers using their own specialized notations. These models are typically not maintained or documented throughout the systems' lifecycle, making it difficult to predict the impact of any emerging changes performed upon the system, which in turn may cut across the systems' overall functionalities. The unanticipated effects of designing approaches or changes are discovered only late in the lifecycle when they are much more expensive to be resolved. A **Model-based Engineering (MBE) approach**, and thus a model-based system architecture, offers a better way to design, develop, analyse, and maintain a system's architecture. By applying an MBE approach, software development teams can: (i) reduce risk through early and repeated analysis of the system architecture, (ii) reduce cost through fewer system integration problems and simplified lifecycle support, (iii) assess system-wide impacts of architectural choices, and (iv) increase confidence since the assumptions made in modelling can be validated in the operational system.

Towards this direction, the **C4 model** was created to help software development teams to describe and communicate their systems' architectures, both during up-front design sessions and when retrospectively documenting an existing codebase. Through this model system, the software development teams can effectively create maps of their code at various levels of detail [74]. In essence, the C4 model is an *abstraction-first approach* for diagramming a system architecture, based upon abstractions that reflect how the software development teams think about and build their system.

In short, the C4 model was built by Simon Brown based on UML [75] and the 4+1 architectural view model [76]. This model breaks down software into smaller units for modelling, distinguishing the level of a diagram into system context, container, component, and code. The name, C4, represents the number of levels in the model that start with the alphabet C. Like the agile methodology, the C4 model is recommended when it requires quick and efficient sharing and constant updates of a software architecture during ongoing software development, as in the case of the FAME research and innovation project.

To efficiently describe a system's architecture, the C4 model indicates two (2) basic concepts: (i) *Abstractions* to describe the architecture, and (ii) *Diagrams* to visualize the abstractions. More specifically, the four (4) supported diagram levels (i.e., Context, Container, Component, Code) act as the visual map of the system with defined levels of abstractions. In turn, these abstractions are designed for different audience types, ranging from non-technical users (high level diagrams) to developer-focused users (low level diagrams) [77].

A summarization of the model's supported *abstractions* is given below:

- **Person**: Any end-user who uses the system.
- **System**: The highest level of abstraction that delivers value to end-users, whether they are human or not.
- **Container**: Any application, datastore, microservice, etc. that makes up the system (being independently deployable and runnable).
- **Component**: Any building block and module that makes up each container of the system.

As for the four (4) supported *diagram types*, since they represent the visual representations of how a system works (exploiting the abovementioned abstractions), they are split up into the levels of:

- **Level 1 [Context Diagram]**: High-level overview of the built system, illustrating how it fits into the world in terms of the people who use it and the other software systems it interacts with. The involved parties are *People* and *Systems*.
- **Level 2 [Container Diagram]**: Zoomed-in view of the built system, showing the containers being executed inside it. The involved parties are *People*, *Systems*, and *Containers*.
- **Level 3 [Component Diagram]**: Zoomed-in view of each container of the built system, depicting the individual components making it up, their responsibilities and their implementation details. The involved parties are *Systems*, *Containers*, and *Components*.
- **Level 4 [Code Diagram]** (*optional*): UML class diagrams of each component, zooming into the component to illustrate how the component is implemented – usually this diagram can be autogenerated from the actual code of each component. The involved parties are *Components*.

It should be noted that there exist supplementary diagrams to fill in information gaps by showcasing views such as deployment information, sequence of events and how systems interact at a higher level, which however are not considered as mandatory into the context of realizing the FAME SA [74].

An overview of the overall C4 Model's supported diagrams is depicted in Figure 15. Essentially, a software system can be made up of one or more containers (applications, services, datastores, etc.), each of which contains one or more components, which in turn are implemented by one or more code elements (classes, interfaces, objects, functions, etc).

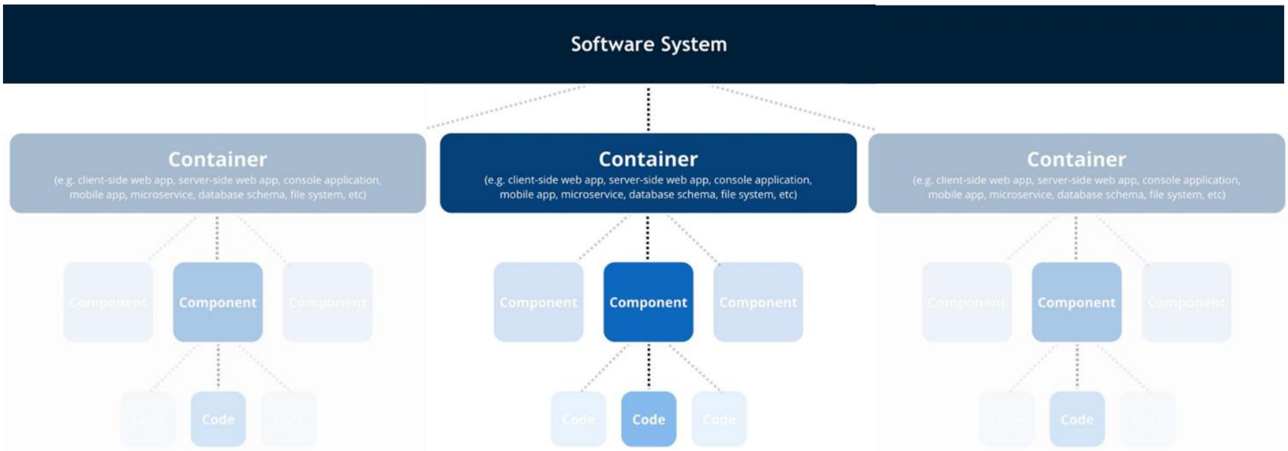


Figure 15 – C4 model diagram types [78]

Finally, as for its *notations*, the C4 model does not prescribe any notation, but instead this can be chosen by each system's architects and developers, making sure that all the chosen notations remain consistent across each level of detail of all the diagrams.

Following all the abovementioned principles and guidelines, FAME has adopted the C4 model for illustrating its designed SA, applying in all the constructed diagrams (i.e., Context, Container, Component) the notations illustrated in Table 2, Table 3, and Table 4 respectively.

Table 2 – FAME SA Context Diagram notations

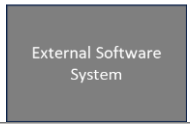
Abstraction	FAME Notation	Notation Description
Person [*EmFi Application*]		Any EmFi application that uses the FAME Federated Data Space.
Person [*User*]		Any end-user that uses the FAME Federated Data Space.
Person [*Data Space*]		Any external data space/data marketplace user that uses the FAME Federated Data Space.
System [*Internal*]		Any internal system included into the overall FAME Federated Data Space.
System [*External*]		Any external system interacting with the overall FAME Federated Data Space.
Relationship		Any relationship between all the abovementioned abstractions.

Table 3 – FAME SA Container Diagram notations

Abstraction	FAME Notation	Notation Description
Person [*EmFi Application*]		Any EmFi application that uses the FAME Federated Data Space.
Person [*User*]		Any end-user that uses the FAME Federated Data Space.
Person [*Data Space*]		Any external data space/data marketplace user that uses the FAME Federated Data Space.

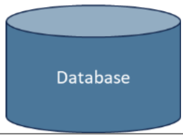
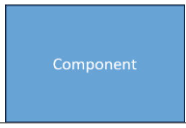
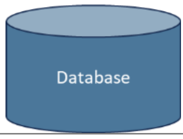
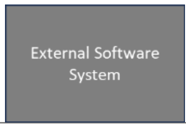
Container [*Any type*]		Any container executed inside the depicted system of the FAME Federated Data Space.
Container [*Database*]		Any database exploited inside the depicted system of the FAME Federated Data Space.
Container [*Web Browser*]		Any web browser (or UI) exploited inside the depicted system of the FAME Federated Data Space.
Relationship		Any relationship between all the abovementioned abstractions.

Table 4 – FAME SA Component Diagram notations

Abstraction	FAME Notation	Notation Description
Component		Any component executed inside the depicted container of the FAME Federated Data Space.
Container [*Any type*]		Any container interacting with the depicted container of the FAME Federated Data Space.
Container [*Database*]		Any database exploited inside the depicted container of the FAME Federated Data Space.
Container [*Web Browser*]		Any web browser (or UI) exploited inside the depicted container of the FAME Federated Data Space.
System [*External*]		Any external system interacting with the depicted container of the FAME Federated Data Space.
Relationship		Any relationship between all the abovementioned abstractions.

7.2 FAME Architecture Overview

7.2.1 Data Spaces Building Blocks

Based on the identified project goals and the collected technical and use case requirements (documented in D2.1 [1]), the initial version of the FAME SA has been compiled, capturing all the FAME systems and components in combination with the various interactions among them, having followed the C4 model's design principles. At the same time, the FAME SA has considered existing different approaches upon the components that a Data Space should have, referring both to the RAs described in Section 5, and more importantly to the specifications of Gaia-X [60] and the IDS-RAM [34], characterised by an **open, reliable, and federated architecture for cross-sector data exchange**.

Figure 16 depicts a taxonomy of the **technology building blocks** that are necessary for realizing a Data Space in a secure and controlled manner, based on the blocks identified in the white paper Design Principles for Data Spaces [79].

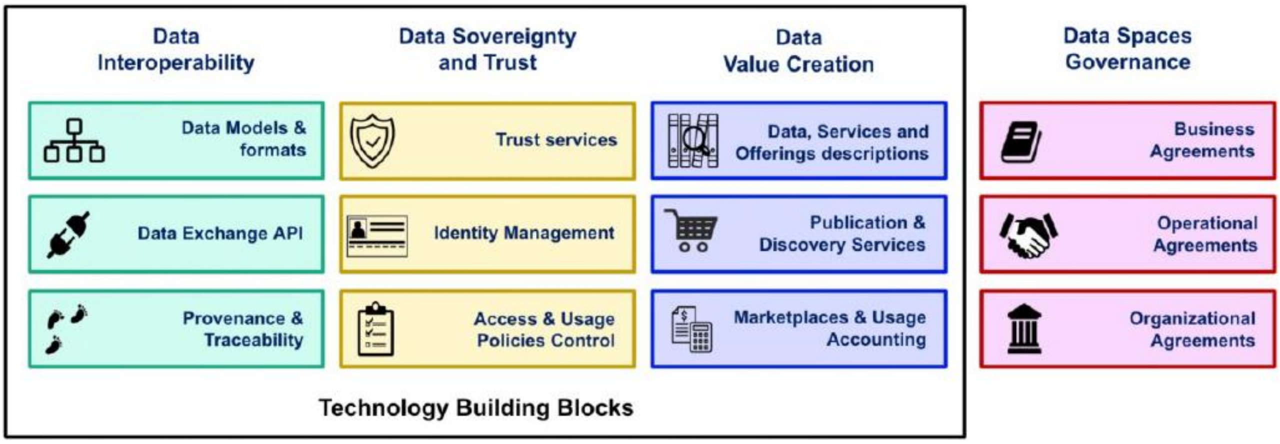


Figure 16 – Technology building blocks towards Data Spaces realization [80]

In short, the technology building blocks that are needed to materialize a Data Space should support:

- *Data Interoperability*: Data interaction between participants should be facilitated effectively through Data Spaces, which should fully support the decoupling of data service providers and customers. This necessitates that common APIs for data interchange and the establishment of common data models to be adopted as a “common lingua” that is used by all participants. Additionally, it is necessary to have common protocols for data provenance and data exchange transaction’s traceability.
- *Data Sovereignty and Trust*: Data Spaces should provide the technical tools that are necessary to ensure that users may trust one another and maintain control over the shared data. This necessitates the establishment of uniform standards for handling participant identities, confirming their veracity, and enforcing agreed-upon data access and usage control measures.
- *Data Value Creation*: Data Spaces should facilitate the development of multi-sided markets where users can profit from the exchange of data (i.e., building data value chains). This necessitates the adoption of standard methods for the description of the terms and conditions (including pricing) associated with data services and data service offerings, the publication and discovery of such offerings, and the accountability of all steps during the lifecycle of the contracts established when a particular participant obtains the rights to access and use a particular data service.

Besides the adoption of a common technology foundation, Data Spaces also require some specific *Governance* building blocks, matching a number of business, operational and organizational agreements among participants. Business agreements, for instance, outline the kinds of terms and conditions that might govern how participants share data and the legal grounds for contracts made through the data space. Operational agreements, on the other hand, regulate policies that must be enforced during a Data Space’s operation like, for example, compliance with GDPR or Payment Services Directive 2 (PSD2) in the finance sector. They might also include guidelines for the tools that administrators of cloud infrastructures or of international services that support Data Spaces must use, allowing for the auditing of certain procedures or the adoption of cyber-security standards. Finally, organizational agreements establish the governance bodies. They deal with identifying precise requirements for goods using technical building blocks in a Data Space, as well as the commercial and operational agreements that need to be adopted.

Figure 17 provides an overview about how the building blocks from Figure 16 can be mapped to a systems' architecture aiming to realize a Data Space.

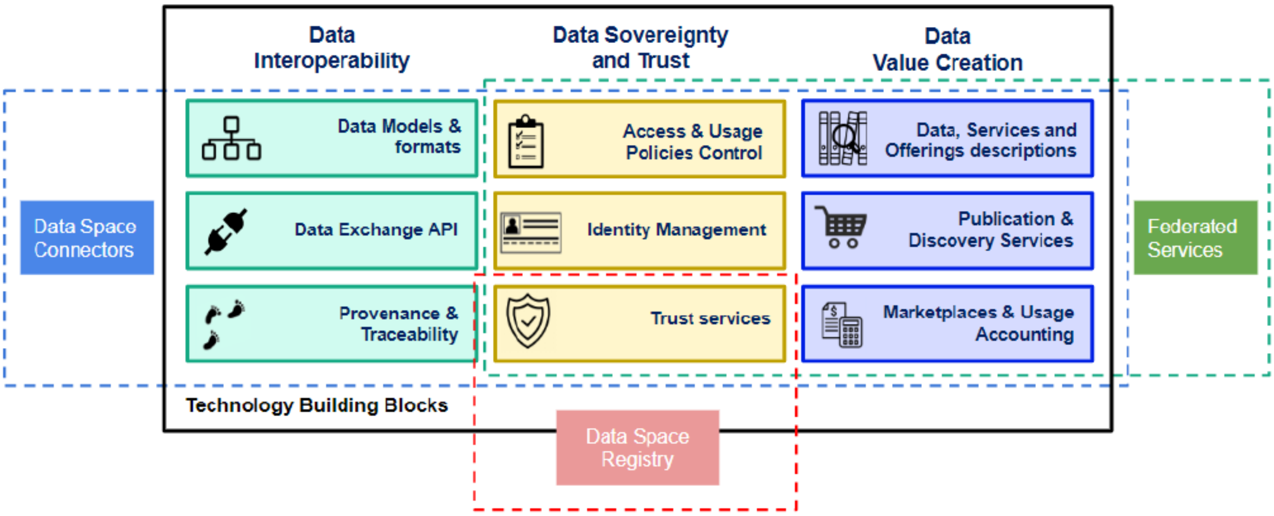


Figure 17 – Mapping of technology building blocks towards Data Spaces creation [81]

All the abovementioned have been considered towards the realization of the 1st version of the FAME SA, where all the identified both technology and governance building blocks are going to be developed into the context of FAME (furtherly described in sub-Section 7.2).

7.2.2 FAME SA Conceptual Overview

Before diving into the FAME C4 model architecture, since as it has been introduced FAME is a quite complex platform integrating existing technical solutions and technologies, as well as implementing new ones, to make sure that the different system/components of the FAME Federated Data Space are simply defined for the external users and systems, the so-called **Conceptual View** of the FAME SA is provided (Figure 18). In more detail, Figure 18 provides a conceptual overview of the FAME architecture that targets on specifying how both the Data Consumers, as well as the Data Producers can benefit from such solution. In deeper detail, following the **top-down approach** in Figure 18, any ***Data Consumer*** (e.g., SMEs, Data Scientists, Research Centre) who wants to access the FAME provided features can be authenticated and authorized access using the appropriate components, being then able to search and identify the desired data asset for further usage. On the other side, following the **bottom-up approach** in the same figure (Figure 18), it can be identified the pathway that can be followed from any ***Data Provider*** (e.g., Data Space, Data Marketplace, Dataset owner, Database owner, etc.), to connect to FAME and index the data assets to be provided, for further exploitation.

In both flows, all the technical and business values as well as the objectives of FAME are considered (e.g., data sovereignty, blockchain tokenization, federated data access, energy efficient analytics). However, for the sake of clarity and to avoid repetitions, in this part of the deliverable, only a high-level description of the FAME flow is described. Additional details follow in sub-Section 7.2.3 along with the FAME C4 model of the 1st (i.e., context diagram) and the 2nd (i.e., container diagram) levels of the architecture, describing the overall flow of FAME, the interactions among the offered systems and their components, as well as the added-value for all of its interacting stakeholders (i.e., end-users).

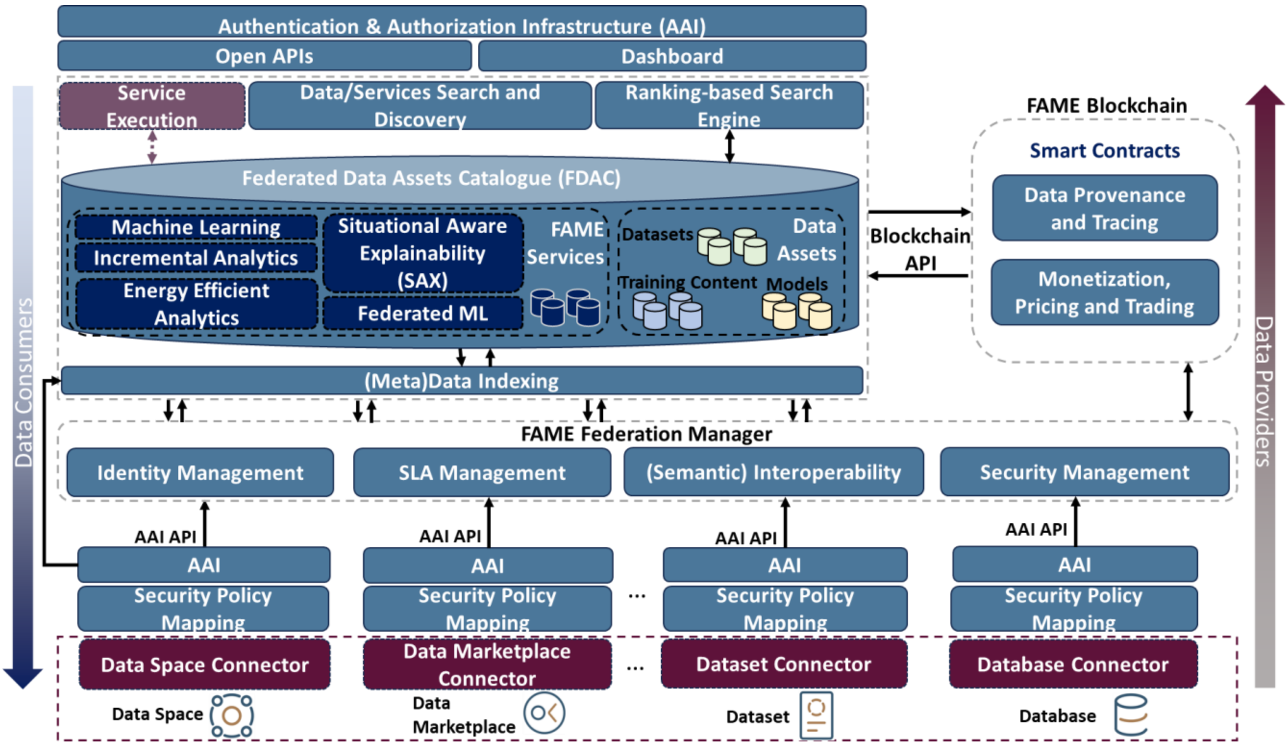


Figure 18 – Conceptual overview of FAME architecture

7.2.3 FAME SA C4 Model

At a broad level, the FAME SA's consists of five (5) discrete layers, namely the: (i) ***Dashboard***, (ii) ***Open APIs***, (iii) ***Federation Manager***, (iv) ***Transactional Operations***, and (v) ***Energy Efficient Analytics Services***. In greater detail, the FAME Federated Data Space supports the interaction and integration with external stakeholders (furtherly outlined in Section 4). All these parties can either contribute their own data assets (e.g., datasets, AI/ML models, analytical models, data insights, data visualizations, software, publications), or access the FAME's existing ones, also having the ability to gain access to all the FAME functionalities via open APIs. To make all this feasible, FAME makes publicly available a user-friendly dashboard that acts as a single-entry point for all the integrated external data assets, facilitating all its registered and authenticated users to discover, access, monetize, price, and possibly trade their underlying assets, which coexist in a federated place, the FAME FDAC. In essence, the latter comprises both semantically annotated data assets and direct points to non-annotated assets of the underlying data sources, offering diverse APIs for accessing, searching, and querying the FDAC's assets, thus ensuring a seamless and platform-agnostic experience for the users of FAME (i.e., users can discover data assets across different data sources like data marketplaces/data spaces). However, for all of this to become feasible, the unique features of each one of the FAME Federated Data Space's layers must put into place, as described in the remaining of this Section.

To begin with, Figure 19 illustrates the Context (i.e., 1st level) diagram of the FAME SA, depicting the diverse groups of stakeholders that are able to interact with FAME (i.e., Data Providers, Data Consumers, Applications Owners, FAME Admin), highlighting the means of how and through which FAME systems (i.e., layers) the involved end-users can exploit the FAME functionalities that are offered via the relevant constructed layers (i.e., Dashboard, Open APIs, Federation Manager, Transactional Operations, and Energy Efficient Analytics Services). To this end it should be mentioned that the notations illustrated in this Figure refer to the ones described in sub-Section 7.1 in Table 2.

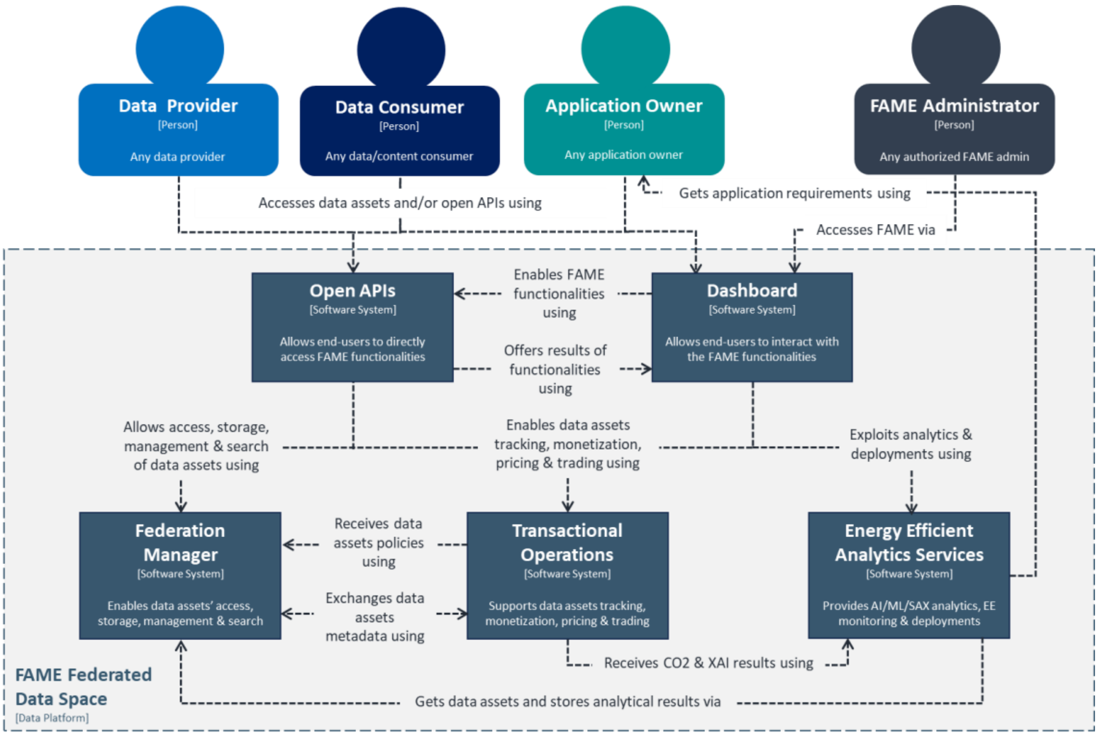


Figure 19 – FAME SA C4 context diagram

Diving deeper into the Container (i.e., 2[nd] level) diagram of the FAME SA, a zoomed-in view of the built platform is provided in Figure 20, showing the containers being executed inside its diverse systems. Hence, in this diagram, all the inner containers of the FAME five (5) systems (i.e., layers) are depicted, referring to the ***Dashboard***, the ***Open APIs***, the ***Federation Manager***, the ***Transactional Operations***, and the ***Energy Efficient Analytics Services***. As in the case of the previous diagram, the notations illustrated in this Figure refer to the ones described in sub-Section 7.1 in Table 3.

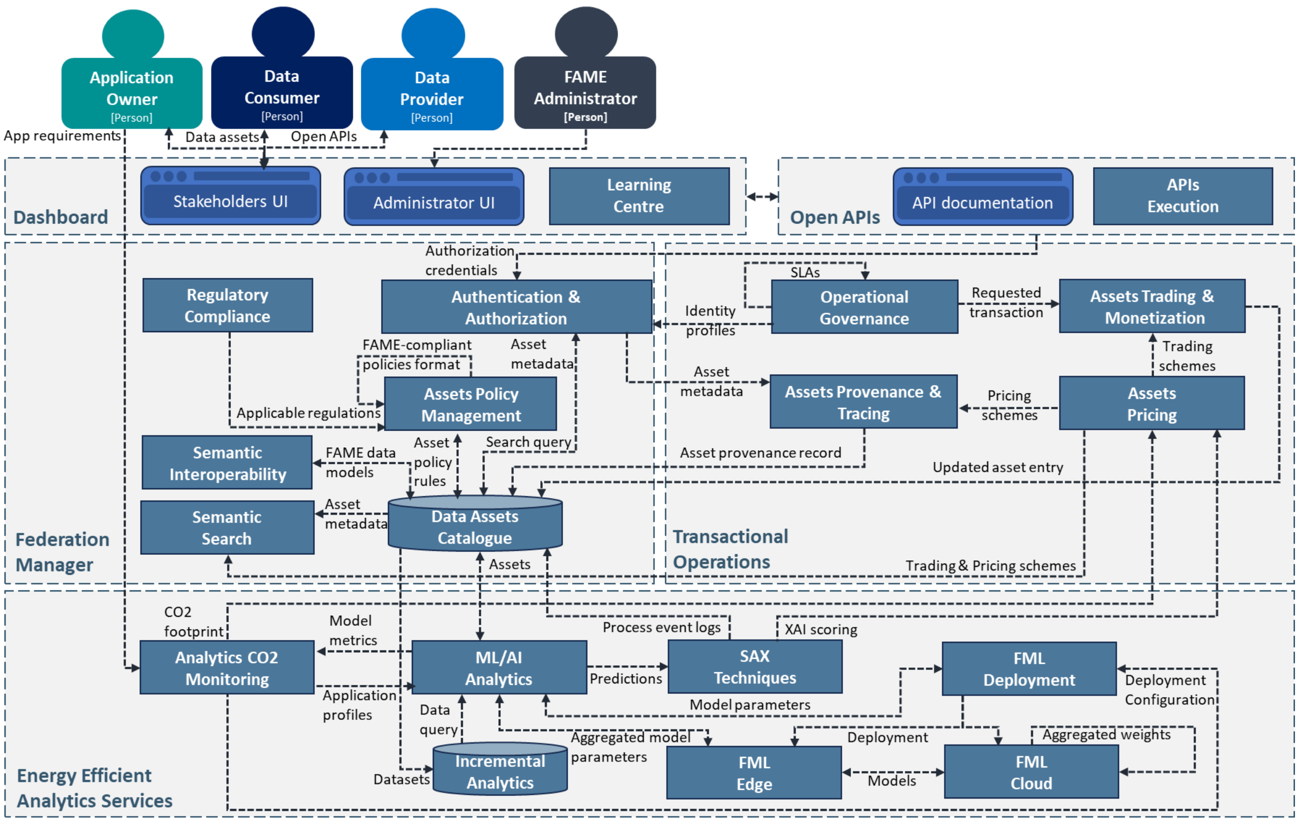


Figure 20 – FAME SA C4 container diagram

In deeper detail, the interaction among the FAME layer's (and their respective systems') containers towards the realization of a **Federated Data Space** is described in the following sub-Sections.

7.2.3.1 Dashboard & Open APIs

The ***Dashboard*** layer contains all the components that are required for **providing access to the FAME end-users** (i.e., stakeholders) regarding the supported functionalities of the FAME Federated Data Space. These end-users can be categorized into the role of a data provider, a data consumer, an application owner, as well as the FAME administrator, all of them having the ability to interact with a variety of ways with FAME. These include diverse UIs for accessing the FAME functionalities, along with a Learning Centre for providing training content and material, all of which can be directly accessed via the ***Dashboard*** layer. What is more, there exists the ***Open APIs*** layer that is closely related to the ***Dashboard*** layer, acting as an additional interacting point with the FAME Federated Data Space. In deeper detail, the components' functionalities and interactions, as well as their information flow within the ***Dashboard*** & ***Open APIs*** layers is as follows:

Whether an external data provider, a data consumer, or an application owner intends to connect and interact with FAME, they have to access the FAME ***Dashboard*** in which they are utilizing the *Stakeholders UI*. The latter is offering a user-friendly and easy-to-use UI for immediate access to the required FAME functionalities to both technical and non-technical users. In the case that the FAME administrator needs to access FAME for administration purposes or for any system parameterization/ maintenance, the *Administrator UI* is offering a related UI for such purposes. The interconnection between these UIs and the FAME functionalities occurs with the corresponding open APIs through the *APIs execution* service (lying into the ***Open APIs*** layer), all of which provide access to the selected operation. The *API documentation* facilitates this process by providing a proper documentation for describing each API along with specific examples, tutorials, and exploitation guidelines.

Finally, to boost its usability and adoption, FAME is also providing a *Learning Centre*, which aims to include two (2) diverse training programs, covering both technical and non-technical users of FAME. The programs will be provided as Massive Open Online Courses (MOOCs) being accessible through the Dashboard for all the interacting stakeholders, while the Learning Centre will emphasize on data assets related with EmFi applications.

7.2.3.2 Federation Manager

This layer contains all the components that are needed for realizing the **federated functionalities** of FAME, ranging from the support of registration, authentication, and authorization of federated users and data sources, to the sharing and indexing of data assets deriving from such federated parties (i.e., external users and data sources). More specifically, the components' functionalities and interactions, as well as their information flow within the ***Federation Manager*** layer is as follows:

Whether an external data provider, or a data consumer, or an application owner or just the FAME admin intends to connect and interact with the FAME Federated Data Space, after accessing the respective FAME *UIs*, the user must be authenticated and granted access through the *Authentication & Authorization* infrastructure. Thus, this infrastructure has the capacity to control access to different types of end-users and their data sources (i.e., data marketplaces, data spaces, databases, etc.), providing them with all the needed authorization credentials, to access all the data assets from the connected (federated) data sources, whilst exploiting all the functionalities of FAME. To this end, it should be noted that in order to create the end-user's identify profiles for his/her successful registration within the platform, the *Operational Governance* component is exploited, which in turn exploits the *Assets Trading & Monetization* component in the context of access control for initializing the registered end-user's monetization and/or trading accounts (furtherly described in the ***Transactional Operations*** layer).

By the time that an interested party is granted access, the latter is able either to directly query the *FDAC* for finding relevant data assets of interest or contribute his/her own assets by providing the metadata that describe the data asset to be uploaded into FAME. In the first case (i.e., data assets' search and discovery), the logged in party is provided with the relevant results and no other functionalities should be invoked by the FAME platform. In the second case (i.e., data assets' index), the logged in party must be guided through the rest of the FAME platform's functionalities, where the actions of the ***Transactional Operations*** layer occur (furtherly described in sub-Section 7.2.3.3).

As soon as these actions are completed for finally creating the required assets' provenance record (i.e., metadata regarding the asset to be indexed) into the *FDAC*, the *Assets Policy Management* takes place to provide the means for accessing the policies of the diverse data assets from the integrated data sources, while exposing them in a FAME compliant format based on standards-based policy languages (e.g., XACML), thus mapping FAME policies to the lower-level policies of the underlying data providers. To initialize such actions, the *Assets Policy Management* is notified by the *FDAC* that a new data asset has been indexed within its catalogue, thus producing by its side the relevant policy rules of the asset. Moreover, the *Assets Policy Management* has the capability to leverage the *Authentication & Authorization* infrastructure, compromising diversities among the different policies for use cases where data assets from different policies should be combined (e.g., specifying the most restrictive policy when needed or the most permissive policy whenever allowed). On top of these, the *Assets Policy Management* exploits the *Regulatory Compliance* component for providing legal support upon the security and the data policies of the assets to be indexed, assuring that the latter will obey to relevant applicable regulations (e.g., PSD2, GDPR, Markets in Financial Instruments Directive (MiFiD), EU taxonomy for Environmental, Social, and Governance (ESG) investments, EU AI Act).

By the time that this process gets complete, the *Assets Policy Management* forwards back to the *FDAC* the new policy information of the asset. However, this is not where the entire asset's management process gets complete. Rather than storing each data asset's information in its raw format, prohibiting its connectivity with other relevant data assets and AI/ML models (furtherly analyzed in ***Energy Efficient Analytics Services*** layer), the *Semantic Interoperability* component provides a semantic middleware to transform all this data from its source formats and semantics (i.e., EmFi applications, data marketplaces/data spaces, users' data) to the FAME ontologies. As a result, this middleware specifies the FAME ontologies and models for EmFi applications and captures the ontologies from other sectors (e.g., retail, smart cities, healthcare) to link them and provide an interoperable view of all the data assets included into the *FDAC*. To successfully perform such task, this middleware includes various data pre-processing functionalities (i.e., data filtering and cleaning) to ease the above listed transformation.

On top of these, this layer offers the functionality of semantically searching the assets lying in the *FDAC*, by putting in place the *Semantic Search* engine. To initiate its functionality, the engine takes as an input the end-users search criteria from the FAME ***Open APIs*** layer (analyzed in sub-Section 7.2.3.1), also presenting the search outcome to the end-users. To produce such results, this engine takes advantage of the assets' metadata as well as their constructed trading and pricing schemes that are produced by the ***Transactional Operations*** layer (furtherly described in sub-Section 7.2.3.3) to provide the relative ranking of the assets' search results and determine their demand-driven price accordingly.

7.2.3.3 Transactional Operations

This layer contains enabling the components that are mostly related to the **daily operation** of a Data Space that supports **trading** – i.e., the online process through which the provider and the consumer of a given data asset can stipulate a legally-binding agreement that determines the **terms and conditions of use**, and possibly includes the **contextual execution of a payment** by means of a digital currency. To this context, the enabling components' functionalities and interactions, as well as their information flow within the ***Transactional Operations*** layer are described below, following the scenario that has been analyzed for the ***Federation Manager*** layer:

Whenever any data asset is requested to be indexed or traded within FAME and the related end-user has been authenticated within the FAME Federated Data Space, the *Assets Provenance & Tracing* component gets as an input the assets' metadata from the ***Open APIs*** layer (analyzed in sub-Section 7.2.3.1) in order to further support the functionality of the *FDAC* (analyzed in sub-Section 7.2.3.2) by extending its metadata model with additional attributes that are stored separately on a blockchain. The most important of these attributes is the catalogue entry's digital fingerprint - i.e., the immutable hash value calculated from the whole entry. By the time that this information is circulated, the *Assets Pricing* component can be applied by the provider of the asset to calculate the cost that will be sustained for its provisioning, adding the produced pricing schemes to the metadata of the asset's entry to the *FDAC*.

As an alternative option, an end-user may be interested in trading hir/her data assets, by exploiting the *Assets Trading & Monetization* component. To perform such action, the latter enables transaction-based trading of data assets by receiving its input from the FAME ***Open APIs*** layer (analyzed in sub-Section 7.2.3.1), whereas to produce its trading schemes it exploits the *Assets Pricing* component. What is more, when a new offering is defined or an existing offering is either modified or dropped, the *Assets Trading & Monetization* gets its input from the *Operational Governance* component (furtherly described below), whereas the updated asset entry's metadata is also forwarded to the *FDAC* for its respective updating. As the *Assets Trading & Monetization* persistent storage is implemented on the same blockchain infrastructure that backs the *Provenance & Tracing* component, the component's output is publicly accessible by any user or system.

All the abovementioned components of this layer are supported by the *Operational Governance* component, which provides the core functionalities for the governance of a federated Data Space that supports online trading of data assets. In particular, it is responsible for the onboarding of traders (i.e., end-users that can engage in trading), the storage of offerings (extending the *FDAC* metadata model), the management of subscriptions, the support of auxiliary functions like billing, and the specification of the relevant Service Level Agreements (SLAs) with the underlying data providers.

7.2.3.4 Energy Efficient Analytics Services

This layer brings to FAME the necessary tools for implementing **analytical functionalities** for EmFi applications. This allows end-users to apply AI/ML models to their own data (or data coming from an external source) to obtain **desired outputs** (i.e., forecasting values, ranking systems, sentiment analysis), and to receive an **explanation of the given results**, thus adding business value to client's companies subscribed to the FAME Federated Data Space. **Energy efficiency**, **smart deployment** and **data-serving efficiency** techniques are implemented to optimize the functionalities described. The components building the ***Energy Efficient Analytics Services*** layer, their interactions and their information flow are stated below, following the scenario that has been analyzed in the above layers:

On top of all the abovementioned functionalities (i.e., data assets indexing, search, discovery, trade, etc.), the FAME Federated Data Space also provides access to a set of analytics that can be applied to existing data assets to transform, reform, and derive data-driven insights as new tradeable assets within FAME. To this context, the ***Energy Efficient Analytics Services*** layer provides a library of different state-of-the-art AI/ML models (e.g., Transformers, RNN, CNN, Linear Regressors), represented by the *AI/ML Analytics* catalogue. The latter is also responsible for performing training, model storing, and inference functionalities alongside the proposed models, whereas to assure that the produced models are the best energy efficient choices for the underlying application, the *AI/ML Analytics* receives as an input from *the Analytics CO_2 Monitoring* component (furtherly described below) the application's produced profile.

Rather than producing those plain models, this layer also the utilizes the models' trustworthiness by leveraging XAI techniques via the *SAX Techniques* component, which offers a framework for scoring the explainability of the different constructed AI/ML models of the *AI/ML Analytics* catalogue, towards comparing alternative approaches, balancing performance and explainability trade-offs. At the same time, this component supports the production of the relevant data assets' trading and pricing schemes, an information that is directly forwarded to the *Assets Pricing* component. The employment of XAI is further enhanced with proprietarily developed techniques for context and process aware enrichments (i.e., SAX), considering the casual sequencing, the constraints, and the broader contextual information (e.g., temporal) behind decisions and AI employments, as well as the inferential associations among subsequent process enactments of the underlying applications – i.e., exploitation of the process event logs.

As soon as a model is produced, the ***Energy Efficient Analytics Services*** layer offers also the capability to capture the consumption of the AI/ML models deployed in the *AI/ML Analytics*, in terms of CPU/GPU usage and CO_2 footprint emitted. The latter is performed by the *Analytics CO_2 Monitoring* component, which captures the abovementioned metrics to create the model's metadata in order to optimise further deployments. At the same time, this component is also in charge of selecting the most suitable AI/ML model according to the running application. For that purpose, it gets as an input from the underlying application its requirements (e.g., type and locality of the data it handles, real-time/low latency/batch processing), it constructs its respective application profile, where for each profile it provides deployment configurations that optimize CO_2 emissions without compromising the functionality and the expected performance of the application. This information is then sent as an input to the *FML Deployment* component (furtherly described below). Additionally, metrics extracted from the running models can be displayed to the end-users (as an additional feature) to raise their awareness about the CO_2 emissions of the FAME processes.

What is more, by applying the *Incremental Analytics* component, this layer employs a designated incremental data handler aiming to support the execution of the *AI/ML Analytics* by providing an abstraction layer in the cases where an external Big Data source is integrated within FAME for providing its data. Among its functionalities, the *Incremental Analytics* component includes the support for mechanisms that incrementally and continually compute (real-time/run-time) analytical results over previously computed snapshots of queries. Since each of these incremental queries performs a small part of the total query operation, I/O and data transfer operations are reduced, improving the overall energy efficiency of the produced models of the *AI/ML Analytics* component.

Finally, on top of all these analytical functions, this layer also offers the ability to execute the developed analytical functions on the edge, optimizing the power efficiency of the application exploiting those functions. To achieve such task, the optimization concept is based on the *Federated Machine Learning (FML) Deployment* component that considers the application profile that was received by the *Analytics CO_2 Monitoring* component, to better select the devices where the analytics layer will be deployed in a federated manner.

To this end, it should be mentioned that considering the **technology & governance building blocks** that are required to **realize a Data Space** (sub-Section 7.2.1 - Figure 16), the following building blocks - and their included functionalities, have been **aligned with the FAME components** (Figure 21), since they are offering the same type of functionalities:

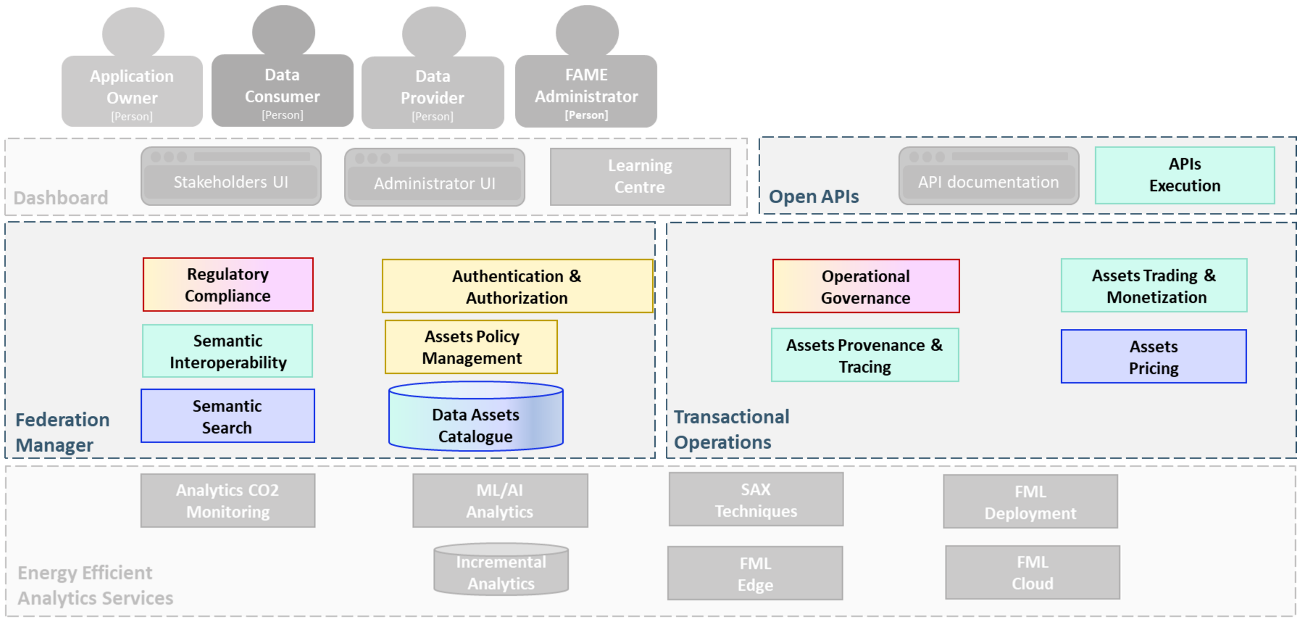


Figure 21 – Mapping of Data Spaces technology building blocks and FAME

Data Interoperability:

¶ The "Data Models & Formats" functionality has been aligned with the *Semantic Interoperability* and the *Data Assets Catalogue* FAME component.

¶ The "Data Exchange API" functionality has been aligned with the *Assets Trading & Monetization* as well as the *APIs Execution* FAME components.

¶ The "Provenance & Traceability" functionality has been aligned with the *Assets Provenance & Tracking* FAME component.

Data Sovereignty and Trust:

¶ The "Access & Usage Policies Control" functionality has been aligned with the *Assets Policy Management* FAME component.

¶ The "Identity Management" functionality has been aligned with the *Authentication & Authorization* FAME component.

¶ The "Trust Services" functionality has been aligned with the *Regulatory Compliance* as well as the *Operational Governance* FAME component.

Data Value Creation:

¶ The "Data, Services and Offerings Descriptions" functionality has been aligned with the *Data Assets Catalogue* FAME component.

¶ The "Publication & Discovery Services" functionality has been aligned with the *Semantic Search* and the *Data Assets Catalogue* FAME component.

¶ The "Marketplaces & Usage Accounting" functionality has been aligned with the *Assets Pricing* FAME component.

Governance:

¶ The "Business Agreements", "Operational Agreements", "Organizational Agreements" functionalities have been aligned with the *Regulatory Compliance* as well as the *Operational Governance* FAME components.

7.3 Specification of Architecture Components

Current sub-Section provides the technical information behind each different component that structures the FAME SA, including its C4 component-level architecture (as it has been defined in sub-Section 7.1), as well as the included underlying techniques and baseline technologies behind its specification and future implementation. To facilitate the overall understanding, for each different component, a short description of its goal is provided as well.

7.3.1 Authorization & Authentication

The FAME *Authentication & Authorization* infrastructure is intended to be implemented over different types of data providers and their data infrastructures (i.e., Data Marketplaces, Data Spaces, Databases, etc.) aiding to control their access to the FAME functionalities and underlying data assets. Towards this direction, the FAME *Authentication & Authorization* leverages the existing i3-MARKET platform [15], which will provide baseline support for self-sovereign identities and access to data from federated marketplaces.

In order to provide authentication and authorization with distributed identity and verifiable credentials, two (2) Node.js microservices will be implemented. The Verifiable Credential microservice that will provide the APIs that implement the core functions to manage verifiable credentials, namely issuing, verifying, and revoking verifiable credentials, and a utility function. The second microservice refers to the OIDC SSI Auth microservice that will provide the API to perform the authorization code flow with PKCE using verifiable credentials as proof method.

The C4 component-level architecture figure of the developed *Authentication & Authorization* infrastructure is illustrated in Figure 22.

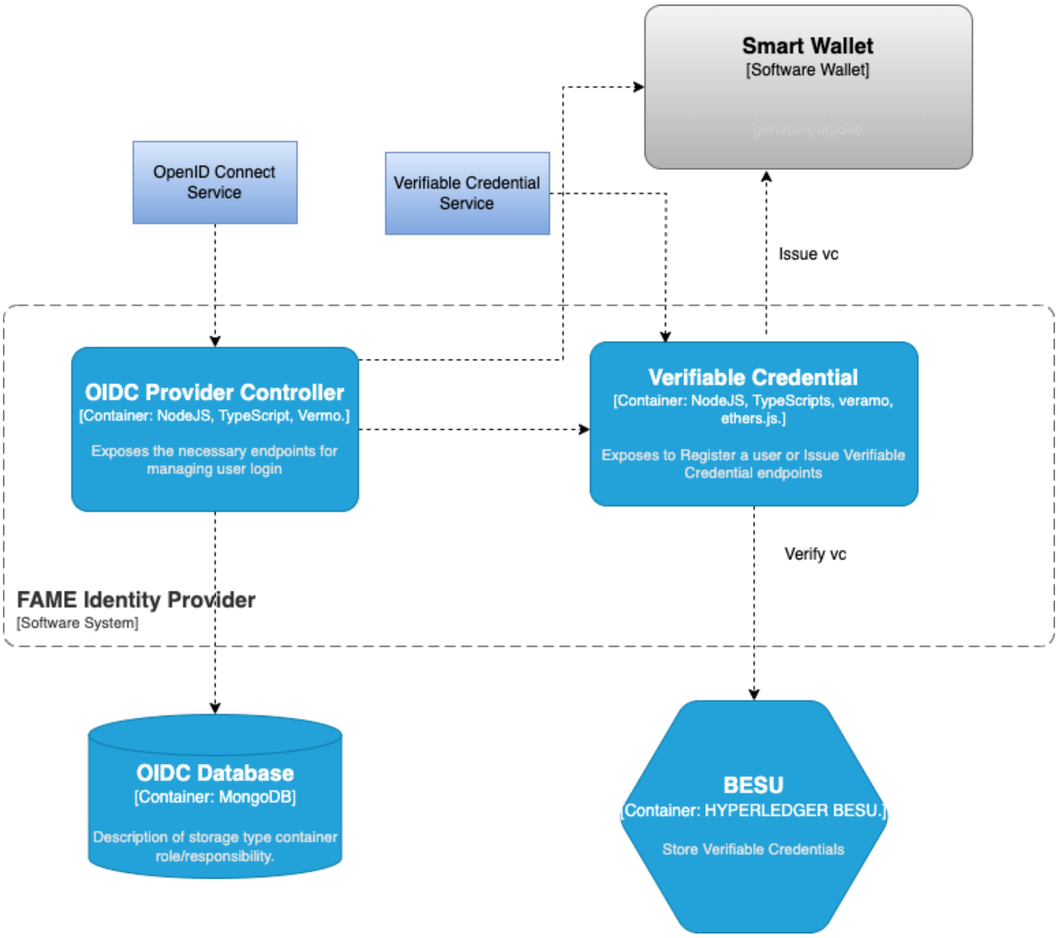


Figure 22 – Authorization & Authentication C4 component-level architecture

To implement this solution, Veramo [82] has been chosen as a framework. Both components (OIDC SSI Auth and Verifiable Credential microservices) integrate the Veramo framework and take advantage of its features to manage DID and verifiable credentials. Each of them has its own OIDC SSI Auth Service and its own Verifiable Credential microservice to generate, verify and revoke verifiable credentials. In relation to the roles of the W3C Recommendations on verifiable credentials [83], the OIDC SSI Auth Service is the verifier, the Verifiable Credential microservice is the issuer (with some extra features), and the user is clearly the holder of his/her verifiable credentials. Each instance of the Verifiable Credential microservice has its own DID [84] and private key used to sign verifiable credentials. In this way, each verifiable credential, has as its issuer the DID of the microservice that generated it. Similarly, for revocation, only the microservice that generated a credential has the privilege to revoke it. Then, the user saves the Verifiable Credentials in his/her wallet and gives an explicit consent to share them with the OIDC SSI Auth Service when requested during the authentication phase.

7.3.2 Assets Policy Management

The *Assets Policy Management* component consists of two (2) distinct sub-components: the Assets Policy Editor and the Assets Policy Engine. Each subcomponent plays a crucial role in enabling end-users to define and enforce access policies for all the assets within FAME.

The first sub-component, namely the Assets Policy Editor, empowers asset owners within FAME to define access policies that regulate who can view and potentially purchase their assets in the platform. This functionality is realized through three (3) levels of access: Confidential, Public, and Restricted. With Confidential access, only the owner of the asset can view it, ensuring utmost privacy and exclusivity. Public access allows all the authenticated users to view the asset, promoting open sharing and collaboration. Restricted access provides more granular control, allowing asset owners to define a set of and/or conditions based on user and organizational attributes (including but not limited to country, organization type, etc.), following the Attribute-Based Access Control (ABAC) model [85]. These conditions act as eligibility criteria for other users to meet viewing the asset. The Assets Policy Editor incorporates a user-friendly UI to facilitate the intuitive definition of policies. Additionally, it provides a REST API that can be leveraged by the UI itself, as well as other components within the FAME platform, such as the *FDAC*, to set policies during external asset indexing processes.

The second sub-component, namely the Assets Policy Engine, acts as the Policy Decision Point (PDP) within the platform. It is responsible for answering two (2) fundamental questions related to asset's visibility and ownership for authenticated users. Firstly, it determines which assets a user can view within the platform, considering the defined access policies and user attributes. This ensures that users are presented with only the assets they are eligible to access. Secondly, the Assets Policy Engine provides information on the assets owned by the user. Ownership includes assets that were uploaded by the user themselves or any other member of their organization, as well as assets that have been acquired with active contracts. To gather information about purchased assets, the Assets Policy Engine interacts with the respective components of the FAME platform. Moreover, to extract the user and organization attributes that are necessary to regulate asset's visibility, the Assets Policy Engine communicates with the *Authentication & Authorization* infrastructure to retrieve the required information before making the appropriate access decisions.

It is important to highlight that comprising the PDP of FAME, any component that needs to present users with asset's information must first contact the Policy Engine to retrieve access eligibility details. Subsequently, these components will act as the Policy Enforcement Points (PEP) of the platform, either allowing or denying access based on the information received from the Policy Engine. This approach ensures consistent and appropriate enforcement of access policies throughout the platform, maintaining a secure and controlled environment for data asset's management and utilization.

By encompassing both the Assets Policy Editor and the Assets Policy Engine, the *Assets Policy Management* ensures that assets' access within FAME is controlled, secure, and aligns with the defined policies. The collaboration between these components allows for user-friendly policy definition and efficient policy enforcement, guaranteeing that users are presented with only the relevant assets that they can access based on their attributes and ownership. The overall idea of the abovementioned approach is depicted in the 3rd level of the C4 diagram (i.e., component diagram) that is presented in Figure 23.

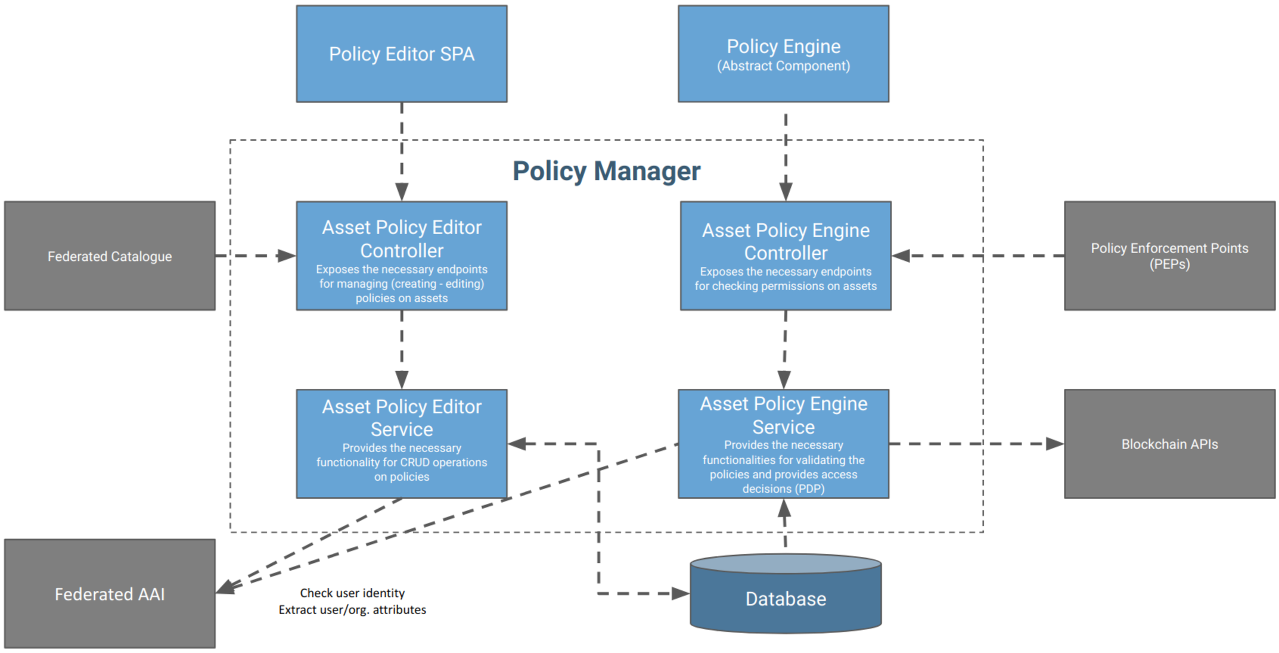


Figure 23 – Assets Policy Management C4 component-level architecture

In terms of implementation technologies, the backend of the Assets Policy Editor and the Engine will be developed based on Spring Boot [86], a Java-based framework, which will be employed to facilitate efficient development and provide robust functionality. The data storage for the backend will be handled by Postgres [87], a reliable and scalable open-source relational database management system. To enable policy definition, storage, and enforcement, both the Assets Policy Editor and the Assets Policy Engine will utilize the JCasbin [88] authorization library. JCasbin offers a comprehensive solution for managing access control policies, providing flexible and rule-based policy management capabilities. Finally, regarding deployment, containerization solutions like Docker will be employed to ensure consistent and reliable deployment across different environments.

7.3.3 Regulatory Compliance

The *Regulatory Compliance* component specifies and implements the security and data policies that will boost the compliance of data assets to applicable regulations in UCs related with EmFi applications. To this end, the security policy management tools of FAME will be used to produce various regulatory support and regulatory compliance tools. The work will be driven by the regulatory requirements that will be specified in the context of the FAME platform and will provide support for regulations like PSDII, GDPR, MiFiD, the EU taxonomy for ESG investments, as well as the emerging EU AI Act. In conjunction with the ethical and legal management part of FAME, this component will also specify how the various tools will be used to support the regulation of the FAME Federated Data Space in-line with applicable laws and directives. Hence, FAME will provide all the technological tools that boost the regulatory compliance of data-driven EmFi applications, while validating ideas for regulating a data market.

However, one of the major drawbacks is that the regulatory landscape of the finance sector has been traditionally very dynamic and volatile. Significant changes in regulations and/or the emergence of new regulations could therefore lead to changes in the FAME Federated Data Space implementation. This could delay the realization of the project's impacts. In response, FAME will offer regulatory compliance tools that will ease the compliance to applicable regulations. It will also provide support for reliable data provenance, which will ease the support for new regulatory rules.

To this context, the *Regulatory Compliance* component exploits a unique tool provided by the European Commission [89] that allows everyone to compare and select open licences based on their content. This tool will be applied to designate relevant licences based on each jurisdiction. This component has already defined the regulations that are relevant to FAME (Figure 24), outlining not only the regulations per se, but also the policies, the standards, and the guidelines for the different stakeholders in the project.

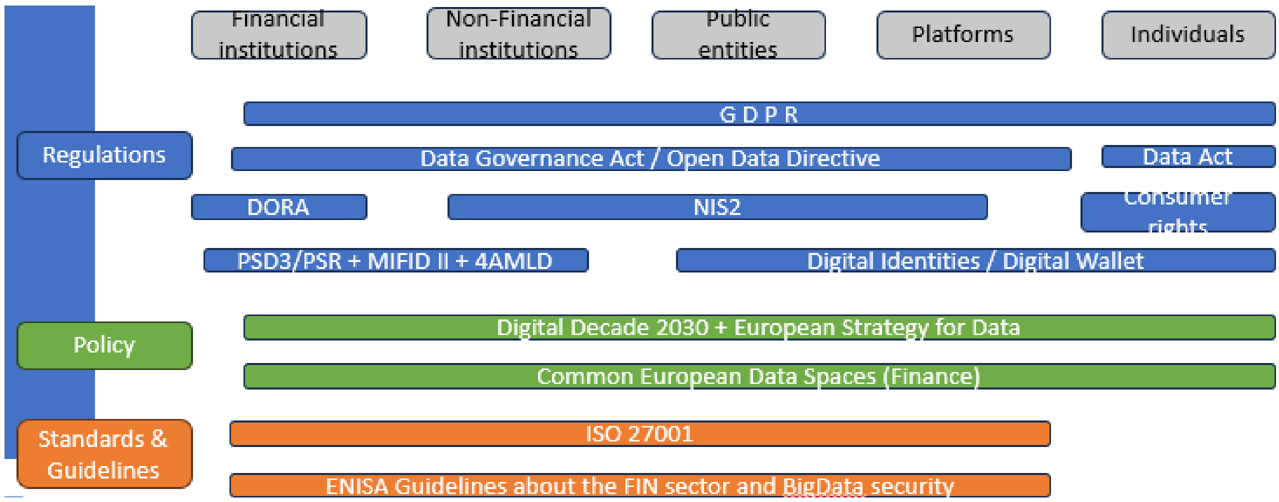


Figure 24 – FAME related laws & regulations

7.3.4 Assets Provenance & Tracing

The *Assets Provenance & Tracing* component provides the means for ensuring the quality of metadata published on the *FDAC*. This is achieved by storing on a blockchain the digital fingerprint of the metadata, so that the integrity of any *FDAC* entry can be verified. Moreover, the same blockchain also maintains a registry of authentic sources, which can be referenced in *FDAC* entries as the provenance of the underlying asset. The overall idea of the *Assets Provenance & Tracing* component is depicted in the 3rd level of the C4 diagram (i.e., component diagram) that is presented in Figure 25.

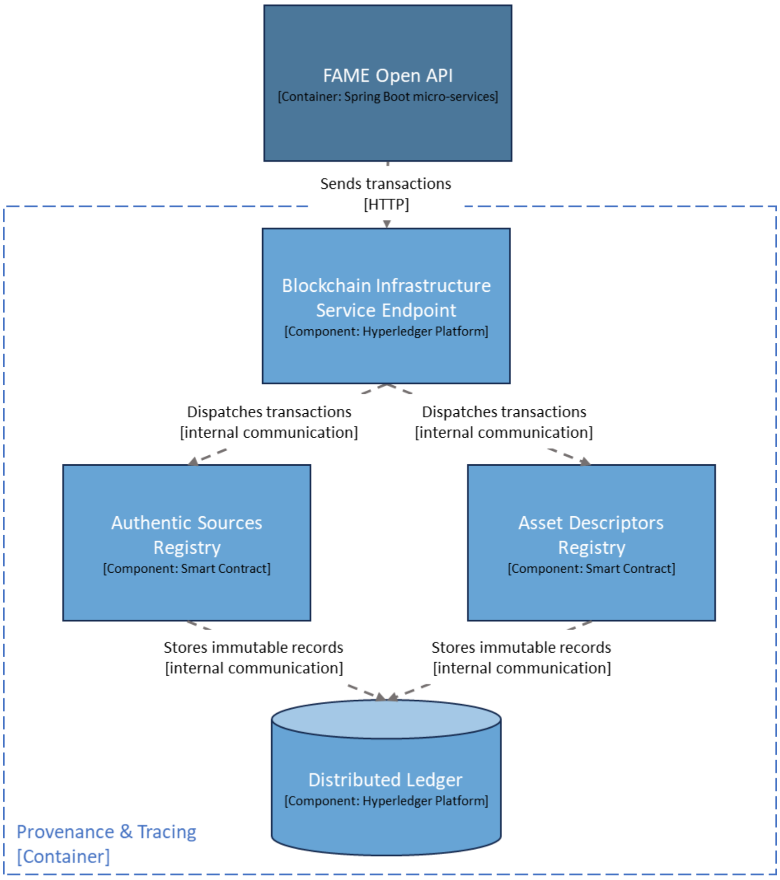


Figure 25 – Assets Provenance & Tracing C4 component-level architecture

The *Assets Provenance & Tracing* component is entirely implemented as Solidity [90] Smart Contracts that are deployed on the FAME's Hyperledger Besu permissioned blockchain network. The network exposes one or more HTTP service endpoints that can be invoked by clients (in the case of FAME through the ***Open APIs*** layer) to execute Smart Contract transactions.

7.3.5 Assets Pricing

The *Assets Pricing* component is closely tight with the *Assets Trading & Monetization* component as well as the *Operational Governance* component of the platform. More specifically, the objective of this tool, which is important for the sustainability of FAME and its business potential, is to price the data assets. Pricing models will be defined for the different types of assets, considering both static variables (based on the nature of the asset/service) and dynamic variables (based on demand), which are automatically extracted from the metadata of each asset, enabled by the *Assets Provenance & Tracing* API.

Objectivity is ensured by the information obtained from the metadata including information about the completeness, the volume, the quality, the timeliness, the CO_2 wastes associated with the asset, the user friendliness, the trustworthiness, the ML-readiness, and additional information related with the offered asset. Input to the *Asset Pricing & Monetization* component will also include weighting factors and values provided by the sellers through questions, which will contribute subjectivity to the final calculation for generating the recommended price of an asset. Then, the output will be a suggested price that will reflect the objective value of the asset and will serve as a reference point for the seller when setting the selling price, being also written to the asset's blockchain information. Finally, this information will be forwarded for use in the *Assets Trading & Monetization* component. The overall idea of the abovementioned approach is depicted in the 3rd level of the C4 diagram (i.e., component diagram) that is presented in Figure 26.

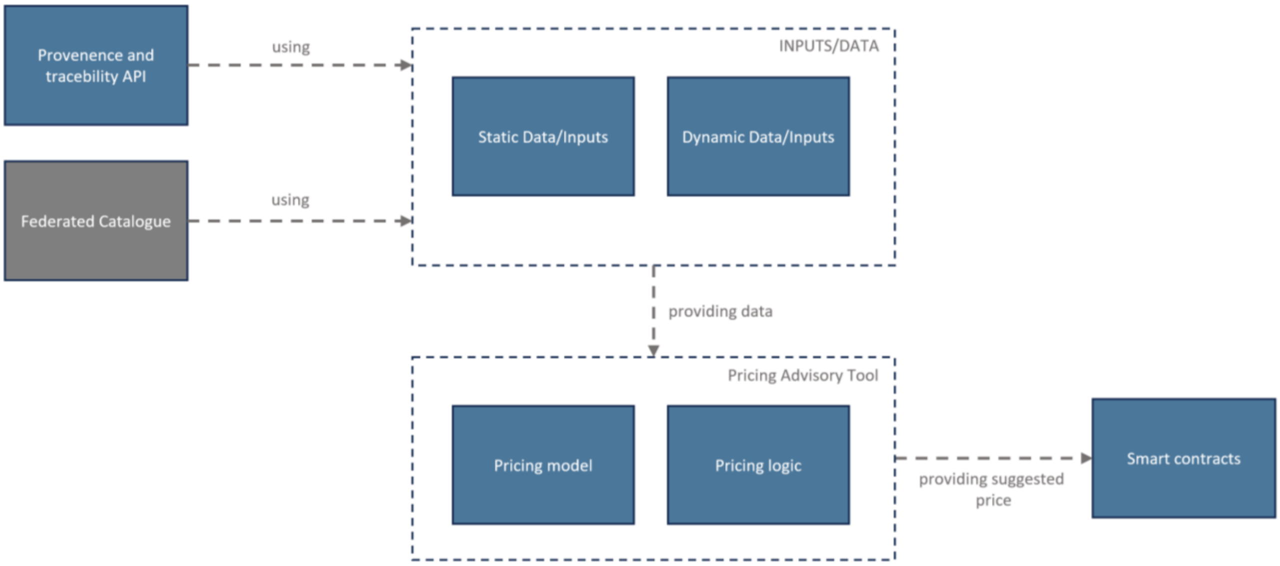


Figure 26 – Assets Pricing C4 component-level architecture

7.3.6 Assets Trading & Monetization

The *Assets Trading & Monetization* component comprises of the "Smart Contract for Trading and Monetization" module, which serves as a key component of FAME, enabling seamless interactions between FAME end-users, who may be providers or consumers of data assets. Users are granted the ability to navigate the *FDAC* and identify data assets that align with their interests. The platform facilitates a trade of access rights for a chosen data asset, which involves an exchange of ERC-20 tokens for ERC-1155 tokens. These ERC-1155 tokens symbolize the right to access the specific data asset. Once the token exchange is successfully completed, the trade is executed, and the consumer is granted permission to access or download the data asset, in compliance with the predefined policies. Importantly, all details pertaining to the trade are meticulously logged on a ledger. This ensures a transparent audit trail, fulfilling essential compliance requirements, allows for the smart contract to run in a tamperproof way and promoting overall transparency in the trading process.

Before jumping into the component-level architecture of this component, it is important to highlight the core containers of this component, which refer to the: (i) REST API that is an open standards API gateway that other FAME actors (i.e., external entities) interact with, (ii) Currency Contracts that is a token payment system that implements the ERC-20 standard, (iii) Data Assets Contracts that refer to the ERC-1155 tokens that provide a representation of data access ownership, (iv) Offerings System that is a ledger for trade offers, implemented using smart contracts, and (v) Trading Contracts that refer to the smart contracts that handle the swapping of currency tokens for data asset tokens.

At the 3rd level diagram (i.e., component diagram) of this component, the goal is to provide greater detail upon the abovementioned containers lying into the Assets Trading & Monetization component, breaking down the containers into its constituent components.

Figure 27 illustrates the diagram for the REST API, including the following components:

- Rest API Router: The main component that handles the routing of requests, where it routes the requests to the appropriate controller based on the request's details. This component is implemented using the Express.js web application framework [91].
- Currency Controller: The component handling the requests related to the Currency Contract.
- Management Controller: The component handling the requests that are related to the Data Access Tokenization.
- Offerings Controller: The component handling the requests related to the Offerings System.
- Trading Controller: The component handling the requests related to Trading Contracts.

Each of these controllers interacts with the Hyperledger Besu blockchain, which is an external system represented as a database in the diagram. The interactions with the blockchain involve various operations related to the specific functionality of each controller.

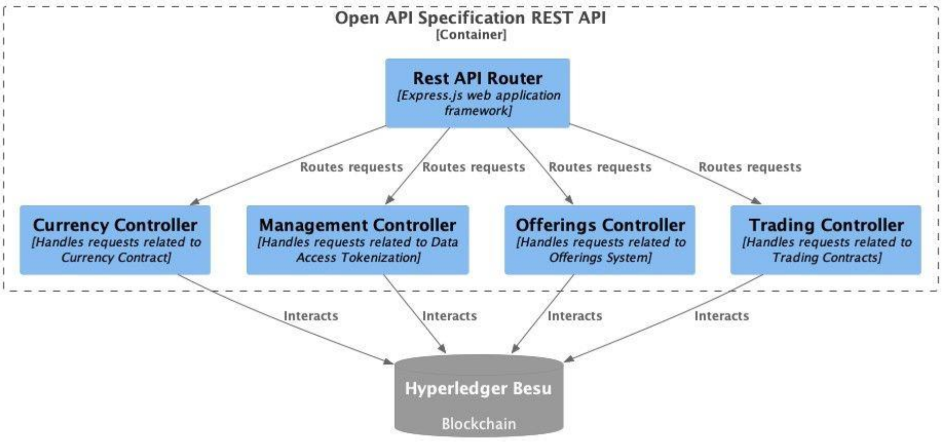


Figure 27 – Assets Trading & Monetization C4 component-level architecture - REST API

Figure 28 depicts the diagram for the Currency Contract, including the following components:

- ERC-20 Token: The main component that implements the standard ERC-20 token for currency.
- Token Transfer: The component that handles the transfer of tokens between addresses.
- Balance Inquiry: The component that provides information about token balances.
- Token Approval: The component that handles token approval for 3rd parties. It allows the Trading Contract to transfer assets.
- Total Supply: The component that provides information about the total token supply.

The REST API interacts with the ERC-20 Token component, which in turn implements the functionality of the other components. Each of these components interacts with the Hyperledger Besu blockchain, which is an external system represented as a database in the diagram. The interactions with the blockchain involve various operations related to the specific functionality of each component.

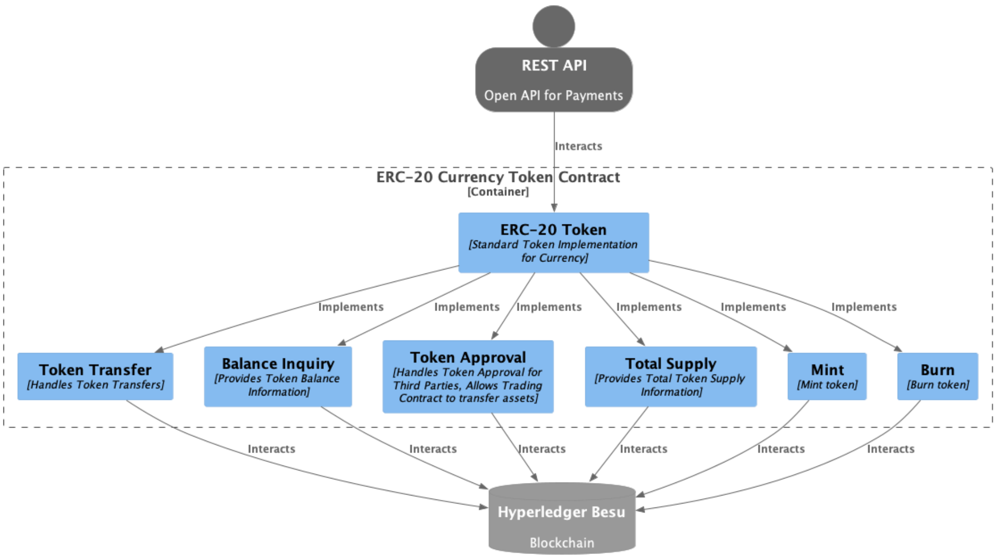


Figure 28 – Assets Trading & Monetization C4 component-level architecture - Currency Contracts

Figure 29 illustrates the diagram for the Data Assets Contracts, including the components of:

- ERC-1155 Token: The component realizing the standard ERC-1155 token for data access.
- Token Minting: The component that handles the creation of new tokens.
- Token Burning: The component that handles the destruction of tokens.
- Token Approval: The component handing token approval for 3rd parties, allowing the Trading Contract to transfer assets.
- Token Transfer: The component that handles the transfer of tokens between addresses.
- Token Balance: The component that provides information about token balances.
- Token Metadata: The component that provides methods for handling data assets metadata.

The REST API and the Proxy Contract interact with the ERC-1155 Token component, which in turn implements the functionality of the other components. Each of these components interacts with the Hyperledger Besu blockchain, which is an external system represented as a database in the diagram. The interactions with the blockchain involve various operations related to the specific functionality of each component. Moreover, the ERC-1155 Token component interacts with the Asset Metadata storage, storing references to assets' metadata and a proof of asset's metadata integrity.

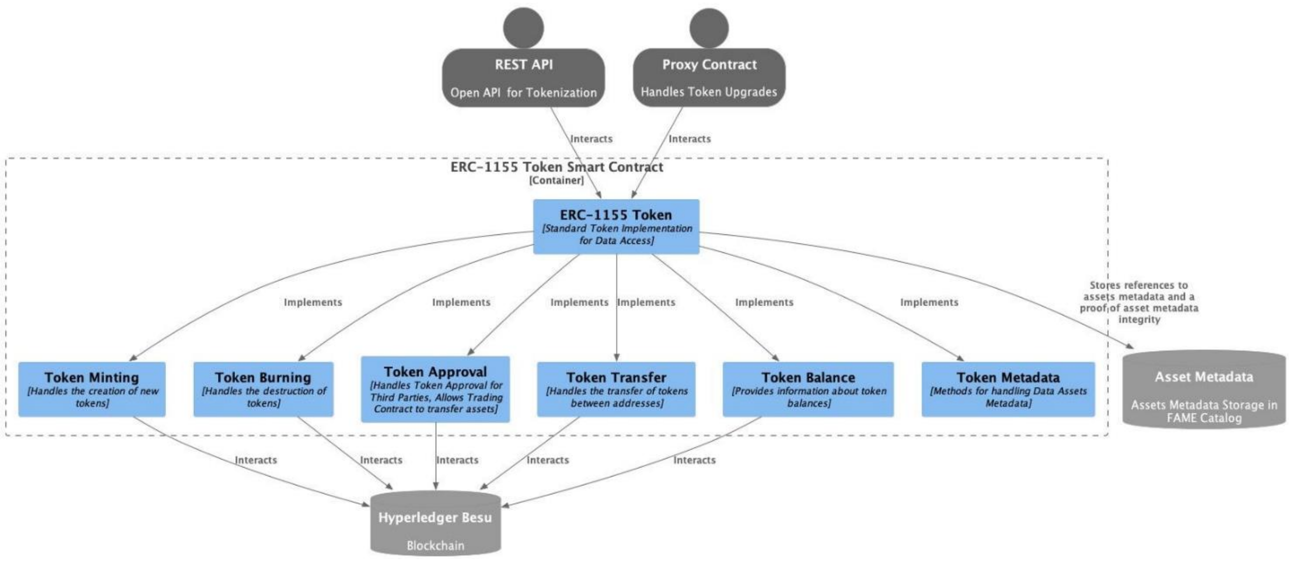


Figure 29 – Assets Trading & Monetization C4 component-level architecture - Assets Contracts

Figure 30 depicts the diagram for the Trading Contracts, including the components of:

- Trading Contract: The component swapping ERC-20 currency and ERC-1155 asset tokens.
- Trade Execution: The component that executes trades based on contract terms.
- Trade Validation: The component that validates the terms of a trade.

The REST API interacts with the Trading Contract component, which in turn implements the functionality of the Trade Execution and the Trade Validation components. Each of these components interacts with the Hyperledger Besu blockchain as in the previous cases. The interactions with the blockchain involve various operations related to the specific functionality of each component. The Trading Contract component also interacts with the Currency Token and the Data Access Token containers, handling the payment and transfer of token ownership respectively.

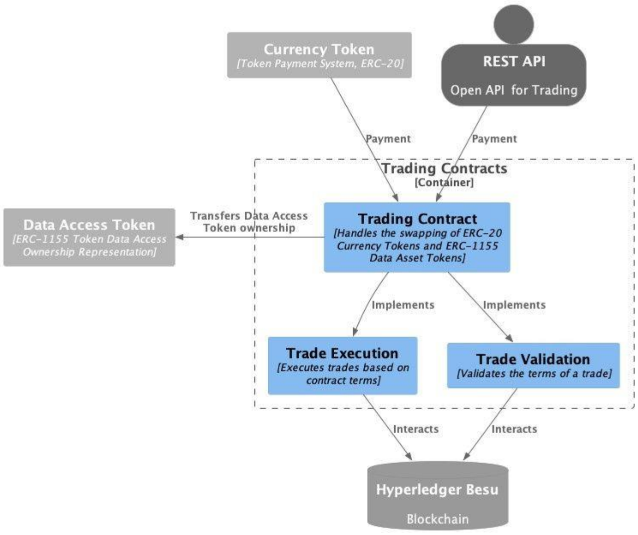


Figure 30 – Assets Trading & Monetization C4 component-level architecture - Trading Contracts

7.3.7 Operational Governance

The goal of the *Operational Governance* component is to provide the necessary services for users' registration, and the support of business models such as pay-as-you-go and data-as-a-service, in particular subscription management, as well as the management of SLAs, as represented in the 3rd level of the C4 diagram (i.e., component diagram) in Figure 31.

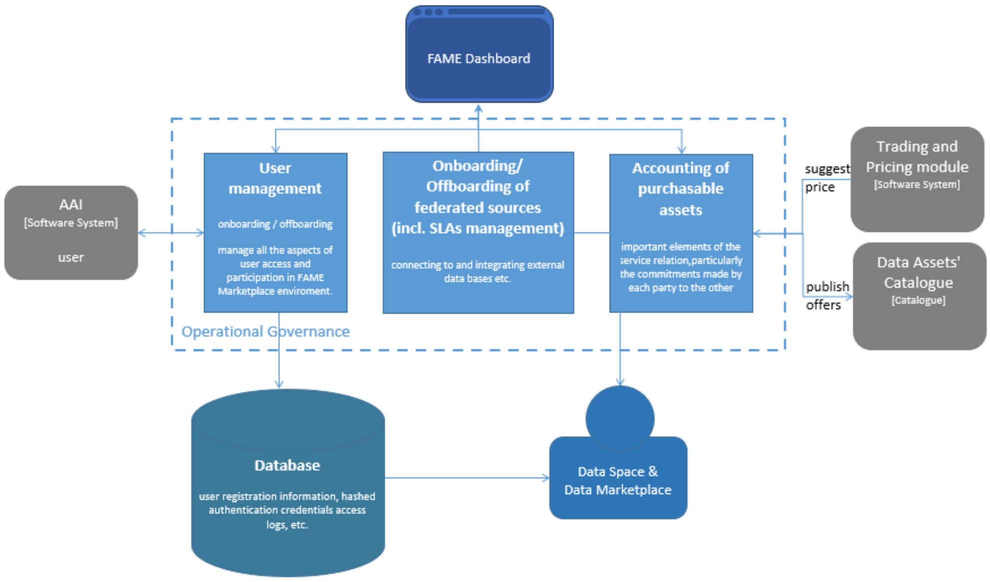


Figure 31 – Operational Governance C4 component-level architecture

The set of technical functionalities of this component can be translated into three (3) different subcomponents (implemented as REST-over-HTTP APIs (microservices)), referring to the ones of:

- User Management/onboarding/offboarding: As a key part of any system, the user management component comprises the onboarding process and provides administrators with a single service to manage all the aspects of the users' access and participation in the FAME platform. The onboarding/offboarding process will be automated and implement different user roles to allow for role-based access/restrictions. To achieve this, it will exchange information with the *Authentication & Authorization* infrastructure to identify and authorize user's registration and login. It will also allow the administrators to easily access user profiles and streamline the onboarding, offboarding and user management processes.
- Onboarding/offboarding/management of federated sources (including SLAs management): In order to leverage access to the external data sources (i.e., Data Marketplaces, Data Spaces, Databases, etc.), the platform needs mechanisms to onboard and offboard these sources as well as to manage the onboarded ones. Service relationship and operational relationship, both will be supported in this component.
- Accounting of purchasable assets: This component deals with the support of selected business models for the monetization of the FAME data assets. This will include important monitoring functions for contractual details like subscription and licensing, but also support EmFi applications like billing/creation of invoices and the execution of user payments and monitoring of payments. This component will be responsible for creating, storing, and managing offers for any type of assets, including setting the terms and conditions, selecting the business model and the price. Consequently, these offers must be published on the *FDAC*. Although the final price will be determined by the asset provider, the *Assets Pricing* component will suggest prices for given assets that a provider supplies and hence, an exchange of information will take place with this component. Subscription business model will be also supported by implementing a subscription management including: (i) Subscription plans: Allow the definition of different subscription plans with corresponding features and pricing tiers for data providers and consumers, (ii) Subscription handling: Enable users to subscribe, upgrade, downgrade, or cancel their subscriptions through the platform, and (iii) Recurring billing: Implement automated recurring billing for subscription fees. This component will also implement licensing and rights management, supporting: (i) Licensing options: Provide options for data providers to specify licensing terms, such as usage restrictions, duration, or pricing models (e.g., per-use, flat fee, royalties), (ii) Licensing agreements: Facilitate the creation and management of licensing agreements between data providers and consumers, and (iii) Rights enforcement: Implement mechanisms to enforce the rights and restrictions specified in the licensing agreements. Subscription management, recurrent billing, invoicing may also be features that are compatible with blockchain-based trading.

7.3.8 Federated Data Assets Catalogue

The *Federated Data Assets Catalogue (FDAC)* is responsible for storing and indexing all the information regarding the data assets of FAME, which represent multiple types of content, ranging from datasets to AI models, services, or relevant documentations. All those assets will be described with a proper level of detail that supports expressive search queries and narrows down the results to the most relevant. This catalogue is able not only to represent assets originated from FAME activities but also to index assets deriving from external sources, such as relevant Data Spaces or Data Marketplaces. All the information about the assets will be made available to any FAME component that requires such information for its operation. Figure 32 presents the 3rd level of the C4 diagram (i.e., component diagram) of all the actions undertake by the *FDAC*.

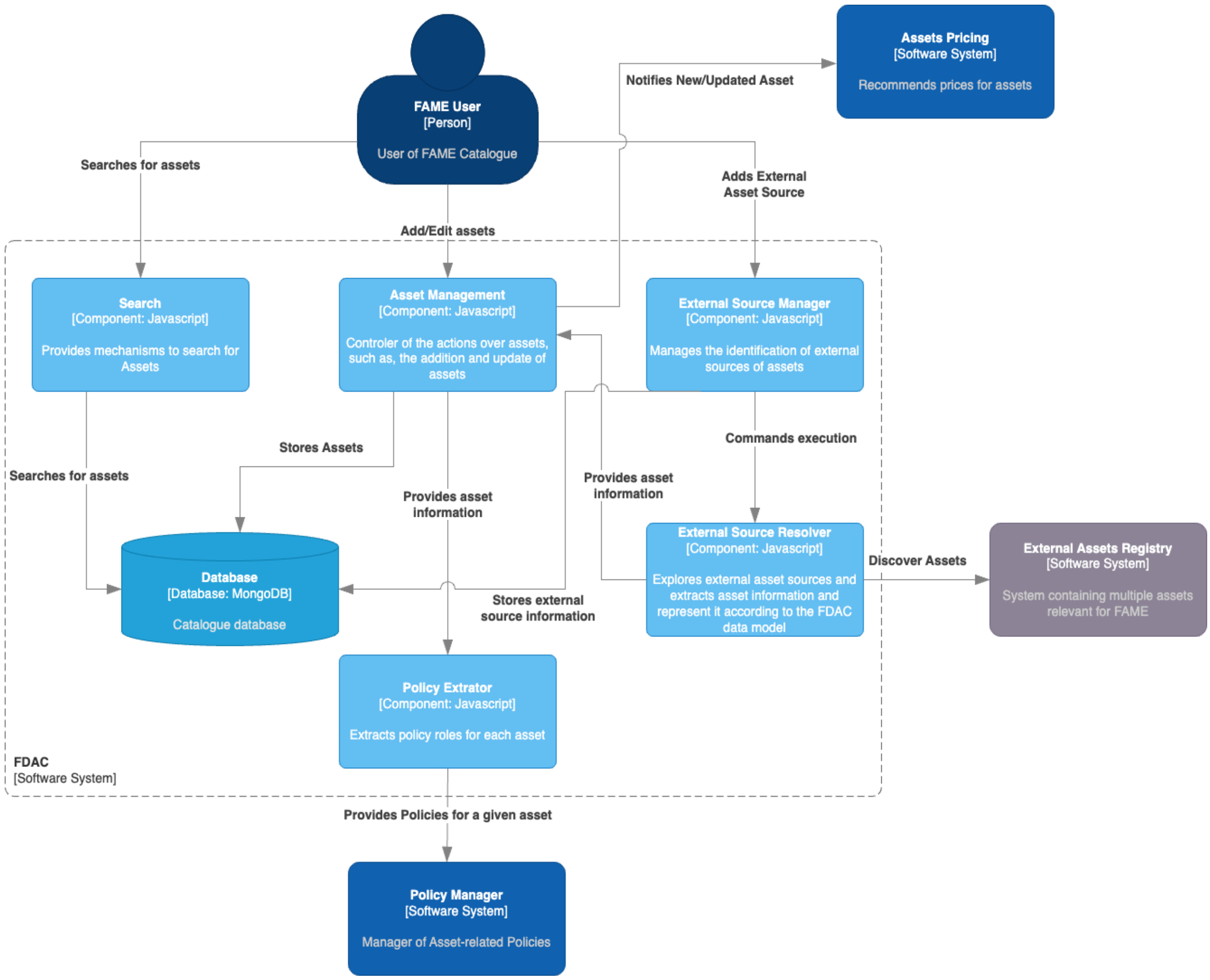


Figure 32 – Federated Data Assets Catalogue C4 component-level architecture

The core component of the *FDAC* is the Asset Management component. This component is responsible for performing controlling edition operations over assets, exposing the related interfaces to allow these operations, and accepting actions from the end-users, also providing the External Asset Resolvers. When an operation is executed, it performs the corresponding actions on the database to persist the information. When a new operation is executed, a notification is generated and sent to the *Assets Pricing* Component, which is responsible for the calculation of the recommended price for that asset. Upon the addition of a new asset, the information is shared with the *Assets Policy Management* component that extracts the policy rules for that asset.

As for the discovery and import of assets' entries from external sources, such actions are conducted by two (2) components: the External Source Manager and the External Source Resolver. The former manages the information about the external sources of assets to be analysed. For each source it collects and stores information about the description and location of sources, any authentication credentials needed to support the connection to the source and specific configurations that need to be considered when getting information from these sources. For each source it also identifies the External Source Resolver that must be executed to process information of that source. External Source Resolvers connect to the associated source of assets, use the available interfaces to explore the available assets, read the metadata of the assets described on the source's data model, extract the information, and add it to the *FDAC*. Among the information extracted by the Resolver are the description of the asset, the location of the asset, any existing pricing information and information about any policy related with the assets' discoverability and purchase conditions.

Since assets' sources may differ on their interfaces and data models, different External Source Resolvers must be used to process different sources. However, when sources comply to some standards, resolvers can be reused to connect to different sources. One example of this is an External Source Resolver developed to process IDS compliant Data Spaces/Data Marketplaces and can be reused to process information on multiple Data Spaces/Data Marketplaces that comply with the IDS architecture.

7.3.9 Semantic Interoperability

The *Semantic Interoperability* component consists of a set of functionalities that enhance the *FDAC* capabilities with semantic functionalities. These functionalities are mainly applied to the asset's metadata, referring to the support of multiple ontologies to describe the concepts used by the *FDAC* API. This allows to expose API methods that support the description of information on other formats than the ones used internally in *FDAC*. For instance, when adding a new asset, the concepts that describe it can be described using DCAT [92] or IDS/Gaia-X Self-Descriptions, adapting to concepts that the user is more familiar with. Responses from the *FDAC* can also be converted to the specified concepts. The components involved on this functionality are represented in Figure 33 in the context of the 3rd level of the C4 diagram (i.e., component diagram) of this component.

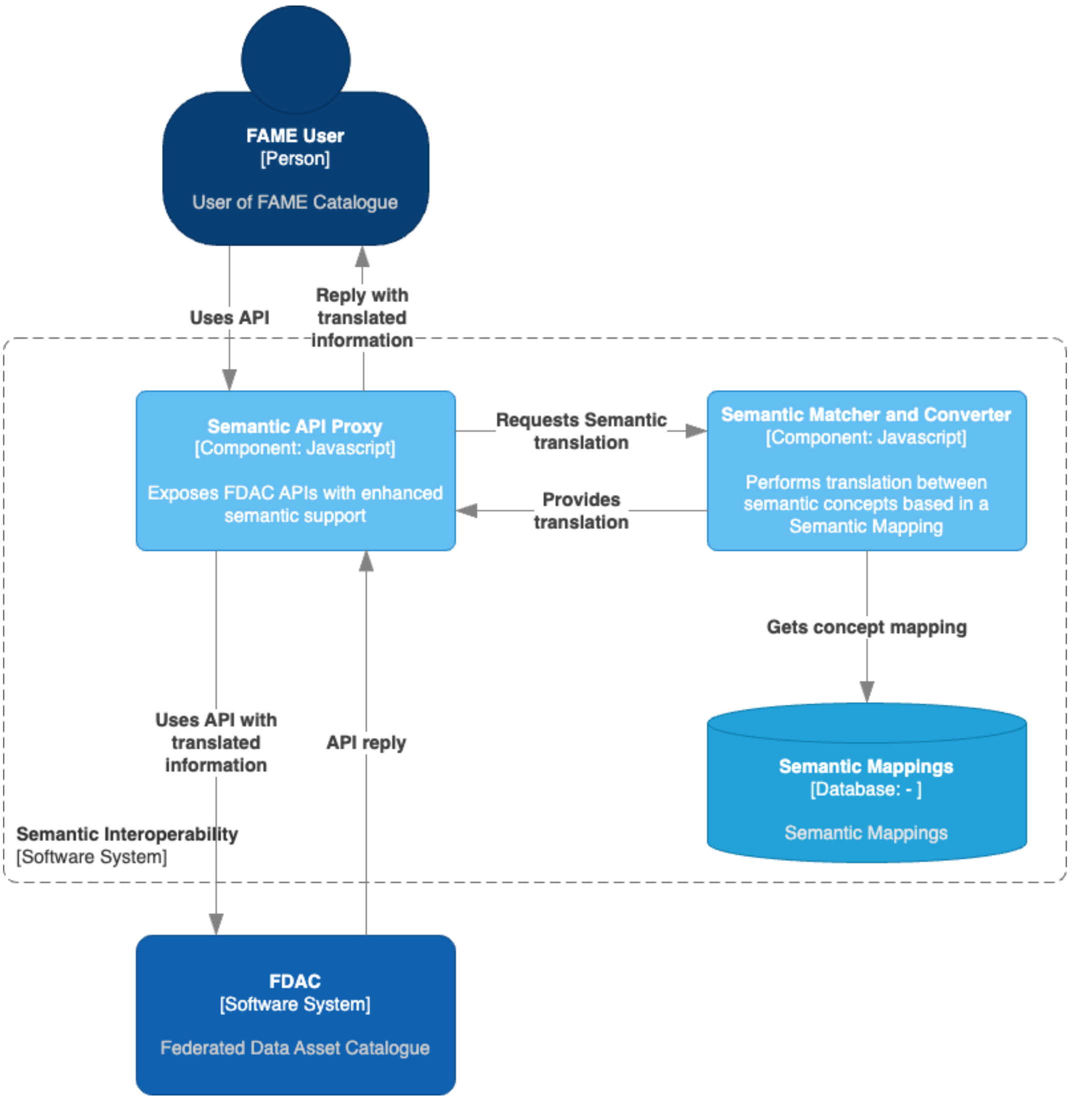


Figure 33 – Semantic Interoperability C4 component-level architecture

In deeper detail, when a user wants to access the *FDAC* API using the Semantic Concepts, his/her requests are routed to the Semantic API Proxy. This proxy exposes the same API endpoints provided by the *FDAC* but supports the description of information using different semantic concepts. When the Semantic API Proxy receives a request expressed in a specific ontology, this component attempts to convert the information received into the concepts understood by the *FDAC* before sending the request to it. To perform such conversion, the Semantic Matcher and Converter component is invoked. This component receives the information to be converted, as well, and the indication of the ontology to be used by the user to describe the information.

With this information, the Semantic Matcher and Converter component accesses the database of the Semantic Mapping and searches for finding mapping specifications that describe how to relate the semantic concepts between the used ontology and the concepts used by *FDAC*. This relation mapping allows the Semantic Matcher and Converter component to perform the conversion between diverse concepts. After the conversion, it sends the information to the Semantic API Proxy, which in turn forwards the request to the *FDAC*. A session is established within the Semantic API Proxy with the first API request to keep track of the ontology used by the user. When the *FDAC* sends the reply to the Semantic API Proxy, it identifies the session related to it and asks the Semantic Matcher and Converter to convert the information provided in the reply to the concepts used by the user.

7.3.10 Semantic Search

The *Semantic Search* engine's goal is to implement semantic search over the *FDAC* based on the data asset consumer specifications, along with schemes for ranking the results according to relevance and value-based attributes of the data assets. This component will take advantage of the pricing schemes developed in the *Assets Pricing* component to provide the relative ranking of the search results and to determine their demand-driven price accordingly. These functionalities are illustrated in the 3rd level of the C4 diagram (i.e., component diagram) in Figure 34.

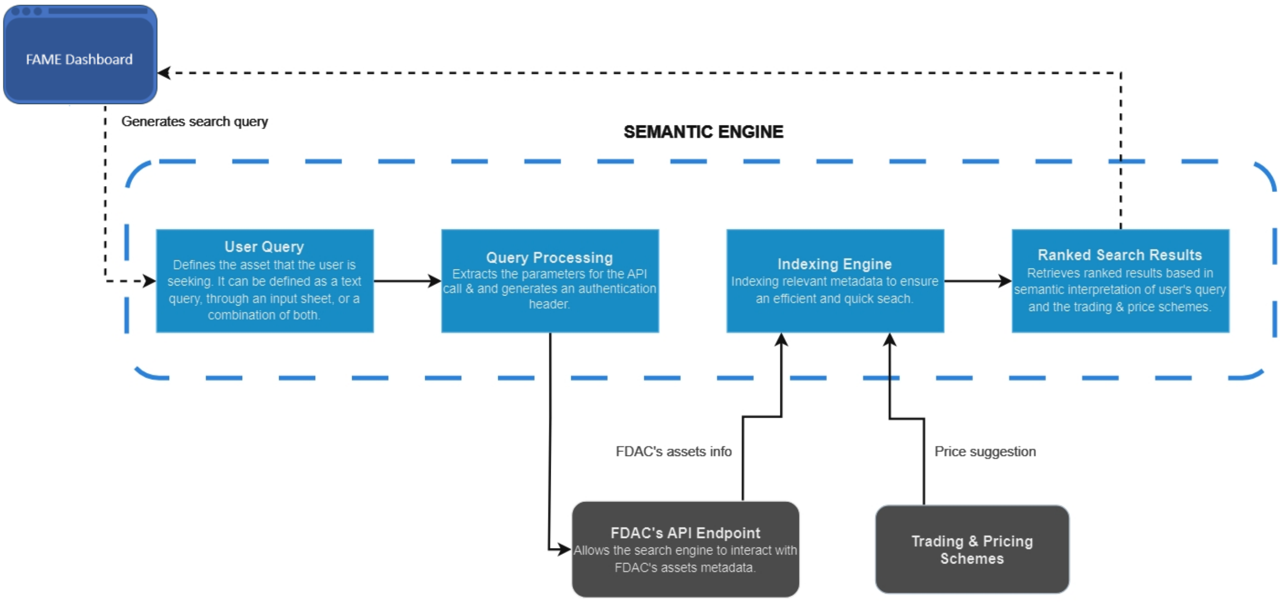


Figure 34 – Semantic Search C4 component-level architecture

¶ Semantic Search Engine: The main component for receiving search queries from the end-users. This engine analyzes the queries to understand the end-user's intent and performs searches in the relevant metadata of the *FDAC*. It utilizes semantic analysis to find assets that best match the user's needs based on the semantics of the query. The development of this engine makes extensive use of Python, which is compatible with crucial components like NLP [93] and Elasticsearch [94] that will facilitate effective semantic analysis of user queries and efficient indexing of metadata, whereas Python's flexibility in processing unstructured data will also play a pivotal role in handling diverse metadata types.

¶ Query Processing: The component that is responsible for analyzing and processing the search queries received from the Semantic Search Engine. It extracts key parameters from the query, such as keywords, filters, and search constraints, to use them in searching for related assets.

¶ FDAC's API Endpoint: This component represents the endpoint of the API of the *FDAC*. It allows the Semantic Search Engine and other components of the platform to interact and make queries to the metadata stored in the *FDAC*. The API endpoint acts as a bridge to access the data stored in the *FDAC*.

¶ Indexing Engine: The Indexing Engine is responsible for indexing the relevant metadata of the *FDAC*. By indexing the metadata, it enables efficient and quick access to relevant information of the assets during search operations.

¶ Ranked Search Results: After processing the search queries and finding relevant assets using the Indexing Engine, the Semantic Search Engine produces ranked search results. These results are not only based on semantics but also consider external trading and pricing schemes. The ranking process considers various factors, including relevance to the query, the defined pricing schemes and others like data volume, file format, etc., to provide end-users with a prioritized list of assets that best match their needs and align with the pricing criteria.

7.3.11 ML/AI Analytics

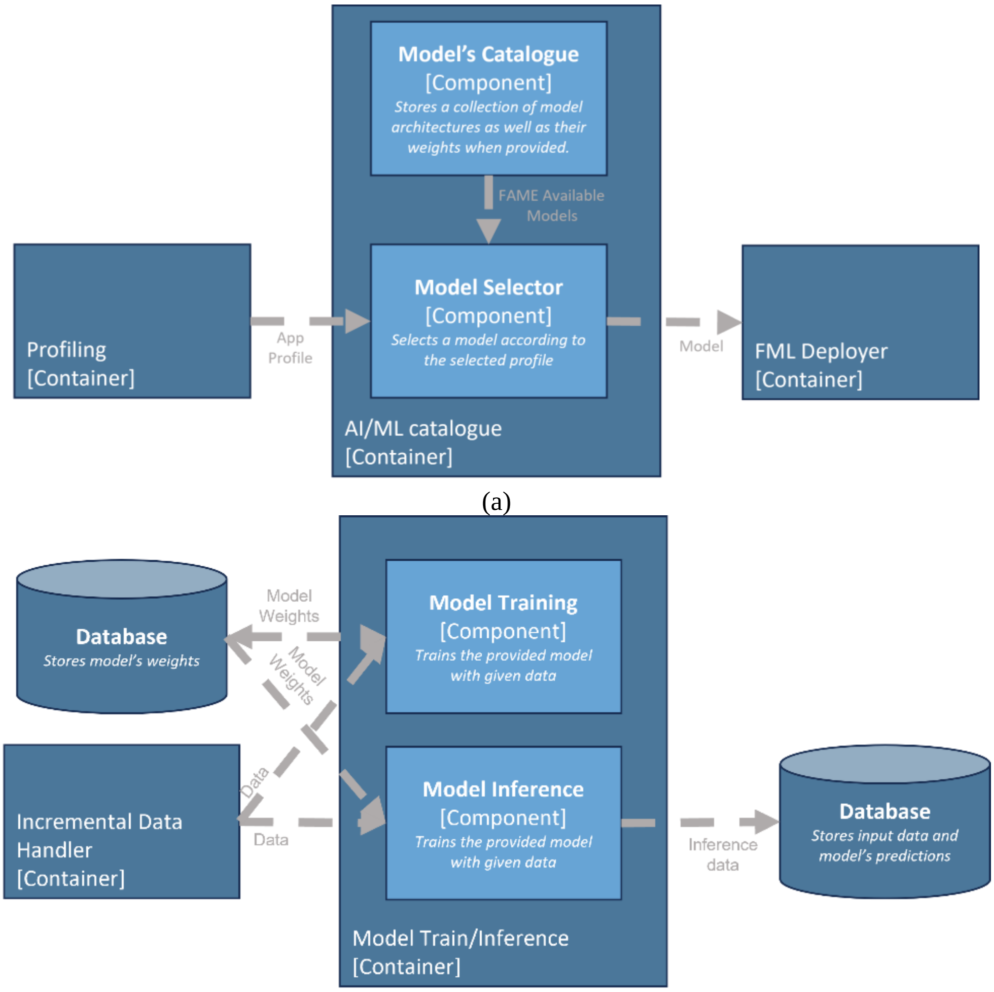
The *AI/ML Analytics* component comprises a comprehensive list of different ML techniques focused on solving problems within the FAME context and the implementation of the necessary functionalities for training the models and for giving predictions from input data.

The above-mentioned *AI/ML Analytics* C4 component-level architecture is shown in Figure 35 (a), where two (2) different components are presented:

¶ A Model Selector component, which is in charge of processing the application's metadata yielded by the Profiling component from the CO_2 monitoring analysis block. This selects the most suitable model for the current running application from the Models' Catalogue.

¶ The Models' Catalogue component, which consists of a list of available models within FAME. This list contains information about the data features needed for training the model or the metrics associated to this model (if it is already trained). This list relays on MLFlow functionalities to record all the said metadata.

Figure 35 (b) shows the C4 component-level architecture for the Model Train/Inference functionalities. The Model Training component implements the algorithms and techniques needed for training the different models available in the *ML/AI Analytics* catalogue. These functions support different ML frameworks such as TensorFlow [95], PyTorch [96] or Scikit-Learn [97]. Data is served from the *Incremental Analytics* component, whereas the model's parameters are taken or stored in a local database when it proceeds. The Model Inference component follows a similar workflow where the model is used to yield estimated data from the input samples. It should be mentioned that the Inference component will be containerized to facilitate the deployment process.



(b)

Figure 35 – ML/AI Analytics C4 component-level architecture

7.3.12 SAX Techniques

The *SAX Techniques* aim to offer the overall packaging for all the SAX functionalities in the context of FAME. As shown in the C4 component-level architecture in Figure 36, to successfully accomplish its goal, the *SAX Techniques* component interacts with: (i) External systems (such as complex event processing engines) for additional services and contextual information, (ii) ML pipelines for accessing the models requiring explainability, (iii) Data assets that constitute the process information (such as event logs), and (iv) XAI scoring framework when requested to assess the explainability of the various constructed models.

As for the end-users of this component, the interaction may take place in two (2) diverse ways: by providing user-preferences for explainability and questions regarding the business process outcomes and decisions on the one hand; and by consuming the situational explanations provided by the component, on the other hand.

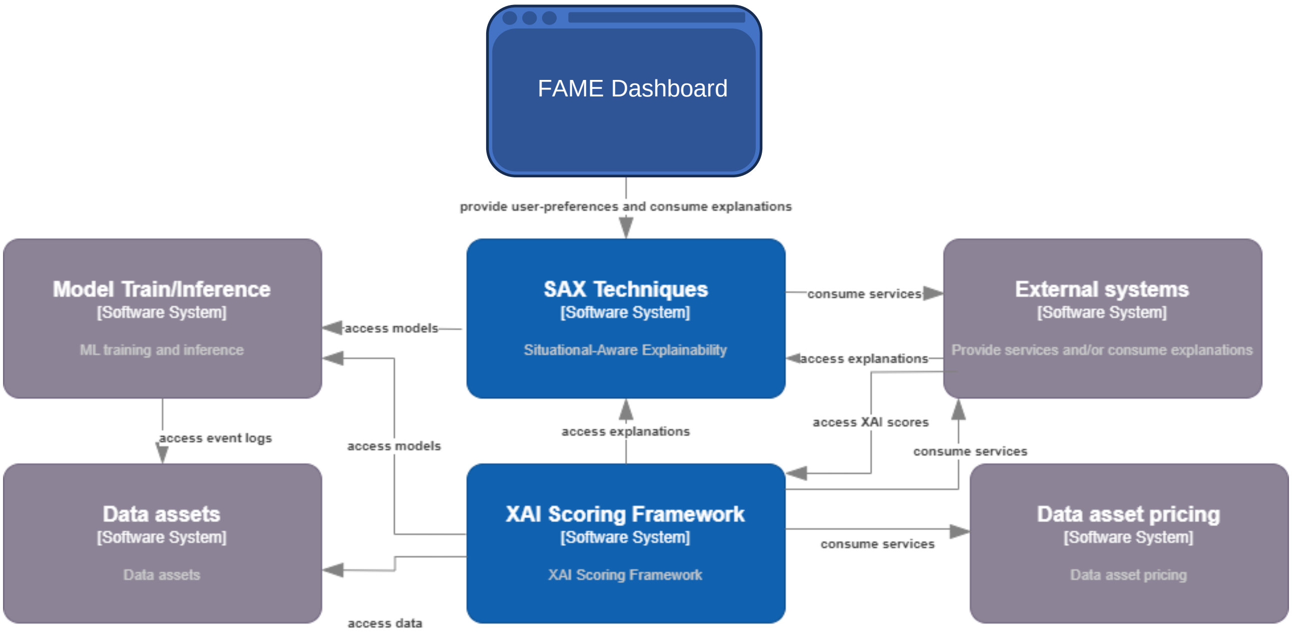


Figure 36 – SAX Techniques C4 component-level architecture

The realization of the *SAX Techniques* is implemented as Python libraries and relies on the following baseline technologies (either by providing wrappers for open-source libraries or by invoking external services): process mining algorithms (e.g., Heuristic miner [98] and Alpha [99]), CEP engines (such as IBM Proactive Technology Online [100]), core causal discovery algorithms (such as LiNGAM [101]) and LLM services (such as ChatGPT [102]), and core XAI techniques (such as LIME [103] and SHAP [104]). These will be consumed by the *SAX Techniques* per demand. By exploiting all these technologies, the developed component will offer FAME capabilities of situation awareness, causal process discovery, and process mining, over given event logs. SAX code will be developed as a set of components and services released as the SAX library to the open-source community by the end of the project.

As for the supplementary *XAI Scoring* framework, this will introduce a comprehensive methodology, considering not only technical facets but also communicative and user-focused aspects of explainability. The realm of XAI is witnessing innovations that extend beyond traditional metrics. User-centric measures, encompassing aspects like satisfaction, mental models, curiosity, and human-AI collaboration dynamics, highlight the need for a more inclusive evaluation approach, as the one provided in the FAME *XAI Scoring* framework. Hence, the FAME *XAI Scoring* framework provides a comprehensive framework for assessing the explainability of various XAI methods and AI-data assets in general, across multiple ML models and XAI methods, with the ultimate goal of creating a unified multidimensional explainability score.

The developed approach will delineate a quantification scheme that considers properties of XAI algorithms including fidelity, stability, simplicity, and coverage. Simultaneously, it will integrate user-centric aspects like satisfaction, trust, and task performance. To ensure a holistic assessment, it will integrate performance metrics such as accuracy, precision, and recall, confirming that the produced explainability benchmarks do not diminish the model's predictive prowess. The intention of this framework is to generate explainability scores that can be contrasted against human expert evaluations, providing a comprehensive and validated measure of explainability. These functionalities are illustrated in the 3rd level of the C4 diagram (i.e., component diagram) in Figure 37.

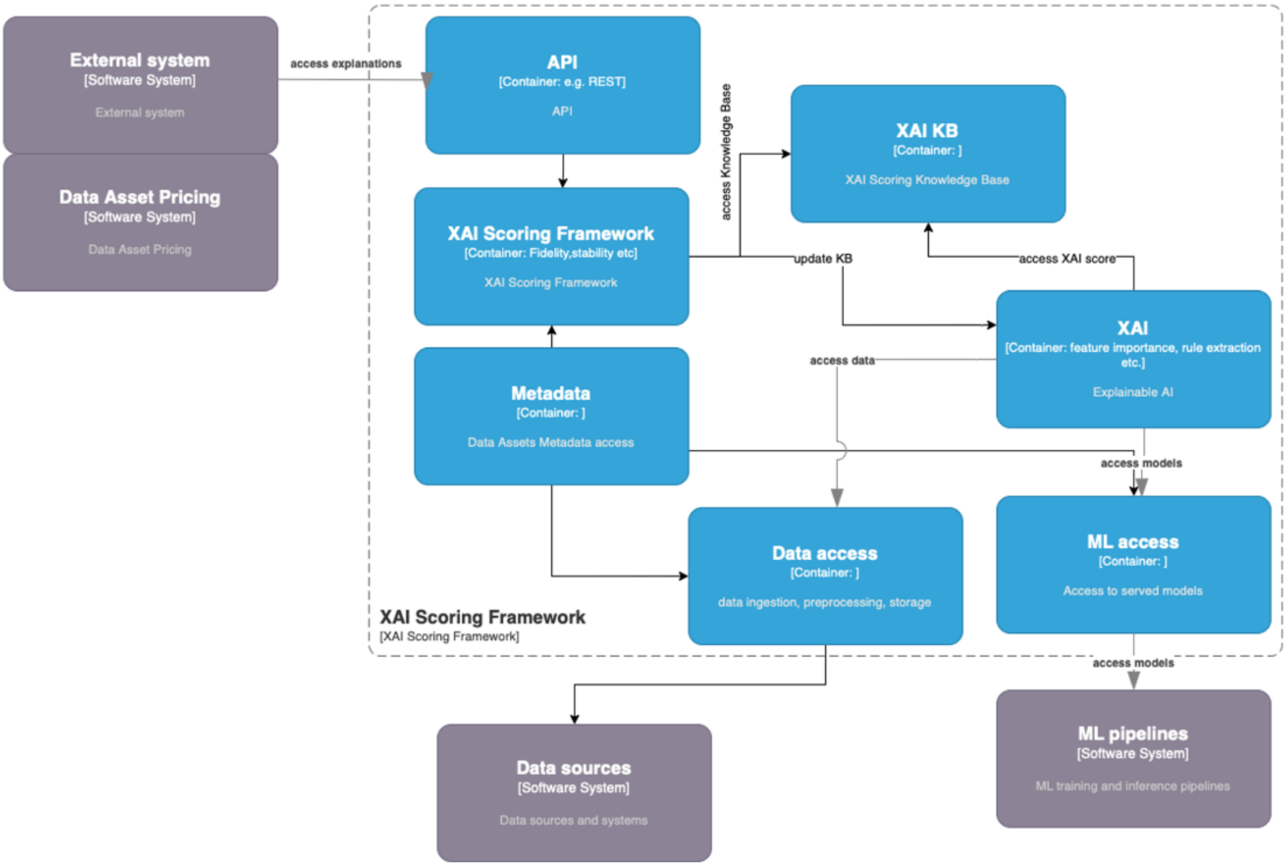


Figure 37 – XAI Scoring Framework C4 component-level architecture

As shown in Figure 27, the *XAI Scoring* framework offers a dynamic online implementation. Its methodology will further evolve into benchmarking XAI techniques, aiming to develop a profound knowledge base for assessing explainability scores across diverse models. This knowledge base, fuelled by metadata from AI/ML models and datasets, will evaluate properties like fidelity, consistency, and comprehensibility, while also accommodating variable weighting preferences, paving the way for a robust and adaptive assessment of AI explainability in all the FAME scenarios.

7.3.13 Incremental Analytics

The *Incremental Analytics* component is responsible to act as the enabler to provide incremental analytics to the ML/AI techniques developed or used within FAME. It is based on the LeanXcale database [105], a hybrid SQL/NoSQL database that allows data ingestion at very high rates and offers a dual SQL and NoSQL interface. With the SQL interface, the data user can treat the LeanXcale database as a relational datastore and submit SQL compatible query statements to retrieve the resulted dataset. However, with the NoSQL interface, the data user can have access to a set of additional capabilities, including the ability to retrieve data in an incremental manner, thus, getting informed of updated result sets, while data is being ingested at the same time. Figure 38 illustrates the internal elements of this component as part of the 3rd level of its C4 diagram (i.e., component diagram).

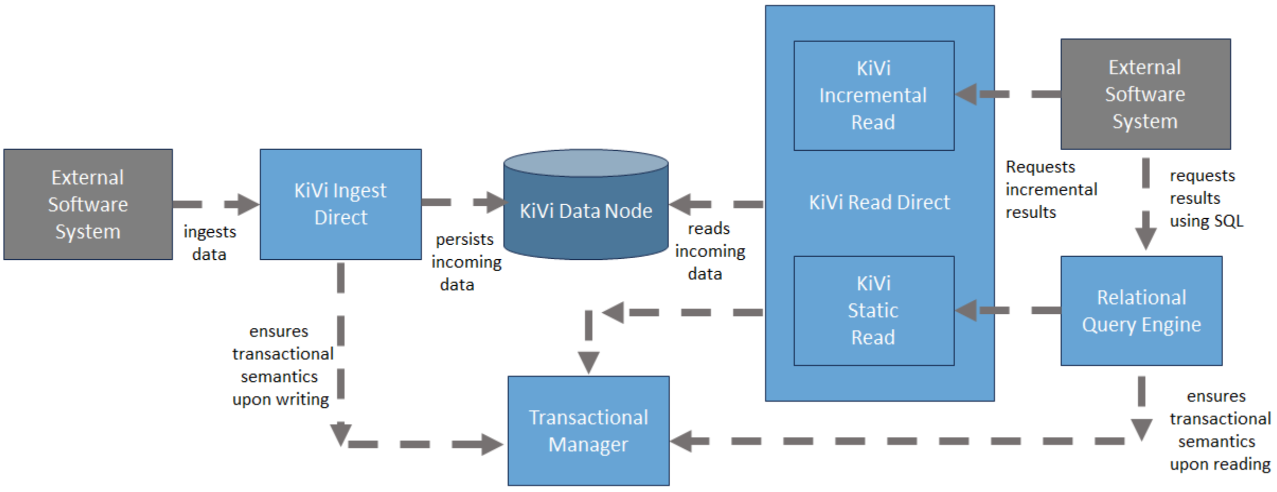


Figure 38 – Incremental Analytics C4 component-level architecture

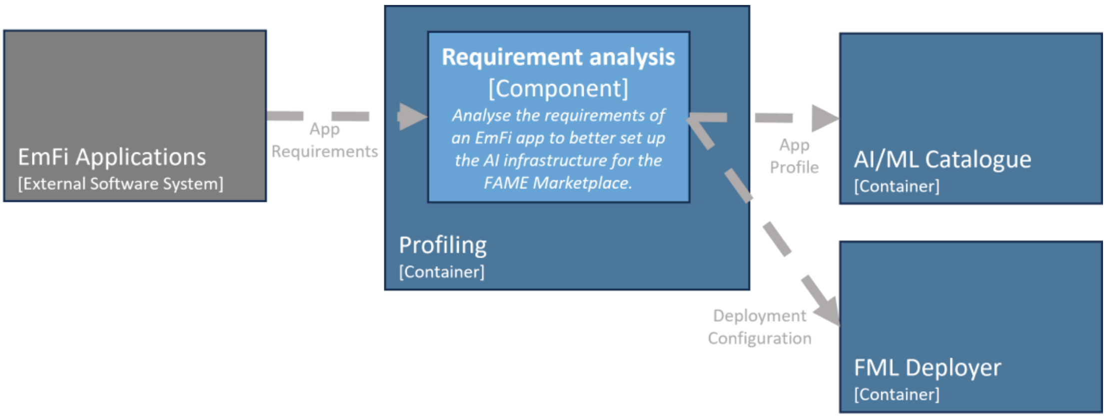
In this figure, two (2) different actors can be distinguished that can be considered as the data users of the component: the data provider that is the software components that ingest data to the database and are suited at the left part of the figure, and the data consumer that is the software components that reads data from the database and are suited at the right part of the figure. The data provider communicates with the system using the NoSQL interface, that is depicted in the figure as the KiVi Ingest Direct. This component is capable to receive data being ingested at very high rates and persistently store them to the physical data nodes of the database. On the other hand, the data consumer can have two (2) options to read data from the database. The first option is to use standard SQL dialect, thus make use of the JDBC driver [106] of the database, which communicates with the relational query engine. The latter is responsible to interpret the SQL statement and transform it to a relational query plan to be executed and retrieve the corresponding results. Each node of the proposed query tree can be considered as a different query operator. Most of these operators are being implemented using the KiVi Read Direct component and more precisely the Kivi Static Read functionalities of the latter. All the relational query operators except for the join have been implemented in this matter. The KiVi Read Direct can be considered as the database driver and provides the interface for data users to submit query statements and retrieve data from the physical data nodes of the database. On the other hand, the data consumer instead of communicating with the database using the JDBC interface, which in fact connects to the relational query engine that makes use of the KiVi Read Direct, he/she can directly interact with the latter and has access to a wider set of functionalities for data processing. By that, the data user can make use of the KiVi Incremental Read, which is capable to retrieve data in an incremental fashion, thus keeping the connection between the data user and the database open and return updated results as data is being ingested at the same time. Finally, the Transactional Manager is an important sub-component of the overall architecture. The *Incremental Analytics* are built upon the LeanXcale database, which is a relational database that ensures transactional semantics. In fact, both the direct interfaces to the physical data nodes and the relational query engine communicate with the Transactional Manager to check for protentional write-write conflicts upon data ingestion or retrieve the correct snapshot of the dataset that needs to be evaluated against a submitted query statement.

7.3.14 Analytics CO_2 Monitoring

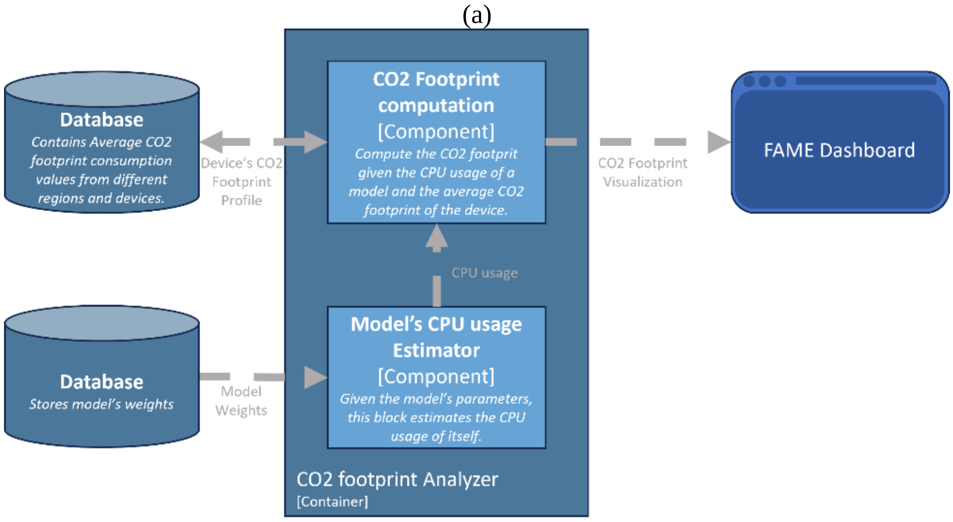
The *Analytics CO_2 Monitoring* component aims to estimate the CO_2 emissions of the ML/AI models used in a FAME application from its CPU/GPU usage. These CO_2-related metrics are later used to deploy ML/AI models in a more efficiently manner in terms of energy consumption so that the CO_2 footprint can be reduced.

Figure 39 (a) shows the C4 component-level architecture that depicts the Requirements' Analysis component. This component gathers the requirements information from the EmFi applications and generates the metadata/profile associated to that application. These profiles are later passed to other components of the platform, referring to the *ML/AI Analytics* and the *FML Deployment* components for further exploitation.

The C4 component-level architecture regarding the CO_2 Footprint Analyzer is shown in Figure 39 (b), where two (2) components are well differentiated: (i) the Models' CPU usage estimator, which estimates the CPU/GPU computation usage of the model in the device where the model is currently running - this process is linked to the Model Train/Infer component from the *AI/ML Analytics* component to compute these metrics, and (ii) the CO_2 Footprint computation component, where by forming the CPU/GPU usage value obtained from the previous component, and leveraging external information regarding the average regional CO_2 g/KWh emissions of the device's location, the ML/AI model's CO_2 footprint is computed. A visualization of these results will be plotted through the FAME *UIs* to raise awareness of the environmental impact of the running process (this is an optional feature within this component).



(a)



(b)

Figure 39 – Analytics CO_2 Monitoring C4 component-level architecture

7.3.15 FML Deployment

The *FML Deployment* infrastructure is destined to support privacy friendly and energy efficient data sharing in federated learning scenarios where many different clients contribute data assets. In the context of FAME, the FML infrastructure will be implemented as an extension of a Machine Learning and Data Science cloud platform for prototyping Big Data Analytics applications. The platform (server side) will be in charge of properly aggregating the weights of the ML/AI models received from each end-user and providing back the aggregated model to each one of them. This infrastructure will therefore support federated use cases involving stakeholders' collaboration across the different data providers and external data sources that will be connected. The 3rd level of the relevant C4 component diagram is defined in Figure 40.

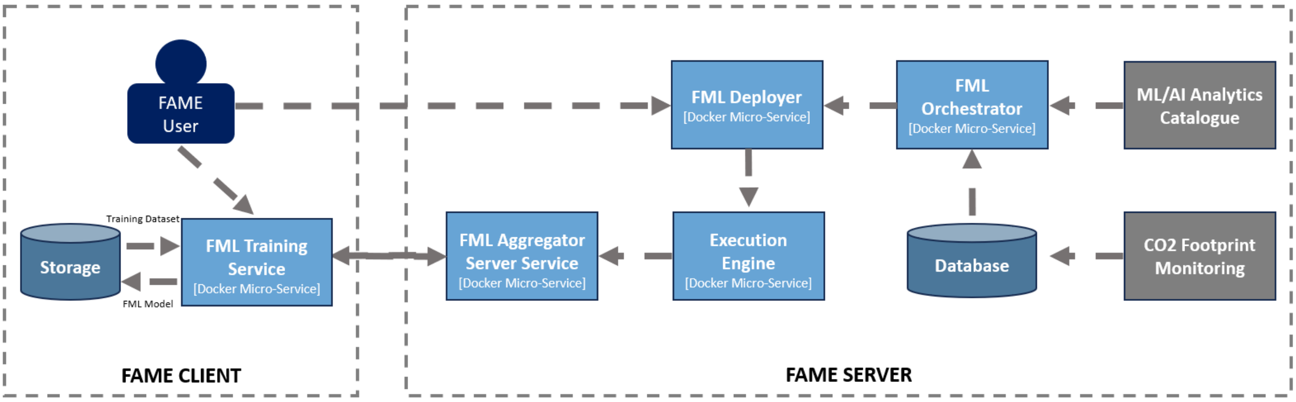


Figure 40 – FML Deployment C4 component-level architecture

As shown in the figure, the FAME client side has been divided from the FAME cloud/server side, assuming on the FAME server a Kubernetes cluster, while on the FAME client side, for instance laptops, virtual machines, located on premises where the FAME end-user's data reside and where it is possible to run containerized applications.

On the FAME server side, there exists the FML Orchestrator component that will interface with the outputs coming from the *ML/AI Analytics* catalogue and the *Analytics* CO_2 *Monitoring* component: on one hand it will process an application service definition coming from the *ML/AI Analytics* catalogue, and on the other hand it will process a set of deployment configurations that we will be used for the energy efficiency purposes in the context of the *Analytics* CO_2 *Monitoring* component. This processed information will then be sent to the FML Deployer microservice that will be responsible to:

- Deploy the model weight aggregator server on the FAME server cloud, which will be exposed using Kubernetes API so that the microservice can be reachable by external clients.
- Provide REST APIs to download in a client node a docker compose file containing a requested specific end-user application coming from the *ML/AI Analytics* catalogue.

The downloaded docker compose file can then be configured to point to the FAME end-user's data and being executed. The FML Training Service and the FML Aggregator Server service will make use of the Flower framework [107] to start a federation training involving multiple FAME edge/client nodes and one aggregator server running on the FAME server cloud. In FAME client, the training dataset and the FML model result will be also able to be loaded/stored in a mounted file system or on other storages where needed (i.e., LXS database).

7.3.16 Learning Centre

The *Learning Center* is intended to provide a single access point to training resources and training content about FAME with emphasis on resources that are related with EmFi applications. Specifically, the Learning Center will offer the core functionalities of: (i) Managing (Creating/Updating) various resources and entities referring to FAME grown courses (i.e., courses developed in FAME) and third-party courses (i.e., courses linked from third party providers) managed within a training catalogue, how-to videos, webinars, tutorials, and demonstrators, and (ii) Searching and accessing the full range of the above-listed resources.

Figure 41 describes the C4 component-level architecture of the *Learning Centre*, which is composed by two (2) core components, referring to the Learning Centre Frontend and the Learning Centre Backend.

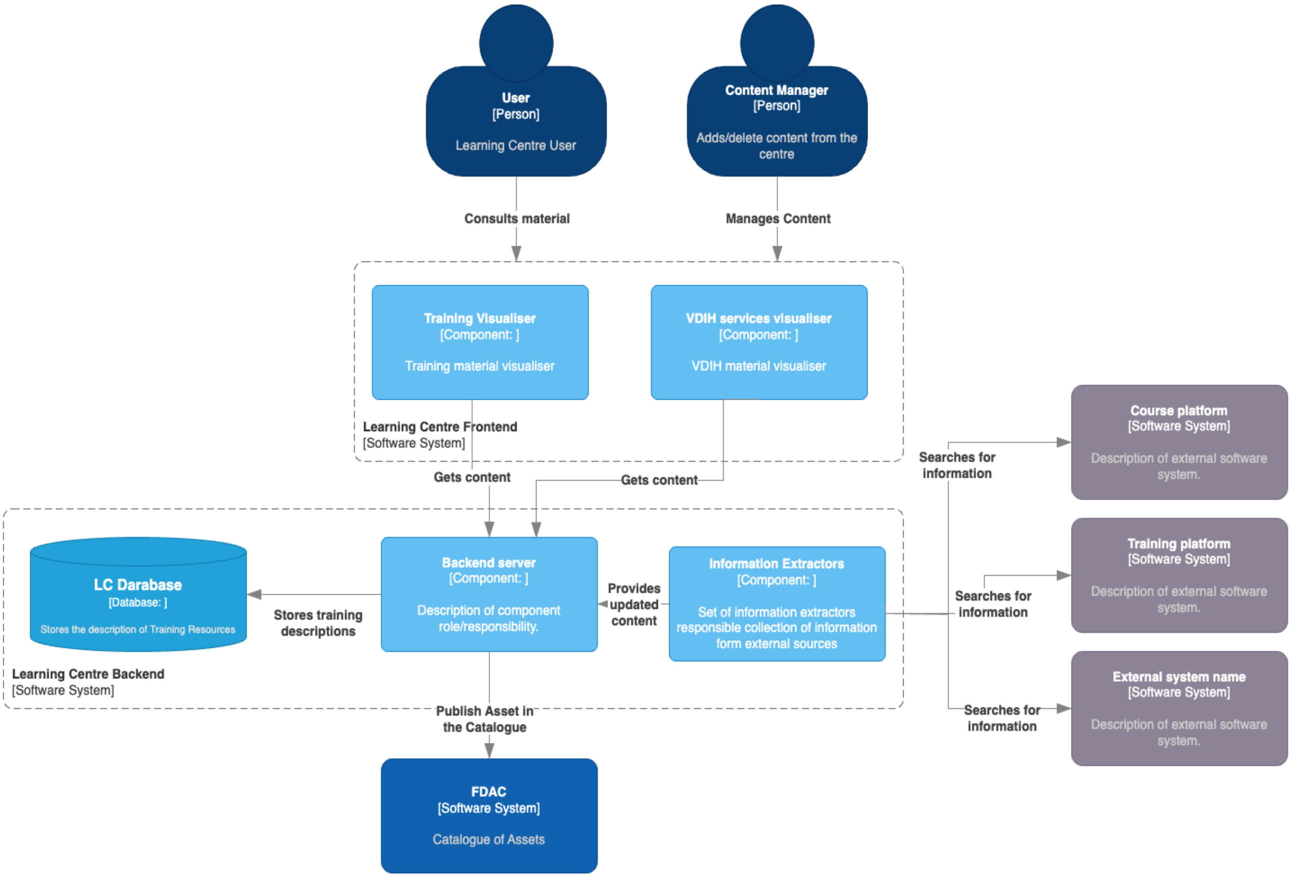


Figure 41 – Learning Centre C4 component-level architecture

In deeper detail, the Frontend has the goal to provide all the required user interfaces to allow the end-users to interact with the provided content. To this context, the Frontend supports two (2) diverse types of users, referring to the Users that are the actual consumers of the content and the Content Managers that are able to add/delete and update the content lying into the Learning Centre. All the required interfaces for collecting inputs from the end-users and to present the information are implemented in this component, providing such interfaces for the different training resources (such as courses, tutorials, webinars, etc.) or services provided in Virtual Digital Innovation Hubs (VDIHs).

As for the Learning Centre Backend, it provides support for the functionalities of the Frontend and handles the logic related with storing and indexing the different training resources and VDIH services' descriptions inside a dedicated database. While some training resources may be entirely stored in the Learning Centre Backend, others may exist on external platforms. In the latter cases, a preview or a description of the information is stored in the Learning Centre for the sake of delivering better training discovery capabilities to the end-users, thus providing both a complete training material and a VDIH service that is only accessed through the external platforms where the actual material is hosted. An example of this case are courses hosted in course-oriented platforms like Udemy [108]. Information in such platforms is subjected to be updated overtime, where a course program may change, further updates will be performed upon a program's price, duration, and lecturers, or a training resource may be totally deleted from the external platform. To cope with these changes that may occur at any time, the Learning Center Backend is equipped with a set of Information Extractors. These extractors will be platform-specific, making use of available APIs or resorting of web crawlers to navigate through the indexed external platforms, look for any existing changes, and collect the relevant information - only information about existing training resources will be collected from these external platforms. To this end, it should be noted that only curated training resources and VDIH services are expected to be added to the *Learning Centre*, requiring human selection.

As soon as either training resources or VDIH services are added within the *Learning Centre*, they are also indexed in the *FDAC*. This action allows to relate training resources with assets already lying into the *FDAC*, such as datasets or AI models. Such relations will improve the discoverability of training resources by identifying, in context, a list of training material relevant for the assets that the *FDAC* end-user is exploring.

To achieve all the abovementioned capabilities, the *Learning Center* will be based on the INFINITECH marketplace platform [109] and more specifically on the VDIH (Virtual Digital Innovation Hub) segment of the platform that manages training resources and content.

8 Conclusions

A SA for a federated Data Space supporting EmFi applications can greatly facilitate stakeholders in structuring, designing, developing, deploying, and operating related solutions. It serves as a stakeholders' communication device, while at the same time providing a range of best practices that can accelerate the development and deployment of effective systems. This deliverable has introduced the 1st version of such a SA for the FAME project. The latter adopts the concept and principles of existing Big Data, Data Marketplaces and Data Spaces RAs, being also in-line with the principles of Gaia-X and IDS as introduced in Section 5, as well as the Data Spaces design guidelines introduced in Section 7. In practice, it extends and customizes technical components of other RAs (e.g., i3-Market) and principles (e.g., GXFS) towards permitting its use for EmFi applications, among its stakeholders - analyzed in Section 3. Current deliverable also defines the structuring principles and technical specifications that will drive the integration and technical implementation of the FAME technical components and technologies.

As part of the deliverable, it has been illustrated how the FAME SA can be used to support a single-entry-point federated Data Space for EmFi applications. FAME's related business and technical value has been introduced, along with related challenges, limitations, and opportunities that are tightly coupled with its interacting stakeholders. Related RAs are also introduced, along with their value towards FAME's vision, continuing to the specification of the main capabilities with which any FAME stakeholder may interact with. All these are technically specified, and their architectural model is designed, acting as a preliminary step towards the actual implementation of them. All these will lead the way towards making FAME's vision a reality and providing a 1st prototype of the FAME federated Data Space following the set milestones and KPIs. These implementations will allow the collection of stakeholders' feedback regarding the SA, based on which the next version of the FAME SA will be produced, including updates in the different views of the architecture. Furthermore, future releases of the FAME SA will possibly consider new functionalities and capabilities for the main FAME components, following current challenges, requirements, and lessons learned.

Finally, as for the achieved KPIs, including the target values and the values that have been achieved until the deliverable submission date, these are depicted in Table 5 (highlighted with yellow colour).

Table 5 – D2.2 Related KPIs

Obj. ID	**KPI ID**	**Measured KPI**	**Expected Value [*M9*]**	**Achieved Value [*M9*]**	**Expected Value [*M18*]**	**Achieved Value [*M18*]**
O1	**1.1**	Standards-based Reference Architectures to be considered	5	8	>=8	N/A

References

1. FAME, D2.1 - Requirements Analysis, Specifications and Co-Creation I (2023).
2. Data Hubs, Data Lakes, and Data Warehouses: How They Are Different and Why They Are Better Together, https://www.gartner.com/en/documents/3980938
3. Loukiala, A., Joutsenlahti, J. P., Raatikainen, M., Mikkonen, T., & Lehtonen, T. (2021). Migrating from a centralized data warehouse to a decentralized data platform architecture. In International Conference on Product-Focused Software Process Improvement, pp. 36-48.
4. Janssen, N. E. (2022). The Evolution of Data Storage Architectures: Examining the Value of the Data Lakehouse (master's thesis, University of Twente).
5. European Data Strategy: Making the EU a role model for a society empowered by data, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en
6. Centralized vs. Federated: State Approaches to P-20W Data Systems, https://nces.ed.gov/programs/slds/pdf/federated_centralized_print.pdf
7. Why Data Mesh is a Must for Federated Data Management, https://www.k2view.com/blog/federated-data-management
8. SGIC Webinar Part 3: The Operational Reality of Statewide Data Ambitions: Federated Models, https://www.appgeo.com/part-3-the-operational-reality-of-statewide-data-ambitions-centralized-and-decentralized-models/
9. Data Federation: Foolproof Guide on Data Management, https://www.g2.com/articles/data-federation
10. Understanding the Value of Reference Architectures, https://doveltech.com/innovation/understanding-the-value-of-reference-architectures
11. Bass, L., Clements, P., & Kazman, R. (2003). Software architecture in practice. Addison-Wesley Professional.
12. Nakagawa, E. Y., Oliveira Antonino, P., & Becker, M. (2011). Reference architecture and product line architecture: A subtle but critical difference. In European conference on software architecture, pp. 207-211.
13. Martínez-Fernández, S., Ayala, C. P., Franch, X., & Nakagawa, E. Y. (2015). A Survey on the Benefits and Drawbacks of AUTOSAR. In Proceedings of the First International Workshop on Automotive Software Architecture, pp. 19-26.
14. Valle, P. H. D., Garcés, L., Volpato, T., Martínez-Fernández, S., & Nakagawa, E. Y. (2021). Towards suitable description of reference architectures. PeerJ Computer Science, 7, e392.
15. i3-MARKET, http://i3-market.eu/
16. i3-MARKET Opensource, www.open-source.i3-MARKET.eu
17. Kruchten, P. B. (1995). The 4+ 1 view model of architecture. IEEE software, 12(6), pp. 42-50.
18. Dragoni, N., Giallorenzo, S., Lafuente, A. L., Mazzara, M., Montesi, F., Mustafin, R., & Safina, L. (2017). Microservices: yesterday, today, and tomorrow. Present and ulterior software engineering, pp. 195-216.
19. INFINITECH, https://www.infinitech-h2020.eu/
20. FIBO: Financial Industry Business Ontology, https://edmcouncil.org/frameworks/industry-models/fibo
21. FIGI: FINANCIAL INSTRUMENT GLOBAL IDENTIFIER, https://www.omg.org/figi/
22. ESTRELLA, http://www.estrellaproject.org/?page_id=5
23. Introduction to Ethereum Tokens, https://coindcx.com/blog/cryptocurrency/ethereum-tokens-erc-20-erc-721-erc-1155
24. HyperLedger Fabric, https://www.hyperledger.org/projects/fabric

25. Finsecurity Solutions, https://finsecurity.eu/
26. European Cluster for Securing Critical Infrastructures, https://www.finsec-project.eu/ecsci
27. EU-CIP, https://www.eucip.eu/
28. What is SOA?, https://aws.amazon.com/what-is/service-oriented-architecture
29. SecureIoT, https://secureiot.eu
30. PolicyCloud, https://policycloud.eu/
31. DataVaults, https://www.datavaults.eu/
32. International Data Spaces Associations, https://internationaldataspaces.org/
33. International Data Spaces, https://www.fraunhofer.de/en/research/lighthouse-projects-fraunhofer-initiatives/international-data-spaces.html
34. IDS Business Layer, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3-1-business-layer
35. IDS Functional Layer, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_2_functionallayer
36. IDS Information Layer, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_3_informationlayer
37. IDS Process Layer, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_4_process_layer
38. IDS System Layer, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_5_0_system_layer
39. IDS Security Perspective, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/perspectives-of-the-reference-architecture-model/4_perspectives/4_1_security_perspective
40. IDS Certification Perspective, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/perspectives-of-the-reference-architecture-model/4_perspectives/4_2_certification_perspective
41. IDS Data Governance Perspective, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/perspectives-of-the-reference-architecture-model/4_perspectives/4_3_governance_perspective
42. IDS Connector, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_5_0_system_layer/3_5_2_ids_connector
43. IDS Data Connector Report, https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Data-Connector-Report-7_May-2023-2.pdf
44. IDS Functional Requirements, https://docs.internationaldataspaces.org/ids-knowledgebase/v/idsa-rulebook/idsa-rulebook/3_functional_requirements
45. IDS Usage Control, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/perspectives-of-the-reference-architecture-model/4_perspectives/4_1_security_perspective/4_1_6_usage_control
46. IDS Usage Control Policies, https://international-data-spaces-association.github.io/DataspaceConnector/Documentation/v5/UsageControl
47. IDS Dataspace Protocol v0.8, https://docs.internationaldataspaces.org/ids-knowledgebase/v/dataspace-protocol/overview/readme

48. Is X-Road a Data Space Technology?, https://www.niis.org/blog/2023/6/21/is-x-road-a-data-space-technology
49. IDS Metadata Broker, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_5_0_system_layer/3_5_4_metadata_broker
50. IDS Vocabulary Hub, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_5_0_system_layer/3_5_6_vocabulary_hub
51. IDSA Information Model, https://github.com/International-Data-Spaces-Association/InformationModel
52. IDS Clearing House, https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_5_0_system_layer/3_5_5_clearing_house
53. IDSA IDS-G, https://github.com/International-Data-Spaces-Association/IDS-G
54. IDS Meta Data Broker, https://github.com/International-Data-Spaces-Association/IDS-G/blob/main/Components/MetaDataBroker/README.md
55. IDS Clearing House (IDS-CH), https://github.com/International-Data-Spaces-Association/IDS-G/blob/main/Components/ClearingHouse/README.md
56. IDSA GitHub, https://github.com/International-Data-Spaces-Association/IDS-G-pre/pull/26
57. The International Data Spaces (IDS) Information Model, https://github.com/International-Data-Spaces-Association/InformationModel/blob/develop/README.md
58. IDS Deployment Scenarios, https://github.com/International-Data-Spaces-Association/IDS-Deployment-Scenarios
59. IDS Reference Testbed, https://docs.internationaldataspaces.org/knowledge-base/ids-testbed-1
60. GAIA-X, https://gaia-x.eu/
61. GAIA-X Federation Services, www.gxfs.de
62. GXFS Authentication Authorization, https://www.gxfs.eu/authentication-authorisation
63. GXFS Organizational Credential Manager, https://www.gxfs.eu/organizational-credential-manager/
64. GXFS Personal Credential Manager, https://www.gxfs.eu/personal-credential-manager/
65. GXFS Trust Services API, https://www.gxfs.eu/trust-services-api/
66. GXFS Data Contract Service, https://www.gxfs.eu/data-contract-service/
67. GXFS Data Exchange Logging Service, https://www.gxfs.eu/data-exchange-logging-service/
68. GXFS Core Catalogue Features, https://www.gxfs.eu/core-catalogue-features/
69. GXFS Continuous Automated Monitoring, https://www.gxfs.eu/continuous-automated-monitoring/
70. GXFS Onboarding Accreditation Workflows, https://www.gxfs.eu/onboarding-accreditation-workflows/
71. GXFS Notarization API, https://www.gxfs.eu/notarization-api/
72. GXFS Tutorial, https://www.gxfs.eu/portal/
73. GXFS Orchestration, https://www.gxfs.eu/orchestration/
74. C4 Architecture Model, https://c4model.com/
75. Unified Modelling Language (UML), https://www.uml.org/
76. Kruchten, P. B. (1995). The 4+ 1 view model of architecture. IEEE software, 12(6), pp. 42-50.
77. C4 model for system architecture design, https://icepanel.io/blog/2022-10-03-c4-model-for-system-architecture-design
78. C4 Model for Software Architecture, https://www.infoq.com/articles/C4-architecture-model/

79. OpenDEI – Design Principles for Data Spaces, https://h2020-demeter.eu/wp-content/uploads/2021/05/Position-paper-design-principles-for-data-spaces.pdf
80. Data Spaces Business Alliance: Unleashing the Data Economy, https://data-spaces-business-alliance.eu/wp-content/uploads/dlm_uploads/Data-Spaces-Business-Alliance-Technical-Convergence-V2.pdf
81. Accelerating Business transformation in the Data Economy, https://data-spaces-business-alliance.eu/
82. Performant and modular APIs for Verifiable Data and SSI, https://veramo.io/
83. Verifiable Credentials Data Model v1.1, https://www.w3.org/TR/vc-data-model/
84. Decentralized Identifiers (DIDs) v1.0, https://www.w3.org/TR/did-core/
85. Yuan, E., & Tong, J. (2005). Attributed based access control (ABAC) for web services. In IEEE International Conference on Web Services (ICWS'05).
86. Spring Boot, https://spring.io/projects/spring-boot
87. Postgre SQL, https://www.postgresql.org/
88. Casbin, https://casbin.org/
89. Joinup Licensing Assistant, https://joinup.ec.europa.eu/collection/eupl/solution/joinup-licensing-assistant/jla-find-and-compare-software-licenses
90. Solidity, https://soliditylang.org/
91. ExpressJS, https://expressjs.com/
92. EU Vocabularies, https://op.europa.eu/en/web/eu-vocabularies/dcat-ap
93. Natural Language Processing with Python's NLTK Package, https://realpython.com/nltk-nlp-python/
94. Elasticsearch Python Client, https://www.elastic.co/guide/en/elasticsearch/client/python-api/current/index.html
95. TensorFlow, https://www.tensorflow.org/
96. PyTorch, https://pytorch.org/
97. SciKit Learn, https://scikit-learn.org/stable/
98. Van der Aalst, W., Weijters, T., & Maruster, L. (2004). Workflow mining: Discovering process models from event logs. IEEE transactions on knowledge and data engineering, 16(9), pp. 1128-1142.
99. Van Der Aalst, W., & van der Aalst, W. (2016). Data science in action, pp. 3-23.
100. Proton, https://github.com/ishkin/Proton/
101. Shimizu, S. (2022). Statistical Causal Discovery: LiNGAM Approach.
102. ChatGPT, https://chat.openai.com/chat
103. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1135-1144.
104. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.
105. Jimenez-Peris, R., Burgos-Sancho, D., Ballesteros, F., Patiño-Martinez, M., & Valduriez, P. (2022). Elastic scalable transaction processing in LeanXcale. Information Systems, 108, 102043.
106. JDBC Drivers, https://www.geeksforgeeks.org/jdbc-drivers/
107. Flower: A Friendly Federated Learning Framework, https://flower.dev/
108. Udemy, https://www.udemy.com/
109. Infinitech Marketplace Training, https://marketplace.infinitech-h2020.eu/vdih/training