

## Federated decentralized trusted dAta Marketplace for Embedded finance



### D2.6 - Technical Specifications and Platform Architecture II

Title	D2.6 - Technical Specifications and Platform Architecture II
Revision Number	1.0
Task reference	T2.2
Lead Beneficiary	UPRC
Responsible	Argyro Mavrogiorgou, Thanos Kiortis
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	2024-06-30 [Month 18]
Delivered Date	2024-06-28
Internal Reviewers	ATOS GFT
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	2024-02-01	UPRC	Table of Contents
0.2	2024-03-02	UPRC	Integrated version with contributions
0.3	2024-04-02	ATOS, BPFI, DAEM, ECO, ENG, GFT, IBM, IDSA IQB, JOT, JRC, KM, LXS, MC, MOH, NOVO, SFS, TRB	Updated version with contributions from all
0.5	2024-04-19	ATOS, BPFI, DAEM, ECO, ENG, GFT, IBM, IDSA IQB, JOT, JRC, KM, LXS, MC, MOH, NOVO, SFS, TRB	Updated version with new contributions
0.6	2024-04-28	ATOS, BPFI, DAEM, ECO, ENG, GFT, IBM, IDSA IQB, JOT, JRC, KM, LXS, MC, MOH, NOVO, SFS, TRB	Updated and upgraded version
0.7	2024-05-20	ATOS, BPFI, DAEM, ECO, ENG, GFT, IBM, IDSA IQB, JOT, JRC, KM, LXS, MC, MOH, NOVO, SFS, TRB	Version for peer review
0.8	2024-05-23	ATOS, BPFI, DAEM, ECO, ENG, GFT, IBM, IDSA IQB, JOT, JRC, KM, LXS, MC, MOH, NOVO, SFS, TRB	Update after peer review
0.9	2024-06-27	UPRC	Version for QA
1.0	2024-06-28	UPRC GFT	Version for submission

Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

## Definitions

<b>Acronyms</b>	<b>Definition</b>
4AML	4th Anti Money Laundering Directive
AAI	authentication authorization infrastructure
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
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 Technologies
EC	European Commission
EE	Energy Efficient
ERC	Ethereum Request for Comments
ESG	Environmental, Social and Governance
EU	European Union
FAIR	Findable Accessible Interoperable Reusable
FAME	Federated decentralized trusted dAta Marketplace for Embedded finance
FDAC	Federated Data Assets Catalogue
FIBO	Financial Industry Business Ontology
FIGI	Financial Instrument Global Identifier
FML	Federated Machine Learning
GDPR	General Data Protection Regulation
GOV	Operational Governance
HTAP	Hybrid transaction/analytical processing
HTML	Hypertext Markup Language
HTTP	HyperText Transfer Protocol
ICT	Information Communication Technologies

ID	Identity
IDM	Identity Management
IDS	International Data Spaces
IDSA	International Data Spaces Association
IEC	International Electrotechnical Commission
IEEE	Institute (of) Electrical (and) Electronic Engineers
ISO	International Organization for Standardization
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
NLTK	Natural Language Toolkit
OIDC	OpenID Connect
OS	Operating System
OSS	Open Source Software
PSD2	2nd Payment Services Directive
PSDII	Second Payment Service Directive
PSR	Project Security Responsible
RA	Reference Architecture
RAM	Random Access Memory
REST	Representational State Transfer
RPC	Remote Procedure Call
SA	Supervisory Authority
SAX	Situation Aware eXplainability
SDK	Software Development Kit
SLA	Service Level Agreement
SOA	Service Oriented Architecture
SQL	Structured Query Language
SSE	Semantic Search Engine
SSI	Server Side Includes

---

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
XAI	Explainable Artificial Intelligence

---

## Executive Summary

The document presents the complete 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 Space for Embedded Finance (EmFi). To this context, the main focus of the deliverable is to set the technical background of the FAME SA, and make clear that the final goal of the project is to realize a practical and mature Federated Data Space to facilitate finance and non-finance organizations to securely access, share, trade, and analyse 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 file snapshot is represented in the current deliverable, considering all the improvements that have been performed over this period.

Following the same design principles of the first version of the FAME SA residing within D2.2, the FAME partners continued working the FAME SA by following the "C4 model" methodology, which is presented in this deliverable as well. 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 implemented within 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 summarized, consisting of a detailed literature review on FAME related state-of-the-art definitions (e.g., data lakes, data warehouses, data marketplaces, data spaces) related with the realization of the overall FAME ecosystem, including its challenges and provided opportunities. The business value, the technical value as well as the European value of FAME is also presented, all of which have led to the specification of the final version of the FAME SA. What is more, this deliverable underlines and thoroughly explains the existing Reference Architectures (RAs) and SAs that have been adopted within the FAME overall technical designs and implementations. By considering those RAs/SAs and the current developments and advancements of all the FAME technical components, the resulting final version of the FAME SA provides a complete schema for building a solid workflow for realizing a Federated Data Space, clarifying the underling communications and interactions among all the FAME building blocks and its interactive end-users.

As a final note, since this deliverable is a continuation of D2.2 (first version of the FAME SA deliverable), it should be stated that there might be text that is repeated, since the current deliverable has been built upon D2.2. Rather than this, the detailed changes that have been performed can be located in Section 1.4 of the current deliverable.

# Table of Contents

1	Introduction.....	6
1.1	Objective of the Deliverable .....	6
1.2	Insights from other Tasks and Deliverables.....	7
1.3	Structure.....	7
1.4	Summary of Changes .....	8
2	Terminology.....	10
3	FAME Baseline Concepts.....	11
3.1	Concepts.....	11
3.1.1	Data Associated Technologies .....	11
3.1.2	Centralized, Decentralized & Federated Data Approaches .....	13
3.1.3	Data Marketplaces Features .....	15
3.2	Challenges.....	17
3.2.1	Technical Challenges .....	18
3.2.2	Business & Organizational Challenges.....	19
3.2.3	Legal Compliance Challenges .....	20
3.2.4	National & Regional Challenges.....	20
4	FAME Value Proposition.....	21
4.1	FAME in a Nutshell .....	21
4.2	FAME Technical Value .....	23
4.3	FAME Business Value.....	25
4.4	FAME European Value.....	26
5	FAME Stakeholders.....	28
6	FAME Technical Background .....	31
6.1	Significance of Reference Architectures.....	31
6.2	Overview of Reference Architectures for Data Marketplaces .....	32
6.2.1	i3-Market.....	32
6.2.2	INFINITECH .....	36
6.2.3	FINSEC.....	38
6.2.4	SecureIoT .....	41
6.2.5	PolicyCLOUD.....	44
6.2.6	DataVaults.....	46
6.2.7	OMEGA-X.....	48
6.2.8	TRUSTS.....	50

6.3	Overview of Reference Architectures for Data Spaces .....	52
6.3.1	GAIA-X .....	52
6.3.2	IDS .....	56
7	FAME Capabilities .....	61
7.1	Operational Governance .....	62
7.2	Authorization & Authentication.....	63
7.3	Federation of External Sources .....	64
7.4	Assets Policy Management .....	64
7.5	Assets Provenance & Tracing .....	65
7.6	Assets Pricing.....	65
7.7	Assets Trading & Monetization .....	66
7.8	Assets Searching .....	66
7.9	ML & AI Analytics .....	67
7.10	SAX & XAI Analytics .....	67
7.11	Incremental & Energy Efficient Analytics.....	68
7.12	Smart Deployment .....	69
7.13	Analytics CO <sub>2</sub> Monitoring .....	69
7.14	FML Deployment.....	69
7.15	Learning Centre.....	70
7.16	Regulatory Compliance.....	70
7.17	Dashboard .....	71
8	FAME Solution Architecture .....	72
8.1	C4 Architecture Model.....	72
8.2	FAME Architecture Overview .....	75
8.2.1	FAME SA Conceptual Overview .....	75
8.2.2	FAME SA C4 Model .....	76
8.2.3	FAME SA Mapping to DSSC Building Blocks.....	82
8.2.4	FAME as a Data Space Intermediary.....	87
8.3	Specification of Architecture Components .....	88
8.3.1	Operational Governance .....	89
8.3.2	Authorization & Authentication.....	90
8.3.3	Assets Policy Management .....	91
8.3.4	Regulatory Compliance.....	93
8.3.5	Assets Provenance & Tracing .....	94
8.3.6	Assets Pricing.....	95
8.3.7	Assets Trading & Monetization .....	97

8.3.8	Federated Data Assets Catalogue.....	100
8.3.9	Semantic Interoperability .....	102
8.3.10	Semantic Search Engine.....	103
8.3.11	ML & AI Analytics .....	104
8.3.12	SAX & XAI Techniques .....	106
8.3.13	Incremental & Energy Efficient Analytics.....	108
8.3.14	Smart Deployment .....	109
8.3.15	Analytics CO <sub>2</sub> Monitoring .....	110
8.3.16	FML Deployment.....	111
8.3.17	Learning Centre.....	112
8.3.18	Dashboard .....	114
9	Conclusions.....	116
	References.....	117

## List of Figures

Figure 1	– Relevance of existing data management associated technologies.....	13
Figure 2	– Difference of centralized, decentralized, and federated data approaches.....	14
Figure 3	– Data Marketplaces’ Features .....	17
Figure 4	– FAME in a Nutshell.....	22
Figure 5	– FAME stakeholders .....	29
Figure 6	– Logical View with i3-MARKET .....	33
Figure 7	– i3-MARKET microservices layout.....	34
Figure 8	– High level view of Infinittech architecture .....	37
Figure 9	– Technical architecture of Finsecurity.eu platform.....	39
Figure 10	– SecureIoT marketplace conceptual architecture.....	41
Figure 11	– PolicyCLOUD data marketplace conceptual architecture.....	45
Figure 12	– DataVaults architecture overview .....	47
Figure 13	– OMEGA-X Marketplace architecture .....	49
Figure 14	– TRUSTS Platform architecture .....	51
Figure 15	– GXFS supported elements .....	53
Figure 16	– General structure of IDS Reference Architecture Model .....	57
Figure 17	– Interaction of technical components.....	58
Figure 18	– DSSC building blocks .....	60
Figure 19	– C4 model diagram types [84] .....	73
Figure 20	– Conceptual overview of FAME architecture.....	76
Figure 21	– FAME SA C4 context diagram .....	77
Figure 22	– FAME SA C4 container diagram .....	78
Figure 23	– Mapping of FAME SA to the DSSC Building blocks.....	83
Figure 24	– FAME Data Space as an Intermediary, providing DSSC Building blocks .....	88
Figure 25	– Operational Governance C4 component-level architecture.....	89
Figure 26	– Authorization & Authentication C4 component-level architecture.....	90
Figure 27	– Assets Policy Management C4 component-level architecture .....	92
Figure 28	– FAME related laws & regulations .....	94

Figure 29 – Assets Provenance & Tracing C4 component-level architecture .....	95
Figure 30 – Assets Pricing C4 component-level architecture.....	96
Figure 31 – Assets Trading & Monetization C4 component-level architecture - REST API.....	98
Figure 32 – Assets Trading & Monetization C4 component-level architecture - Currency Contracts .....	98
Figure 33 – Assets Trading & Monetization C4 component-level architecture - Assets Contracts ..	99
Figure 34 – Assets Trading & Monetization C4 component-level architecture - Trading Contracts .....	100
Figure 35 – Federated Data Assets Catalogue C4 component-level architecture.....	101
Figure 36 – Semantic Interoperability C4 component-level architecture .....	102
Figure 37 – Semantic Search C4 component-level architecture .....	103
Figure 38 – ML & AI Analytics C4 component-level architecture .....	105
Figure 39 – SAX Techniques C4 component-level architecture .....	106
Figure 40 – XAI Scoring Framework C4 component-level architecture.....	107
Figure 41 – Incremental Analytics C4 component-level architecture .....	108
Figure 42 – Smart Deployment C4 component-level architecture. ....	110
Figure 43 – Analytics CO <sub>2</sub> Monitoring - Requirements’ Analysis C4 component-level architecture .....	110
Figure 44 – Analytics CO <sub>2</sub> Monitoring - CO <sub>2</sub> Footprint Analyzer C4 component-level architecture .....	111
Figure 45 – FML Deployment C4 component-level architecture.....	112
Figure 46 – Learning Centre C4 component-level architecture.....	113
Figure 47 – Dashboard end-users’ visual sitemap .....	114
Figure 48 – Dashboard administrators’ visual sitemap.....	115

## List of Tables

Table 1 – D2.6 Updates from D2.2.....	8
Table 2 – Data Marketplaces’ Features.....	15
Table 3 – FAME Overall Capabilities .....	61
Table 4 – FAME SA Context Diagram notations .....	74
Table 5 – FAME SA Container Diagram notations .....	74
Table 6 – FAME SA Component Diagram notations .....	75
Table 7 – D2.6 Related KPIs .....	116

# 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 [1] serving as its primary tangible product. The Data Governance Act [2] 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 [3], 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 fourteen (14) common European Data Spaces: Agriculture, Cultural Heritage, Energy, Finance, Green Deal, Health, Language, Manufacturing, Media, Mobility, Public Administration, Research and Innovation, Skills, and Tourism [4]. 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) [5] 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, summarizing 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), Machine Learning (ML), Blockchain). FAME’s main goal is to provide EmFi with Europe’s first standard-based, secure, interoperable, and federated Data Space. To this context, as an updated version of D2.2, this deliverable provides the final version of the FAME SA that makes use of already existing reference models for constructing a Data Space, connecting various external data sources (e.g., external Data Marketplaces, Datastores), also providing interfaces for pricing, trading, and managing data assets. FAME also offers 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 report the final version of 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, Infinittech) and Data Spaces (e.g., IDS, Gaia-X). The deliverable outlines the roadmap for realizing this final version of the FAME SA, including its value proposition, stakeholders, addressed challenges, and provided opportunities. FAME SA’s involved components are also thoroughly described and technically specified, providing their added-value to the overall

SA. It is worth mentioning that the presented FAME SA has driven the technological developments of the project, as well as the design and the integration of the underlying use cases and scenarios.

## 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 have driven 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 the final version of the platform's requirements and specifications deliverable of WP2, namely the D2.5 - Requirements Analysis, Specifications and Co-Creation II [6], where this deliverable considers the final extracted both business and technical requirements of the underlying components, use cases and platform stakeholders, along with the EmFi application domain's needs, towards providing the final version of the SA.
- Finalized based on all the current and future implementations of the platform, also reported in the relevant released technical deliverables or to be reported in the upcoming technical deliverables deriving from WP2, WP3, WP4, WP5, and WP7. To this context, the current deliverable provides insights and technical specifications for the detailed design of the technical components of all those WPs, which are thoroughly described within this deliverable. Overall, this 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 understandability.
- *Section 3* includes the basic concepts and overall idea behind the FAME project, analysing 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.
- *Section 5* introduces all the potential stakeholders/end-users of the overall FAME ecosystem.
- *Section 6* provides all the necessary information regarding existing RAs, principles, and functionalities that are 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 7* 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 8* 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 9* contains the conclusions that have been derived from the FAME SA and its overall components' technical specifications, concluding to the achieved KPIs.

## 1.4 Summary of Changes

Compared with the previous version of this deliverable (D2.2 - Technical Specifications and Platform Architecture I [7]), the Sections/sub-Sections that have been updated/moved/added within D2.6 are depicted in Table 1.

Table 1 – D2.6 Updates from D2.2.

Section	Status	Description of Update/Addition
<b>Section 1</b>		
1.2 - Insights from other Tasks and Deliverables	Updated	Alignment of D2.6 input regarding the technical activities and progress of the project.
1.4 - Summary of Changes	Added	Depiction of updates between D2.6 and D2.2.
<b>Section 2</b>		
2 - Terminology	Updated	Update of current terminologies' list, to also provide the "Analytics" as well as the "(FAME) Analytics" terms that are widely used within the project.
<b>Section 3</b>		
3.1.1 - Data Associated Technologies	Updated	Update of Figure 1 to better depict the connection among the diverse data concepts and their relevance with FAME.
3.1.3 - Data Marketplaces Features	Added	Description of features that need to be present in any Data Marketplace, and thus being also covered through the FAME's Data Marketplace functionalities.
<b>Section 4</b>		
4.1 - FAME in a Nutshell	Updated	Update of Figure 4 to depict the updated FAME strategy and the overall FAME value proposition towards realizing a Federated Data Space with its own business, legal and technology framework. Also, the concepts of the EmFi application domain are outlined, to make clear the domain that FAME targets to contribute.
4.2 - FAME Technical Value	Updated/Added	Update of components provided within FAME, based on its current offerings and implementations. Also, addition of all the cutting-edge technologies leveraged by FAME and its developed components.
4.3 - FAME Business Value	Updated	Update of text to better reflect the current vision as well as the business orientation of FAME.
4.4 - FAME European Value	Added	Analysis of FAME's contribution in the context of the European Union's data strategy.
<b>Section 5</b>		
5 - FAME Stakeholders	Moved	Addition of FAME stakeholders and their offered added-value to a new Section (part of Section 4 in D2.2).
<b>Section 6</b>		
6.2 - Overview of Reference Architectures for Data Marketplaces	Updated/Added	Study and description of two (2) additional RAs that were considered during the finalization of the FAME SA, referring to the RAs of the OMEGA-X project (sub-Section 6.2.7) and the TRUSTS project (sub-Section 6.2.8). Also, the "Relevance to FAME & Added-Value" parts were updated for all the described RAs, to provide the performed final alignments and contributions between FAME and the chosen platforms/systems.

6.3 - Overview of Reference Architectures for Data Spaces	Updated/ Added	Update of IDS concepts to depict the advancements within IDS and the larger Data Spaces' ecosystem since the writing of D2.2. Also, the "Relevance to FAME & Added-Value" parts were updated both for the GAIA-X and the IDS exploited concepts and guidelines to depict their current exploitation and contribution within FAME.
<b>Section 7</b>		
7 - FAME Capabilities	Updated	Update of Table 3 and its related text to depict the overall FAME supported capabilities based on its final offerings/implementations. Also, all the included sub-Sections (sub-Section 7.1 – 7.17) were updated to depict the final goal and supported functionalities of each module offering such capabilities.
7.12 - Smart Deployment	Added	Description of the "Smart Deployment" capability as it was introduced a new capability of the platform.
<b>Section 8</b>		
8.1 - C4 Architecture Model	Updated	Update of Tables 4, 5 and 6 to depict the final supported "Person" Abstractions of the platform.
8.2.1 - FAME SA Conceptual Overview	Updated	Update of Figure 20 to illustrate the final overall concept support the FAME, outlining its offerings both to the data consumers and the data providers.
8.2.2 - FAME SA C4 Model	Updated	Update of Figures 21 and 22 to depict the final SA of FAME across different levels of C4 diagrams (Context and Container diagrams). Also, the relevant texts have been updated to capture the final status of the FAME SA in accordance with its included technological layers.
8.2.3 - FAME SA Mapping to DSSC Building Blocks	Updated/ Added	Description of how the FAME implementation will comply with the DSSC guidelines and building blocks, provided detailed texts and figures for clarifying this mapping.
8.2.4 - FAME as a Data Space Intermediary	Added	Study and description of how FAME can be positioned as a Data Space intermediary, considering different points of view.
8.3 - Specification of Architecture Components	Updated/ Added	Update of all the included sub-Sections (sub-Section 8.3.1 – 8.3.17) to depict the final FAME internal interconnections among each component, each component's internal architecture (through its updated C4 component-level diagram), its offered functionalities, as well as its exploited technologies. Also, sub-Section 8.3.18 was added to outline the abovementioned information for the FAME Dashboard that was not implemented during the release of D2.2.
<b>Section 9</b>		
9 - Conclusions	Updated	Update of the overall conclusions of the deliverable to reflect the final findings and outcomes of the deliverable, accompanied by the update of the KPI numbers depicted in Table 7 to depict the real numbers captured until the FAME SA's finalization.

## 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 (2) 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 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.
- **Analytics:** A data-processing pipeline/process that improves the user's understanding of an asset by extracting structures, patterns or metrics from data and models.
- **(FAME) Analytics:** In the context of FAME the analytics can refer to AI models, XAI/SAX processes, data handling or energy efficiency processes, among others.
- **(FAME) Dashboard:** A user interface for FAME stakeholders to interact with the overall FAME Federated Data Space and its functionalities.

## 3 FAME Baseline Concepts

The current Section includes the basic concepts and overall idea behind the FAME project. In essence, some of the most critical data associated technologies are analysed, whereas some related state-of-the-art methodologies, data approaches and features 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 [8]. Gartner's research illustrated how much confusion there is about what the different concepts entail, which is also intensively spotted in the research community [9][10]. 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 BI 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's data* (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 goal is to provide a complete 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 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 data from various industries and domains. Thus, the goal is to serve as a hub where data is curated, organized, and made available for end-users to be 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 (data platform(s), data marketplace(s)) 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.
- **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.

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 [11]. 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.

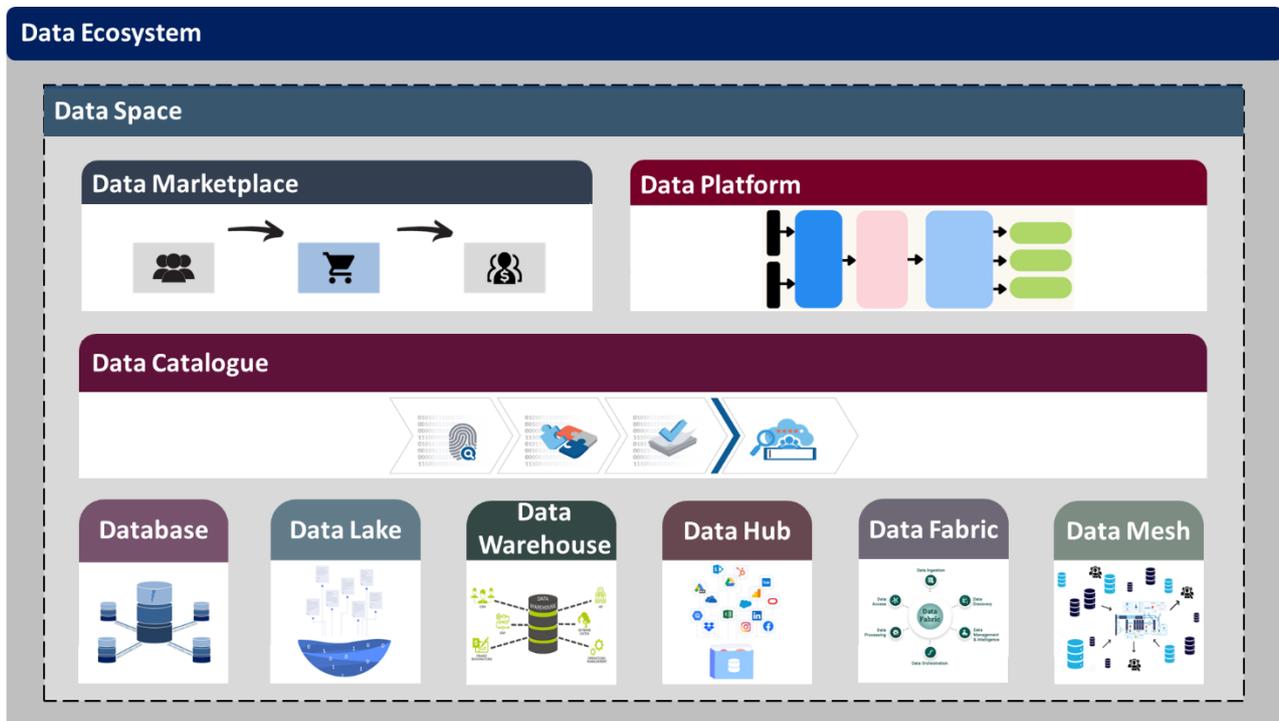


Figure 1 – Relevance of existing data management associated technologies

Through Figure 1, it should be made clear that FAME targets at delivering a Data Space that includes its own Data Marketplace as its fundamental technology framework.

### 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. [12][13][14].

- **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 [15]. All 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.

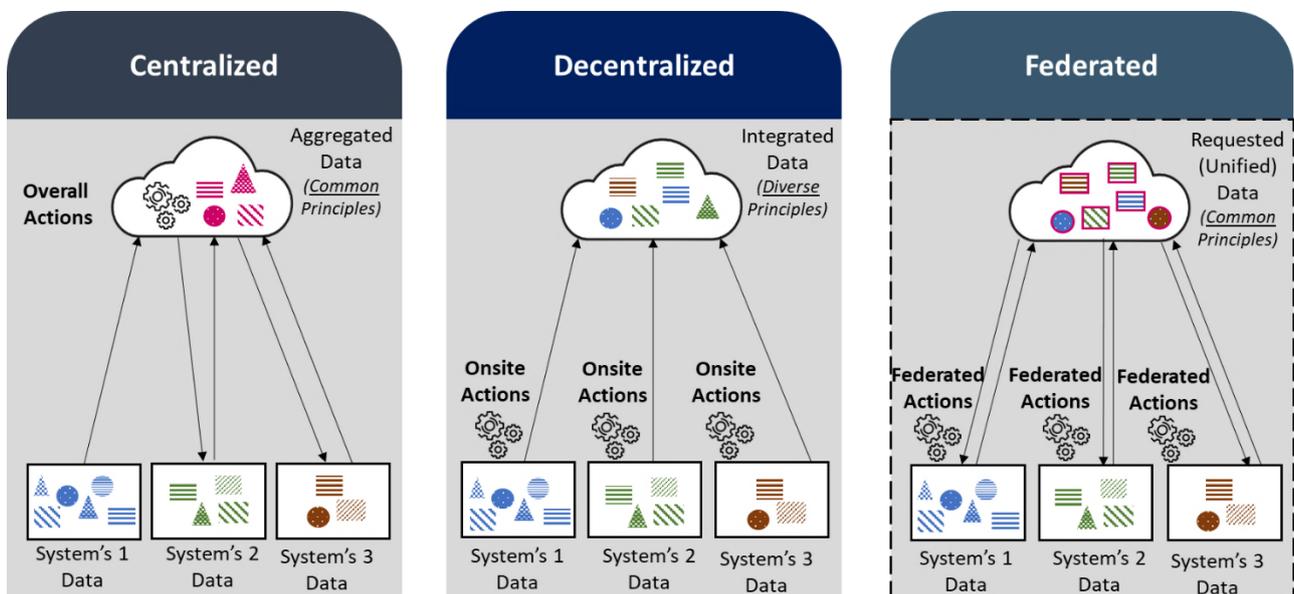


Figure 2 – Difference of centralized, decentralized, and federated data approaches

### 3.1.3 Data Marketplaces Features

As mentioned above, the Data Marketplace is the main technical component that supports the FAME Federated Data Space. This Section enhances the understanding of what functions Data Marketplaces possess and how they differ from generic marketplaces (e.g., Amazon, eBay, Aliexpress). Based on the literature that was extracted from four (4) Data Marketplaces studies that discuss the high-level view/business model of Data Marketplaces [16][17][18][19], there exist specific features that are most relevant to Data Marketplaces (Table 2), categorized under the market orientation, independent ownership, and many-to-many matching mechanisms. Other Data Marketplace categories may only provide these features partially.

Table 2 – Data Marketplaces’ Features

ID	Feature	Category
1	Assign <b>user verification and certification</b>	Onboarding
2	Upload <b>(meta)data</b> for data providers	Dataset Discovery
3	Provide an <b>electronic catalogue</b> that displays the current data, product listings, and provider directories that are available	
4	Allow data buyers to submit selection criteria (such as a search keyword or query) using <b>advanced search</b> , also <b>providing filtering</b> and sorting features	
5	Provide reliable <b>matching algorithms</b> to link consumers and producers of data	
6	Enable data suppliers to choose who can view or purchase their data products, giving <b>visibility management</b>	Trading Arrangements
7	Offer <b>pricing models</b> to ascertain the business plan for a Data Marketplace, where prior to a transaction, <b>pricing discovery</b> must establish the monetary value of the data (data price)	
8	Establish licensing or <b>contract terms</b> pertaining to the data, such as data ownership and usage	
9	Establish <b>communication channels</b> to let data suppliers and buyers negotiate and communicate.	
10	View data assets before being purchased to determine their value ( <b>pre-purchase testability</b> )	
11	Validate contractual condition agreements using <b>smart contract</b> mechanisms	Transaction Workflow
12	Offer <b>billing systems</b> that let data purchasers pay data suppliers	
13	Transfer data from data suppliers to data buyers by executing <b>transaction workflow</b> . Should there be any disruptions, the transaction can be completed at the same location	
14	Participants in the data markets are able to assess data assets and present the total reputation of the data providers (via <b>review system</b> )	Review System
15	Provide the necessary infrastructure for <b>transaction security</b> , such as HTTPS and encryption protocols, to enable safe data exchanges	Security
16	Support smart contracts implemented on <b>blockchain</b> technology	

17	Use technologies like two-factor authentication to provide <b>profile security</b> and guarantee that only verified and authorized individuals may access their profiles and responsibilities	
18	Utilize <b>anonymization</b> to safeguard individual data (removal of personal identity/information)	Privacy Preserving
19	Provide <b>encryption</b> to protect data assets	
20	Provide a defined set of interoperability guidelines (such as an API description) and an access type for data exchange via <b>APIs</b> between data buyers and suppliers	Interoperability
21	Provide either dynamic or static <b>data streams</b>	
22	Support <b>data provenance mechanisms</b> that track data from its origin to its destination	Data Governance
23	Offer (verifiable) <b>metadata management</b> to provide information on the data content, origin, rights, collection, etc.	
24	Provide <b>semantic representation</b> models	
25	Support additional data analysis services, like <b>data visualization</b> , aiding in the simplification of complicated data into graphical displays (e.g., scorecards, diagrams, heat maps)	Data Analysis
26	Provide <b>data normalization</b> to identify data that has been traded to predefined formats, data models, and attributes	Data Transformation
27	Provide <b>data aggregation</b> to perform combinations of multiple datasets	
28	Support <b>multiple formats of data outputs</b>	
29	Offer tools for trust management to monitor data consumers' <b>contractual compliance</b>	Data Monitoring
30	Guarantee <b>data quality</b>	
31	Provide <b>purchase history</b> of data buyers	User Management
32	Make <b>dashboards</b> available to users so they can manage offerings, requests, transactions, and other operations in a consolidated manner	
33	Offer <b>notification control</b> and support feature (e.g., helpdesk, chatbot)	
34	Provide the <b>purchase trend</b> and trade leads	News Service
35	Provide <b>notification channels</b> to enable notification to users about related and recent news (e.g., social networks, emails)	

The above Table (Table 2) constitutes a collection of functionalities that are used in the state-of-the-art Data Marketplace initiatives. Analyzing all those 35 features, it has been derived that all of them are fully compatible with the requirements and vision of the FAME Federated Data Space and its Data Marketplace to be released, as these are defined in D2.5 and furtherly explained in the current deliverable. In addition, these features are being depicted in Figure 3, mapping the four (4) generic building blocks of Data Marketplaces (i.e., core process, technological backbone, data service ecosystem, and account management), into the categories presented in Table 2.

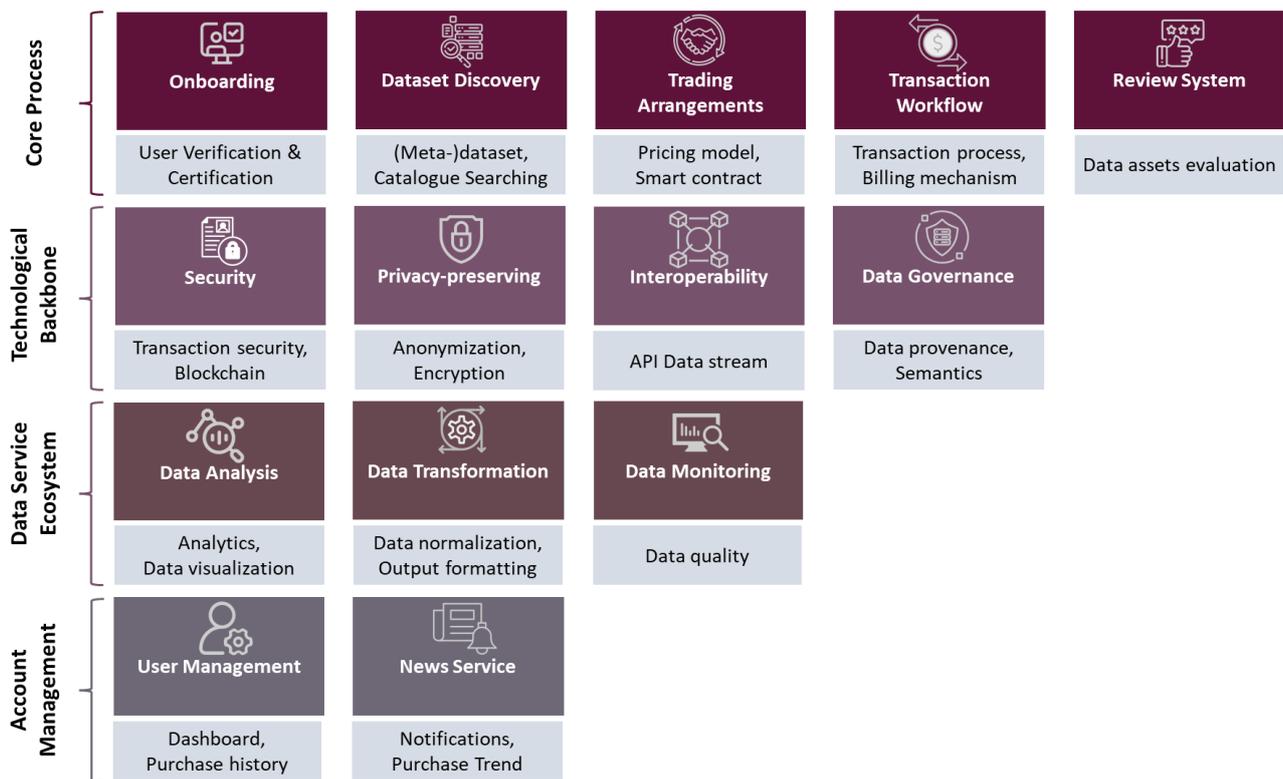


Figure 3 – Data Marketplaces' Features

### 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 identified. 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 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 to 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, whilst 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. Thus, the issue is to assess investment in data-centric businesses and guarantee that 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.

## 4 FAME Value Proposition

FAME purpose is to provide a regulated data assets exchange and ecosystem, compliant with EU regulations and strategies, enabling organizations to leverage on the FAME solutions (i.e., the Data Marketplace and all the underlying technology components), solving the problem of data sovereignty, data security, data trading, data trustworthiness by federating with other participants that share common values and rules to interoperate. This Section presents the current state of related activities and identifies the challenges that exist in the mission to fill in the gap of current Data Marketplaces for realizing a Data Space ecosystem to serve applications with financial services. A high-level overview of FAME is provided, detailing how its contributions are designed to tackle the challenges described in Section 3. Then, the Section highlights the technical and the business values of FAME.

### 4.1 FAME in a Nutshell

Considering the current status of Data Spaces and their cross-cutting challenges (analysed 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. To go one step beyond this problem, 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, ML, blockchain).

When referring to the *EmFi application domain*, it should be outlined that this is related with the *integration of financial services*, such as lending, payment processing and insurance, *into non-financial offerings* with examples of the related application domain including an e-commerce merchant providing insurance, a coffee shop application that offers one-click payments, or a department store's branded credit card. This process is done without the need to route these services through traditional financial institutions and uses companies such as Affirm, Afterpay, Klarna and Uplift to finance a product purchase instead. EmFi is not a new concept. Nonbanks have offered private-label credit cards for retail stores and airlines for years. Another example includes sales financing at car dealerships. These arrangements offer a channel for the banks behind them to reach the customer without direct contact. Customers see examples of EmFi on digital interfaces they interact with such as digital wallets, shopping cart platforms, loyalty applications and even social media e-commerce platforms. EmFi started with the changes in consumer behaviour and technology, whereas with the digitalization of e-commerce and business management, the opportunity to offer nonfinancial customer experiences has increased. Having EmFi options brings financial services to a customer right when they need it instead of searching for services separately. To this context, FAME will represent a practical and mature Federated Data Space that will be accessible through a single-entry point (its offered Data Marketplace) to facilitate finance and non-finance organizations, external Data Marketplaces/Data Spaces, and single users (thoroughly described in Section 5) towards securely accessing, sharing, trading, and analysing their data in a federated manner for facilitating the realization of EmFi applications by exploiting such FAME features. 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 4, 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).

It should be noted that the Data Marketplace offered by FAME is not just another Data Marketplace, as it is specifically engineered to support the burgeoning field of Data-Driven finance applications. This focus allows FAME to cater to a niche yet rapidly expanding segment of the financial industry. By providing a unified platform, FAME facilitates the discovery and utilization of diverse data assets tailored to meet the specific needs of various entities within the financial sector, ranging from banking institutions to fintech startups, enabling them to innovate and deliver services efficiently. On top, for realizing the FAME Federated Data Space, the existence of a *Legal* as well as a *Business Framework* are also of great importance, and that is why such frameworks are put into practice and harmonized with the *Technology Framework* (i.e., Data Marketplace) of FAME (as shown in Figure 4).

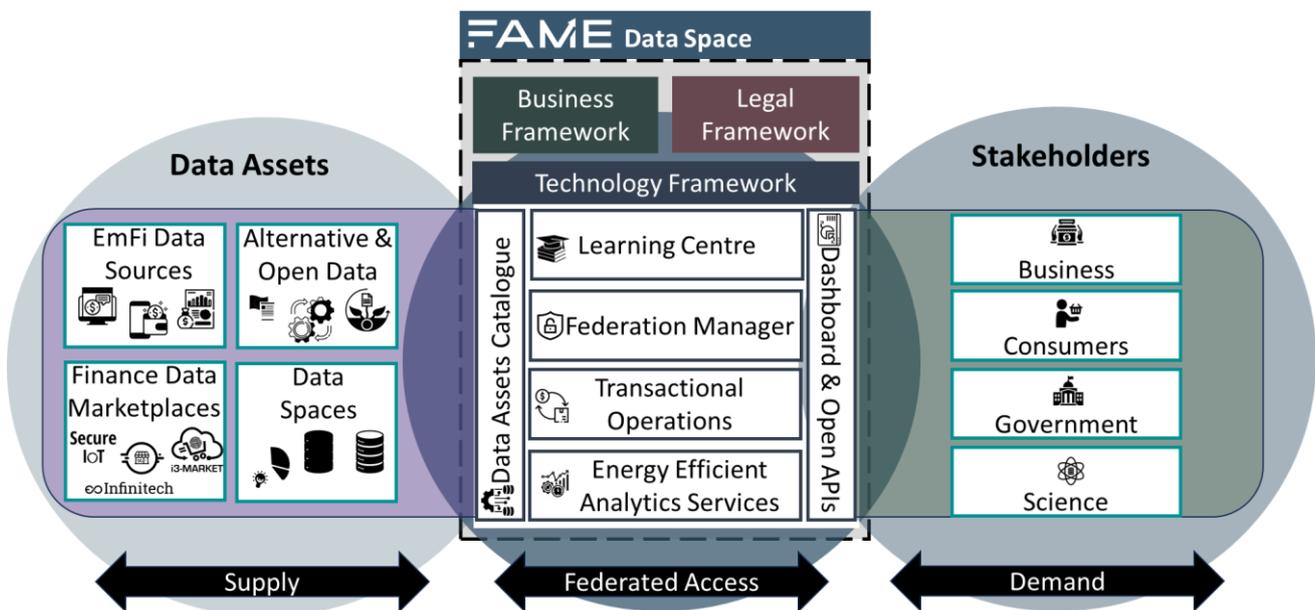


Figure 4 – FAME in a Nutshell

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 technology framework of FAME is the Data Marketplace that heavily relies on its developed Federated Data Assets Catalogue (FDAC) 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 Marketplace 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 assets. Towards this direction, FAME provides novel functionalities in three (3) core directions: (i) Secure, interoperable and regulatory compliant data exchange across multiple federated data providers in-line with emerging EU 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. Further details on these functionalities of the FAME Technology Framework are provided in Section 6, whereas for the Business Framework additional details are provided in sub-Section 4.3.

## 4.2 FAME Technical Value

By actively following all the abovementioned core directions, and all the analysed challenges (Section 3), the implementation of the following FAME technical components (furtherly analysed in Section 8) has taken place, which represent the fundamental building blocks of FAME's overall functionality:

- **Authorization & Authentication:** It addresses the challenges of “Security” through providing the possibility of securely granting access to either an end-user or an administrator of the platform, acting as a single-entry point to the FAME Data Marketplace.
- **Federation of External Sources:** It addresses 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 addresses 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 addresses the challenges of “EU Values”, “Data Protection”, “Regulatory Compliance”, and “EU-Wide Policies” through ensuring the secure management of the FAME indexed data assets, regulating who can view and have access to them.
- **Assets Provenance & Tracing:** It addresses 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 addresses 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 addresses 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 addresses 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 addresses the challenges of “Free-Flowing Data”, “EU Values” and “Global Competition” providing the ability to the FAME end-users to locate and execute specific ML/AI models for their preferred scenarios.
- **SAX & XAI Analytics:** It addresses 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<sub>2</sub> Monitoring:** It addresses 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 addresses the challenges of “Dynamic Ecosystems” through giving to the FAME end-users the opportunity for data sharing in federated learning scenarios.
- **EmFi Training:** It addresses 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 and Data Marketplaces (focusing on EmFi applications), by offering a plethora of dedicated training programs and a related learning centre.

- **Dashboard:** It addresses 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.

For successfully achieving the implementation of the abovementioned components, FAME leverages cutting-edge technologies to enhance functionality and user experience:

- **Secured and Trusted, Adaptive Identity & Access:** FAME’s Authentication and Authorization is managed in a secured manner through Digital Identities (DIs). Adaptive trust using decentralized technologies between infrastructures and services dynamically for ensuring the protection of confidentiality in federated security domains for different applications are also used. This ensures adaptive policies enforcement for restricting unauthorized access.
- **Semantic Interoperability:** FAME promotes adoption of the extended Data Catalog Vocabulary (DCAT) standards and financial sector-specific ontologies like FIBO/FIGI, ensuring that data can be seamlessly exchanged across different platforms and marketplaces without semantic discrepancies.
- **AI & ML:** FAME provides advanced AI technologies for analytics and decision-making support, including Quantitative Explainable AI and Situation Aware Explainability (SAX), which enhance the analytical capabilities of the platform. It also offers other AI-related solutions, such as a Federated Learning approach to guarantee data privacy while training models, or a smart deployment service that considers CO<sub>2</sub> emissions to adapt the deployments of the AI models.
- **Blockchain:** Technologies rooted in Blockchain are used to achieve two (2) goals: (i) Enhance transparency and trust in the FAME ecosystem, by tracking the provenance of data assets and securing the integrity of the data assets catalogue, and (ii) Enable the decentralized trading of data assets, by means of automated provider/consumer transactions (Smart Contracts), asset tokenization (Non-Fungible Tokens), and an internal digital currency (Stablecoin).
- **Self-Sovereign Identities:** FAME has a privacy-friendly, user-centric approach to identity management, as it adopts the European Digital Identities framework as the foundation of its online access control system.
- **Regulatory Compliance and Law:** FAME’s ambition is to deliver Europe’s first standards-based, secure, regulatory compliant, interoperable, and Federated Data Marketplace for Embedded Finance (EmFi) applications. Apart from the unique feature of the FAME project, its federated access control, the project provides a unified access to all the related regulations in the field. In that sense, it provides a harmonious ecosystem of laws and regulations according to the needs of different stakeholders.

These technologies not only boost the operational efficiency of FAME, but also ensure that the platform remains at the forefront of the Data Spaces/Data Marketplaces industry, capable of meeting the sophisticated demands of modern financial services.

### 4.3 FAME Business Value

As a federated data exchange encompassing various data assets (e.g., datasets, algorithms, runtime integrable IT services, tutorials), FAME offers specialized functionalities that facilitate the discovery and utilization of data and technological components by different entities, such as assets providers, assets consumers, and application owners. Addressing the specific needs and use cases and the semantics of the applications that embed financial services, FAME can be considered as 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 and “composable finance 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 and associations (e.g., IDSA and Gaia-X among others like DSSC). 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:

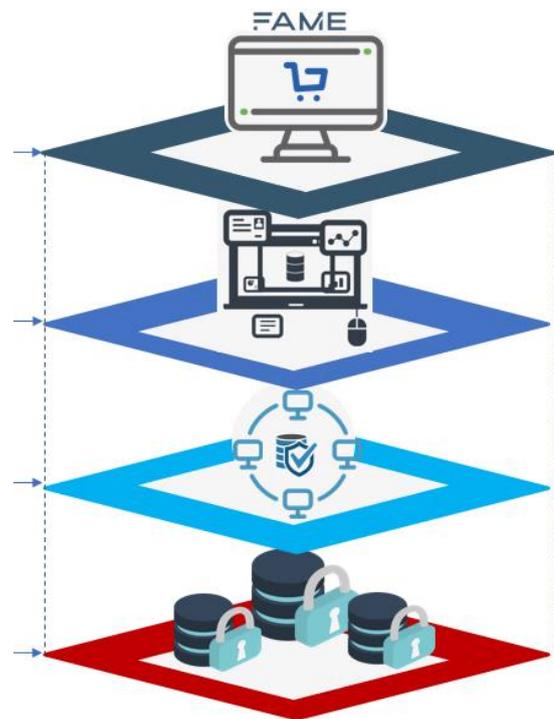
**FAME Federated Data Space** will be a **functional platform to federate data and technology assets** for EmFi data-driven applications in compliance with the EU data strategy and regulations.

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** provides:

- **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.



#### 4.4 FAME European Value

The European Union's data strategy, formulated in 2020, aims to establish a unified European Data Space by 2030, promoting economic growth while prioritizing environmental sustainability. This initiative encompasses legislative measures, governance frameworks, and investment in digital infrastructure and competences to ensure secure data exchange and compliance with European values.

Sector-specific common European Data Spaces will be developed to cater to industry needs and promote innovation. At the same time, legislative actions, including the Data Governance Act and the Data Act, the Data Service Act, and the AI Act have been taken that are crucial for regulating data marketplaces, ensuring transparency, trustworthiness, and data sovereignty.

As outlined in the Introduction Section, the EU Commission has already identified fourteen (14) common European Data Spaces, one of which is Finance. Key features of these Data Spaces include secure infrastructure, FAIR access, adherence to EU rules and values, and the ability for data holders to grant access or share data within the Data Space.

These infrastructures (Data Spaces) ensure reliable data sharing and exchange based on agreed principles. This entails numerous challenges:

- **Business/Organizational Challenges:** Companies must build trust, adhere to EU rules, and keep pace with market changes.
- **Legal Compliance Challenges:** GDPR and other legal rules protect privacy and determine data ownership, aligning with EU-wide policies.
- **Technical Challenges:** Interoperability, provenance, quality assurance, and scalability issues in data sharing solutions require attention.

Considering all these, FAME is such kind of a Data Space, providing a Technology Framework, its own Data Marketplace, customized for buying and selling federated data assets in the financial sector. In addition to this Data Marketplace, FAME also offers a Federative Business/Governance Framework to ensure sovereignty and the sharing and exchange of trusted data assets resources within the digital ecosystems, all grounded on commonly agreed principles.

Overall, the FAME project is expected to contribute to the European Union's data strategy in several ways:

- **Market Focus:** FAME is tailored for the financial sector, offering a platform for using customized data assets, fostering innovation, and improving service delivery.
- **Product Diversity:** FAME offers a variety of data assets, including datasets, AI models, analytics, algorithms, services, and educational content.
- **Governance Model:** FAME uses a federated governance model to ensure data transactions are trustworthy, private, and secure, following EU regulations.
- **Cutting-edge Technologies:** FAME employs advanced technologies like semantic interoperability, AI, ML, blockchain, and adaptive authentication to enhance data analytics, ensure secure transactions, and manage dynamic identities.
- **Regulatory Compliance:** FAME provides tools to ensure compliance with laws like GDPR, Data Act, Payment Service Regulation (PSR), Markets in Crypto-Assets (MiCA) Regulation and 4th AntiMoney Laundering directive (4AML). These tools help to ensure that FAME's functionalities meet both security and regulatory requirements, following EU regulations and security-by-design principles.

## 5 FAME Stakeholders

As indicated in the previous 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 (and their Data Marketplaces representing their core technical implementations), 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

mentioned categories, for whom a plethora of opportunities can be identified to have a broader socioeconomic value. The overall involved stakeholders are depicted in Figure 5.

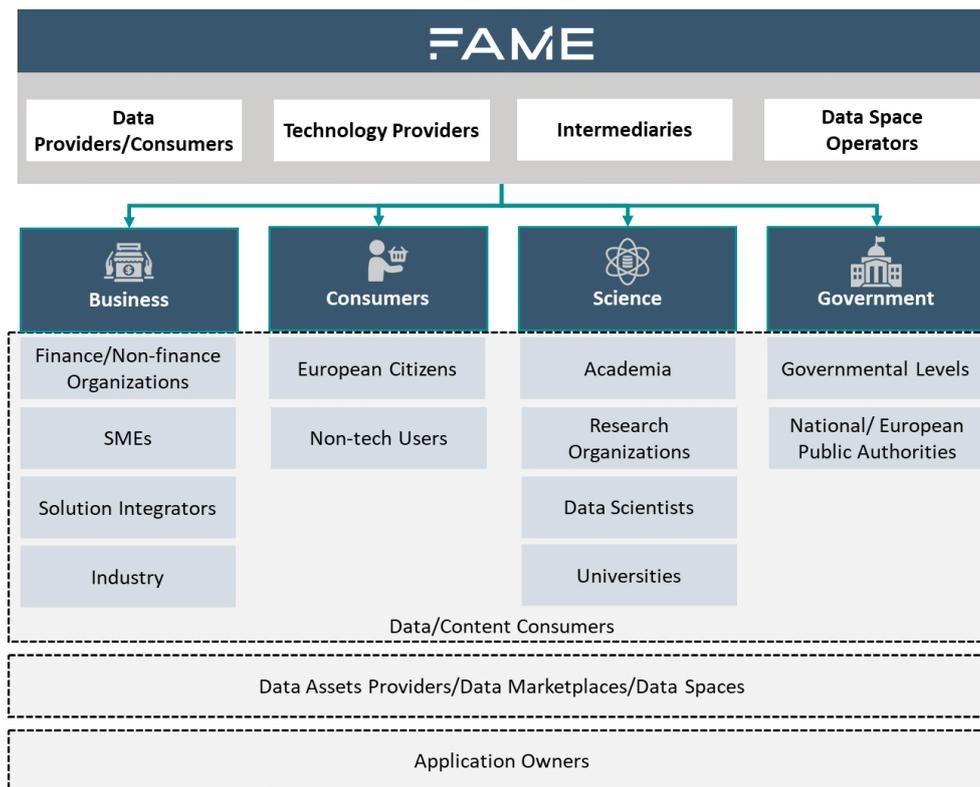


Figure 5 – FAME stakeholders

Concerning the opportunities that have been identified for the roles of Figure 5, 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 social impacts of research data:** It will become simpler to find, integrate, or collaboratively process and analyse data in diverse scientific disciplines as standards for data/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 analysing 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 analysing 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.

## 6 FAME Technical Background

The current Section provides all the necessary information regarding existing RAs, principles, and functionalities that are exploited in the context of the FAME SA. In more detail, RAs related with Big Data Management, 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 development of FAME. The relevance of these RAs and their added-value to FAME SA is furtherly described, concluding to the FAME related technical advancements.

### 6.1 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 specific 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 RA should be technology and domain-independent, abstract, and flexible, being defined at various levels of detail, from high level principles to detailed implementation guides [20][21].

Summarizing, a RA can serve as a *template 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 [22][23].

Nowadays, there exists a variety of different types of RAs, deriving from diverse application domains [24], 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 8).

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.

## 6.2 Overview of Reference Architectures for Data Marketplaces

### 6.2.1 i3-Market

The i3-MARKET project [25] 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*” [26] that addresses the growing demand for connecting multiple Data Spaces and Data Marketplaces in a secure and federated manner.

#### 6.2.1.1 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 [27], 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 6) 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).

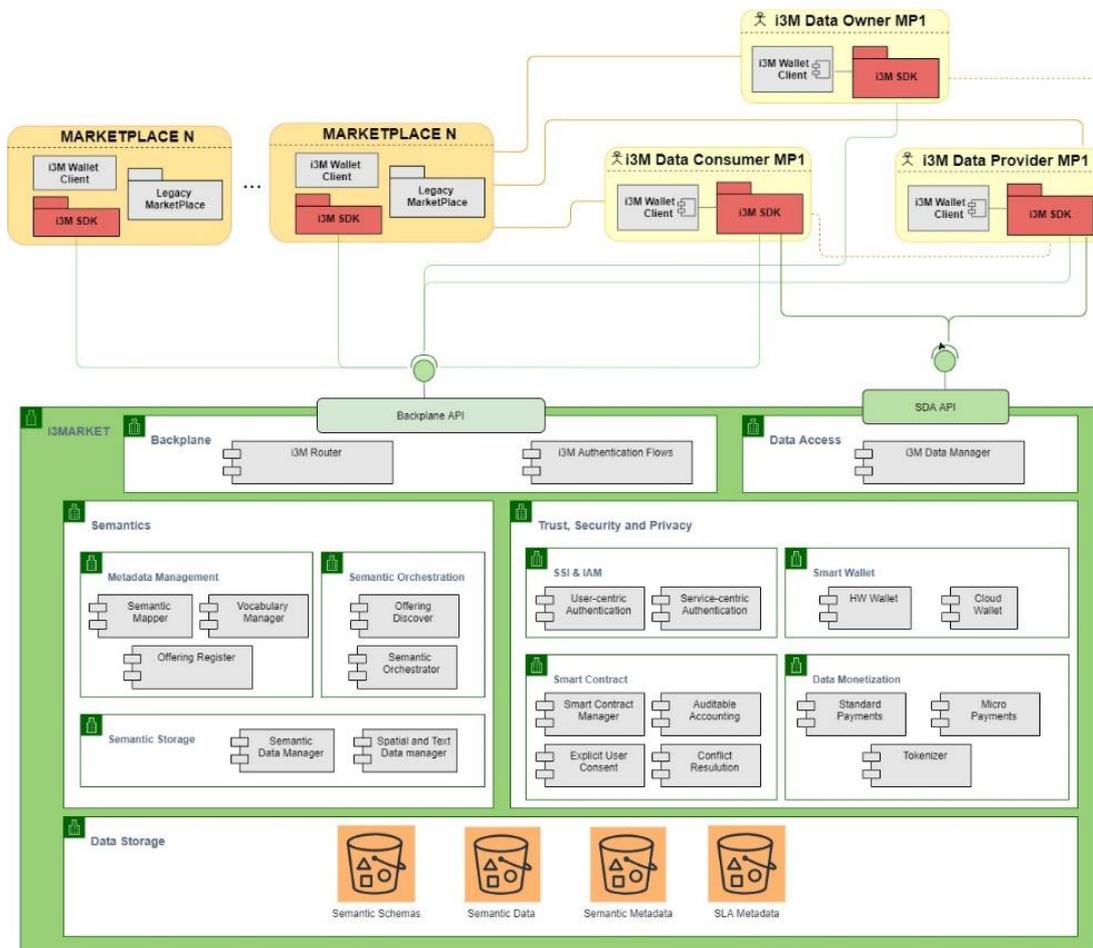


Figure 6 – Logical View with i3-MARKET

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 6), 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 6), 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 are 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 [28] to the i3-MARKET system.

Figure 7 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.

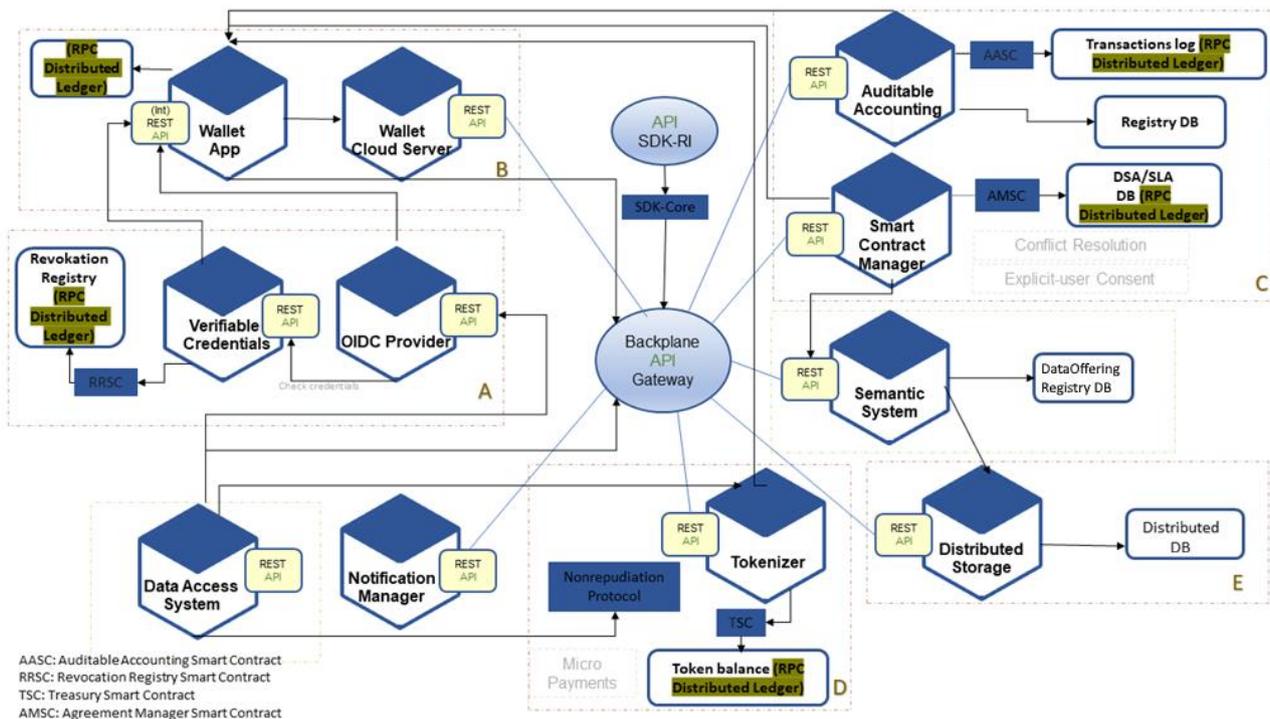


Figure 7 – i3-MARKET microservices layout

As Figure 7 shows, there exist the following dependencies among the i3-MARKET components:

- *SDK System*
  - *SDK-core* libraries for making easier the development of applications that make use of the Backplane API. It interfaces with the Backplane Gateway.
  - *SDK-RI* common pilots-driven complex workflows based on the Backplane services. It interfaces with the SDK-core library.
- *Trust, Security and Privacy System*
  - *SSI and IAM Subsystem (label A)*
    - *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.
    - *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.

### 6.2.1.2 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 leverages 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 provides 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 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 aids 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 are 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 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.

## **6.2.2 INFINITECH**

The INFINITECH project [29] 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.

### **6.2.2.1 Architecture Overview**

The overall high-level view of the INFINITECH RA is illustrated in Figure 8, providing a generic perspective in order to leave great level of flexibility depending on the implementation.

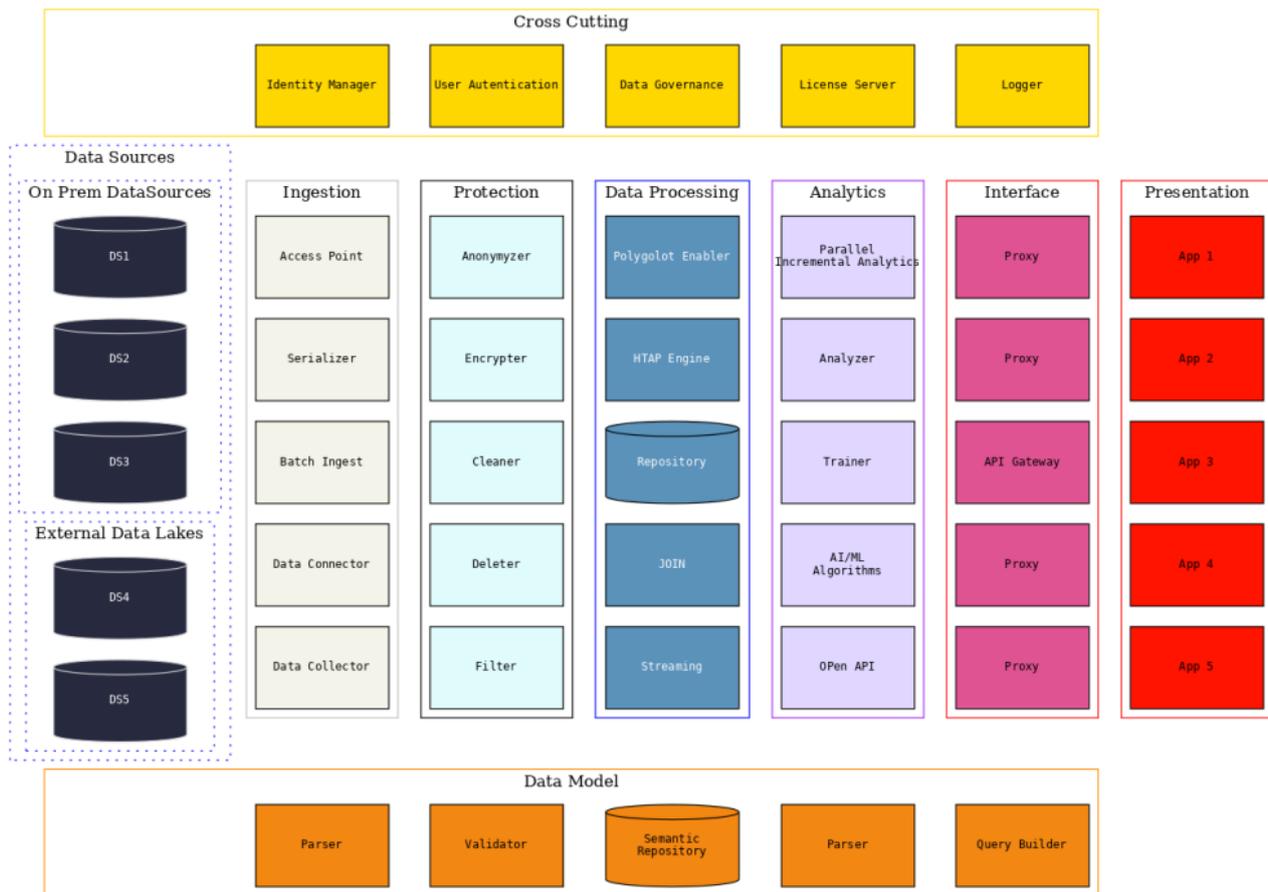


Figure 8 – High level view of Infinitech architecture

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.

### 6.2.2.2 *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 extends 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 is:

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

FAME specifies 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) [30], the Financial Instrument Global Identifier (FIGI) [31], and the Legal Knowledge Interchange Format (LKIF) [32] that are considered in the context of FAME and its semantic interoperability functionalities.

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

FAME learns from the experience and leverages a readily available high performance and energy efficient blockchain infrastructure that supports tokenization, which has been developed in INFINITECH project. FAME reuses and extends the readily available permissioned blockchain infrastructure of the INFINITECH project, which provides support for ERC-20, ERC-721 and ERC-1155 tokens [33]. Specifically, the INFINITECH blockchain has extended the Hypeledger Fabric blockchain [34] 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 is integrated in a single platform. The integration is driven by FAME's architecture. FAME also leverages 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.

### 6.2.3 FINSEC

The FINSEC project developed the Finsecurity.eu platform [35], 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 [36] and linked to the knowledge hub of the Horizon Europe Coordination and Support Action (EU-CIP CSA) [37]. 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.

#### 6.2.3.1 *Architecture Overview*

The technical architecture of the platform is illustrated in Figure 9. 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.

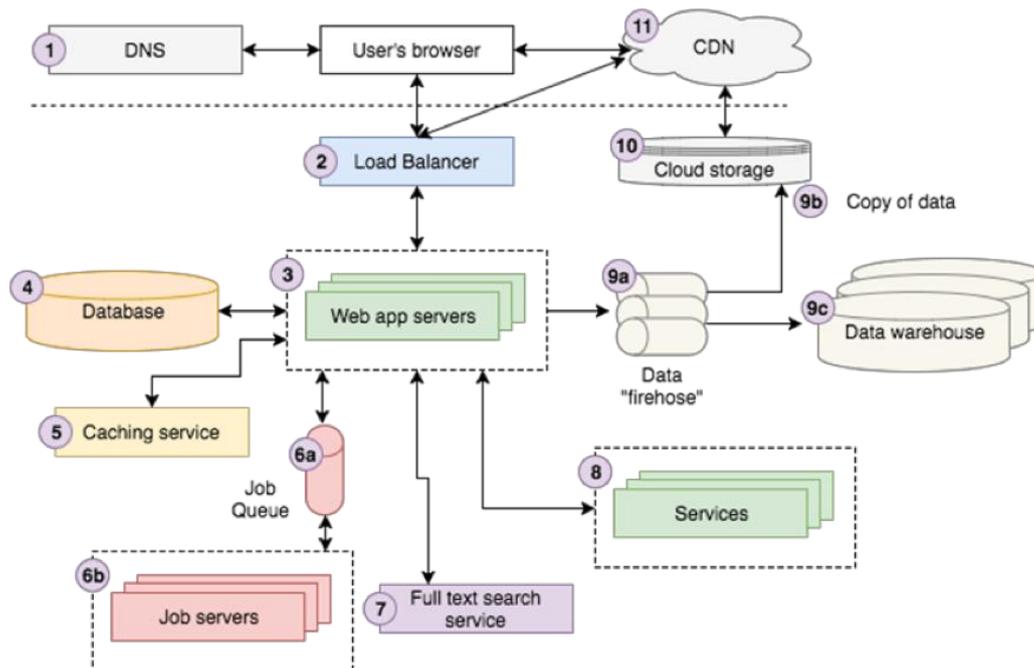


Figure 9 – Technical architecture of Finsecurity.eu platform

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 [38], 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.

#### 6.2.3.2 *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 provides access to content items and data assets about CIP and cybersecurity, fostering the engagement of the CIP community in the FAME Federated Data Space, and subsequently increases both the exploitation of the overall FAME functionalities and the demand for the FAME assets. The Finsecurity.eu data assets are indexed, and the resulting metadata are integrated into the FDAC.

**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 are also adopted within FAME. Nevertheless, it is also acknowledged that other existing Data Marketplaces may offer access to similar types of data assets and content items. The Finsecurity.eu platform and owners are federated in around the FAME marketplace, following the decentralised governance model that has been specified in the context of WP4.

**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 provides FAME with opportunities to demonstrate how the FAME Federated Data Space can interface and communicate with external platforms. The FAME AAI interface demonstrates how FAME stakeholders can easily access the contents of Finsecurity.eu without needing to deal with the above-listed low-level architecture details. Finsecurity.eu is integrated based on the AAI of the FAME platform. Note however that the platform provides a very simple layer of security and authentication without sophisticated roles and permission, and as a result Finsecurity.eu provide a rather simple demonstration of the AAI.

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

The Finsecurity.eu platform provides 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 is accomplished through the FAME assets policies manager that is able to shape such assets-related policies. It is however acknowledged that the public access to the FAME data assets do not offer opportunities for implementing complex security schemes (e.g., location-based security, content-based security) over the Finsecurity.eu integration.

## 6.2.4 SecureIoT

The main goal of the SecureIoT project [39] 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.

### 6.2.4.1 Architecture Overview

Figure 10 illustrates the conceptual architecture of the SecureIoT Marketplace, outlining its supported functionalities in accordance with the external users that are able to exploit them.

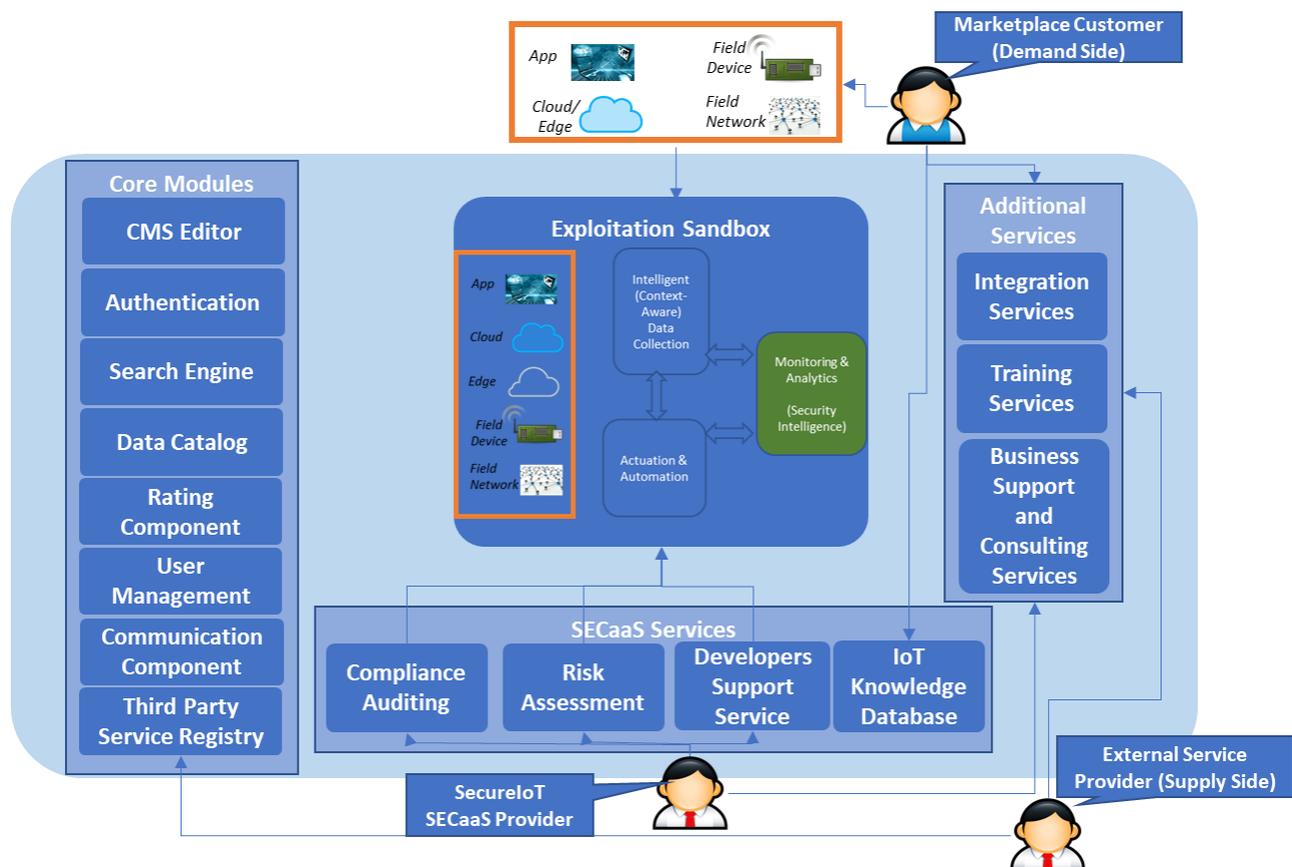


Figure 10 – SecureIoT marketplace conceptual architecture

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 3<sup>rd</sup> Party Authentication Services*: Refers to the support for authentication with 3<sup>rd</sup> 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 - Industry 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 3<sup>rd</sup> party integration.

#### 6.2.4.2 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 provide its design and implementation principles towards specifying the characteristics and the supported capabilities of the FDAC.

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

SecureIoT offers the technical knowhow with respect to its additional services that target on integration and proper skillset training, to build the FAME Learning Centre that aims 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 provide proper guidelines and best practices for constructing privacy data assets' policies based on existing regulations (e.g., GDPR), towards the 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, acts as a backbone for building the decentralized pricing and trading schemes of the data assets within FAME.

## 6.2.5 PolicyCLOUD

The PolicyCLOUD project [40] 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.

### 6.2.5.1 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 11) 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 11, 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.

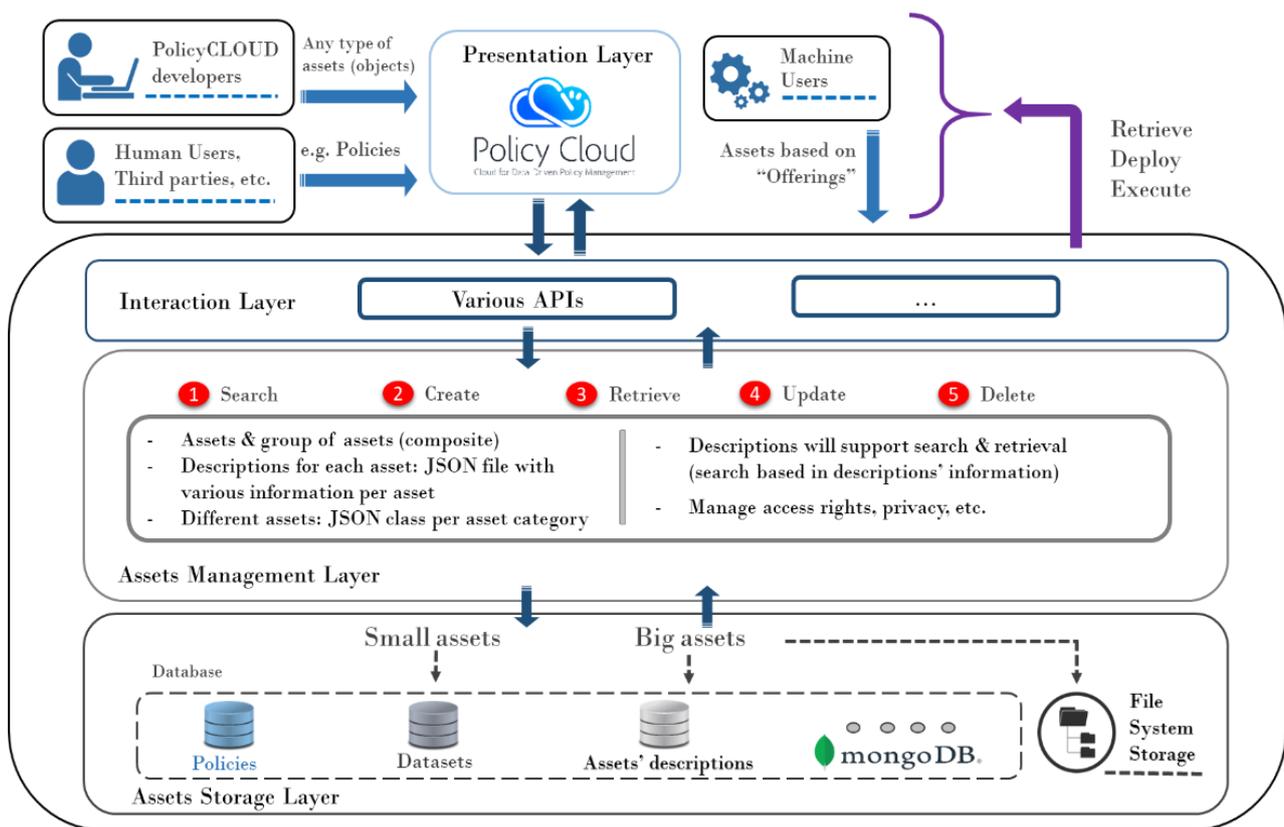


Figure 11 – PolicyCLOUD data marketplace conceptual architecture

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.

### 6.2.5.2 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 can 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 is developed and implemented supporting similar functionalities with the PolicyCLOUD Data Marketplace enabling its end-users both to search and retrieve the indexed assets of the platform, also being able to provide their own assets (with relevant access policies). Such core functionalities are offered through the FAME Dashboard and its integrated backend services, that exploit the FDAC that acts 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 supports diverse types of data assets, including tools, AI/ML models, datasets, as well as training material (e.g., webinars, tutorials, and documents), towards the facilitation of data exchange, repurposing, and reuse of a variety of data assets. On top, FAME also adopts 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 provides 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<sub>2</sub> and wastes, location specific characteristics (e.g., locality of data capture), timeliness, demand (e.g., number of searches and more), algorithmic metadata, and more.

### 6.2.6 DataVaults

The DataVaults Cloud Platform [41] 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.

#### 6.2.6.1 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 12.

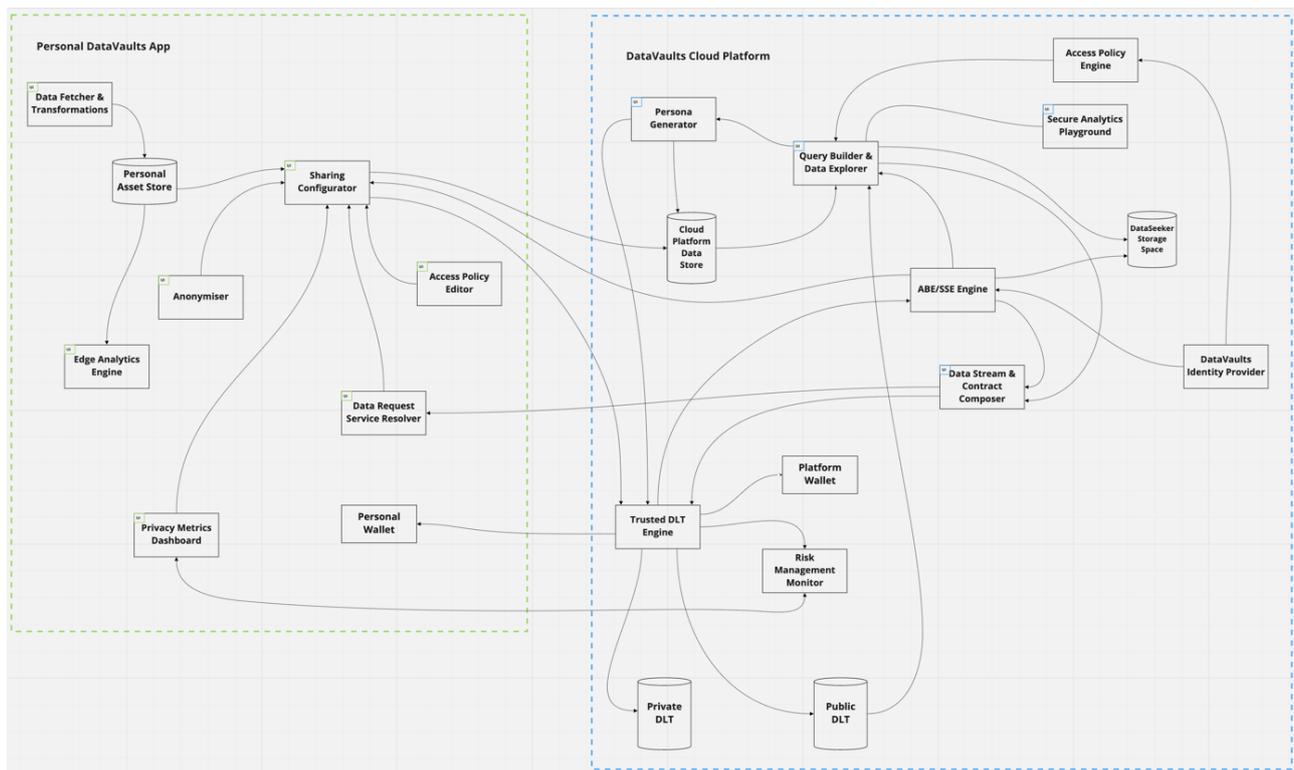


Figure 12 – DataVaults architecture overview

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.

#### 6.2.6.2 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 is used into FAME, to boost up the searching ability of the search engine with regards to the FDAC.

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 offers 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.

#### 6.2.7 OMEGA-X

Relying on European common standards, the OMEGA-X project [42] aims to implement and deliver an Energy Data Space. This will include a federated infrastructure, a data marketplace and service marketplace, involving data sharing between different stakeholders and demonstrating its value for concrete energy use cases while guaranteeing scalability and interoperability with other Data Space initiatives. OMEGA-X has set as a main goal to develop an Energy Data Space that enables multiple actors sharing data and services while ensuring privacy, security, and sovereignty. This specifically addresses the current problem of low availability of data for innovative uses in the energy sector and beyond. Towards this direction, OMEGA-X collaborates with stakeholders to identify where energy-based service improvements and innovation are required, and how OMEGA-X could potentially be used and adopted to address the following needs:

- Guarantee companies and organizations that they can share their data safely. At the same time, it helps existing market actors (including SMEs and start-ups) to have access to a variety of datasets to improve their AI models, and thus be able to upgrade existing services and/or bring innovative services that otherwise could not be developed.
- Provide high availability of data to empower new participants and market roles such as aggregators and local energy community managers. This facilitates the large-scale penetration of renewables in the local grid without significant investments in grid infrastructure and also creates an opportunity for new business models to emerge.
- Put a prominent focus on developing and promoting inclusive and collaborative behaviours, which lead to a multitude of societal and economic benefits, such as, an increase in energy autonomy and a reduction in CO<sub>2</sub> emissions.

### 6.2.7.1 Architecture Overview

Figure 13 illustrates the conceptual architecture of the OMEGA-X marketplace, focusing especially on presenting the building blocks of the Data & App Marketplace and the Data Exchange Services, whereas to facilitate convergence with other initiatives within the Data Space ecosystem, the architectural design is heavily influenced by the Blueprint of Data Spaces Support Center (DSSC).

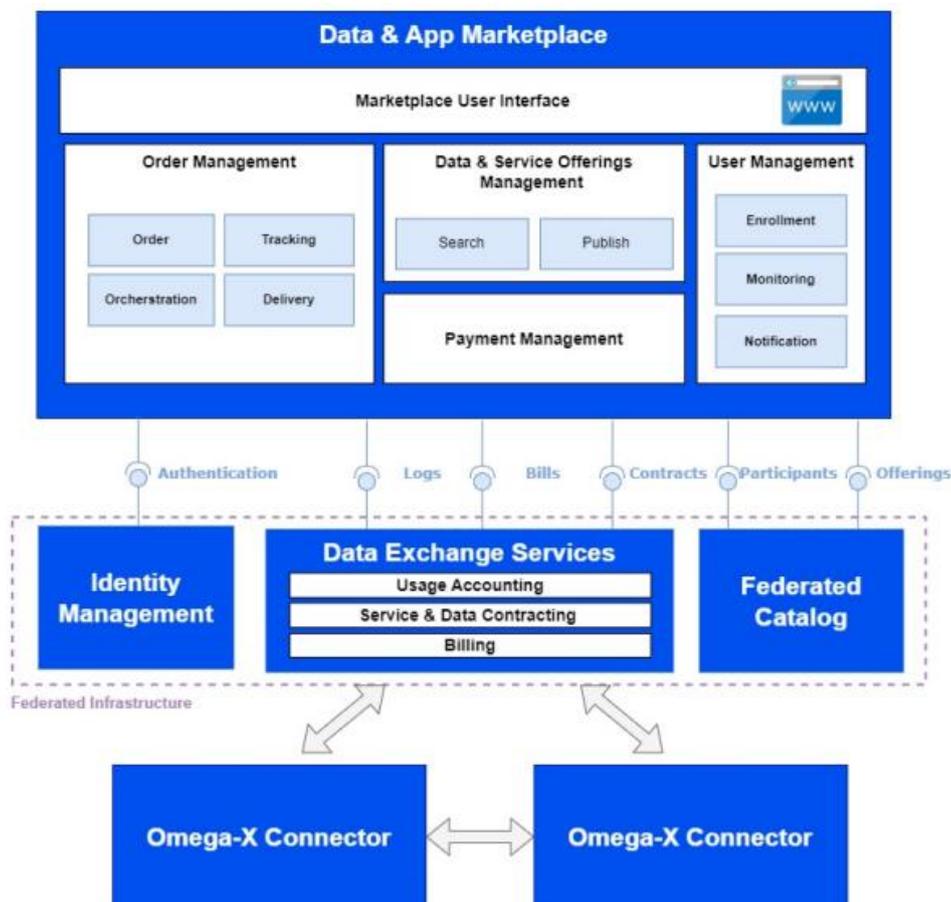


Figure 13 – OMEGA-X Marketplace architecture

In deeper detail, the Data & App Marketplace is comprised of the following elements:

- *Marketplace User Interface (UI)*: A Graphical enabling the intuitive communication of the user and the OMEGA-X Data Space. It allows users to search available offerings (Service and Data) in the OMEGA-X Data Space, create and manage offerings/contracts, and manage financial-related activities in the Marketplace. It also provides a channel for users to share their feedback.

- *Order Management*: This element is in charge of the orchestration of an offering (either Service or Data) order, enabling the tracking of the different steps until delivery.
- *Data & Service Offerings Management*: This element offers users the ability to add, edit and delete Data and Service offerings. The component should provide the ability to set different visibility options for an offering (product) and pricing models.
- *User Management*: This module enables and orchestrates the registering procedure of organisations and the onboarding procedure of new users. It is also responsible for managing all participants of the Data Space as well as their “offboarding”.

In addition, the Data Exchange Services are comprised of the following elements:

- *Service & Data Contracting*: It is responsible for managing the contracting of Data and Service offerings among users and providers. It provides a process to allow the tracking the different steps of finalizing or amending a contract.
- *Usage Accounting*: This module tracks Data transmitted/received with respect to a contract, as well as that the honouring of the usage policies has been respected.
- *Billing and Settlement*: This element is responsible for managing the settlement and billing of transactions taking place in the ecosystem.

#### 6.2.7.2 Relevance to FAME & Added-Value

The integration of the OMEGA-X Marketplace 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**]

With respect to the FAME Dashboard, the backbone of the OMEGA-X Marketplace is considered for designing a user-friendly marketplace UI in addition with its elements for supporting the ability to discover, monetize, and trade the available data assets, reusing the concept and state-of-the-art analysis of the OMEGA-X Marketplace Federated Catalogue. All these, equally provide technical knowledge and baseline technologies for the core functionalities of the main layers of the FAME SA.

Authentication and Authorization Infrastructure (AAI) for Secure & Regulatory compliant integration of data spaces and marketplaces [**Obj 1**]

OMEGA-X Marketplace offers its implementations on providing related authentication and identity management schemes that facilitate the process of specifying and implementing the AAI capabilities of FAME, supporting data assets trading.

#### 6.2.8 TRUSTS

TRUSTS [43] aims to develop a data sharing platform for secure, trustworthy, and GDPR-compliant data exchanges. The TRUSTS platform acts independently and as a platform federator, while investigating the legal and ethical aspects that apply on the entire data valorisation chain, from data providers to consumers, aiming to: (i) set up a fully operational and GDPR-compliant European Data Marketplace for personal related and non-personal related data targeting individual and industrial use by leveraging existing data marketplaces and enriching them with new functionalities and services to scale out, and (ii) demonstrate and realise the potential of the TRUSTS platform in various use cases targeting the industry sectors of corporate business data in the financial and operator industries while ensuring it is supported by a viable, compliant and impactful governance, legal and business model.

### 6.2.8.1 Architecture Overview

As illustrated in Figure 14, the TRUSTS platform consists of a set of interconnected nodes, which can be of three (3) distinct types: Central Node, Corporate Node, and User Portal Node. Each node is owned and operated by different organizations, where all together form the basis for a trusted exchange of digital assets. Each node is a computing environment (e.g., a virtual machine), where each computing environment runs standardized components to support the various functionalities of the platform. Components deployed within a node are interconnected via a network environment provided and secured by the node's operator. Both the nodes and the standardized components running on these nodes are illustrated in Figure 14. By using a standardized set of components in a node, a consistent set of functionalities is provided to the TRUSTS platform participants. The functionalities that the platform should provide are defined by the functional requirements. The standardized set of components can also be used to enforce restrictions that affect, for example, trading and access to assets within nodes. Along the components, common to all nodes, organizations can execute their in-house developed applications and services, which can then be traded through the TRUSTS platform, and can also, in turn, consume and produce other assets through the platform.

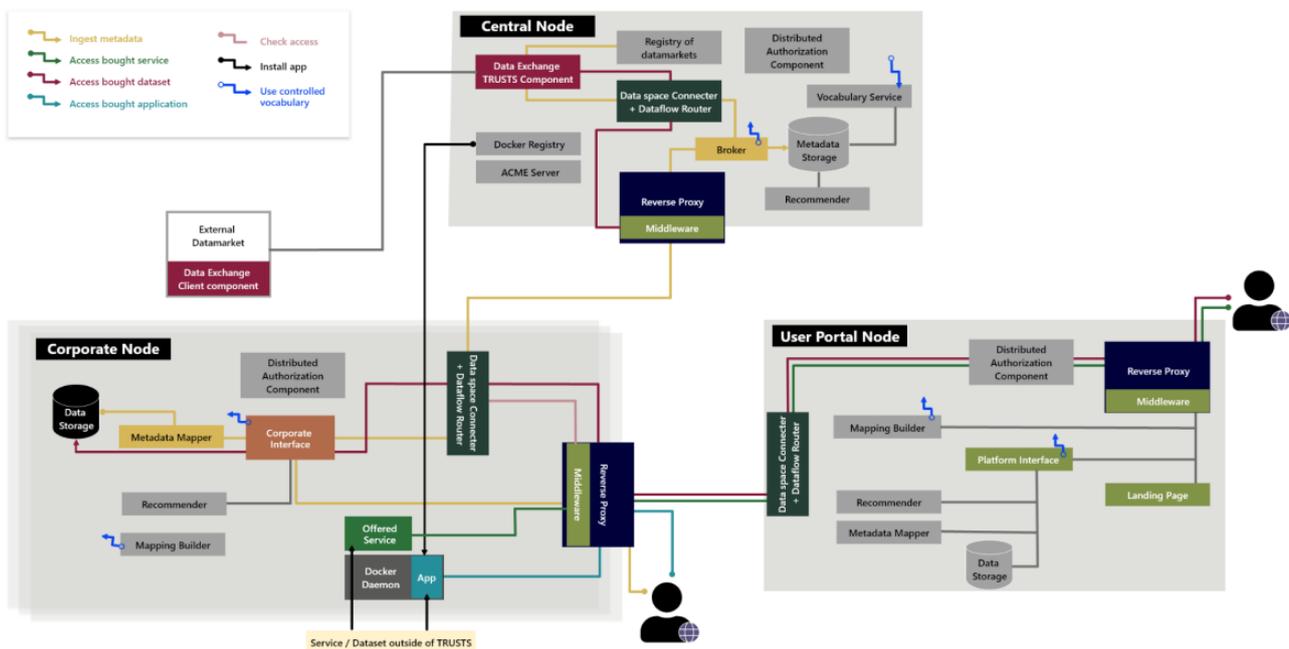


Figure 14 – TRUSTS Platform architecture

In deeper detail, the nodes shown in Figure 14 refer to the User Portal Node, Central Node, and Corporate Node. The two (2) nodes that must be hosted by a specific organization, referred to as the TRUSTS Operator, are the User Portal Node and the Central Node. These nodes host a set of services necessary for the operation of the platform as a whole. Accordingly, the TRUSTS Operator occupies a special position of trust with respect to the other organizations hosting nodes. Any other node is referred to as a Corporate Node.

- Corporate Node:** It is used by partners of TRUSTS that have complex infrastructure and participate in TRUSTS. They can have many users and many services and applications that are provided or consumed via the TRUSTS platform. They can set any authorization system and communication structure inside their nodes. The TRUSTS Operator is not responsible for setting up or supporting instances of corporate nodes but will provide detailed instruction manuals and all necessary software. Among the components being set up, is the Corporate Interface, a web application through which the organization's users can interact with the TRUSTS platform.

- *User Portal Node*: It is created to cover the needs of individual users of the TRUSTS platform. It is set up and maintained by the TRUSTS Operator. By accessing this node, individual users can benefit from some of the functionalities of the TRUSTS platform, without the complexity of setting up a corporate node. Additionally, it is the entry point for all organizations intending to join the TRUSTS platform, as it provides landing pages, legal information, and setup instructions. One of the components included in this node is the platform Interface, which acts as the main point of access to the above-mentioned functionalities. In principle, the TRUSTS platform could have more than one User Portal Node.
- *Central Node*: Exists to support the operation of the whole TRUSTS platform, playing the role of authorization, monitoring, smart contract executor, catalogue, application repository, among others. This node is created and maintained by the TRUSTS Operator

#### 6.2.8.2 Relevance to FAME & Added-Value

TRUSTS is going to provide the below added-value to FAME's general set objectives:

##### Federated Data Marketplace Platform [Obj 1]

As in the case of the TRUSTS, FAME supports a plethora of users that can authenticate themselves to be able to interact with the FAME platform, supporting its federation concept. To successfully support such functionalities, within its User Portal Node, TRUSTS offers a single-entry point platform through which its users can interact with, facilitating and supporting all their different requirements and needs. This interaction is efficient enough through related landing pages, legal information, and setup instructions, that has been considered as the core backbone of the design of the FAME Data Marketplace, and its required UIs residing within the frontend Dashboard.

### 6.3 Overview of Reference Architectures for Data Spaces

#### 6.3.1 GAIA-X

Gaia-X [44] 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 [45] 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. Hence, the primary goal of GXFS is to support the development of federated digital ecosystems. Through the latter, various participants can connect with each other, having the ability to develop new innovative products and services, optimize existing processes and exploit previously untapped potential through data. Such digital ecosystems consist of interconnected data and infrastructure ecosystems that are grouped together as federations and individually orchestrated and operated by federated services.

##### 6.3.1.1 GXFS & XFSC Elements

To be more specific, GXFS consists of several components (Figure 15) 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.

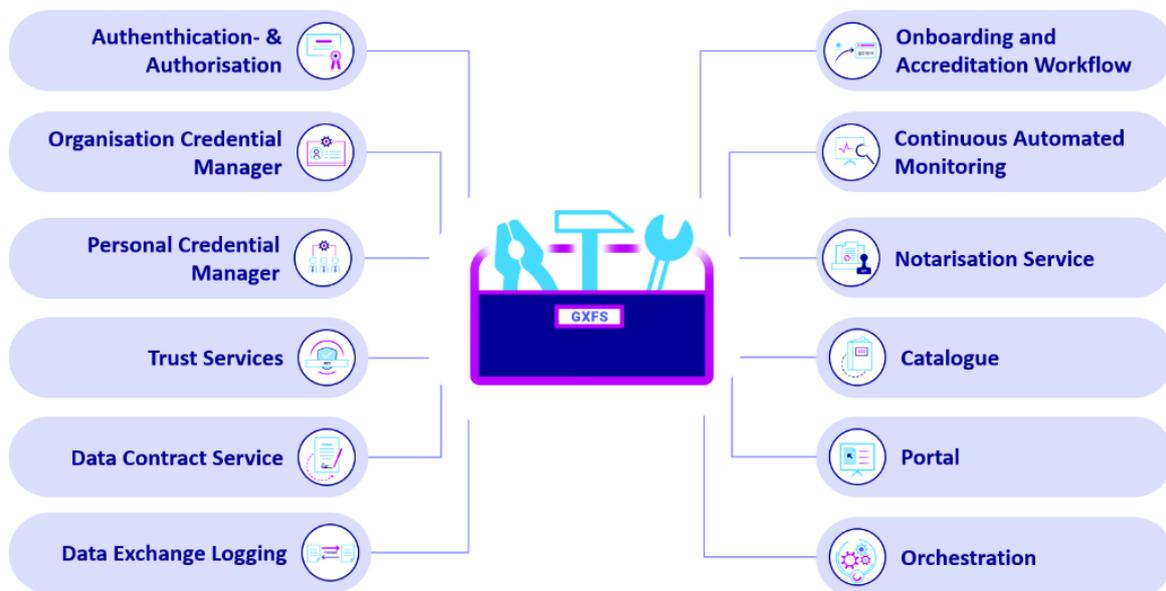


Figure 15 – GXFS supported elements

- *Identity and Trust Elements*
  - *Authentication & Authorization* [46]: 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* [47]: 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* [48]: 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 Service's API* [49]: 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* [50]: 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* [51]: 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* [52]: 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* [53]: 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* [54]: 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* [55]: 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* [56]: 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* [57]: 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.

To better support the development of the federated ecosystems described above (representing Phase 1 of GXFS), the so-called Cross Federation Service Components (XFSC) toolbox was developed as part of GXFS (representing Phase 2). The XFSC toolbox comprises the abovementioned GXFS federation services (Figure 15), offering free and open-source code for all the interested parties. The XFSC specification process has been built upon the groundwork laid in GXFS specification Phase 1 [58] and aligns with the principles of the Gaia-X Trust Framework 22.10 [59]. Further details on the XFSC specification process can be located in the official documentation in [60].

### 6.3.1.2 Relevance to FAME & Added-Value

All the GXFS/XFSC services have 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/XFSC 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 has analyzed and examined 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. Based on this study, FAME has identified and prioritized all the related GXFS/XFSC services that enable federations and provide interoperability for building federated data ecosystems, and has provided the basic blueprints and guidelines towards contributing and enhancing these services with its added value gained from the implemented backend services. As stated above, most of the FAME backend services are based upon existing technical specifications/implementations of the previously identified RAs, that can collectively contribute into an upcoming GXFS specification phase, with knowhow and lessons learned from a wide variety of previous/ongoing EU funded projects.

### 6.3.2 IDS

The International Data Spaces Association (IDSA) [61] 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.

#### 6.3.2.1 Architecture Overview

IDS-RAM was originally defined as part of the research activities conducted in the Industrial Data Space project by Fraunhofer [62] 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 IDS.

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 16 illustrates the general structure of the IDS-RAM that uses five (5) layers and three (3) perspectives.

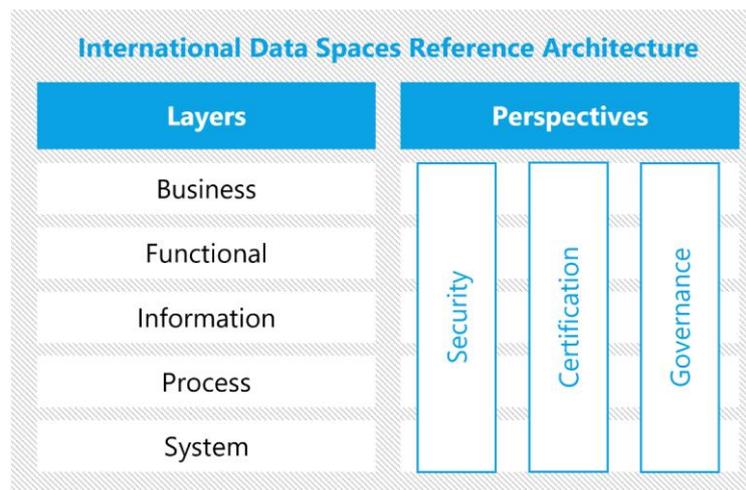


Figure 16 – General structure of IDS Reference Architecture Model

- *Business Layer* [63]: 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* [64]: 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* [65]: 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* [66]: 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* [67]: Is concerned with the decomposition of the logical software components, considering aspects such as integration, configuration, deployment, and extensibility of these components.

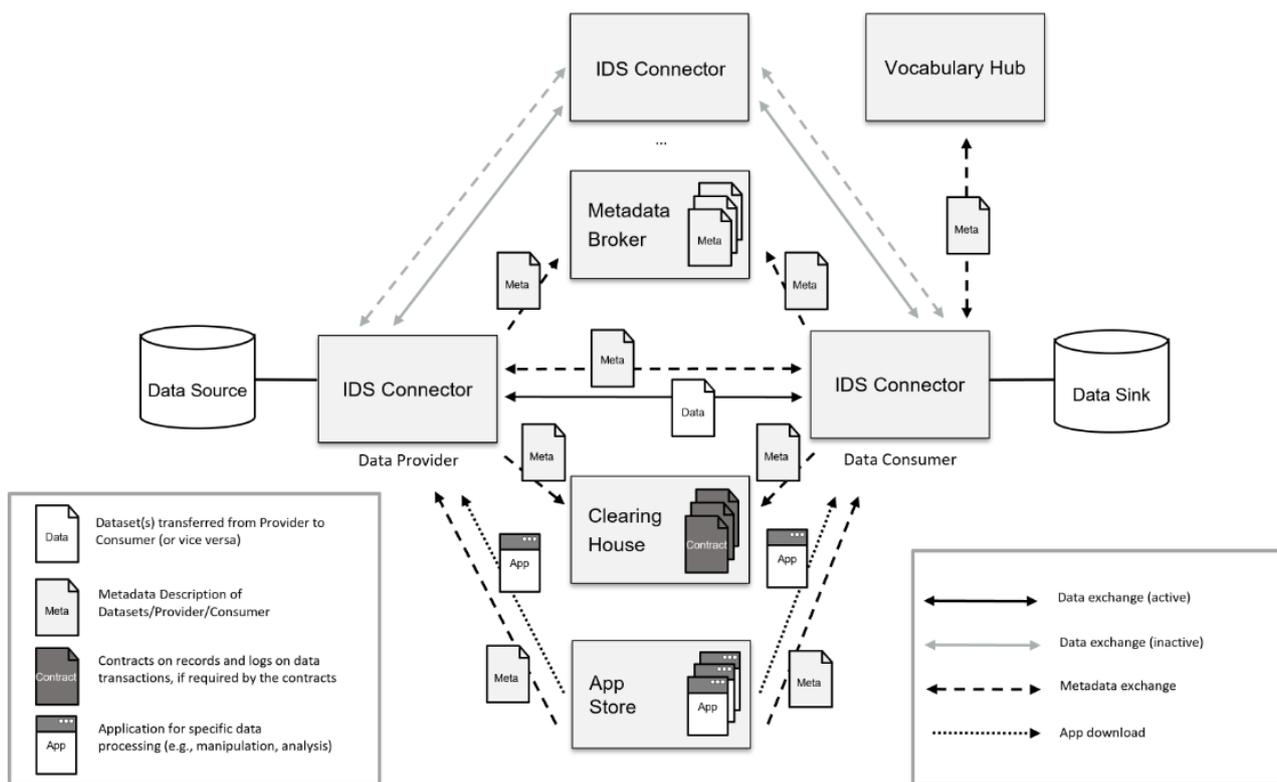


Figure 17 – Interaction of technical components

Figure 17 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** [68]: 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** [69]: 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** [70]: The Governance Perspective defines the roles, functions, and processes of the International Data Spaces from a governance and compliance point of view.

### 6.3.2.2 Relevance to FAME & Added-Value

Within D2.2, the IDS-RAM and its Rulebook were deeply investigated to be adopted within the FAME SA, bringing value to FAME in the two (2) diverse ways of: (i) architectural elements, and (ii) building blocks for technical implementation. Nevertheless, there have been many advancements within the IDSA and the larger Data Spaces' ecosystem since the writing of the first version of this deliverable (D2.2), which the FAME project team has been following closely. Such advancements refer to the updated version of the *IDS-RAM* (i.e., *IDS RAM 5*), the *Dataspace Protocol*, and the *Data Spaces Support Center (DSSC)*.

### 6.3.2.2.1 IDS RAM 5

IDSAs have started working on a new version of the IDS-RAM [71] in collaboration with its member organizations. This new version (IDS RAM 5) will be aligning with the latest developments in IDSA and Data Spaces, such as the DSSC, the new Rulebook and the Dataspace Protocol. It also aims to be more modular and flexible, providing architectural guidance and support for all roles building a Data Space, participating in one, or providing added-value Data Space services.

#### 6.3.2.2.2 Dataspace Protocol

Interoperability is a big challenge in any given context that is concerned with data sharing. The Dataspace Protocol [72] is a set of specifications aiming to address technical interoperability in Data Spaces. It enables interoperable data sharing between entities governed by usage control and based on web technologies. These specifications define the schemas and protocols required for entities to publish data, negotiate agreements, and access data as part of a federation of data spaces.

The Dataspace Protocol is a specification on its way to becoming an international standard under the name “ISO/IEC AWI 20151: Cloud computing and distributed platforms, Dataspace concepts and characteristics”, which has been officially registered within the TC/SC work program at ISO/IEC JTC1 SC38.

#### 6.3.2.2.3 Data Spaces Support Center (DSSC)

DSSC [73] is an initiative of the EC to support the development of Data Spaces in many sectors. It explores the needs of Data Space initiatives, defines common requirements, and establishes best practices to accelerate the formation of sovereign Data Spaces as a crucial element of digital transformation in all areas.

The DSSC Blueprint is a consistent and comprehensive set of guidelines to support the development cycle of Data Spaces. It includes the *Data Space conceptual model*, the *Data Space building blocks*, and the *Recommended standards and specifications* [74].

As for the *DSSC building blocks* [75], they help to break down Data Spaces into manageable smaller pieces. These are basic units or components that can be implemented and combined with other building blocks to achieve the functionality of a Data Space.

- The essential DSSC building blocks are represented by the *business and organizational components*, which are structured around the three (3) main pillars of Business, Governance, and Legal. These pillars determine the functionalities that are accessible to end-users, guiding developers in crafting robust data environments and allowing end-users to effectively utilize the available services. Each of the eight (8) organizational and business components offers a distinct capability that is not replicated by the others (Figure 18).
- To deploy Data Spaces, specific technological capabilities are required. The purpose of the *technical components* is to pinpoint these necessary capabilities and establish universal standards for each. This approach aids data governance bodies and end-users in making well-informed technical decisions. Additionally, widespread open standards and specifications are established for most components, facilitating the reuse of technical solutions and enhancing technical interoperability. The technical components correspond to the required capabilities within a data environment, where each capability involves specific design choices. To this context, a technical building block is not directly equivalent to a single software component or role within the data environment, but instead, multiple building blocks may interact with various software components. The technical capabilities (Figure 18) are structured according to the three (3) pillars of: (i) *Data interoperability* - The capabilities needed for data exchange referring to semantic models, data formats and interfaces, as well as functionalities for

provenance and traceability, (ii) *Data sovereignty and trust* - The capabilities needed for the identification of participants and assets in a Data Space, the establishment of trust, and the possibility to define and enforce policies for access and usage control, and (iii) *Data value creation* - The capabilities used to enable value creation in a Data Space (e.g., by registering and discovering data offerings or services and provide added-value services).

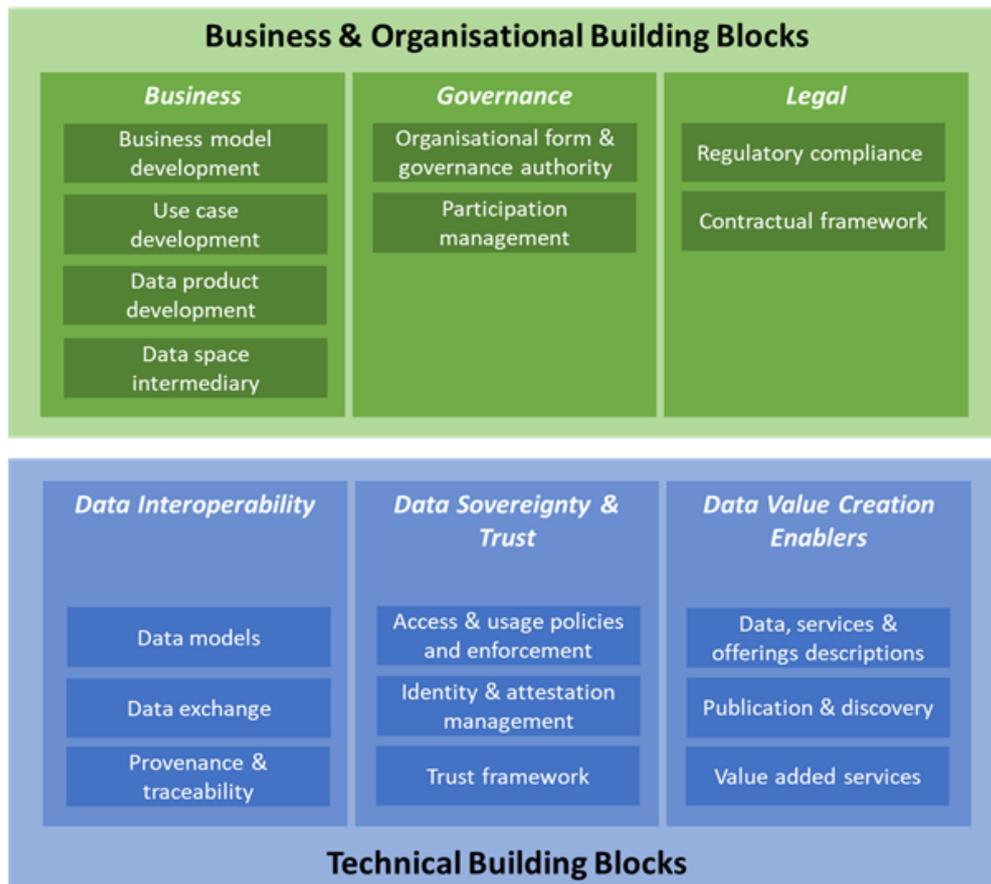


Figure 18 – DSSC building blocks

By exploiting all the aforementioned guidelines and principles, the IDS RAM 5, the Dataspace Protocol, and the DSSC aim at contributing into the achievement of the following FAME objective:

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

As a key starting point for successfully being aligned with the existing IDS guidelines and principles, the final version of the FAME SA has been specified considering the IDS RAM 5 for providing federated and secure data exchange and data sharing. On top, since the Dataspace Protocol is a key enabler towards achieving interoperability across Data Spaces, its adoption is also pivotal within FAME. Since it has only been released recently, FAME has not yet had the chance to align with this specification. However, by the final release of the overall platform, such protocol will have been fully investigated and adopted with the FAME technical implementation to fully support the Data Spaces notion in the EmFi application domain. Finally, as for the DSSC Blueprint, this was also still underworked when the FAME SA was finalized. Nevertheless, it still had an impact on the overall design for realizing a mature Federated Data Space, with a goal of being aligned with the existing status of all the DSSC building blocks. By mapping its SA to the DSSC guidelines, FAME will ensure that it will deliver a cohesive and efficient ecosystem that is both accessible and secure for all the involved stakeholders, ranging from financial institutions to individual data users. A detailed mapping of the FAME SA components and the relevant DSSC building blocks is provided in Section 8.2.3.

## 7 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 offers novel decentralized programmable pricing and trading of data, among others. Hence, the FAME's capabilities are 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 tools for applying trusted and energy-efficient analytics. All the underlying capabilities are presented in this Section (also summarized in Table 3), while Section 8 provides details for the mechanisms/components that realize these capabilities.

Table 3 – FAME Overall Capabilities

#	Capability	Short Description
1	 <b>Authorization &amp; Authentication</b>	FAME allows to authenticate the interested stakeholders and authorize them to access the respective FAME's Data Marketplace functionalities and the existing content.
2	 <b>Federation of External Sources</b>	FAME supports 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	 <b>Assets Policy Management</b>	FAME ensures the secure management of data assets, regulating who can view and purchase the assets in the platform, also providing stakeholders with a list of assets they own accompanied by their surrounding regulations.
4	 <b>Assets Provenance &amp; Tracing</b>	FAME ensures 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	 <b>Assets Pricing</b>	FAME provides a tool to calculate the cost of an asset to be published, considering both static (based on the nature of the assets) and dynamic (based on demand) variables.
6	 <b>Assets Trading &amp; Monetization</b>	FAME grants to its stakeholders the ability to navigate the FDAC and identify the data assets that are aligned with their interests, supporting the trading of access rights for a chosen asset.
7	 <b>Assets Searching</b>	FAME allows the stakeholders to perform data queries upon the existing data assets, based on their provided semantic models.

8		<b>ML &amp; AI Analytics</b>	FAME allows the stakeholders to locate in the FDAC, select, and execute an ML/AI model that is suitable for their needs, being also able to improve the overall energy efficiency of the produced model by applying incremental analytics and energy efficient capabilities.
9		<b>SAX &amp; XAI Analytics</b>	FAME supports 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		<b>Analytics CO<sub>2</sub> Monitoring</b>	FAME estimates the CPU usage of the ML/AI models, which then are translated into the energy consumed by each model so as to be optimized in terms of consumption.
11		<b>Smart Deployment</b>	FAME provides the pipelines for the deployment of various services, enabling stakeholders to implement optimised workflows in terms of CO <sub>2</sub> consumption.
12		<b>FML Deployment</b>	FAME supports energy efficient data sharing in federated learning scenarios where various stakeholders contribute their data assets.
13		<b>EmFi Training</b>	FAME provides a pool of training assets' resources (focusing especially in EmFi applications), enabling the stakeholders to search, access, as well as comment and rate all the available resources, whereas integrating new training content.
14		<b>Dashboard</b>	FAME provides an integrated user-friendly end-to-end UI facilitating the interaction of the FAME stakeholders with the involved FAME services, components, and processes.

## 7.1 Operational Governance

By providing an extensive range of administration and governance functions necessary for the effective operation of the federation, the Operational Governance (GOV in brief) module considerably increases the FAME federation's overall functionality. End-users can benefit significantly from these additional capabilities, which enhances their experience within the federation.

The GOV module, first and foremost, facilitates the accession of new members and the withdrawal of existing members from the federation, streamlining the procedures for both. This guarantees simple and trouble-free maintenance of memberships. Without complicated processes, new members can join right away, and departing members can do so without facing any red tape. The federation's ability to smoothly transition members ensures that it will always be dynamic and flexible, able to meet the changing needs of its members.

The module also has a strong support system that improves user experience by quickly and effectively handling user questions and help tickets. The prompt resolution of customer complaints by this committed support team enhances user pleasure and builds platform confidence. Quick support is available to help prevent problems before they get worse, which builds user confidence and dependability.

Moreover, the GOV module enhances thorough and safe federation transactions. The capacity to handle transaction accounts, interact with other trading parties in an efficient manner, and obtain informative and historical data is advantageous to end-users. This improves their trading ability and all-around strategic planning by providing them with the tools they need to make wise selections. Additionally, the safe transaction features guarantee that all financial activities are protected, instilling confidence in the safety of their operations.

Overall, the FAME federation is more effective, safe, and intuitive thanks to the GOV module. It guarantees solid commercial operations, effective user administration, simple integration processes, and adaptable support services. The GOV module's provision of these all-inclusive management and governance functionalities enhances the end-user experience, while streamlining administrative processes, so rendering the FAME federation a more appealing and efficient milieu for all its constituents.

## 7.2 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 is the process of FAME for verifying the identity of an external user or system. Its primary function is to ensure that the entity requesting access is who or what it professes to be. Such functionality is of high importance for the FAME stakeholders 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 is able to deliver a personalized user experience, including preferences, settings, and other personalized data.

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

Hence, through authentication, the FAME stakeholders are assured that their accounts and data are protected from unauthorized access, whilst through authorization, they always 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 provides 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 allows the creation of credentials that can be issued and verified without the need of a central certification authority and can be owned by the end-users and directly shared with third parties without involving the credential issuer.

### 7.3 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) provides such functionality through a developed mechanism to describe and execute asset metadata importers - called Resolvers. The latter are 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 refers to 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 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 IDS 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.

### 7.4 Assets Policy Management

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

Firstly, the Assets Policy Management enables the complete lifecycle management of policies associated with the federated data assets that are discoverable through FAME, determining who is eligible to view and potentially acquire specific assets within FAME. Leveraging the Attribute Rule-Based Access Control (RuBAC) model, the component considers various user and organizational attributes to combine them in boolean expression formatted rules to make informed policy decisions. By fulfilling its role as a Policy Decision Point (PDP), the Assets Policy Management ensures that other components of the platform always display the appropriate data assets to authenticated and authorized individuals or organizations. In essence, it acts as a central authority, facilitating the enforcement of access controls throughout the platform.

Secondly, the component provides end-users with a comprehensive list of the data assets they own. This encompasses 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 offers end-users a clear and consolidated view of their assets, enhancing their ability to manage and track their assets portfolio effectively. This functionality also enhances transparency and accountability within the platform, allowing end-users to have a clear view of the assets under their control.

Overall, the Assets Policy Management plays a critical role within FAME, since it serves as the central authority for managing data asset security policies and ensuring that the appropriate data assets are displayed throughout the system. By acting as a PDP, it enables effective access control enforcement by other components, whilst it provides to the end-users a comprehensive overview of their owned data assets, enabling better asset management and control. With its pivotal functionalities, the Assets Policy Management reinforces the FAME's security posture, fosters transparency, and promotes efficient assets' management practices.

## 7.5 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 three (3) things: firstly, providing metadata attributes that unambiguously identify the nature, meaning the provenance of the underlying data asset; secondly, managing additional attributes that add market-related information to the asset (e.g., terms & conditions and pricing); lastly, 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.

## 7.6 Assets Pricing

FAME is implementing a data-driven pricing advisory mechanism leveraging issuer-provided data, stakeholder survey responses, and historical pricing realization analytics. 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 the FAME Data Marketplace.

Firstly, it suggests an objective value (i.e., price recommendation) for different data assets and services leveraging the metadata of the data assets. To this context, it specifies and implements a set of different pricing schemes for different data assets, using metadata information, including the asset's completeness, volume, quality, timeliness, CO<sub>2</sub> wastes, user friendliness, trustworthiness, and more. This solution is based on a questionnaire filled by the seller based on the characteristics of the various data assets, whereas to implement the above listed schemes, this component leverages the Assets Provenance & Traceability API of the blockchain infrastructure. Secondly, the proposed price also reflects the results from the analysis of the transaction price of similar data assets identified by the Similarity Analysis.

Overall, all the weighting factors and values obtained through the questions from the sellers are important for asset valuation. These 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. In addition, the seller can choose the best price that would reflect all costs associated with acquiring the provided data asset. Hence, the result of this component is a price recommendation for the end-users

(sellers), being targeted to help them better determine the price of the asset they are offering. The sellers can either lean towards the recommended price or set their own price, supporting a seller-oriented approach.

## 7.7 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 within the offered Data Marketplace. 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 offers a robust and secure mechanism for monetizing their data assets. By utilizing smart contracts, it ensures that providers are fairly compensated with ERC-20 tokens for the value they bring to the Data 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. Different smart contract transaction models (pay-as-you-go, subscription, pay-as-you-use) are implemented to allow payment schemes that fit best the nature of the data asset.

For data asset consumers, the component simplifies the acquisition process. Consumers are 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 allows 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 also plays a crucial role in maintaining the integrity and transparency of the Data Marketplace. All the trade details are logged on a ledger, providing a clear audit trail that promotes trust among end-users and ensures compliance with regulatory requirements. In essence, it streamlines the trading process, ensures fair compensation for providers, simplifies asset acquisition for consumers, and fosters a transparent and trustworthy marketplace.

## 7.8 Assets Searching

The semantic search engine provides a unified search interface fully integrated in the FAME Dashboard enabling real-time data availability, and enhanced discoverability. In that sense the end-users are able to quickly find relevant data assets according to their needs/requirements, for analysis and decision-making. Thanks to the single-entry point for finding data assets' information through the developed searching mechanism, FAME not only ensures data security and governance to its end-users, but it also promotes collaboration and knowledge sharing while optimizing resource efficiency.

The semantic search engine has access not only to the FDAC but it also sends the user request to other Data Marketplaces supporting the federation approach of the project like European Data and Google Dataset Search, among others. In addition to that, the semantic search engine offers the requested data assets ranked not only considering semantic criteria, but also considering additional criteria such as the assets' price, size, and format. Lastly, this service also suggests related searches to offer the end-users complementary options based on semantic analysis of the principal search query. Hence, the main goal is to provide the data asset consumers the most relevant offer depending on their specific needs and requirements for data exploitation.

## 7.9 ML & AI Analytics

The ML & AI Analytics 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 needs that the clients have. 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, is also implemented. State-of-the-art models such as Transformer, Recurrent Neural Network (RNN), Convolutional Neural Network (CNN) or classical ML models are provided to address different EmFi applications' tasks, which comprise solving time series, sentiment analysis, ranking system, or recommendation system problems.

Hence, the ML & AI Analytics catalogue provides 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. What is worth mentioning, is that all these models can be accessed in two (2) different ways:

- **Analytics as a Service:** where the end-user is provided with an endpoint to a third-party server, where he/she can query his/her data. This option is the most suitable one when the end-user does not have neither any ML-related knowledge, nor the necessary computing capabilities to deploy the AI/ML assets.
- **Analytics as an Asset:** where the end-user downloads the analytics he/she needs and runs them on his/her own premises. This option is the most suitable one when the end-user does not have ML expertise, but he/she has the DevOps/infrastructure needed to execute the asset.

According to the requirements coming from the different existing FAME pilots, the current catalogue of these models can be located in D5.1 - Trusted and Explainable AI Techniques I [76].

## 7.10 SAX & XAI Analytics

The main functionalities of the FAME developed SAX capability revolves around process-aware explainability services, aimed at providing comprehensive insights into analysed processes given their event logs as input. These capabilities include the discovery of causal execution dependency views, allowing the end-users to understand cause-and-effect relationships among events and decisions. SAX also enriches the event logs with contextual information, focusing on significant situations to gain a deeper understanding of the processes. The component offers 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 synthesizes and streamlines explanations, ensuring they are sound and interpretable, empowering end-users to make informed decisions and gain a clear understanding of the underlying processes. The SAX analytics is materialized via a set of services in the SAX4BPM library to be released as open source by the end of the project.

It is important to note that the SAX/XAI techniques are tightly coupled with the ML models (explained in sub-Section 7.9). Therefore, the SAX4BPM library can be exposed in the FAME platform (alongside the ML models) and consumed in the two (2) following ways, as in the case of the ML & AI Analytics.

- **Analytics as a Service:** where the end-user is provided with an endpoint to a third-party server, where he/she can run the analytics on his/her data. This mode of consumption requires the monetization of “service” assets in the FAME platform.
- **Analytics as an Asset:** where the end-user downloads the analytics he/she needs and runs them on his/her own premises. Being one of the data assets indexed within the FDAC, this mode is subject to FAME’s Data Marketplace manipulation on assets.

As for the XAI Scoring framework, it provides an additional capability for scoring the explainability of the different models towards comparing alternative approaches. The provided explanations are easily interpretable by human-users who can also express their preferences over the explanations based on the considered attributes. In essence, the XAI Scoring framework serves as a transformative tool that evaluates the explainability of the produced models, offering to end-users the benefits of: (i) Trust gauge, since by understanding AI decision-making processes, allows end-users to better trust and rely on the system’s outputs, (ii) Human-focused user experience, since the framework prioritizes the user experience, ensuring that AI is not just technically sound but also user-friendly, and (iii) Benchmarking, enabling users to compare different AI techniques, making informed decisions about which models best align with their objectives. Additionally, the output of this framework aids 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 are 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.

Moreover, XAI Scoring framework is developed in a way that it is use case agnostic in terms of usability, but it takes into consideration the underlying sector or industries, such finance, where clear model explainability is crucial for regulatory compliance and stakeholder trust. This practical utility highlights the framework’s potential to enhance decision-making processes across different sectors in the future, underscoring its significant impact.

### 7.11 Incremental & Energy Efficient Analytics

The purpose of this component is twofold: firstly, to provide the capability of incremental processing of analytical operations, and secondly to perform analytical query processing with an energy efficient manner.

Regarding the incremental analytics part of this artefact, it is important to highlight that with this term we refer to the analytical query operations that can be found in a relational database. These can be used by AI analytical processing in order to push such operations down as close to the storage. The benefit for the end-user, the data analyst or application developer, is that they do not have to bother on how to execute such complex operations, but instead, leave the database to perform those on their behalf. By doing so, the data analyst or application developers do not need to read enormous volumes of data, transmit them to the application or analytic layer and do the process there, but retrieve calculated results to be later used. This reduces significantly the resources needed for such operations, along with the observed latency, thus being as much closer as possible to offer near real time analytics. However, the analytical operations include an inherent complexity. In order for any data management system to calculate the results, this implies the scanning (or access) of a plethora of data elements that reside in a big dataset. Each invocation of such operation is time consuming and demands the consumption of significant resources, even if they are performed close to the storage itself. Given this innovative characteristic, the Incremental & Energy Efficient Analytics component can calculate these results incrementally, that is, as data modification operations occur in parallel. This means that the end-user can observe minimum latency, as the result is pre-calculated and modified incrementally. This can happen while database transactional semantics are ensured on the other side, thus, having the results being consistent in case of parallel modifications of the data. This allows for the provision

of real-time analytics (as latency is minimized) while data can be concurrently ingested at the same time (as database transactions are enforced).

The second innovation of this component is its ability of performing advanced query processing in an energy efficient fashion. Firstly, by exploiting the incremental analytics part, it does not have to calculate every time the analytic result. This results in reduced energy consumption. Furthermore, based on the novel indexing mechanism and the implementation of the query engine, the module can offer reduced energy consumption by managing data in a more sophisticated manner. End-users can take advantage of this feature while designing their applications or AI data pipelines as their overall solution can result in reduced carbon footprint compared to the greedy consumption of resources that complex AI analytic often require.

### 7.12 Smart Deployment

The Smart Deployment component provides end-users with deployment templates for analytics components so that they can be deployed standalone or in pre-defined pipelines. This allows FAME end-users to run complex analytics services without requiring in-depth knowledge of the various components that compose the workflow.

These pipelines are monitored for CO<sub>2</sub> emissions and optimised to promote the reduction of these emissions. This means that the workflow of a complex service evolves according to the metrics collected by the CO<sub>2</sub> Monitoring component, with the objective of minimizing it.

### 7.13 Analytics CO<sub>2</sub> Monitoring

Estimating the CPU usage of the AI/ML models used within the end-users' applications exploiting FAME is the main functionality of the Analytics CO<sub>2</sub> Monitoring component. This metric can then be translated into the energy consumed by the model. Using public information or public databases on the average kilograms of CO<sub>2</sub> emitted per kilowatt-hour per nation, this component then optimises 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<sub>2</sub> emissions, hence the main impact is to raise awareness of the end-users. 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.

### 7.14 FML Deployment

When data is collected in a massively distributed manner, with many sensors geographically distributed, there may be circumstances in which gathering all the data in a central server is not desired, mostly due to two (2) reasons: (i) data privacy must be preserved, either to protect users' anonymity, intellectual property, or both, and (ii) sensors often have limited communication resources, for example, if they are battery-powered or they are connected to low data rate networks (e.g., see the Internet of Things).

However, despite the distributed and private nature of this data, there still are incentives to build ML analytics based on it, since correlation is likely to exist across the different datasets. Therefore, if a common model could be trained without sharing the data, all parties could benefit from better models rather than individually training models on the end devices.

Federated Learning (FL) provides the mechanisms to train aggregated models from clients collecting data in a decentralized manner [77]. Instead of transmitting the raw data to a central server, the clients send the model updates obtained with the local data. As a result, FL achieves higher data privacy, since the raw data is maintained in the clients, and improves energy efficiency, since only the updates

must be transmitted. The transmitted model updates are protected with differential privacy [78], which ensures the privacy of the raw data and minimizes any data leakage.

FAME's FML Deployment implements the required components to provide such a privacy-friendly, energy-efficient, and decentralized model training mechanism, where end-users can locally gather data and train their models in the edge, whilst the aggregated model is generated at the central server deployed in the cloud. Every training round finishes once the central server distributes the aggregated model to the clients. Following this notion, the FML deployment allows end-users to access FL Analytics that enable them to make decisions based on their own data without any FL-specific knowledge. Moreover, FL Analytics as a Service is also offered, such that the end-users can cooperate with third parties in the data ecosystem built in FAME, while preserving the privacy of their own data.

### 7.15 Learning Centre

FAME offers to its end-users with access to a pool of training assets' resources, notably resources related with EmFi applications. To this context, the EmFi Training implemented functionalities of FAME 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) Integrate new EmFi training resources, and (v) Integrate with the FDAC to ensure that the resources of the Learning Centre are accessed as FAME data assets and are part of the FAME Data Marketplace, effectively yielding the Learning Centre as one more platform that will be federated around FAME. Some of the above-listed will be accessible only to properly authorized end-users, while others will be made available to all the end-users of FAME. Deliverable D7.4 - Training Programs and Learning Centre [79] provides a description of an initial mass of those FAME data assets and the approach followed for their integration in the FAME Data Marketplace via the FDAC.

### 7.16 Regulatory Compliance

Recently, various advanced Data Marketplaces have been developed in Europe, providing functionalities for data catalogues, search, analytics, trading, and accounting. Marketplaces such as i3-MARKET, DataVaults, MOSAICrOWN, MUSKETEER fall within these implementations, providing added-value features for integrating, accessing, and trading data assets, such as data assets monetization, data sovereignty, personal data protection, compliance to regulations (e.g., to GDPR (General Data Protection Regulation)). As stated above, FAME's ambition is to deliver Europe's first standards-based, secure, regulatory compliant, interoperable, and federated Data Marketplace for EmFi applications. Apart from the unique feature of the FAME Data Marketplace of the federated access control, FAME also provides a unified access to all related regulations in the field. In that sense, it provides a harmonious ecosystem of the laws and regulations according to the need of the different interested stakeholders.

The FAME Regulatory Compliance tool is based on the definition and enforcement of policies that ensure compliance with applicable laws and regulations such as GDPR, PSD2 (2nd Payment Services Directive), and 4AML (4th AntiMoney Laundering directive). Hence, the main objective of this tool is to ensure compliance of the FAME functionalities with related laws and regulations that is to comply with the security and regulatory requirements deriving from EmFi applications. In doing so, the Regulatory Compliance tool is set according to the prominent regulations of the sector (i.e., PSD2, MiFIDII, 4AML) in addition to general regulations (e.g., GDPR AI Act). Therefore, the ultimate regulatory compliance asset is a tool applicable for unified data policies management in-line with security-by-design and regulatory compliance by design principles.

## 7.17 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 has already been designed, specified, and implemented in such a way that it enhances user experience, driving user engagement, and ultimately determining the success of FAME. It focuses on providing an interface that is aesthetically pleasing, simple to use, inviting all the FAME stakeholders (either end-users or platform administrators) to further explore all the associated capabilities. As a result, user experience is 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 is to maintain consumers' interest throughout time. To successfully achieve such planning, the implemented Dashboard has visually appealing components, engaging tools, and considerate animations to produce a captivating and enjoyable user experience. As a result, the interacting stakeholders are more likely to use FAME capabilities, discover its features, and spread their positive experiences, which will promote in general FAME acceptance and growth. To ensure that all the stakeholders can connect effectively and to make FAME accessible to a wider audience, it also considers accessibility elements including appropriate colour contrast, font sizes, and sensible navigation options. This in turn improves 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.

## 8 FAME Solution Architecture

The current Section includes the overall FAME SA that has been designed considering all the related activities of Section 3, Section 4, Section 5, Section 6, and Section 7 that have led the design of the final 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.

### 8.1 C4 Architecture Model

Modelling and validating systems are 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 [80]. 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 [81] and the 4+1 architectural view model [82]. 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) [83].

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 [84].

An overview of the overall C4 Model's supported diagrams is depicted in Figure 19. 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).

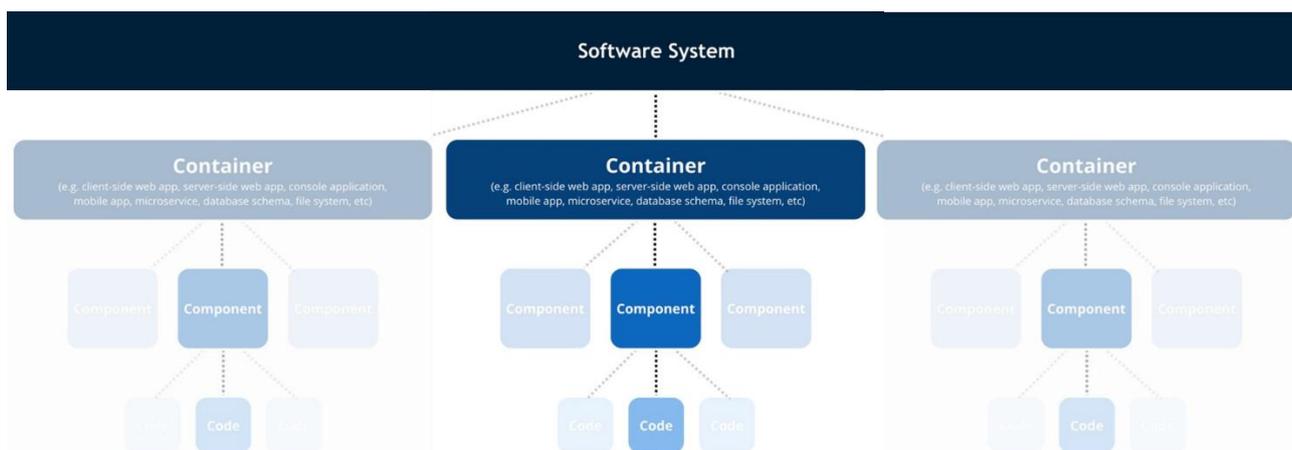


Figure 19 – C4 model diagram types [84]

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 4, Table 5, and Table 6 respectively.

Table 4 – FAME SA Context Diagram notations

Abstraction	FAME Notation	Notation Description
<b>Person</b> [ <i>Data Provider</i> ]		Any end-user that provides data assets to the FAME Federated Data Space.
<b>Person</b> [ <i>Data Consumer</i> ]		Any end-user that consumes data assets from the FAME Federated Data Space.
<b>Person</b> [ <i>FAME Administrator</i> ]		Any administrator that has access to the admin functionalities of the FAME Federated Data Space.
<b>System</b> [ <i>Internal</i> ]		Any internal system included into the overall FAME Federated Data Space.
<b>System</b> [ <i>External</i> ]		Any external system interacting with the overall FAME Federated Data Space.
<b>Relationship</b>		Any relationship between all the abovementioned abstractions.

Table 5 – FAME SA Container Diagram notations

Abstraction	FAME Notation	Notation Description
<b>Person</b> [ <i>Data Provider</i> ]		Any end-user that provides data assets to the FAME Federated Data Space.
<b>Person</b> [ <i>Data Consumer</i> ]		Any end-user that consumes data assets from the FAME Federated Data Space.
<b>Person</b> [ <i>FAME Administrator</i> ]		Any administrator that has access to the admin functionalities of the FAME Federated Data Space.
<b>Container</b> [ <i>Any type</i> ]		Any container executed inside the depicted system of the FAME Federated Data Space.

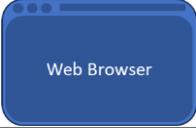
<b>Container</b> [ <i>Database</i> ]		Any database exploited inside the depicted system of the FAME Federated Data Space.
<b>Container</b> [ <i>Web Browser</i> ]		Any web browser (or UI) exploited inside the depicted system of the FAME Federated Data Space.
<b>Relationship</b>		Any relationship between all the abovementioned abstractions.

Table 6 – FAME SA Component Diagram notations

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

## 8.2 FAME Architecture Overview

### 8.2.1 FAME SA Conceptual Overview

Based on the overall identified project goals and the final collected technical and use case requirements (documented in D2.5 [6]), the final 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 6, especially focusing on the IDS RAM 5, the Dataspace Protocol, as well as the DSSC, towards realizing an **open, reliable, and federated architecture for cross-sector data exchange**.

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 end-users and systems, the so-called **Conceptual View** of the FAME

SA is provided (Figure 20). In more detail, Figure 20 provides a conceptual overview of the FAME architecture that targets on specifying how both the Data Consumers, as well as the Data Providers can benefit from such solution. In deeper detail, following the **top-down approach** in Figure 20, any **Data Consumer** (e.g., SMEs, Data Scientists, Research Centre, Financial Organizations) 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 20), it can be identified the pathway that can be followed from any **Data Provider** (e.g., Data Space, Data Marketplace, 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 8.2.2 along with the FAME C4 model of the 1<sup>st</sup> (i.e., context diagram) and the 2<sup>nd</sup> (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).

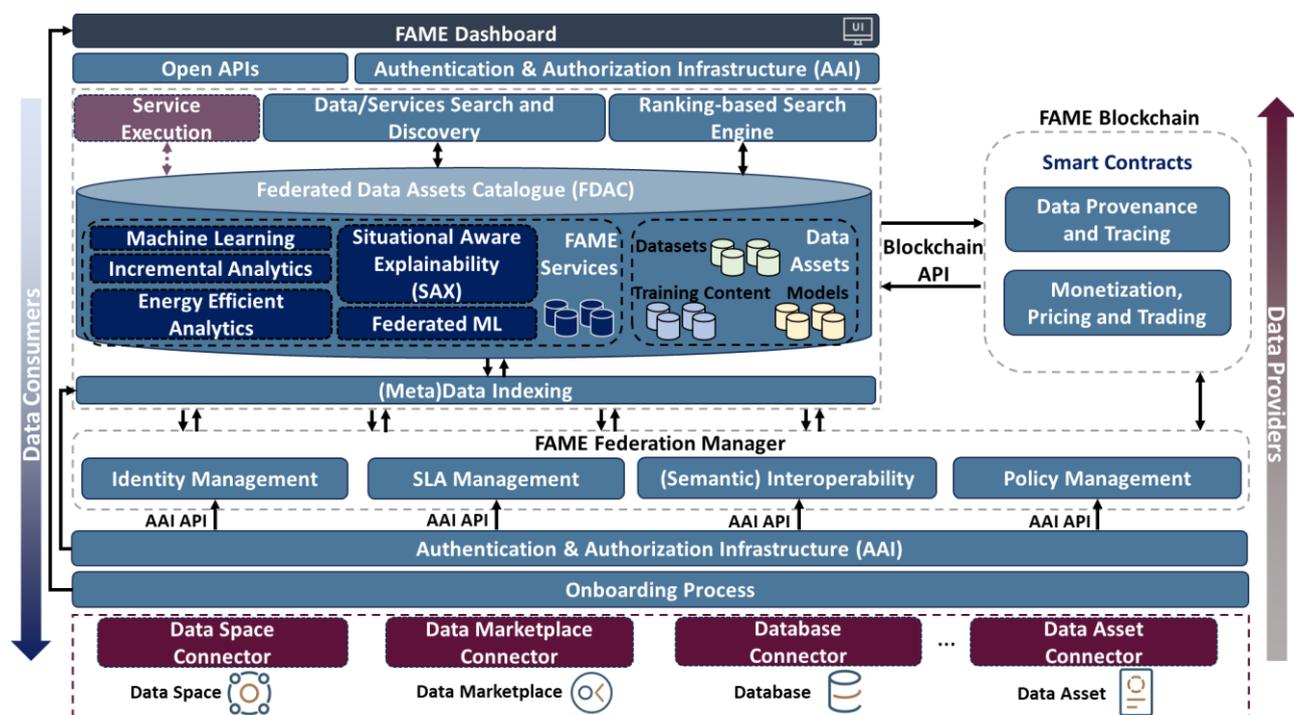


Figure 20 – Conceptual overview of FAME architecture

### 8.2.2 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 5). 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 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 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 end-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 21 illustrates the Context (i.e., 1<sup>st</sup> level) diagram of the FAME SA, depicting the diverse groups of stakeholders that are able to interact with FAME (i.e., Data Provider, Data Consumer, FAME Administrator), highlighting the means of how and through which FAME systems (i.e., layers) the involved stakeholders 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 8.1 in Table 4.

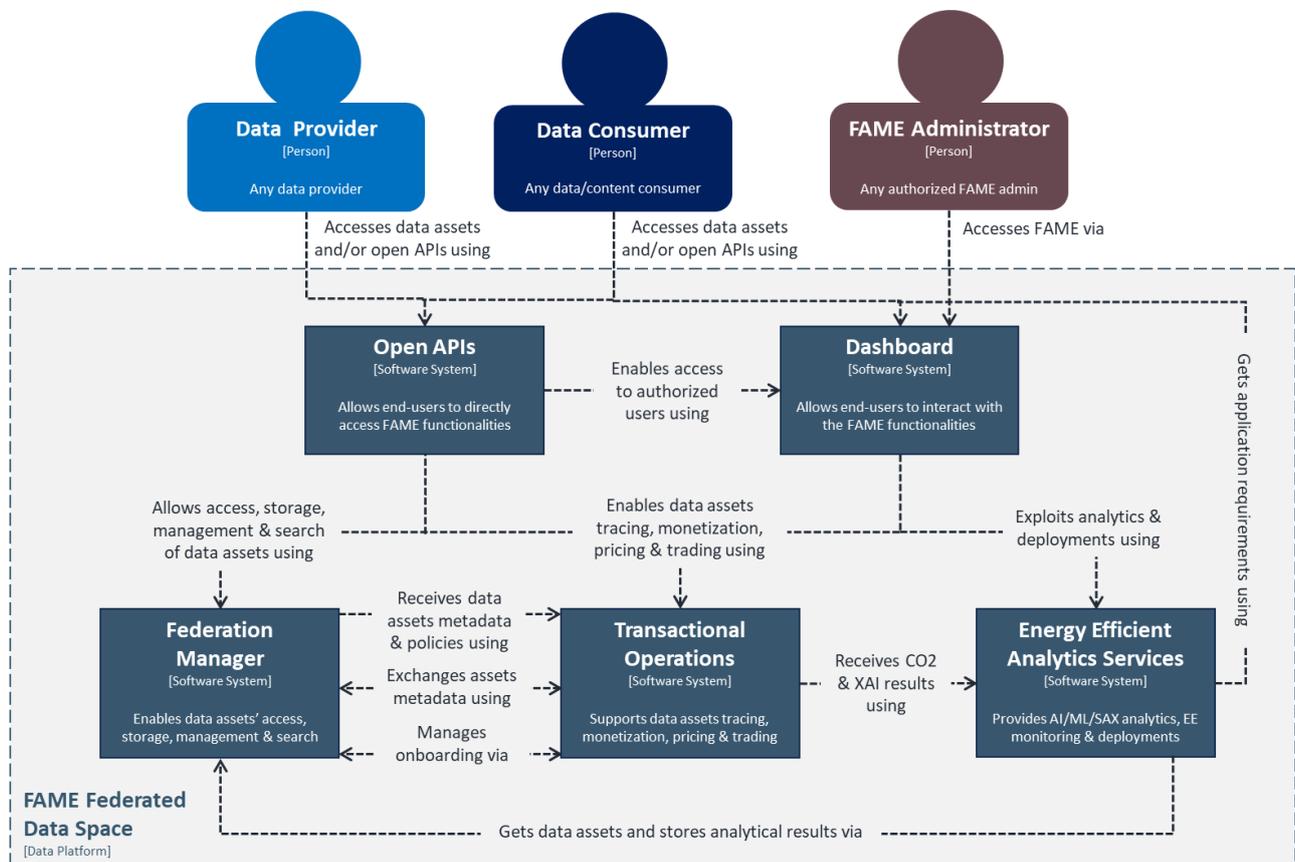


Figure 21 – FAME SA C4 context diagram

Diving deeper into the Container (i.e., 2<sup>nd</sup> level) diagram of the FAME SA, a zoomed-in view of the built platform is provided in Figure 22, 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 8.1 in Table 5.

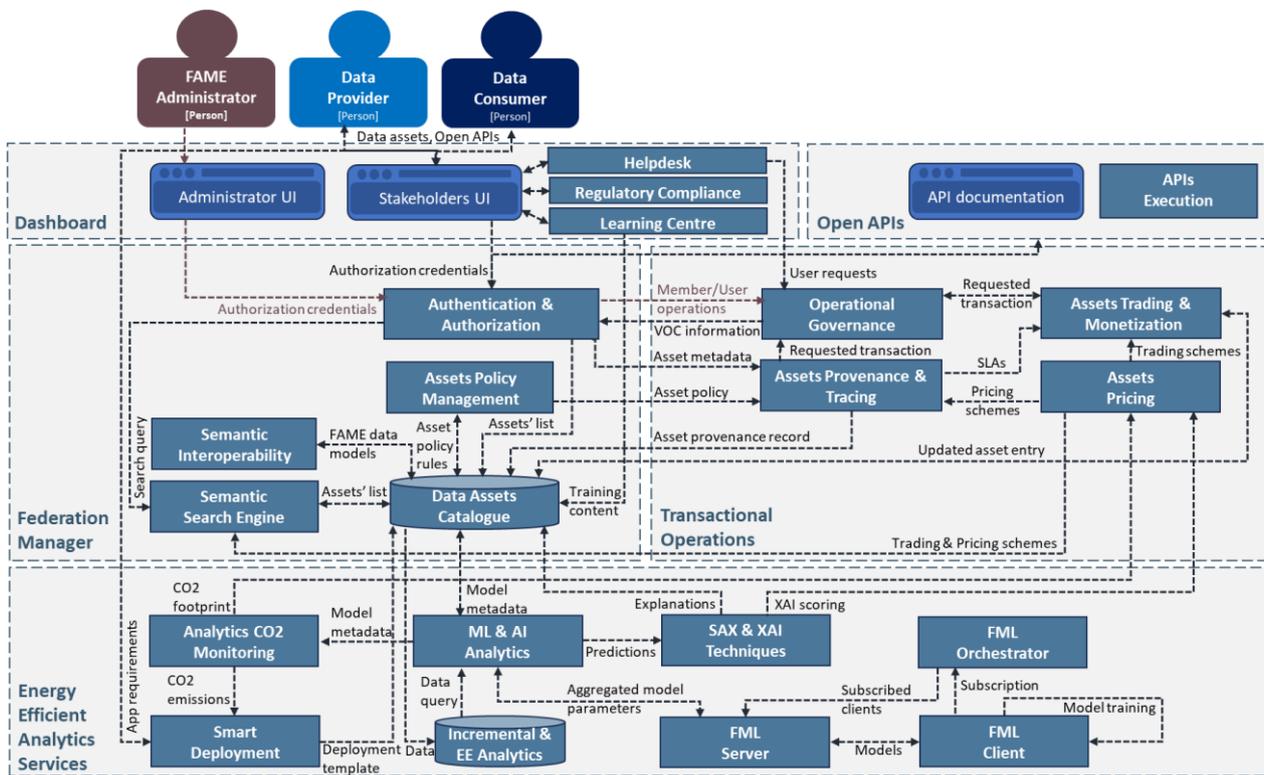


Figure 22 – FAME SA C4 container diagram

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.

### 8.2.2.1 Dashboard & Open APIs

The **Dashboard** layer contains all the components that are required for **providing access to all the FAME stakeholders** (i.e., **end-users** (data consumers, data providers) and **administrators**) regarding the supported functionalities of the FAME Federated Data Space. Hence, all of these stakeholders have the ability to interact with a variety of ways with FAME. These include diverse UIs for accessing the FAME core functionalities, referring to the *FDAC*, the *Semantic Search Engine*, the *Assets Provenance & Tracing*, the *Learning Centre*, the *Regulatory Compliance* tool, the *Helpdesk*, as well as the *Authentication & Authorization* functionality, which are directly accessible from the Dashboard Stakeholders UI residing within the **Dashboard** layer. All of these functionalities are furtherly described below, whereas it should be mentioned that for fully exploiting those capabilities, an end-user must be firstly authenticated and authorized to gain his/her access to FAME. 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 or a data consumer 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 end-users are authenticated and authorized such access is fully available, whilst for non-authenticated and non-authorized end-users such access is limited. Within this fully availability, the end-users are also able to exploit the existing open functionalities of the platform through the *APIs execution* service (lying into the **Open APIs** layer), having thus direct access to the selected underlying operation. The *API documentation* facilitates this process by providing a proper documentation for describing each API

along with specific examples, tutorials, and exploitation guidelines. Rather than these, the end-users have also access to the *Helpdesk* component, an auxiliary component of the *Dashboard* layer, to aid the end-users towards reporting and resolving any operational or technical identified issues, acting in essence as a ticketing system of the platform. Also, to boost its usability and adoption, in the *Dashboard* layer, FAME is also providing a *Learning Centre*, which aims to provide a pool of training assets' resources, notably resources related with EmFi applications. Moreover, in this layer the end-users can locate the *Regulatory Compliance* component for providing legal support upon the laws and the regulations of the assets to be indexed, assuring that the latter will obey to the relevant applicable ones (e.g., PSD2, GDPR, Markets in Financial Instruments Directive (MiFiD), EU taxonomy for Environmental, Social, and Governance (ESG) investments, EU AI Act). Finally, in the case that the FAME administrator needs to access FAME for administration purposes to manage the platform members, users and their relevant requests (as described in D4.3 - Operational Models, Business Models and Governance I [85]), the *Administrator UI* is offering a related UI for such purposes.

#### 8.2.2.2 Federation Manager

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

Whether an external end-user (either data provider or a data consumer) or just the FAME administrator 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 stakeholders 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 successfully complete such pipeline, the *Operational Governance* component is exploited for sharing the stakeholders' Verifiable Onboarding Credential (VOC) (as described in D4.3).

By the time that an interested end-user is granted access, the latter is able either to directly query through the *Semantic Search Engine* 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 end-user 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 end-user 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 8.2.2.3).

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* 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. Towards this direction, the *Assets Policy Management* serves as a critical component that ensures proper policy access, exposure, creation, and enforcement across the diverse data assets and integrated sources, also harmonizing different policies, and ensuring that data usage complies with defined policies, thereby supporting secure and efficient data management.

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 analysed in *Energy Efficient Analytics Services* layer), the *Semantic Interoperability* component provides a semantic middleware to transform 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*.

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 *Stakeholders UI*, 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 the following sub-Section) to provide the relative ranking of the assets' search results and determine their demand-driven price accordingly.

### 8.2.2.3 *Transactional Operations*

This layer contains 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 analysed 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 *Dashboard* layer in order to further support the functionality of the *FDAC* 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 *Dashboard* layer, 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* and the *Assets Provenance & Tracing* component (being also responsible for managing the offering's Service Level Agreement (SLA)), whereas the updated asset entry's metadata is also forwarded to the *FDAC* for its respective update. 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, as well as the support of auxiliary functions like billing.

#### 8.2.2.4 *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 analysed 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 either to existing or to newly indexed 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. It also provides functionalities to train, score, store and perform inference with these models within the catalogue. Access to these ML analytics is provided through two (2) different approaches:

- **Analytics as a Service:** FAME end-users are provided with a third-party server endpoint to connect to and perform model inference. This approach is suitable for developers without ML or DevOps expertise.
- **Analytics as an Asset:** FAME end-users can download the ML analytics and implement them on their own premises. This approach would give them greater control over what, how and where the models run. As such, developers will need to have a high level of expertise in both ML and DevOps.

Rather than producing those plain models, this layer also ensures the models' trustworthiness by leveraging XAI techniques via the *SAX & XAI 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 also offers 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<sub>2</sub> footprint emitted. The latter is performed by the *CO<sub>2</sub> Monitoring* component, which captures the above metrics to create the model's metadata, which is then used by the *Smart Deployment* component to optimise further deployments. The latter takes as input from the underlying application its requirements (e.g., type and locality of the data it handles, real-time/low

latency/batch processing), constructs its respective application profile, and for each profile it provides deployment configurations that optimise CO<sub>2</sub> emissions without compromising the functionality and the expected performance of the application. 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<sub>2</sub> emissions of the FAME processes.

Furthermore, by applying the *Incremental & Energy Efficient (EE) 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 & EE 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, as an additional functionality to the complexity of the given analytics' capabilities, the *Federated Machine Learning (FML) Deployment* component provides a federated learning infrastructure to support those scenarios where FAME end-users have their data distributed across different databases (silos), or where they want to leverage collaboration between companies to extract insights from their data in a federated manner.

### 8.2.3 FAME SA Mapping to DSSC Building Blocks

Considering the **DSSC building blocks for realizing a Data Space** (sub-Section 6.3.2.2.3 - Figure 18), in the context of the FAME and its supported technical functionalities, the following **technical building blocks** have been **aligned with the FAME components** (Figure 23). This analysis demonstrates how FAME integrates these foundational elements to enhance its capabilities within the federated Data Space framework. By systematically aligning with DSSC standards, FAME ensures robust interoperability, enhanced security, and optimal governance across its operations. This mapping not only underscores FAME's commitment to following best practices but also highlights its role as a key enabler within the data ecosystem. By mapping its SA to the DSSC guidelines, FAME ensures that it will deliver a cohesive and efficient service that is both accessible and secure for all stakeholders involved, ranging from financial institutions to individual end-users.

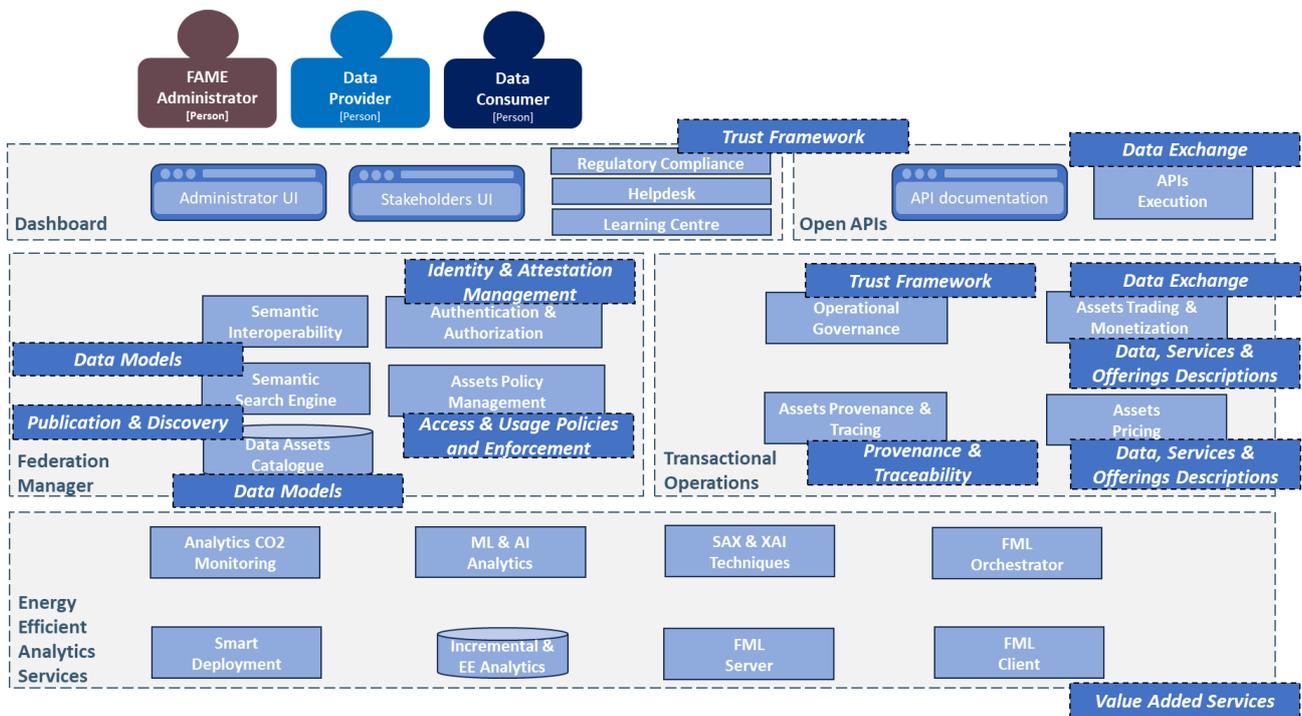


Figure 23 – Mapping of FAME SA to the DSSC Building blocks

### 8.2.3.1 Data Interoperability

This pillar is foundational to ensuring that data formats, structures, and exchanges within the FAME platform are consistent and meaningful across various systems, thereby facilitating seamless interactions and data fluidity across boundaries.

- The “Data Models” DSSC building block has been aligned with the *Semantic Interoperability*, the *Data Assets Catalogue* and the *Semantic Search Engine* FAME component. In deeper detail, “Data Models” establishes a common format for data model specifications and serialization of data in data exchange payloads. Combined with the “Data Exchange” building block, it ensures full interoperability among participants.

**FAME Implementation:** (i) The *Semantic Interoperability* ensures that the data formats and structures used by different systems can be understood and processed in a meaningful way, thus allowing various systems to interpret and use the data consistently, (ii) The *Semantic Search Engine* implements semantic search over the FDAC based on the data asset consumer specifications, (iii) The *FDAC* serves as a centralized repository for all the data assets within the FAME ecosystem, including datasets, AI models, and services. This catalogue is meticulously detailed to support advanced search functionalities, making it easy to find the most relevant data assets. It also incorporates data assets from external sources, enhancing the breadth and depth of the repository, thus it not only catalogues data assets originated from FAME, but also indexes external data sources. This comprehensive cataloguing supports the semantic interoperability by providing detailed descriptions and metadata that adhere to standardized data models, facilitating cross-platform data usage and discovery.

- The “Data Exchange” DSSC building block has been aligned with the *Assets Trading & Monetization* and the *APIs Execution* FAME component. More specifically, “Data Exchange” manages the mechanisms through which data assets are exchanged, including the monetization aspects and the execution of APIs. This involves various systems that allow end-users to trade, price, and access data assets via standardized APIs.

**FAME Implementation:** (i) The developed *REST APIs* act as open standards-based gateways that enable external actors to interact with FAME components, (ii) The *Currency Contracts & Data Assets Contracts* implement tokenization of financial exchanges and data access rights, respectively, using Ethereum standards (ERC-20 and ERC-1155), (iii) The *Offerings System & Trading Contracts* utilize smart contracts to manage and execute trade offers and transactions, ensuring secure and transparent data exchange processes. All these are core functionalities provided by the *Assets Trading & Monetization* module of FAME.

- The “*Provenance & Traceability*” DSSC building block has been aligned with the *Assets Provenance & Tracing* FAME component. In short, “*Provenance & Traceability*” ensures the integrity and traceability of data assets by maintaining a verifiable record of their origins and changes over time. This component is crucial for establishing trust in the data provided and used across Data Space platforms.

**FAME Implementation:** (i) The *Blockchain-based Traceability* utilizes blockchain technology to store digital fingerprints of metadata, ensuring that any entry in the FDAC can be verified for authenticity. It also keeps a registry of authentic sources to confirm the provenance of assets, (ii) The *Smart Contracts* are implemented using Solidity on Hyperledger Besu, a permissioned blockchain network. These contracts manage and verify the provenance and traceability of data assets, accessible via HTTP service endpoints for transactions.

### 8.2.3.2 Data Sovereignty & Trust

This pillar ensures that data within FAME is managed under strict governance protocols that protect data integrity and user privacy, thereby fostering trust and compliance.

- The “*Access & Usage Policies and Enforcement*” DSSC building block has been aligned with the *Assets Policy Management* FAME component. In deeper detail, “*Access & Usage Policies and Enforcement*” ensures data sovereignty and secure data sharing by enforcing granular control over who can access data based on predefined policies.

**FAME Implementation:** (i) The *Assets Policy Editor* allows data asset owners to define access policies for their data assets within FAME, (ii) The *Assets Policy Engine* acts as the Policy Decision Point (PDP) to manage and enforce access based on the policies set by the *Assets Policy Editor*. It determines asset visibility and ownership for users by evaluating user attributes and ownership status. By exploiting both the *Assets Policy Editor* and the *Assets Policy Engine*, the *Assets Policy Management* component in FAME effectively implements the DSSC’s “*Access & Usage Policies and Enforcement*” functionality by providing a comprehensive system for defining and managing access policies. This setup not only secures sensitive data assets, but also ensures that data utilization across the platform adheres to organizational and regulatory standards. By integrating advanced technology and robust policy management tools, FAME establishes a controlled environment where data access is systematically regulated to protect data integrity and privacy. This implementation exemplifies a sophisticated approach to data governance within the broader context of Data Spaces, aligning with DSSC objectives to foster secure and sovereign data ecosystems.

- The “*Identity & Attestation Management*” DSSC building block has been aligned with the *Authentication & Authorization* FAME component. In short, “*Identity & Attestation Management*” manages digital identities, authenticates users, and ensures that credentials are valid and verifiable. This component is crucial for establishing trust within Data Spaces by providing secure, verified access to various stakeholders, thereby maintaining the integrity and security of the data ecosystem.

**FAME Implementation:** (i) The *Self-Sovereign Identities (SSI)* allow users to manage their own identity data securely and interact with federated data marketplaces, (ii) The *Verifiable Credentials Microservice* utilizes Node.js to create a microservice dedicated to the management of verifiable credentials, (iii) The *OIDC SSI Auth Microservice*, also built with Node.js, focuses on authorization using OpenID Connect (OIDC) with SSI as the proof mechanism, (iv) The *Use of Veramo Framework* within microservices (Verifiable Credential and OIDC SSI Auth) and for managing Distributed Identifiers (DIDs) and verifiable credentials efficiently, (v) The *User Interaction and Credential Management* where users store their verifiable credentials in personal digital wallets, maintaining control over their identity data. All the abovementioned are part of the overall *Authentication & Authorization* FAME functionality.

- The “*Trust Framework*” DSSC building block has been aligned with the *Regulatory Compliance* as well as the *Operational Governance* FAME component. More particularly, “*Trust Framework*” is an integral component for establishing and maintaining trust within Data Spaces by ensuring robust governance, compliance with regulatory standards, and secure management of data assets. It involves a systematic approach to handling user interactions, asset transactions, and compliance with legal and regulatory requirements, crucial for fostering a reliable and secure data ecosystem.

**FAME Implementation:** (i) The *Operational Governance* provides the foundational services for user registration, management of business models such as pay-as-you-go and data-as-a-service, and subscription management, (ii) The *Regulatory Compliance* focuses on ensuring that all the operations within FAME comply with relevant regulatory standards such as PSD2, GDPR, MiFID, and the EU taxonomy for ESG investments.

### 8.2.3.3 Data Value Creation Enablers

This pillar emphasizes the enhancement of data assets to deliver greater value to the users, crucial for maintaining a competitive edge in a Data Marketplace.

- The “*Data, Services & Offerings Descriptions*” DSSC building block has been aligned with the *Assets Trading & Monetization* and the *Assets Pricing* FAME component. In short, “*Data, Services & Offerings Descriptions*” is designed to accurately catalog and describe the data assets available within a Data Space. This includes providing detailed metadata that supports the discovery, valuation, and proper utilization of data services and offerings. It is vital for enabling users to understand what is available within the platform, under what conditions, and how these assets can be beneficial for their specific needs.

**FAME Implementation:** (i) The *Assets Trading & Monetization* facilitates the interaction between data providers and data consumers to trade data access rights using the blockchain technology, which is crucial for the operationalization of data asset exchanges within the FAME ecosystem, (ii) The *Smart Contracts* exploited for trading utilizes ERC-20 and ERC-1155 tokens to represent monetary exchanges and data access rights respectively, which allows secure and transparent transactions, (iii) The exploited *REST API* serves as the gateway for external and internal interactions with the trading system, enabling other components of FAME to interact with the trading functionalities, (iv) The *Assets Pricing* determines the pricing of data assets based on comprehensive metadata that includes both static and dynamic variables. This component is essential for establishing a fair market value for data assets, influencing both selling and purchasing decisions, (v) The applied *Dynamic Pricing Models* utilize various data points from data assets’ metadata to formulate pricing, considering factors such as data completeness, quality, and market demand, (vi) The *Metadata Usage*, which includes detailed descriptions of the data assets, is crucial for pricing and is obtained via the *Assets Provenance & Tracing API*, (vii) The *Blockchain*

*Storage* where pricing information, once computed, is recorded on the blockchain associated with the data asset, which ensures transparency and trust in the pricing mechanism, (viii) The *Operational Synergy* where the trading component is tightly integrated with the pricing component to ensure that data assets' valuations are readily available during trade transactions, enhancing the user experience and decision-making process, (x) The *Data Usage and Monetization Strategy* where by linking these components, FAME effectively leverages its data assets, ensuring that end-users have access to well-described and appropriately valued data services. This setup promotes a sustainable business model and supports the platform's strategic objectives.

- The “*Publication & Discovery*” DSSC building block has been aligned with the *Semantic Search Engine* and the *Data Assets Catalogue* FAME component. In deeper detail, “*Publication & Discovery*” is critical for enabling users to efficiently locate and access diverse data assets within a federated Data Space. In FAME, this functionality is intricately aligned with the *Federated Data Assets Catalogue (FDAC)* and the *Semantic Search Engine*, providing a robust infrastructure for managing, discovering, and utilizing data assets with enhanced search capabilities. The alignment of “*Publication & Discovery*” with the *Semantic Search Engine* and the *FDAC* in FAME exemplifies a well-structured implementation of DSSC principles, enhancing the platform's utility and user experience. By leveraging advanced technologies and thoughtful integration, FAME ensures that its data assets are not only discoverable and accessible, but also contextually relevant to the users' needs, driving more informed decision-making and efficient data utilization across the ecosystem.

**FAME Implementation:** (i) The *FDAC* serves as the central repository for all the data assets within FAME, including datasets, AI models, services, and documentation. It handles the storage and indexing of detailed asset descriptions which support expressive search capabilities, (ii) The *Semantic Search Engine* implements advanced semantic search capabilities over the *FDAC*, utilizing NLP and ML to interpret user queries and match them with relevant data assets, (iii) The *FDAC's API Endpoint* acts as the communication bridge between the *Semantic Search Engine* and the *FDAC*, enabling seamless query operations and access to stored metadata, (iv) The *Integrated Approach* between the *FDAC* and the *Semantic Search Engine* ensures that FAME can offer efficient discovery, enhanced accessibility and dynamic pricing integration.

- The “*Value Added Services*” DSSC building block has been aligned with the *Energy Efficient Analytics Services* FAME layer. More specifically, “*Value Added Services*” within the DSSC framework aim to enhance the utility and appeal of Data Spaces by providing additional capabilities that increase the value of the data. In the context of FAME, this is exemplified by the *Energy Efficient Analytics Services* layer, which integrates advanced analytical tools and technologies to offer significant business value to end-users. This layer focuses on applying AI/ML models to data for generating actionable insights, with an emphasis on energy efficiency and explainability of AI decisions. This specific FAME implementation significantly enhances the FAME Federated Data Space by providing tools that not only meet the functional requirements of various EmFi applications, but also address critical aspects such as energy efficiency and sustainability. The integration of AI/ML analytics with an emphasis on CO<sub>2</sub> efficiency and explainability aligns with contemporary needs for responsible AI usage and environmental consciousness. By implementing these advanced analytical services, FAME not only improves the decision-making capabilities of its end-users, but also ensures that these capabilities are sustainable and transparent. This approach effectively increases the intrinsic value of the data assets managed within FAME, making them more marketable and useful for a wide range of applications. This strategic implementation of “*Value Added Services*” exemplifies how FAME is positioning itself as a leader in innovative and responsible Data Space management.

**FAME Implementation:** (i) The *ML & AI Analytics* catalogue provides a comprehensive library of state-of-the-art AI/ML models, including Transformers, RNNs, CNNs, and Linear Regressors, to perform data analysis tasks such as forecasting, sentiment analysis, and ranking systems, (ii) The *Analytics CO<sub>2</sub> Monitoring* monitors and captures CPU/GPU usage and the CO<sub>2</sub> footprint of the AI/ML models during deployment, contributing to the creation of metadata that aids in optimizing future model deployments, (iii) The *SAX & XAI Techniques* utilize explainable AI techniques to score and improve the explainability of the AI/ML models, enhancing trustworthiness and transparency, (iv) The *Incremental & EE Analytics* support real-time or run-time analytical processing by handling data incrementally, which reduces I/O and data transfer operations, thereby improving energy efficiency, (v) The *FML Deployment* provides a federated learning infrastructure to support those scenarios where FAME end-users have their data distributed across different databases (silos), or they want to leverage collaboration between companies to extract insights from their data in a federated manner, (vi) The *Smart Deployment* provides deployment configurations that optimize CO<sub>2</sub> emissions without compromising the functionality and the expected performance of the application.

#### 8.2.4 FAME as a Data Space Intermediary

FAME's vision positions it as a pivotal intermediary within the Federated Data Space, specifically tailored to the EmFi application domain. As a platform, FAME aims to act as a single-entry point, simplifying the complexity typically associated with accessing, sharing, trading, and analysing data. By leveraging cutting-edge technologies such as AI/XAI, ML/DL, and blockchain, FAME facilitates a comprehensive suite of services that underline the core functionalities of Data Space intermediaries. Towards this direction, the role and the impact of FAME as a Data Space intermediary can be identified from:

- **Business Perspective:** From a business standpoint, FAME's structure as a single intermediary centralizes the provision of enabling services, which could otherwise be a substantial overhead cost. This centralization allows FAME to streamline funding through targeted fees and other revenue models, ensuring the economic sustainability of the platform. By providing direct services to finance and non-finance organizations, Data Marketplaces, Data Spaces, and individual users, FAME fosters a competitive and innovative environment that supports the overall business model of the federated Data Space.
- **Technical and Operational Perspective:** Technically, FAME significantly lowers the barriers to entry for various organizations and individual users by providing an integrated platform that handles the complexities of modern data exchange and analytics. This facilitation is crucial for maintaining adherence to Data Space standards and ensuring operational consistency across diverse data environments, thereby enhancing the ease of integration and compliance.
- **Governance Functions:** In terms of governance, FAME operates as a critical intermediary by offering tailored services that enhance the management, security, and accessibility of data. These services are not only crucial for efficient data transactions, but also support the governance authorities of the Data Space in enforcing policies and managing the ecosystem effectively.
- **Facilitating Onboarding and Transactions:** FAME excels in simplifying the onboarding process for new participants, whether they are organizations or natural persons. By acting as a transaction facilitator and providing a clear, manageable entry point, FAME ensures that data exchanges within the Data Space are conducted smoothly and transparently, which is essential for building trust and reliability among participants.

- **Attracting New Participants:** Moreover, FAME enhances the attractiveness of the Data Space by reducing technical and operational hurdles, thereby encouraging broader participation. Its role as an intermediary reassures potential new entrants of the Data Space's viability and robustness, promoting an enriched data ecosystem.
- **Role within Data Space Governance:** As a service provider and a participant in the Data Space, FAME adheres to a governance framework that dictates its functions, responsibilities, and the value-added services it can provide. This governance framework ensures that FAME not only supports the Data Space effectively, but also adheres to the highest standards of ethical, legal, and operational integrity.

In conclusion, incorporating FAME as an intermediary within the data space framework (Figure 24) exemplifies a model where strategic design and implementation of intermediary functions align with broader goals of effective data utilization and ecosystem health. By providing essential services, facilitating entry and transactions, and enhancing governance, FAME stands out as a cornerstone component that drives the dynamic and sustainable environment of the data space, making it an indispensable part of the data economy.

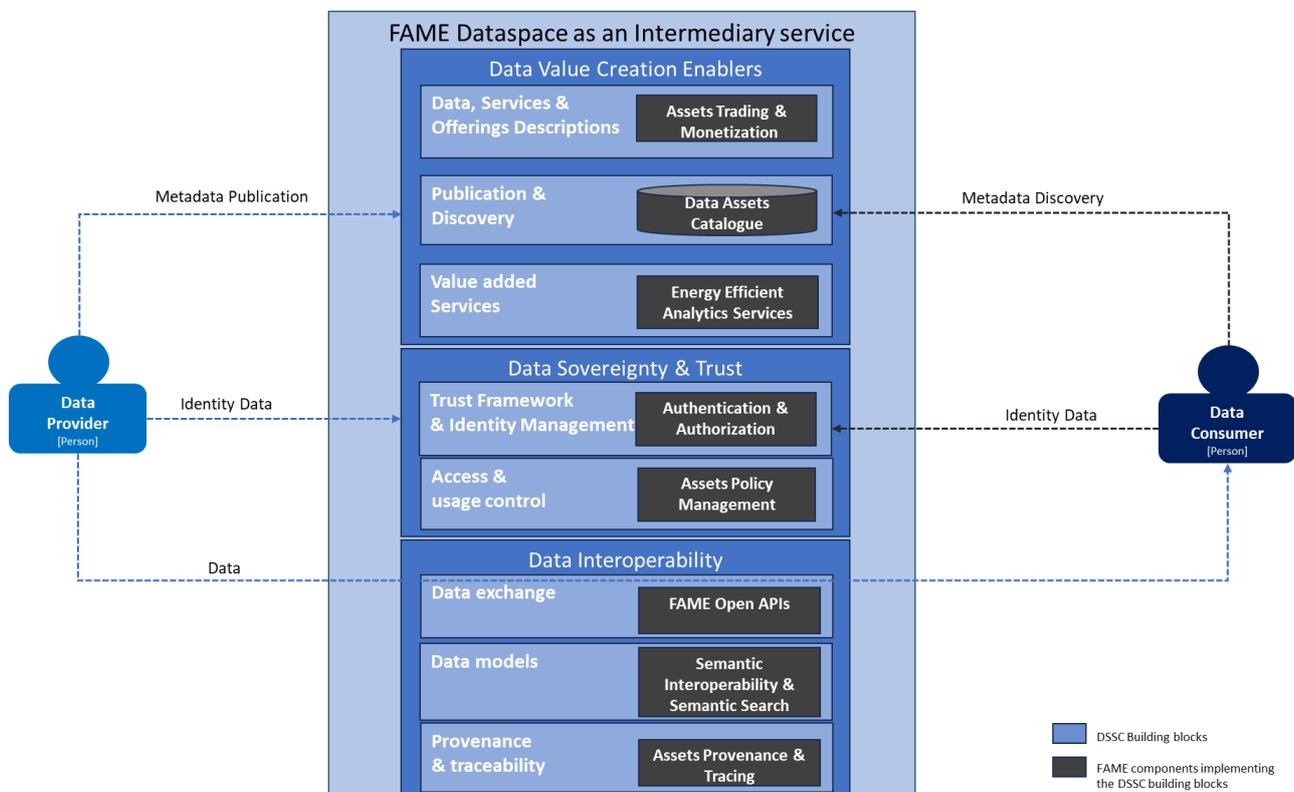


Figure 24 – FAME Data Space as an Intermediary, providing DSSC Building blocks

### 8.3 Specification of Architecture Components

Current sub-Section provides the technical information behind each different component that structures the final FAME SA, including its C4 component-level architecture (as it has been defined in sub-Section 8.1), as well as the included underlying techniques and baseline technologies behind its specification and current implementations. To facilitate the overall understanding, for each different component, a short description of its goal is provided as well, accompanied by the respective deliverable where additional information can be located.

### 8.3.1 Operational Governance

The goal of the *Operational Governance* (GOV) component is to provide the necessary services for the management and governance of the FAME federation. The component diagram, according to the C4 standard, is shown in Figure 25.

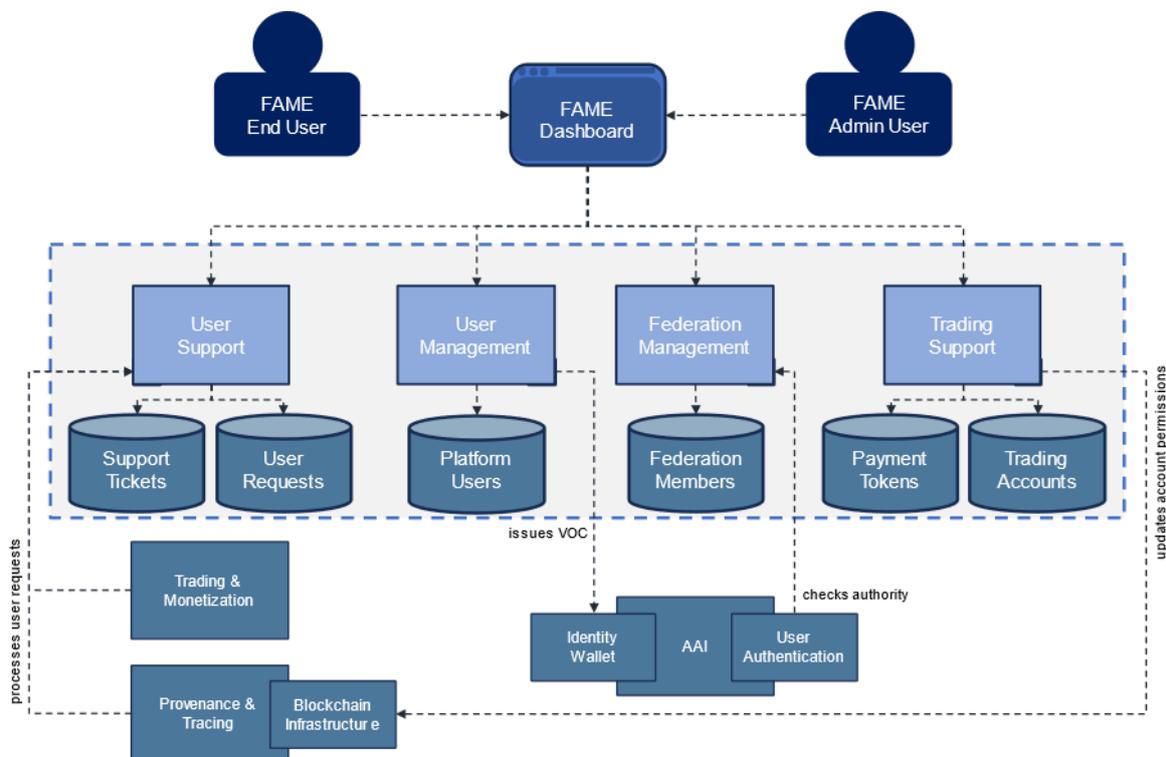


Figure 25 – Operational Governance C4 component-level architecture

The technical functionalities of the GOV component are provided by four (4) subcomponents:

- **Federation Management:** Deals with the onboarding, offboarding and record management of the members of the FAME federation, enabling them to act as Onboarding Authorities for the FAME ecosystem. When performing user authentication, the *Authentication & Authorization Infrastructure* component receives the information required to check the validity of Verifiable Onboarding Credentials from this subcomponent.
- **User Management:** Deals with the onboarding, offboarding and record management of a special category of users, referring to the platform operators. Thanks to this subcomponent, administrators can issue Verifiable Onboarding Credentials that grant special privileges to their holder, enabling them to access to the *Administrator UI* and its provided management dashboard.
- **User Support:** Implements two (2) ancillary functions of the FAME platform: (i) a ticketing system that allows the helpdesk of the FAME federation to support end-users, and (ii) a user request queue that is leveraged by the *Trading & Monetization* and the *Provenance & Tracing* components for the asynchronous processing of *Open API* operations.
- **Trading Support:** Enables blockchain accounts, owned by end-users, to engage in asset trading on the FAME platform. To this goal, the subcomponent implements a registry where trading accounts are mapped to their owners. Moreover, it interacts with the *Blockchain Infrastructure* from the *Provenance & Tracing* component for granting blockchain permissions to registered accounts, and with the *Trading & Monetization* component for provisioning them with payment tokens.

All subcomponents are implemented using the TypeScript programming language [86], as a set of RESTful microservices that run on the Node.js platform [87] and use the NestJS framework [88]. Persistence of data is managed with the MySQL [89] and MongoDB [90] databases. Blockchain-related functionalities are implemented in the Trading Support subcomponent, by smart contracts developed with the Solidity language [91] and deployed on the Hyperledger Besu platform [92] (which is managed by the *Provenance & Tracing* component). Finally, the User Management subcomponent leverages the Sphereon SDK [93] for the implementation of the issuing process of Verifiable Onboarding Credentials.

<b>Detailed deliverable:</b>	D4.3 - Operational Models, Business Models and Governance I
------------------------------	---

### 8.3.2 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 [25], which provides baseline support for self-sovereign identities and access to data from federated marketplaces.

To provide authentication and authorization with distributed identity and verifiable credentials, three (3) Typescripts and Node.js microservices have been implemented and one (1) mobile digital identity wallet. The detailed C4 component-level architecture figure of the developed *Authentication & Authorization* infrastructure is illustrated in Figure 26.

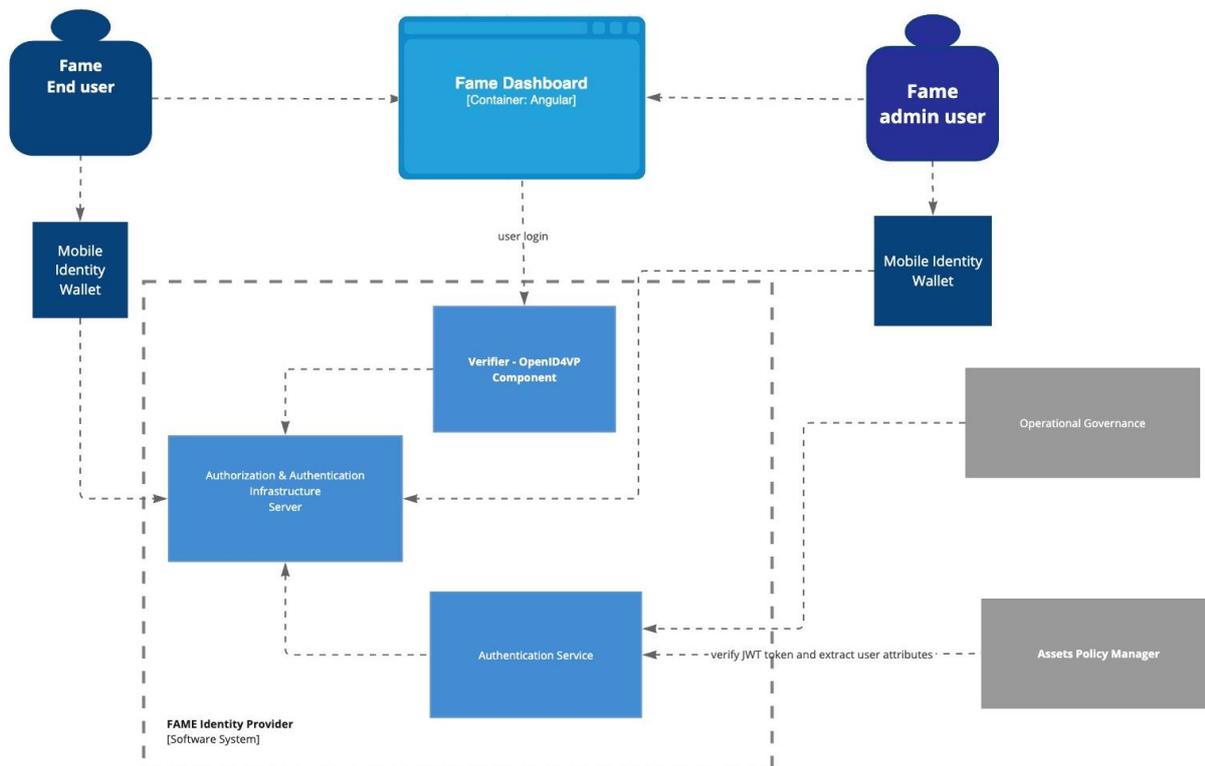


Figure 26 – Authorization & Authentication C4 component-level architecture

To implement this solution, the second version of the component adopts OpenID specifications for verifiable credentials and presentations. The *Authentication & Authorization* system uses OpenID for Verifiable Credentials (OpenID4VC [94]) and OpenID for Verifiable Presentations (OpenID4VP [95]). This allows for the secure and standardized issuance of verifiable credentials, which users can

receive from trusted sources or self-issue and store in digital wallets. These credentials are compatible with various formats, including those defined by the W3C.

As for the exploited technologies of the *Authentication & Authorization* system, OpenID4VCI [96] that is used, defines an API for issuing verifiable credentials using OAuth 2.0 [97], enabling both existing and new services to act as Credential Issuers. This supports interactions between holders/wallets and OpenID4VCI Issuer systems. Also, SIOP v2 is used to allow users to act as their own identity providers, authenticating and presenting claims directly to relying parties without third-party involvement. This enables a fully self-sovereign, peer-to-peer process. Additionally, it supports Verifiable Presentations, allowing relying parties to request and validate specific credentials from users using the Presentation Exchange.

The technical functionalities of the *Authentication & Authorization* infrastructure are defined below:

- **Verifier-OpenID4VP component:** This component initiates an authentication request when the user clicks the Login button. It generates a URI with a correlation ID to display a QR code, which requests the authentication from the Authentication & Authorization service.
- **Authentication & Authorization Service:** This core service is responsible for generating a “vp\_token” and a JWT based on the user’s verifiable onboarding credentials. It assesses the verifiable presentation to ensure it meets the presentation definition criteria. After the verification, it sends the assessment token to the verifier component, which uses it to continue the user’s session initiated by the QR code scan.
- **Authentication Service:** This service is used to validate the JWT token and extract user attributes. It is employed by internal components when a JWT is passed as a bearer token. Other services will use this service’s API to verify the validity of the JWT token.
- **Mobile Identity wallet APP:** The Digital Identity Wallet stores the user data exclusively on the user’s phone, ensuring that no one else has access unless the user chooses to share it. Hence, the user by himself/herself decides if and when to share his/her data with others. The wallet is built around W3C Decentralized Identifiers [98] and can receive W3C Verifiable Credentials [99] from issuers and present them to verifiers.

It should be also mentioned that the wallet is developed using the Apache2 open-source licensed SSI-SDK, whereas the rest of the components have been developed using Typescripts [100], React Native [101] and the Sphereon-opensource library [102] for OpenID4VCI and OpenID4VP.

<b>Detailed deliverable:</b>	D3.4 - Secure Federated Data Management II
------------------------------	--

### 8.3.3 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 sub-component plays a crucial role in enabling end-users to define and enforce access policies for all the data 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 data asset can view it, ensuring utmost privacy and exclusivity. Public access allows all the authenticated users to view the data 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 Rule-Based Access Control (RuBAC) model. These conditions act as eligibility criteria for other users to meet viewing the data asset. The

Assets Policy Editor incorporates a user-friendly UI that is part of the *Stakeholders 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* and the *Assets Provenance & Tracing* component, 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 data asset's visibility and ownership for authenticated users. Firstly, it determines which data 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 data assets they are eligible to access. Secondly, the Assets Policy Engine provides information on the data 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 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 assets' management and utilization.

By encompassing both the Assets Policy Editor and the Assets Policy Engine, the *Assets Policy Management* ensures that data assets' access within FAME is controlled, secure, and aligned with the defined policies. The collaboration between these components allows for user-friendly policy definition and efficient policy enforcement, guaranteeing that the end-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 3<sup>rd</sup> level of the C4 diagram (i.e., component diagram) that is presented in Figure 27.

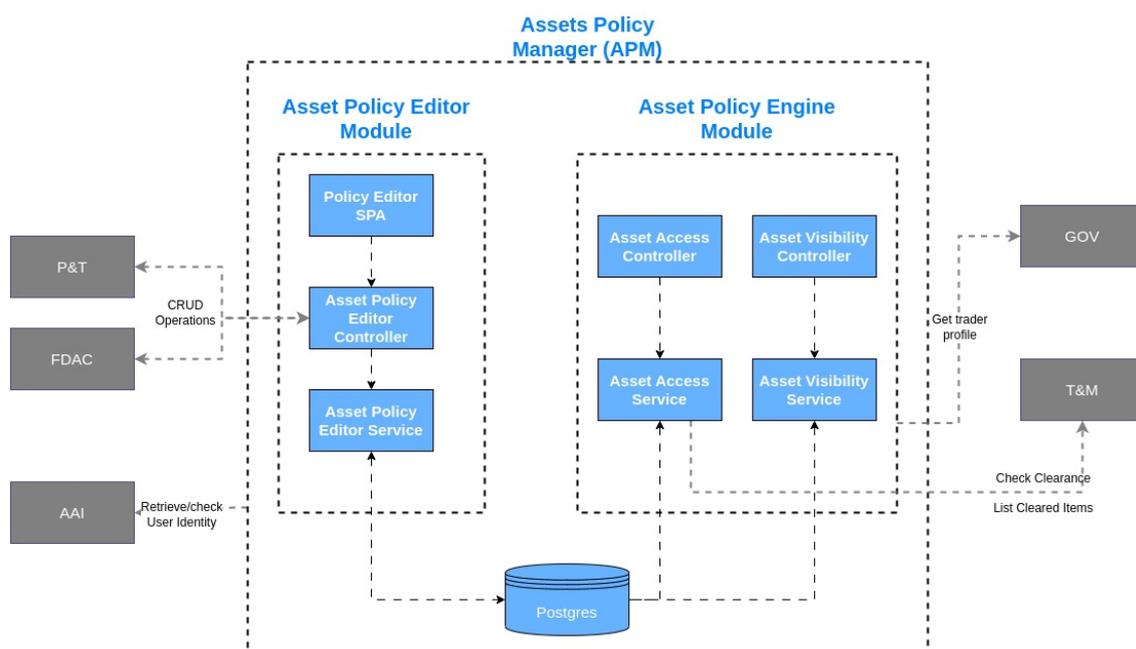


Figure 27 – Assets Policy Management C4 component-level architecture

In terms of implementation technologies, the backend of the Assets Policy Editor and the Engine is developed based on Spring Boot [103], a Java-based framework, which is employed to facilitate efficient development and provide robust functionality. The data storage for the backend is handled by Postgres [104], 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 utilize the EvalEx [105] library. EvalEx is a Java expression evaluator library that enables dynamic evaluation of mathematical and logical expressions at runtime. This makes it particularly useful for applications that require the execution of dynamic logic. EvalEx provides a flexible and lightweight alternative for policy evaluation, especially in scenarios where dynamic and custom logic is essential. It is a valuable tool for specific access control needs, allowing for the creation and evaluation of custom policies tailored to unique application requirements. Finally, regarding deployment, containerization solutions like Docker is employed to ensure consistent and reliable deployment across different environments.

<b>Detailed deliverable:</b>	D3.1 - Secure Federated Data Management I
------------------------------	---

### 8.3.4 Regulatory Compliance

The *Regulatory Compliance* component specifies and implements the security and data policies that boost the compliance of the FAME data assets to applicable regulations in UCs related with EmFi applications 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 also specifies how the various tools should be used to support the regulation of the FAME Federated Data Space in-line with applicable laws and directives. Hence, FAME provides 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 offers this *Regulatory Compliance* tool to ease the compliance to applicable regulations, also providing 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 [106] that allows everyone to compare and select open licences based on their content. This tool is applied to designate relevant licences based on each jurisdiction. This component has already defined the regulations that are relevant to FAME (Figure 28), outlining not only the regulations per se, but also the policies, the standards, and the guidelines for the different stakeholders in the project.

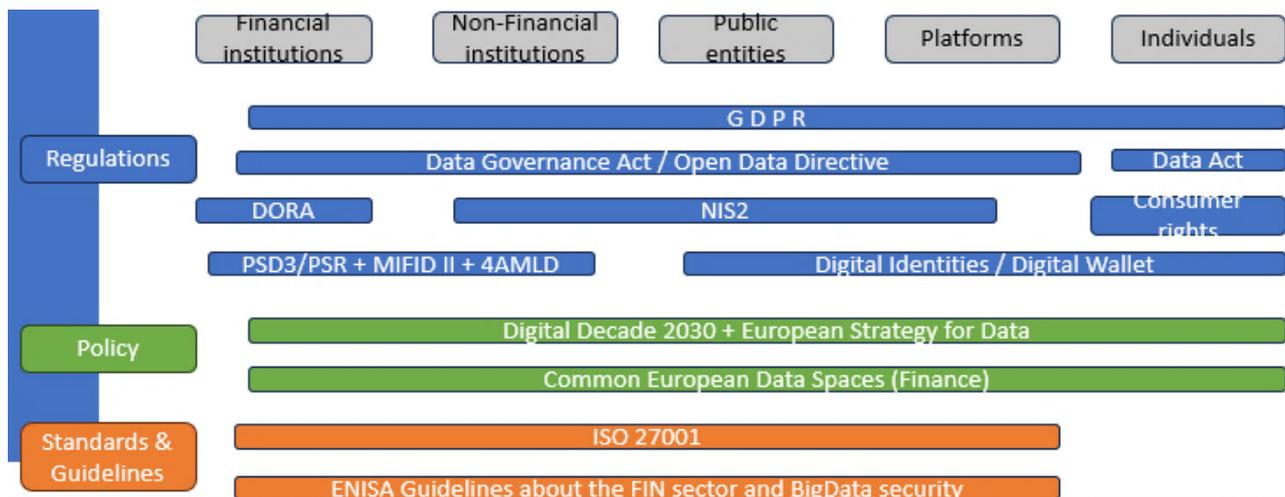


Figure 28 – FAME related laws &amp; regulations

In deeper detail, the *Regulatory Compliance* component is represented by an online tool that is already operationalized and is accessible through the *Stakeholders UI*, providing all the necessary information about background study on laws and regulations. In essence, this tool represents a regulatory policy framework portal as a window to all related laws and regulation in the FAME ecosystem. To this context, the tool encompasses an information services section, where the end-users can find information related to data protection for the financial sector. Recent news is also listed as a subsection of the tool, which provides update on regulatory ecosystem of FAME for different stakeholders. What is more, the tool provides EU data protection regulations, where the most relevant policies and regulations on Data Protection and Data Governance are listed. On top of all these, it includes an interactive tool for stakeholders to browse into the regulatory framework, providing a download link for retrieving related laws and regulations.

<b>Detailed deliverable:</b>	D3.3 - Mechanisms and Tools for Regulatory Compliance I
------------------------------	---

### 8.3.5 Assets Provenance & Tracing

The *Assets Provenance & Tracing* component provides the means for ensuring the quality of metadata published on the *Federated Data Assets Catalogue (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 on a Tracing Ledger. Moreover, the same blockchain also maintains a registry of authentic sources, the Provenance Ledger, which can be referenced in *FDAC* entries as the provenance of the underlying asset.

A separate database, the Offering Catalogue, supports the *FDAC*, the *Operational Governance* and the *Assets Trading & Monetization* components by linking offering definitions - i.e., the formal description of the terms and conditions under which access to digital content is sold through the FAME Platform - to the assets published on the catalogue.

The overall picture of the *Assets Provenance & Tracing* component and its relationships is depicted in the 3<sup>rd</sup> level of the C4 diagram (i.e., component diagram) that is presented in Figure 29.

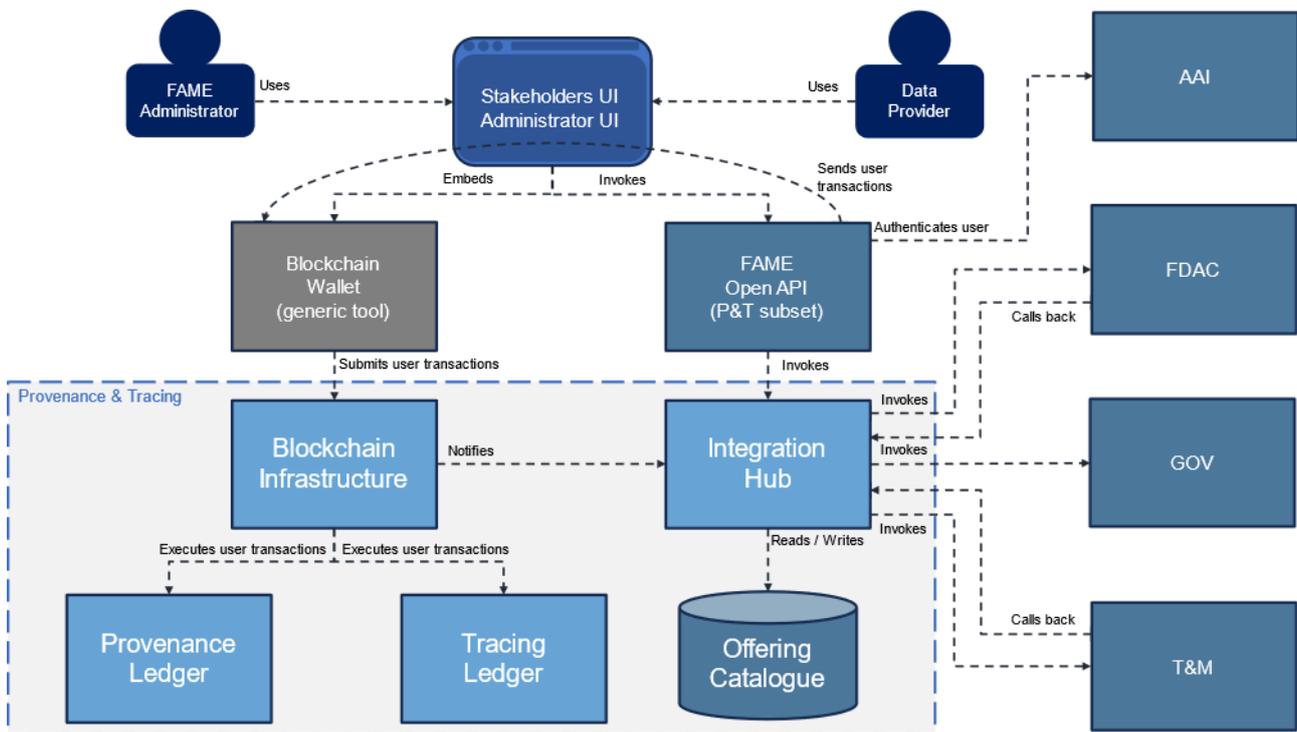


Figure 29 – Assets Provenance & Tracing C4 component-level architecture

The blockchain-related part of the component is entirely implemented as Solidity [91] Smart Contracts that are deployed on the FAME’s Hyperledger Besu [107] permissioned blockchain network. The network exposes one or more HTTP service endpoints that can be invoked by clients to execute Smart Contract transactions. The Offering Catalogue is a plain MongoDB no-SQL database [90]. The Integration Hub, which provides integration hooks for the other components of the FAME Platform, is a set of JavaScript microservices hosted in a Node.js [87] server instance.

<b>Detailed deliverable:</b>	D4.1 - Blockchain-based Data Provenance Infrastructure I
------------------------------	--

### 8.3.6 Assets Pricing

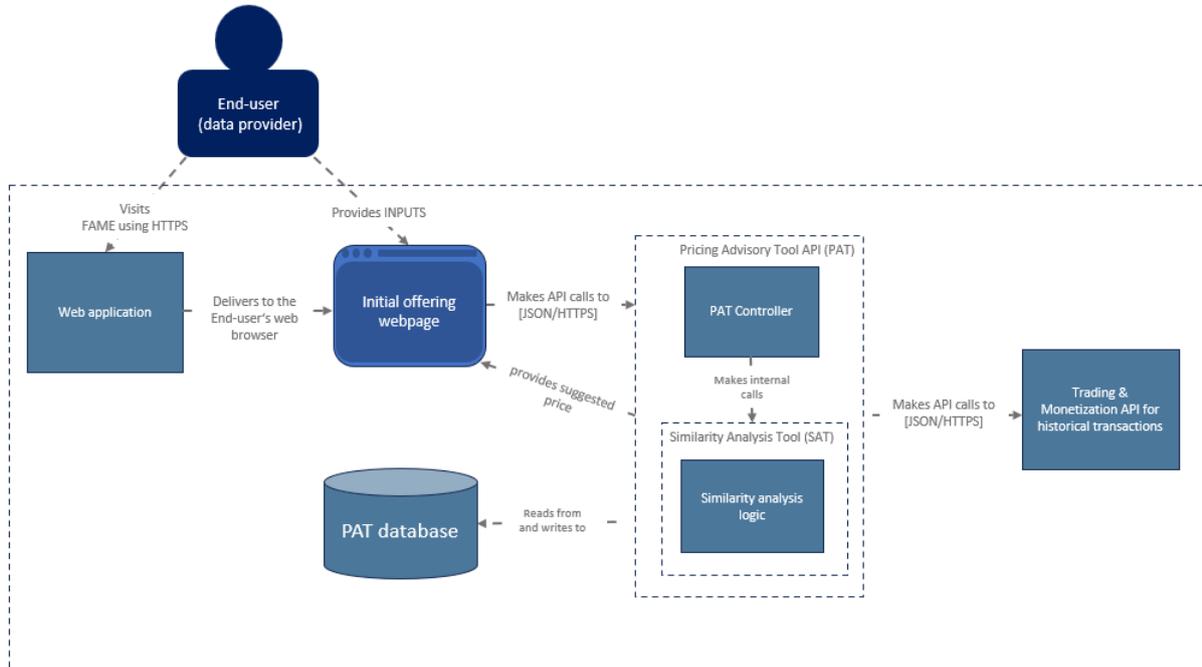
The *Assets Pricing* component is closely tied with the *Assets Provenance & Tracing* as well as the *Assets Trading & Monetization* components of the platform. More specifically, the objective of this tool, which is important for the sustainability of FAME and its business potential, is to price the data assets. To this context, pricing models are defined for the different types of data 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 data asset, enabled by the *Assets Provenance & Tracing* API.

Objectivity is ensured by the information obtained during the offering process through *FDAC*, where information is received about the asset, such as descriptions, titles, or summaries (in the case when a data asset is not available in the *FDAC*, this step is skipped). This information is later used in the Similarity Analysis process of the *Assets Pricing* component and is supported by historical transactional data. Input to the *Asset Pricing* component also includes weighting factors and values provided by the sellers through questions, which contribute subjectivity to the final calculation for generating the recommended price of a data asset.

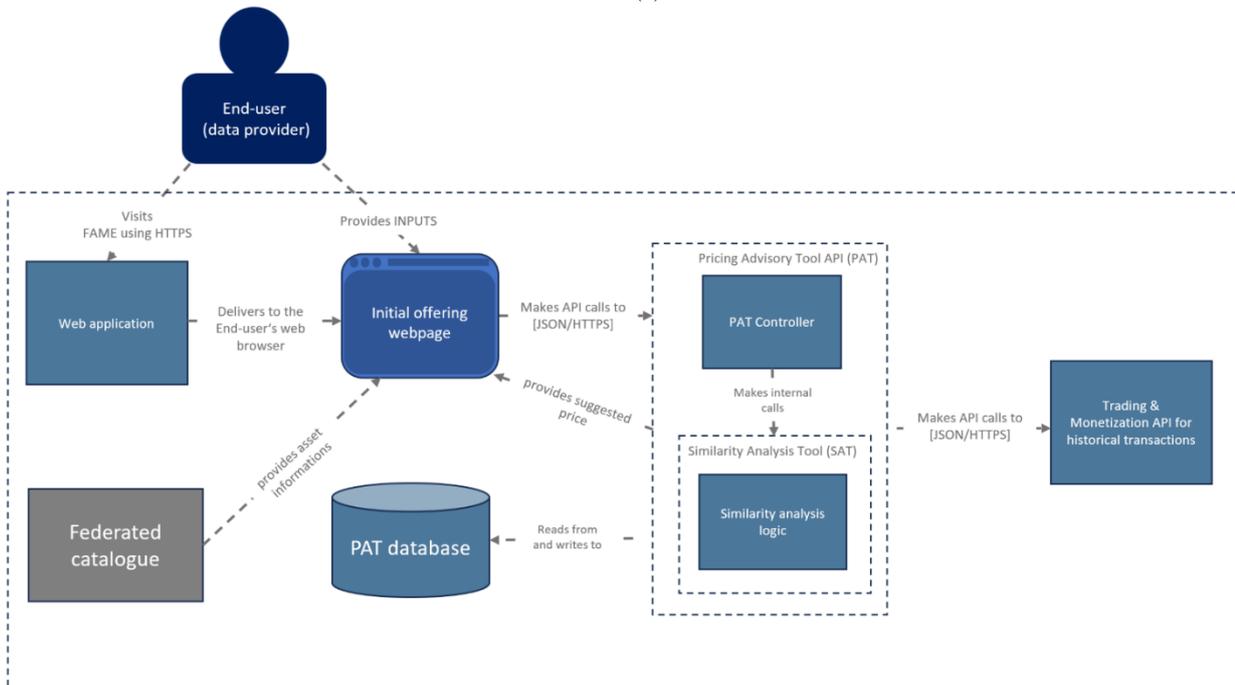
Both sets of information are used during the Similarity Analysis, where all the obtained data are used to provide the best match for similar sold data assets. In this process, transactional information of similar assets is used - not the suggested one, which improves the suggestion process since real data is exploited to propose the suggested price. The assumption that should be stated to this part, is that

there should always exist historical transactional data, (that are needed for the price proposals), which actually can be received through the *Trading & Monetization API*. Then, the output is a suggested price that reflects the objective value of the data asset, serving as a reference point for the seller when setting the selling price, being also written to the asset’s blockchain information. Finally, this information is forwarded for use in the *Assets Trading & Monetization* component.

The overall idea of the abovementioned approach is depicted in the 3<sup>rd</sup> level of the C4 diagram (i.e., component diagram) of the *Assets Pricing* component that is presented in Figure 30(a) – for the cases that FDAC data is not available, Figure 30(b) – for the cases that FDAC data is available.



(a)



(b)

Figure 30 – Assets Pricing C4 component-level architecture

<b>Detailed deliverable:</b>	D4.2 - Pricing, Trading and Monetization Techniques I
------------------------------	---

### 8.3.7 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 either data providers or data 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.

Also important to mention is that there are different payment models that can be performed. Depending on the type of data asset exchanged, payment happens with the download of the digital asset. To support these payment schemes, a smart contract allows to temporarily store the ERC-20 token in an escrow account, to be released and sent to the data asset seller once the data asset has been consumed. The smart contract also holds the notion of an expiry date, enabling subscriptions to data assets.

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 3<sup>rd</sup> 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 31 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 [108].
- 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.

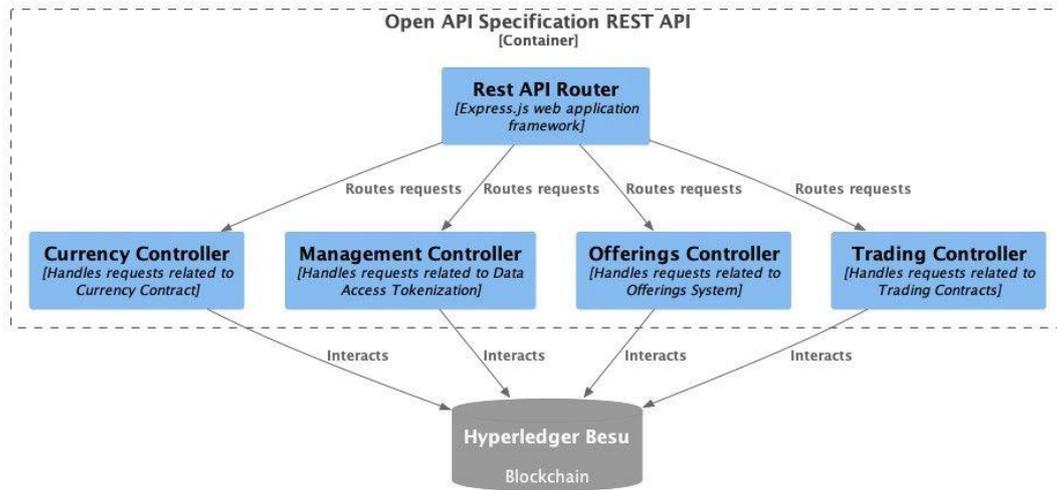


Figure 31 – Assets Trading & Monetization C4 component-level architecture - REST API

Figure 32 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 3<sup>rd</sup> 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.

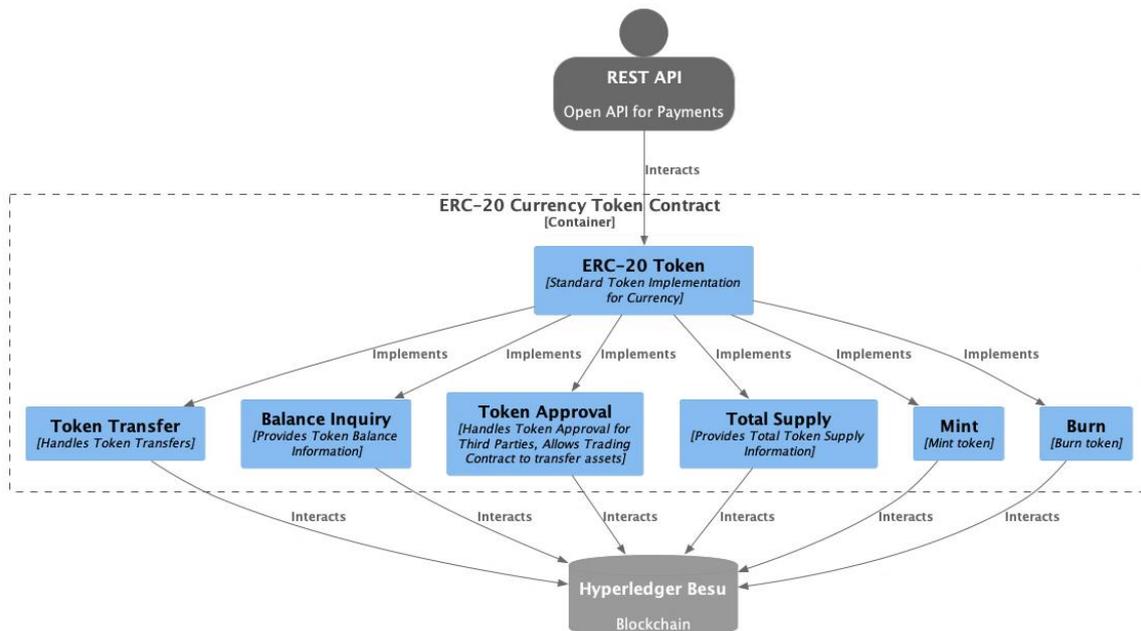


Figure 32 – Assets Trading & Monetization C4 component-level architecture - Currency Contracts

Figure 33 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 handling token approval for 3<sup>rd</sup> 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.

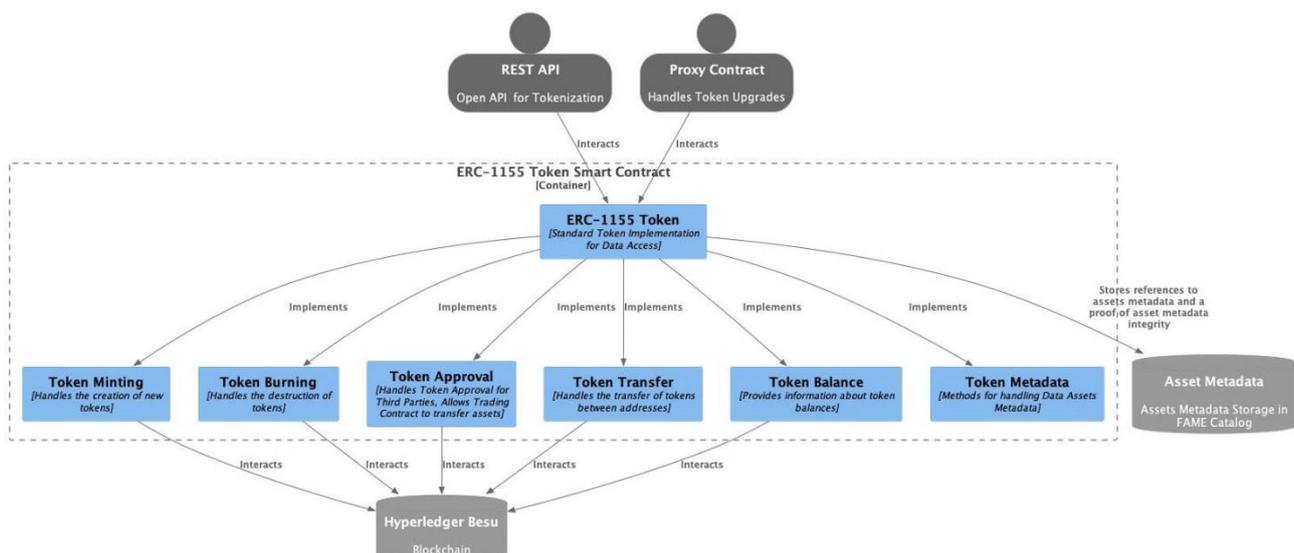


Figure 33 – Assets Trading & Monetization C4 component-level architecture - Assets Contracts

Figure 34 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.

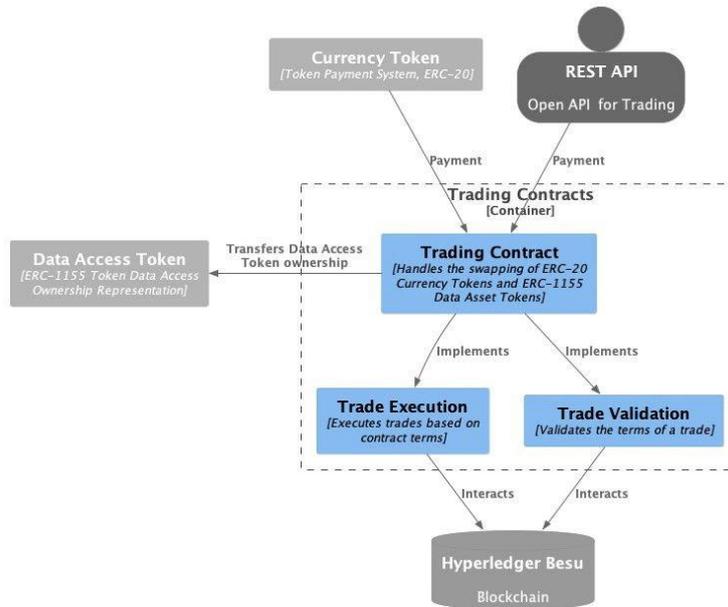


Figure 34 – Assets Trading & Monetization C4 component-level architecture - Trading Contracts

<b>Detailed deliverable:</b>	D4.2 - Pricing, Trading and Monetization Techniques I
------------------------------	---

### 8.3.8 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 are 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 Data Spaces or Data Marketplaces. All the information about the assets is made available to any FAME component that requires such information for its operation. Figure 35 presents the 3<sup>rd</sup> level of the C4 diagram (i.e., component diagram) of all the actions undertaken by the *FDAC*.

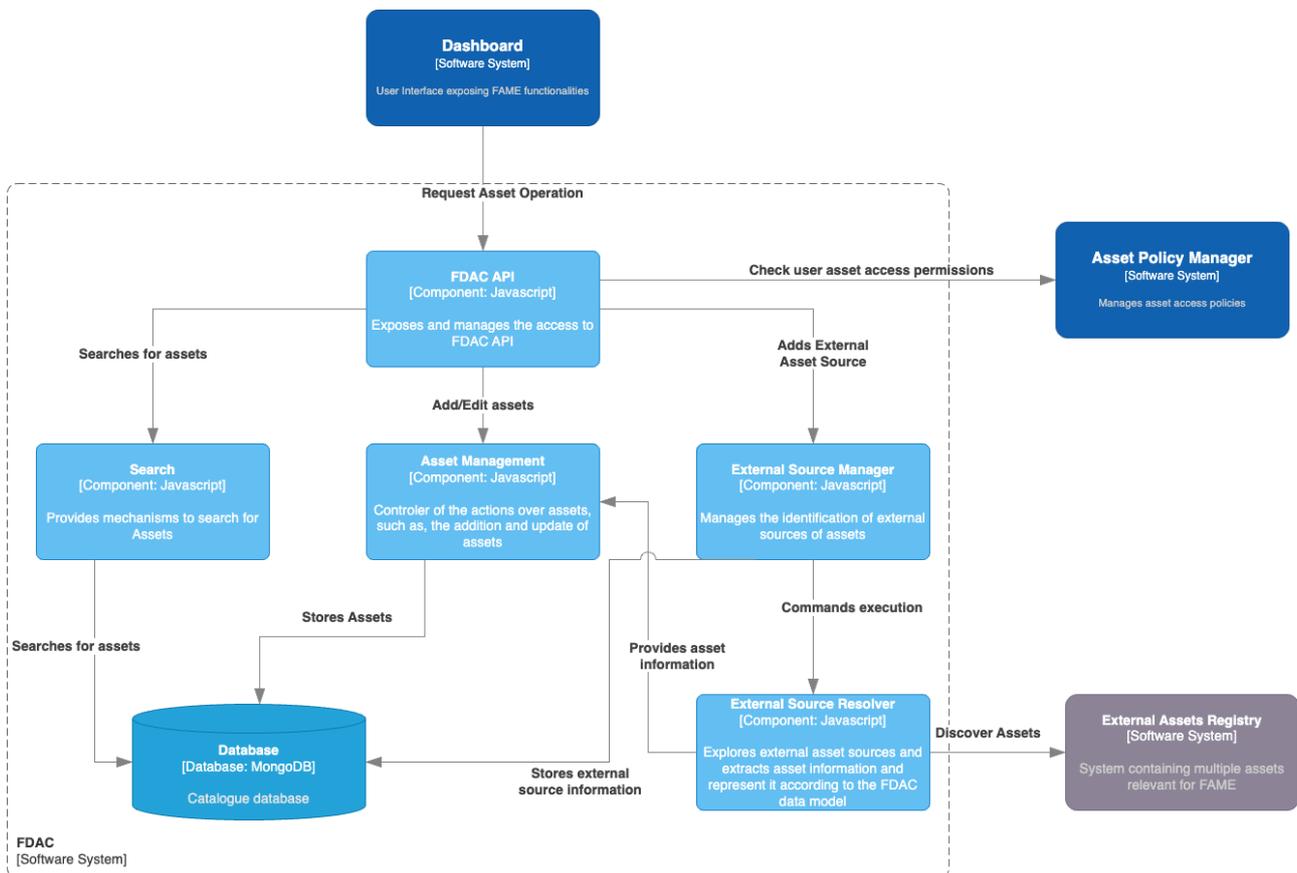


Figure 35 – Federated Data Assets Catalogue C4 component-level architecture

The core component of the *FDAC* is the Asset Management component. This component is responsible for controlling the edition operations over the data assets, providing the mechanisms to add, delete and edit the assets. When an operation is executed, it performs the corresponding actions on the database that persists the information.

As for the discovery and import of the data 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 the data assets to be analysed. For each source it collects and stores information about the description and location of the 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 the data 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 data asset, the location of the asset, any existing pricing information, as well as information about any policy related with the data 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.

End-users are able to interact with the *FDAC* through the *Stakeholders UI* that integrates the functionalities of the *FDAC* by providing the adequate UIs. Such UIs communicate with the *FDAC* through the *FDAC API* component that exposes a set of REST APIs to give access to the functionalities. During the usage of the exposed APIs, the access permissions to the data assets is verified by checking with the *Assets Policy Management* component (through the *Asset Policy Manager*) if the user has permissions to perform the required actions over each data asset.

<b>Detailed deliverable:</b>	D3.2 - Federated Data Assets Catalogue I
------------------------------	--

### 8.3.9 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 data 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 the *FDAC*. For instance, when adding a new data asset, the concepts that describe it can be described using DCAT [109] 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 having in place this functionality are represented in Figure 36 in the context of the 3<sup>rd</sup> level of the C4 diagram (i.e., component diagram) of this component.

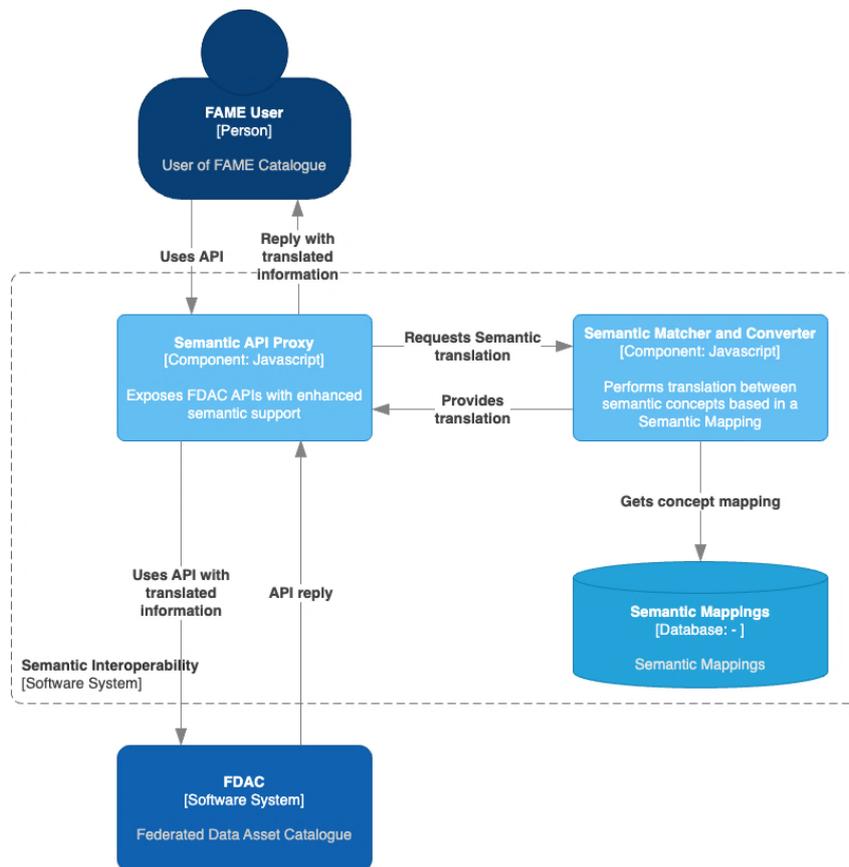


Figure 36 – Semantic Interoperability C4 component-level architecture

In deeper detail, when an end-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 end-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 end-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 end-user.

<b>Detailed deliverable:</b>	D3.2 - Federated Data Assets Catalogue I
------------------------------	--

### 8.3.10 Semantic Search Engine

The *Semantic Search Engine's* goal is to implement semantic search over the *FDAC*, and other data platforms based on the data asset specifications filled in by the user in the *Stakeholders UI*, along with schemes for ranking the results according to relevance and value-based attributes of the data assets. This component takes 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 3<sup>rd</sup> level of the C4 diagram (i.e., component diagram) in Figure 37.

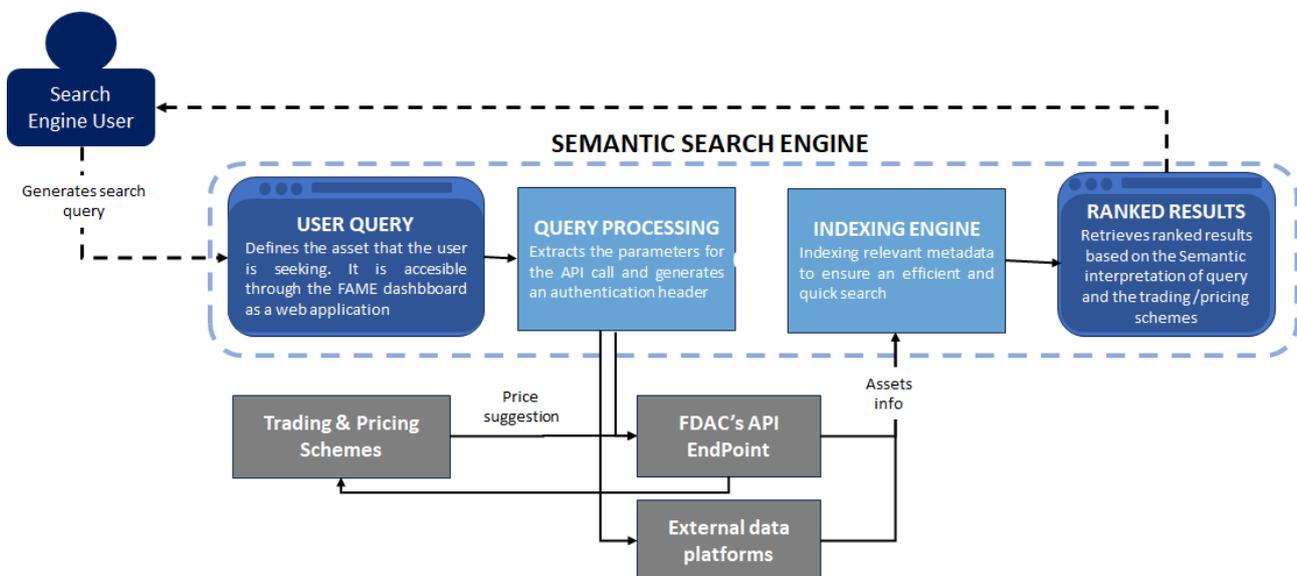


Figure 37 – Semantic Search C4 component-level architecture

- **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* and other data platforms like Google, European

Commission and Kaggle. It utilizes semantic analysis to find data assets that best match the end-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 [110] and Elasticsearch [111] that facilitate the effective semantic analysis of user queries and the efficient indexing of metadata, whereas Python's flexibility in processing unstructured data also plays 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*. This endpoint also retrieves the information about the pricing and trading schemes based on the specific query.
- **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 data 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.

The semantic search is an interface that is integrated with the *Stakeholders UI* of the FAME Dashboard using an iFrame. This setup embeds the Vue.js application within the FAME Dashboard, ensuring that it functions as a part of the larger environment. The iFrame serves as a container for the Vue.js application, allowing it to run within the FAME Dashboard while maintaining the existing layout and design. This method provides a straightforward way to incorporate the Vue.js interface into the FAME Dashboard, enabling interaction between different components and simplifying the deployment and management of the application within the existing infrastructure.

<b>Detailed deliverable:</b>	D4.2 - Pricing, Trading and Monetization Techniques I
------------------------------	---

### 8.3.11 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 aforementioned component consists of three (3) different blocks:

- **AI Model Trainer:** This block provides users with a set of interfaces and general functions necessary to train pre-built models with their own data. These functions support different ML frameworks such as TensorFlow [112], PyTorch [113] or Scikit-Learn [114], whilst data is served from the *Incremental & EE Analytics* component.
- **AI Model Catalogue:** This block provides end-users with the necessary functionality to index and download models to both the *FDAC* component of FAME and model repositories such as MLFlow.
- **AI Model Serving:** This block provides end-users with the necessary resources (i.e., an endpoint) to make inferences with a model hosted on a third-party server.

Figure 38 shows how these three (3) components can be combined depending on the role. Figure 38(a) shows a data provider schema that wants to train an ML/AI model and then publish it to the *FDAC* and a model repository. Figure 38(b) shows the configuration of the components when the end-user wants to use a model as an asset, where he/she first locates the model in the repository and then downloads the asset from it. Finally, Figure 38(c) shows the scheme following the analytics as a service approach, where the inference block provides the endpoint to connect to a third-party server.

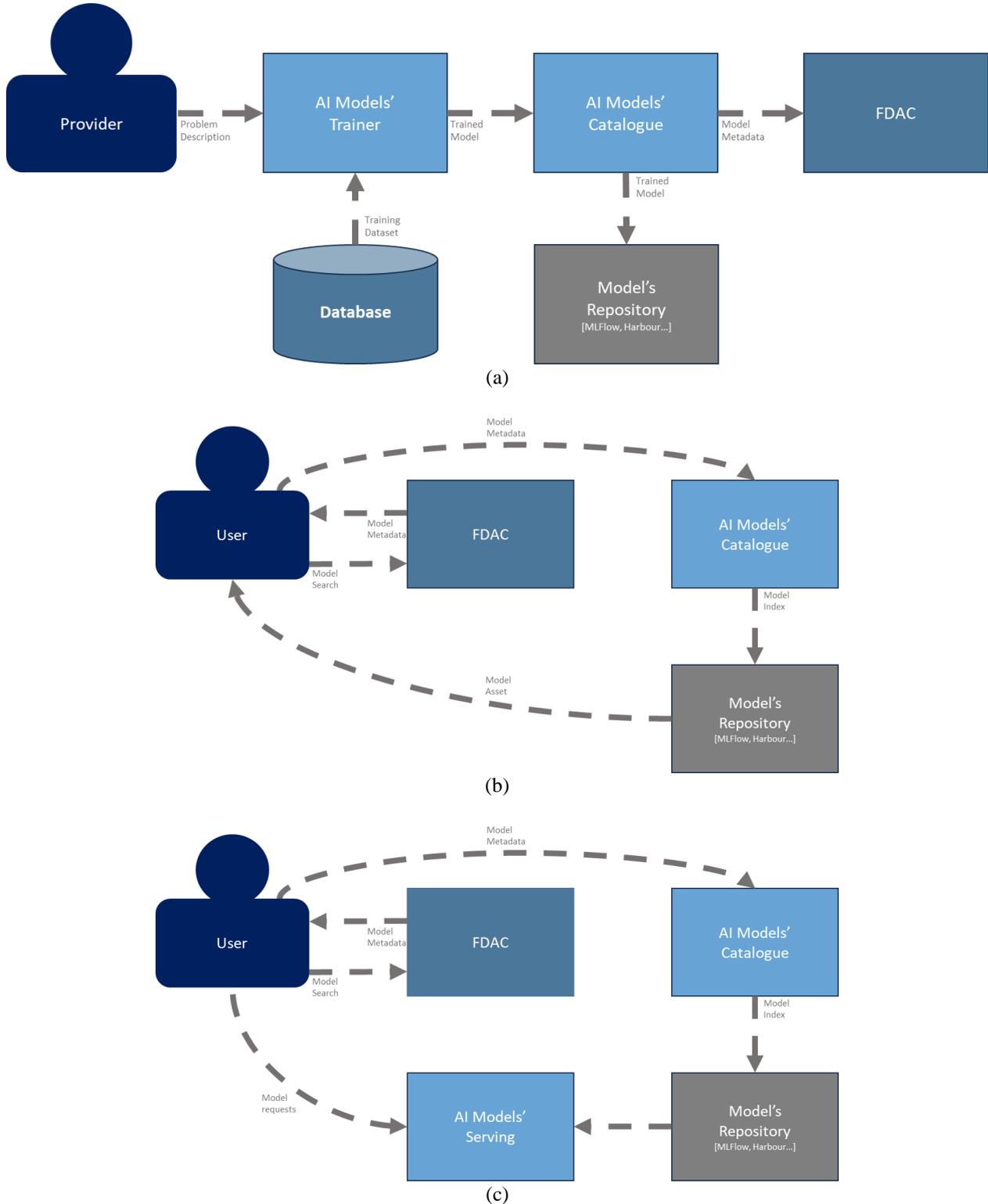


Figure 38 – ML & AI Analytics C4 component-level architecture

<b>Detailed deliverable:</b>	D5.1 - Trusted and Explainable AI Techniques I
------------------------------	--

### 8.3.12 SAX & XAI Techniques

The *SAX & XAI Techniques* aim to offer the overall packaging for all the SAX and XAI functionalities in the context of FAME. More specifically, the realization of the *SAX Techniques* is implemented as Python libraries. SAX code has been developed as a set of components and services, as the SAX4BPM library will be released to the open-source community by the end of the project. Among its supported functionalities, the SAX4BPM library offers the following services:

- Mining4Process: Generates the process model knowledge ingredient.
- Causal4process: Generates the causal process model knowledge ingredient.
- ContextEnrichment: Enriches the event log with any contextual information associated with the process and the environment in which it executes.
- X4Process: First, it generates and determines importance ranking over the set of features that are used in the process to predict the condition of interest while adhering to the constraints of the process model. Second, it uses the ContextEnrichment service to enrich the event log with broader, context-relevant attributes, about the same condition.
- NLP4X: Interweaves the various knowledge ingredients and act as a facade for the interaction with the user, prompting this composition to a Large Language Model (LLM).
- LLMsSentimentXplain: Derives explanations from sentiment analysis.

Figure 39 depicts the high-level architecture of the SAX4BPM library.

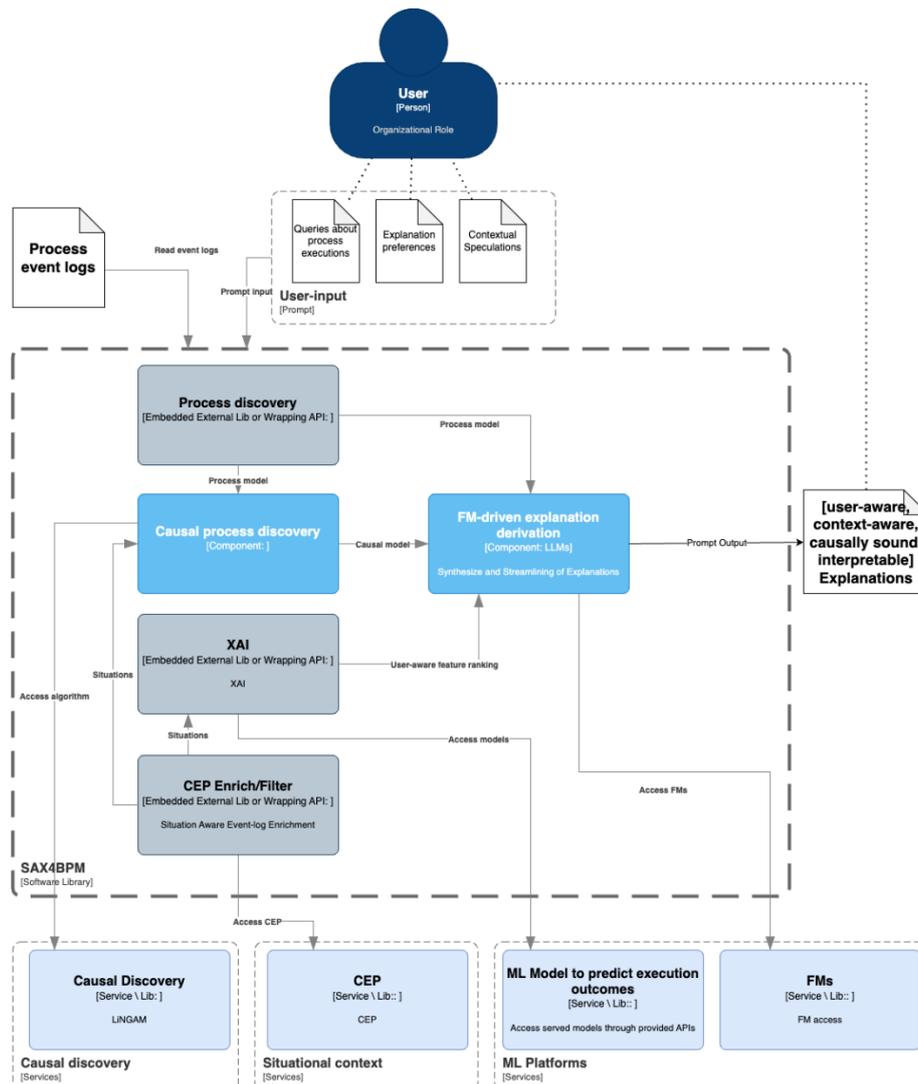


Figure 39 – SAX Techniques C4 component-level architecture

As for the supplementary *XAI Scoring* framework, this introduces 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 delineates a quantification scheme that considers properties of XAI algorithms including fidelity, stability, simplicity, and coverage. Simultaneously, it integrates user-centric aspects based on user-satisfaction. To ensure a holistic assessment, it integrates 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 3<sup>rd</sup> level of the C4 diagram (i.e., component diagram) in Figure 40.

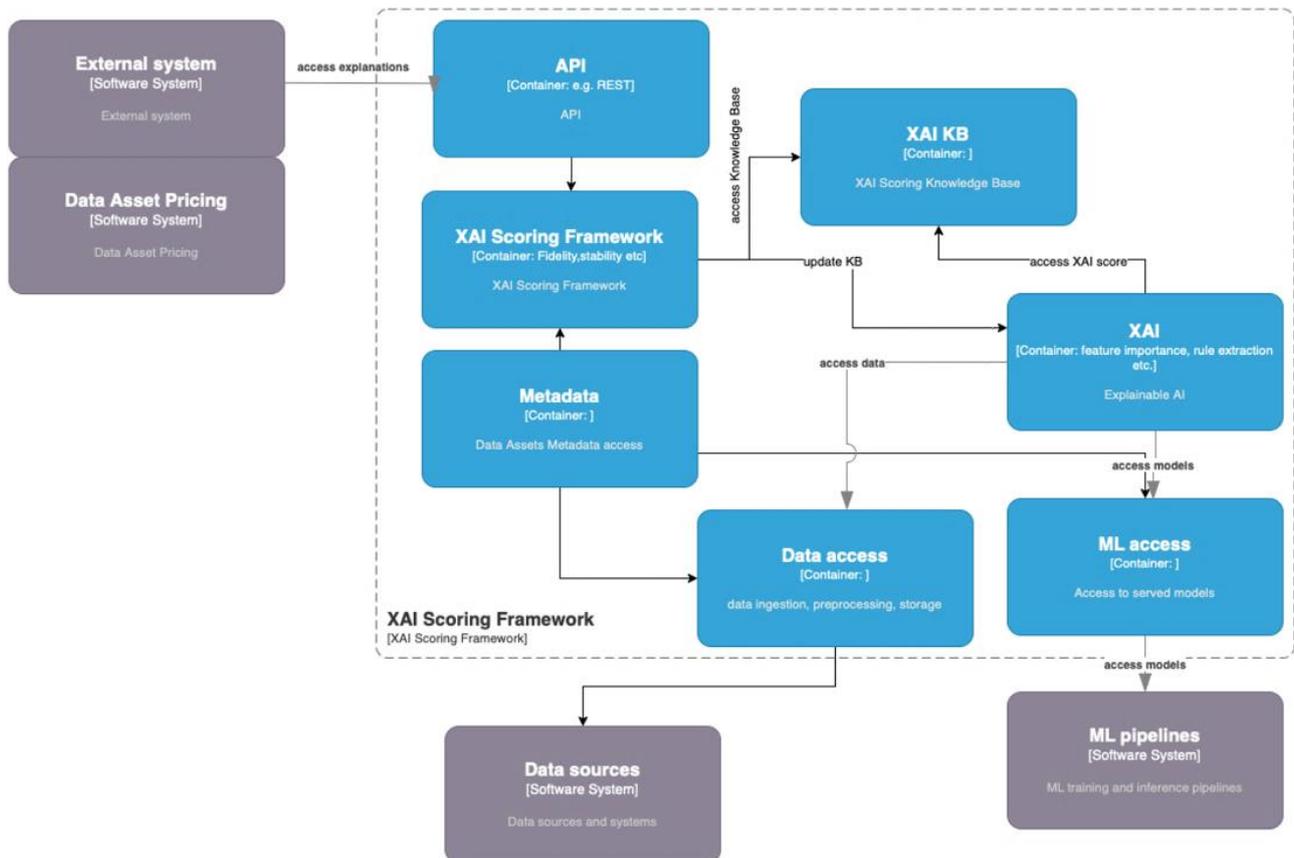


Figure 40 – XAI Scoring Framework C4 component-level architecture

As shown in Figure 40, the *XAI Scoring* framework offers a dynamic online implementation. Its methodology evolves 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, evaluates 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.

For evaluating and delivering explainability scores, a robust REST API has been designed and implemented. This API serves as the backbone for the FAME *XAI Scoring* framework, supporting a wide range of functionalities essential for assessing the explainability of AI models and datasets, referring to the ones of:

- **Calculation Endpoints:** The API includes dedicated endpoints for computing various explainability metrics, such as fidelity, stability, simplicity, coverage. These endpoints are designed to accept parameters that specify the model and dataset under evaluation.
- **Result Delivery:** Upon computation completion, the API delivers detailed results to the user. Results are provided in structured JSON format, ensuring compatibility with various client applications and ease of integration with other systems.
- **Inference Endpoint:** The API utilizes metadata from AI/ML models and datasets to evaluate properties like fidelity, consistency, and comprehensibility. This is managed on an ML model that is trained on the dataset that is the output of the benchmarking process.
- **Benchmarking Support:** Users can benchmark the explainability scores of various XAI techniques against established benchmarks, ensuring continuous improvement and validation of model explainability.

<b>Detailed deliverable:</b>	D5.1 - Trusted and Explainable AI Techniques I
------------------------------	--

### 8.3.13 Incremental & Energy Efficient Analytics

The *Incremental & Energy Efficient (EE) Analytics* component is responsible to act as the enabler to provide incremental analytics to the *ML & AI Analytics* component developed or used within FAME, to be performed in an energy efficient manner. It is based on the LeanXscale database [115], 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 LeanXscale database as a relational datastore and perform complex analytical query processing by submitting SQL compatible query statements to retrieve the resulted dataset. In addition, with the NoSQL interface, the data user can have access to a set of advanced capabilities, like the ability for data ingestion at very high rates, taking advantage of its internal novel indexing mechanism. Last but not least, this component includes the ability to read or retrieve data in an incremental manner, thus, getting informed of updated result sets, while data is being ingested at the same time, ensuring database transactional semantics at the same time. Figure 41 illustrates the internal elements of this component as part of the 3<sup>rd</sup> level of its C4 diagram (i.e., component diagram).

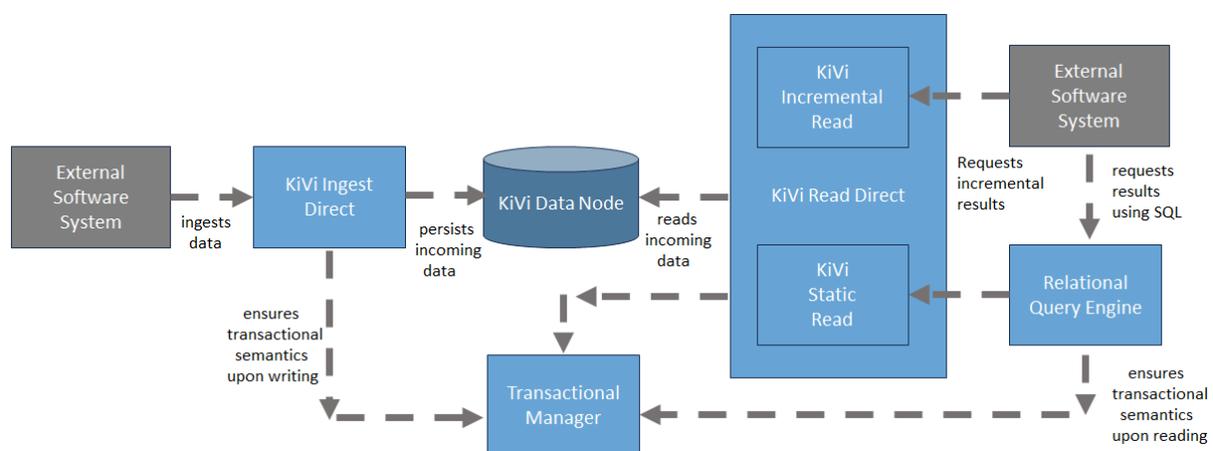


Figure 41 – Incremental Analytics C4 component-level architecture

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 Figure 41, and the data consumer that is the software components that reads data from the database and are suited at the right part of Figure 41. The data provider communicates with the system using the NoSQL interface, which is depicted in Figure 41 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 [116] 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 have 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. This means that data user can submit queries that involve analytical query operations (such as min, max, sum, count, average), whose result are being calculated incrementally as data is being modified at the same time. In the second period of the project, this capability is now offered also via standard SQL, that is the end-user can make use of the standard JDBC interface to exploit the incremental analytics offering of the database.

Finally, the Transactional Manager is an important sub-component of the overall architecture. The *Incremental & EE 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. It is important to highlight that the Transactional Manager plays a crucial role in the *Incremental & EE Analytics* component, as the latter can provide incrementally calculated analytical results, while the data is being modified at the same time, ensuring though the consistency of the results as the transactions are forced by the former

<b>Detailed deliverable:</b>	D5.2 – Energy Efficient Analytics Toolbox I
------------------------------	---

#### 8.3.14 Smart Deployment

The *Smart Deployment* component focuses on providing pre-defined service pipelines to FAME end-users. This component gathers the requirements of a number of different EmFi applications (mainly based on the requirements of the FAME pilots) and defines and publishes deployment templates to enable these services via the FAME federated Data Space.

The data pipelines consist of ML/AI analytics trained with common ML frameworks (i.e., TensorFlow, Scikit-Learn, or Pytorch) and deployed on Docker containers available as a service or as an asset. These containers can be deployed in Kubernetes and handled by Kubeflow. When deployed as a service, the end-user will be able to have full control of the deployment. On the contrary, when deployed as a service, these pipelines are scaled to meet the application requirements and optimised to reduce CO<sub>2</sub> emissions. This is done using the CO<sub>2</sub> metrics extracted by the *CO<sub>2</sub> Monitoring* component, so that each pipeline has an associated consumption profile. New, more optimised versions of these pipelines are generated and made available to the FAME end-users.

Figure 42 shows the C4 component-level architecture, where the components used to implement the abovementioned functionalities can be observed. These blocks are:

- **Pipeline Profiler:** This block updates (or generates if there are no previous versions) the profile associated with the pipeline of a service, focusing on the metrics related to CO<sub>2</sub> emissions collected by the *CO<sub>2</sub> Monitoring* component.
- **Pipeline Generator:** Considering the predefined requirements of a service pipeline and the emissions associated with previous releases, this block generates a new, more optimal service deployment template and indexes it into the *FDAC*.

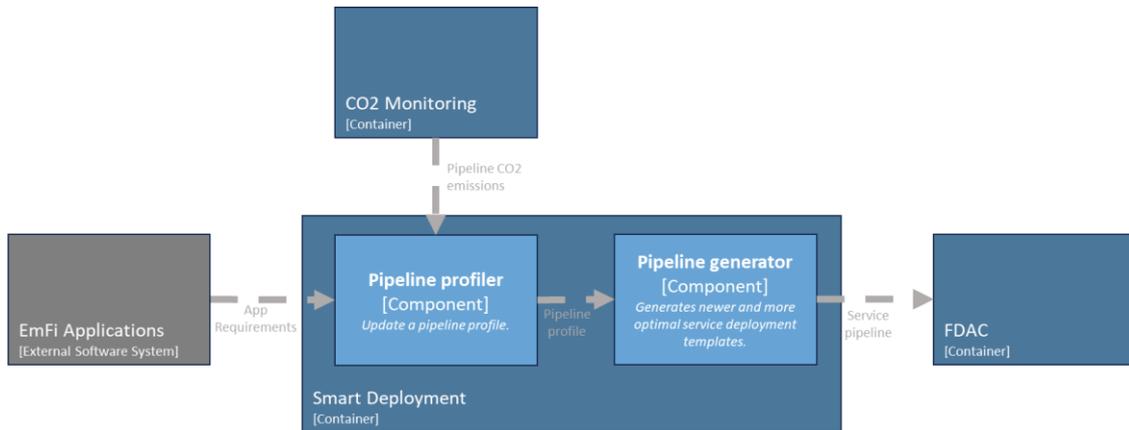


Figure 42 – Smart Deployment C4 component-level architecture.

<b>Detailed deliverable:</b>	D5.2 - Energy Efficient Analytics Toolbox I
------------------------------	---

### 8.3.15 Analytics CO<sub>2</sub> Monitoring

The *Analytics CO<sub>2</sub> Monitoring* component aims to estimate the CO<sub>2</sub> emissions of the ML/AI models used in a FAME application from its CPU/GPU usage. These CO<sub>2</sub>-related metrics are later used to deploy ML/AI models in a more efficiently manner in terms of energy consumption, thereby reducing the CO<sub>2</sub> footprint.

Figure 43 shows the C4 component-level architecture that depicts the Requirements' Analysis sub-component of the *Analytics CO<sub>2</sub> Monitoring* component. This sub-component gathers the requirements information from the EmFi applications and generates the metadata/profile associated with 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.

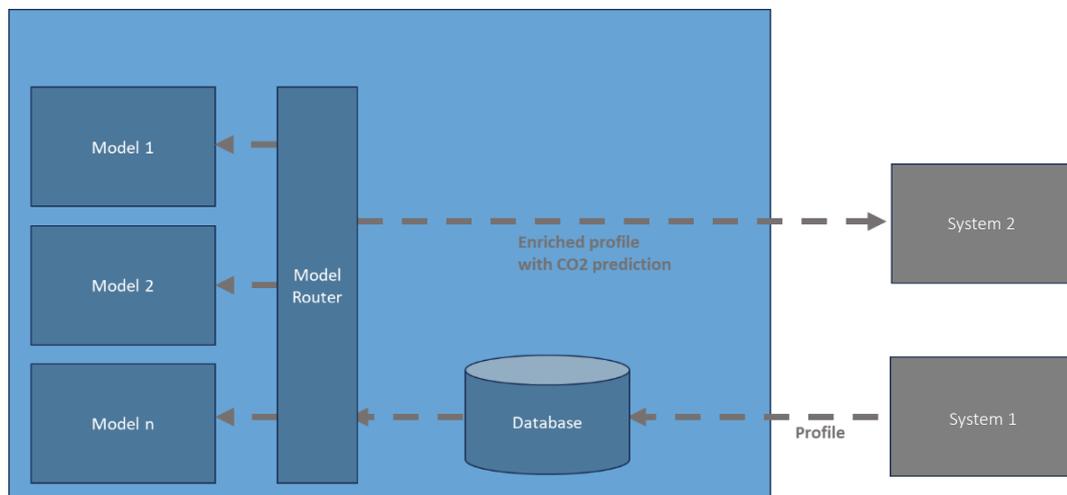


Figure 43 – Analytics CO<sub>2</sub> Monitoring - Requirements' Analysis C4 component-level architecture

Currently, for executing this part of the component, a custom-made mini-training infrastructure is used. This is necessary to generate training data to learn the AI/ML models that can predict CO<sub>2</sub> consumption. By executing various experiments, the component's used database is populated with examples (i.e., profiles/metadata). Based on these examples, a set of models will derive and learn, which will be used in the final version of the component.

The C4 component-level architecture regarding the CO<sub>2</sub> Footprint Analyzer sub-component of the *Analytics CO<sub>2</sub> Monitoring* component is shown in Figure 44, 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<sub>2</sub> 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<sub>2</sub> g/KWh emissions of the device's location, the ML/AI model's CO<sub>2</sub> footprint is computed. A visualization of these results will be plotted through the provided component to raise awareness of the environmental impact of the running process (this is an optional feature within this component).

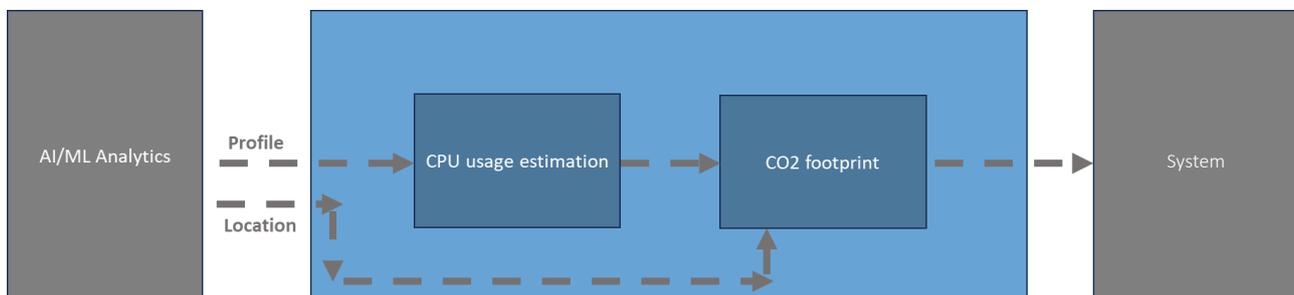


Figure 44 – Analytics CO<sub>2</sub> Monitoring - CO<sub>2</sub> Footprint Analyzer C4 component-level architecture

<b>Detailed deliverable:</b>	D5.2 - Energy Efficient Analytics Toolbox I
------------------------------	---

### 8.3.16 FML Deployment

The *FML Deployment* in FAME is designed to provide privacy-preserving and energy-efficient training of ML/AI analytics in a federated fashion. Clients can retain ownership of the locally gathered data to ensure its privacy, whereas they only communicate the model updates to the server over a secure channel. The server aggregates these updates into a global model, which is later broadcasted to the clients. These components are containerized, so that the FAME end-users will be able to run them on the provider's cloud (i.e., analytics as a service) or download and execute them on either their own infrastructure or any other cloud provider of their choice (i.e., analytics as an asset).

Figure 45 depicts the C4 component-level architecture of the current version of the *FML Deployment*. The FML server aggregates model updates from the clients and broadcasts the aggregated model back to the clients after each round. Once the training is done, the FML server indexes the resulting model into the *FDAC* and stores the model in the repository. Conversely, the FML client receives the model from the server, trains it with the data stored in a local database and sends the trained model back to the FML server. The communication between the FML server and FML clients is mediated by the FML orchestrator, which provides an asynchronous communication channel and client discovery mechanism. The discovery mechanism enables clients to asynchronously subscribe to the server at any point in time of the federated training process, also enabling fault tolerance against idle clients or unexpected breakdowns.

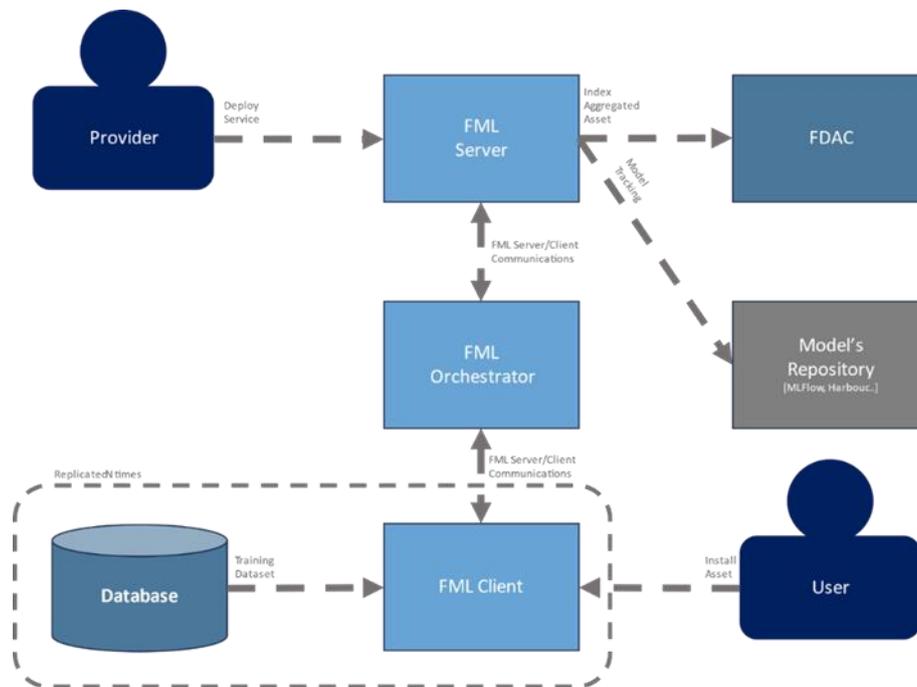


Figure 45 – FML Deployment C4 component-level architecture

The *FML Deployment* is based on a proprietary fully extensible federated learning framework developed by EVIDEN, which implements the FL functionalities mentioned above on top of existing ML frameworks such as TensorFlow, Scikit-Learn or Pytorch. This framework refers to FLEVIDEN, which follows the pipes and filters programming paradigm, so that the FL functionalities are implemented in the so-called Pods, whose inputs and outputs can be connected to other Pods to obtain the desired high-level behaviour. As a result, FLEVIDEN provides high versatility and flexibility to the users, allowing them to build powerful, robust, and scalable Federated Learning solutions.

The clearest example of this flexibility is the range of supported architectures in the aforementioned framework. Depending on the user's needs, the server, client, and orchestrator components in Figure 45 can be arranged in the traditional client-server architecture, but also in the hierarchical and swarm ones. As its name implies, in the hierarchical architecture, servers can act as clients for other servers. Thus, the servers located at the centre of the hierarchy split the FL into different domains, also known as silos. On the contrary, the swarm architecture represents a flat hierarchy in which every node acts as client and the server is rotated among them. Every component will be provided as a Docker image that the user can execute locally (i.e., asset) or in a Kubernetes cluster in the cloud (i.e., service).

<b>Detailed deliverable:</b>	D5.2 - Energy Efficient Analytics Toolbox I
------------------------------	---

### 8.3.17 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* offers 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 46 describes the C4 component-level architecture of the *Learning Centre*, which is composed by two (2) core components, referring to the Course Player and Courses Backend to handle training

content produced by FAME, which is backed by the *FDAC* for the indexing of external training resources and for allowing the linking between training resources and assets.

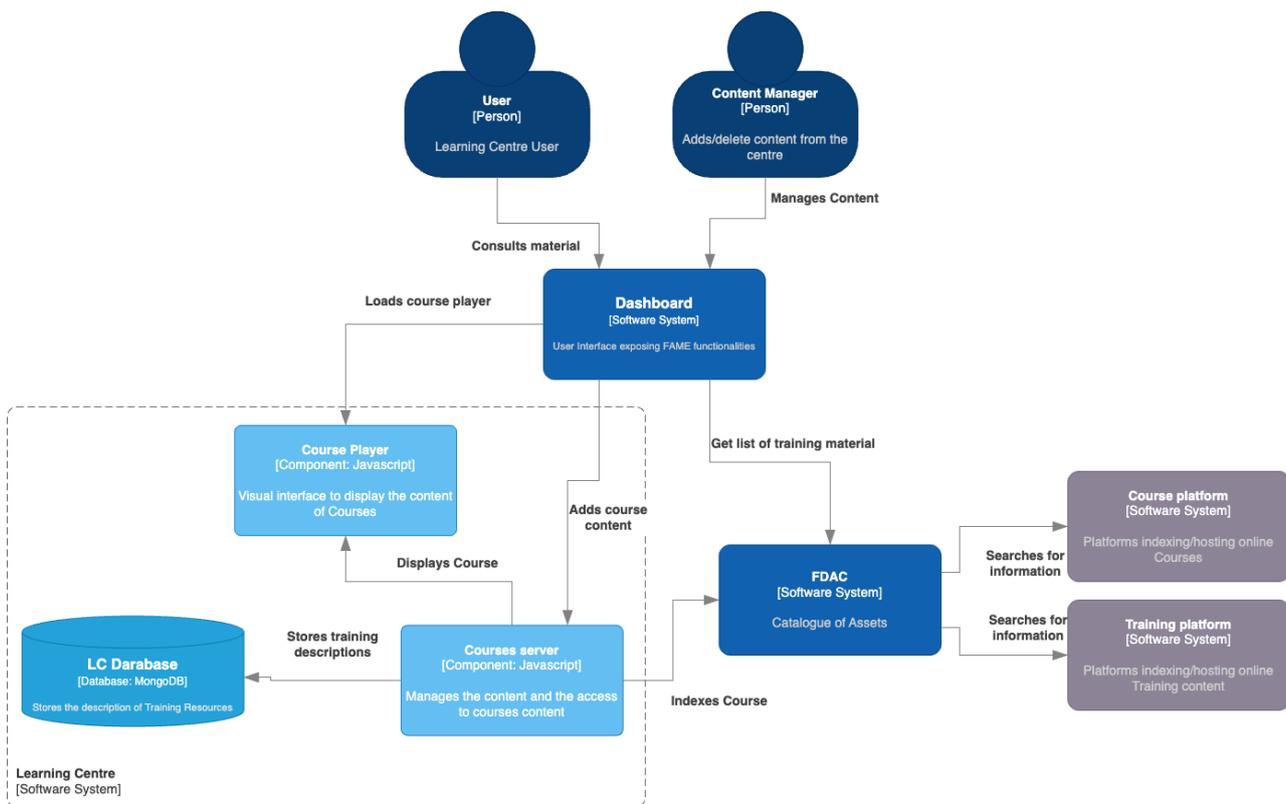


Figure 46 – Learning Centre C4 component-level architecture

The *Learning Centre* supports two (2) diverse types of users, referring to the end-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*, where end-users can access the content and the functionalities of the *Learning Centre* through the FAME Stakeholders UI (residing with the FAME Dashboard). This includes a set of UIs that allow users to navigate across the training material created in FAME as well as training material created by external stakeholders and identify how they are related with existing assets. Also, interfaces are provided for collecting inputs from the end-users providing such interfaces for the different training resources (such as courses, tutorials, webinars, etc.) or services provided in Virtual Digital Innovation Hubs (VDIHs). Moreover, the *Learning Centre* provides a dedicated Course Player to visualize the training content, providing an easy to navigate interface with progress history capabilities to allow end-users to keep their reading progress across multiple sessions.

As for the *Learning Centre* Backend, it provides support for the functionalities of the abovementioned functionalities, 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 *FDAC* 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 [117]. Information in such platforms is subjected to be updated overtime, where when 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 *FDAC* is used along with dedicated resolvers to continuously collect and update the relevant information. These resolvers are 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 is collected from these external platforms. To this end, it should be noted that only curated training resources and VDIH services are expected to be indexed in the *FDAC*, requiring human selection. 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 data assets that the *FDAC* end-user is exploring.

The core functionalities of the *Learning Centre* were developed during the Horizon 2020 STAR project [118]. In the context of FAME, the *Learning Centre* has been extended, specifically referring to the implementation of the APIs needed to achieve the integration with the *FAME Dashboard* and the *FDAC*.

<b>Detailed deliverable:</b>	D7.4 - Training Programs and Learning Centre
------------------------------	--

### 8.3.18 Dashboard

The *Dashboard* component provides an integrated user-friendly end-to-end UI facilitating the interaction of the FAME stakeholders with all the involved FAME backend services, components, and processes. Through this component, various types of stakeholders can have access to the FAME Data Marketplace, where in the cases that an external stakeholder may be an end-user representing either a data provider, or a data consumer intending to connect and interact with FAME, he/she has access to the *Stakeholders UI*. Otherwise, in the cases that the external stakeholder is a FAME administrator needing to access FAME for administration purposes or for any system parameterization/maintenance, he/she is navigated through the *Administrator UI* that offers a related UI for such purposes.

Since the *Dashboard* in essence is a frontend component not including any internal architecture, its overall design and its included UIs are depicted through the respective visual sitemaps that have been drafted both from the end-user point of view (Figure 47) and from the platform administrator point of view (Figure 48).

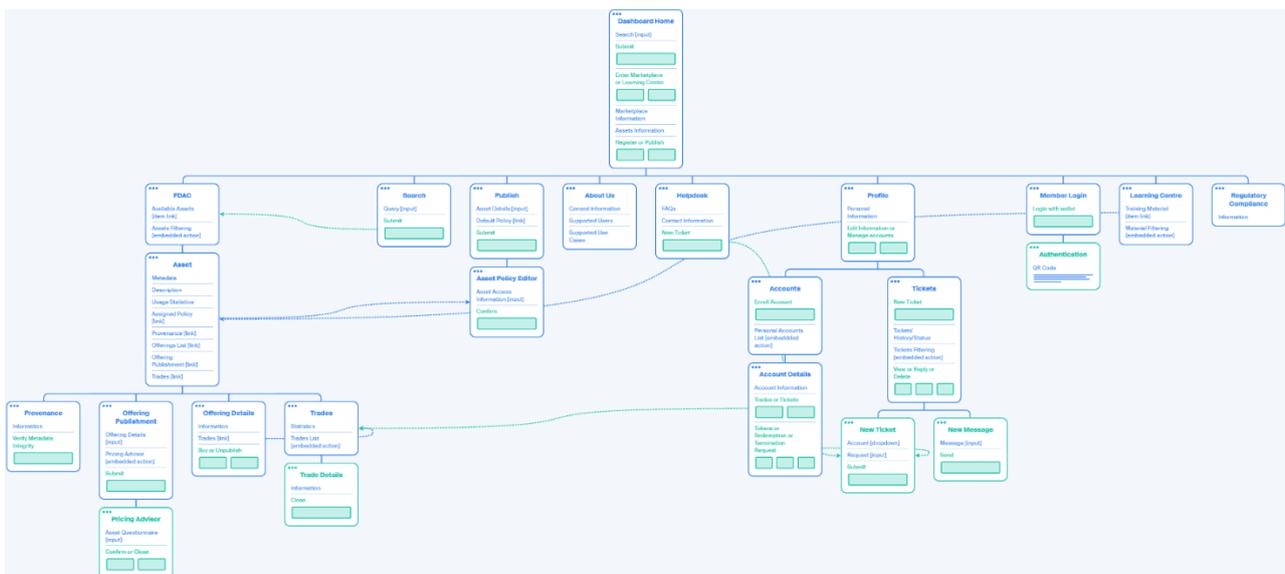


Figure 47 – Dashboard end-users' visual sitemap

As for the end-users Visual Sitemap (Figure 47), it can be seen that the landing page of the end-user is the Dashboard Home from which the end-user can navigate to the different 1<sup>st</sup> level backend services/functionalities of the FAME Data Marketplace, referring to the FDAC, the Search Engine, the Asset Publishment, the Learning Centre, the Regulatory Compliance Tool, the Login (Authentication/Authorization) and the Profile ones, which are directly accessible either from the Dashboard Home menu bar or the provided call-for-action buttons residing within the Home page. The rest of the backend services (i.e., Asset Policy Editor, Asset Provenance, Asset Trade, Asset Pricing) are also accessible through the *Dashboard*, with the difference that they can be reached through the respective internal pages of the 1<sup>st</sup> level services, since their functionalities are implemented as part of those 1<sup>st</sup> level services' procedures. Also, through the Dashboard Home, the end-users can access the FAME platform general information pages (i.e., About Us, Helpdesk) for locating details upon the platform, its overall scope, supported functionalities, etc. To this end, it should be noted that Figure 47 also depicts the actions that can be performed to each page from the end-user side, as well as the connections among the different pages that interact with the underlying backend services.

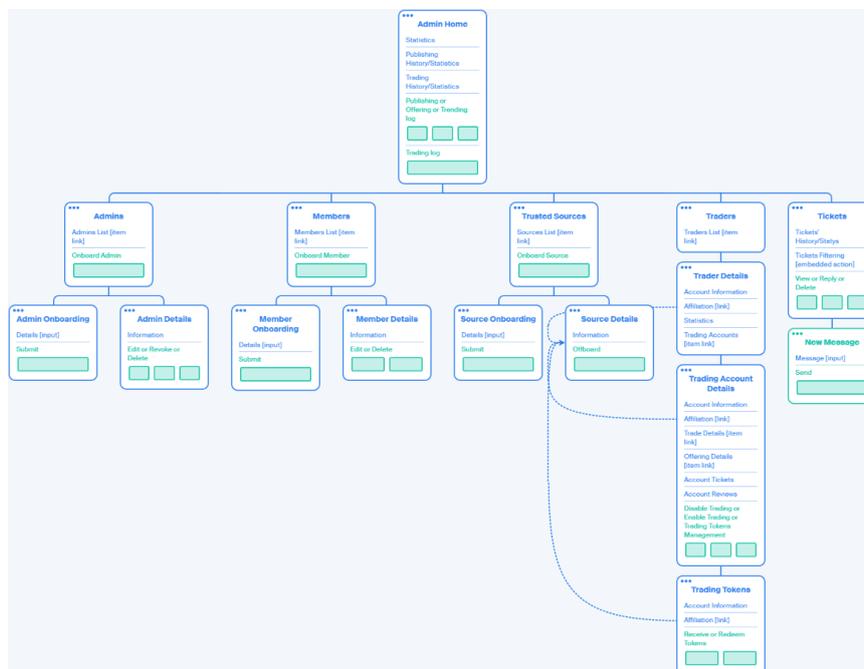


Figure 48 – Dashboard administrators' visual sitemap

As for the administrators' Visual Sitemap (Figure 48), it can be observed that the landing page of the administrator is the Admin Home from which the administrator can monitor the platform general statistics (e.g., federation members, trusted sources, catalogue entries) and navigate to the different backend services/functionalities of the administrative part of the FAME Data Marketplace, referring to the Administrators', the Members', the Trusted Sources', and the Traders' management pages, as well as the related Tickets' overview page. To this end, it should be also stated that Figure 48 additionally depicts the actions that can be performed to each page from the administrator side, as well as the interactions among the different pages.

Finally, as for the technologies exploited for realizing the complete *Dashboard* environment, these refer to Angular [119] along with the technologies of HTML [120], SCSS [121], TypeScript [122], and Angular Material [123], furtherly analyzed in D2.3 - Integrated FAME Data Marketplace I [124].

<b>Detailed deliverable:</b>	D2.3 - Integrated FAME Data Marketplace I
------------------------------	---

## 9 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 reported the final 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 IDS the current advancements as introduced in Section 6, as well as the Data Spaces design guidelines outlined in Section 8. The 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, towards realizing a fully functional Data Marketplace that can be exploited by a plethora of stakeholders (Section 5).

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 by delivering its own Technology Framework (i.e., Data Marketplace), accompanied by its proposed Business as well as Legal Frameworks. To this context, 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 have been also outlined, 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 core step towards the actual implementation and integration of them. All these have led the way towards both making FAME's vision a reality, and also realizing its 1st prototype, ensuring that all the set milestones and KPIs are fully covered. Until the realization of this prototype, several updates had been performed towards the construction of the final version of the FAME SA, based on the collection of the stakeholders and the components feedback.

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 7 (highlighted with yellow colour).

Table 7 – D2.6 Related KPIs

<b>Obj. ID</b>	<b>KPI ID</b>	<b>Measured KPI</b>	<b>Expected Value [M9]</b>	<b>Achieved Value [M9]</b>	<b>Expected Value [M18]</b>	<b>Achieved Value [M18]</b>
<b>O1</b>	<b>1.1</b>	Standards-based Reference Architectures to be considered	5	8	<b>&gt;=8</b>	<b>10</b>

## References

1. European data strategy, <https://digital-strategy.ec.europa.eu/en/policies/strategy-data>
2. European Data Governance Act, <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act>
3. Data Act, <https://digital-strategy.ec.europa.eu/en/policies/data-act>
4. Common European Data Spaces, <https://digital-strategy.ec.europa.eu/en/policies/data-spaces>
5. Big Data Value Association, <https://bdva.eu/>
6. FAME, D2.5 - Requirements Analysis, Specifications and Co-Creation II (2024).
7. FAME, D2.2 - Technical Specifications and Platform Architecture I (2023).
8. Data Hubs, Data Lakes, and Data Warehouses: How They Are Different and Why They Are Better Together, <https://www.gartner.com/en/documents/3980938>
9. 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.
10. Janssen, N. E. (2022). The Evolution of Data Storage Architectures: Examining the Value of the Data Lakehouse (master's thesis, University of Twente).
11. 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](https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en)
12. Centralized vs. Federated: State Approaches to P-20W Data Systems, [https://nces.ed.gov/programs/slds/pdf/federated\\_centralized\\_print.pdf](https://nces.ed.gov/programs/slds/pdf/federated_centralized_print.pdf)
13. Why Data Mesh is a Must for Federated Data Management, <https://www.k2view.com/blog/federated-data-management>
14. 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/>
15. Data Federation: Foolproof Guide on Data Management, <https://www.g2.com/articles/data-federation>
16. P. Koutroumpis, A. Leiponen, & L. D. W. Thomas (2020). Markets for data. In *Industrial and Corporate Change*, vol. 29, no. 3, pp. 645-660.
17. M. Fruhwirth, M. Rachinger, & E. Prlja (2020). Discovering Business Models of Data Marketplaces. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.
18. L. Meisel & M. Spiekermann (2019). Datenmarktplätze: Plattformen für Datenaustausch und Datenmonetarisierung in der Data Economy. In *ISST Work Report*.
19. M. Spiekermann (2019). Data Marketplaces: Trends and Monetisation of Data Goods. In *Intereconomics*, vol. 54, no. 4, pp. 208-216.
20. Understanding the Value of Reference Architectures, <https://doveltech.com/innovation/understanding-the-value-of-reference-architectures>
21. Bass, L., Clements, P., & Kazman, R. (2003). *Software architecture in practice*. Addison-Wesley Professional.
22. 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.
23. 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.

24. 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.
25. i3-MARKET, <http://i3-market.eu/>
26. i3-MARKET Opensource, [www.open-source.i3-MARKET.eu](http://www.open-source.i3-MARKET.eu)
27. Kruchten, P. B. (1995). The 4+1 view model of architecture. *IEEE software*, 12(6), pp. 42-50.
28. 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.
29. INFINITECH, <https://www.infinitech-h2020.eu/>
30. FIBO: Financial Industry Business Ontology, <https://edmcouncil.org/frameworks/industry-models/fibo>
31. FIGI: FINANCIAL INSTRUMENT GLOBAL IDENTIFIER, <https://www.omg.org/figi/>
32. ESTRELLA, [http://www.estrellaproject.org/?page\\_id=5](http://www.estrellaproject.org/?page_id=5)
33. Introduction to Ethereum Tokens, <https://coindcx.com/blog/cryptocurrency/ethereum-tokens-erc-20-erc-721-erc-1155>
34. HyperLedger Fabric, <https://www.hyperledger.org/projects/fabric>
35. Finsecurity Solutions, <https://finsecurity.eu/>
36. European Cluster for Securing Critical Infrastructures, <https://www.finsec-project.eu/ecsci>
37. EU-CIP, <https://www.eucip.eu/>
38. What is SOA?, <https://aws.amazon.com/what-is/service-oriented-architecture>
39. SecureIoT, <https://secureiot.eu>
40. PolicyCloud, <https://policycloud.eu/>
41. DataVaults, <https://www.datavaults.eu/>
42. OMEGA-X, <https://omega-x.eu/>
43. TRUSTS, <https://www.trusts-data.eu/>
44. GAIA-X, <https://gaia-x.eu/>
45. GAIA-X Federation Services, [www.gxfs.de](http://www.gxfs.de)
46. GXFS Authentication Authorization, <https://www.gxfs.eu/authentication-authorisation>
47. GXFS Organizational Credential Manager, <https://www.gxfs.eu/organizational-credential-manager/>
48. GXFS Personal Credential Manager, <https://www.gxfs.eu/personal-credential-manager/>
49. GXFS Trust Services API, <https://www.gxfs.eu/trust-services-api/>
50. GXFS Data Contract Service, <https://www.gxfs.eu/data-contract-service/>
51. GXFS Data Exchange Logging Service, <https://www.gxfs.eu/data-exchange-logging-service/>
52. GXFS Core Catalogue Features, <https://www.gxfs.eu/core-catalogue-features/>
53. GXFS Continuous Automated Monitoring, <https://www.gxfs.eu/continuous-automated-monitoring/>
54. GXFS Onboarding Accreditation Workflows, <https://www.gxfs.eu/onboarding-accreditation-workflows/>
55. GXFS Notarization API, <https://www.gxfs.eu/notarization-api/>
56. GXFS Tutorial, <https://www.gxfs.eu/portal/>
57. GXFS Orchestration, <https://www.gxfs.eu/orchestration/>
58. GXFS Specifications, <https://www.gxfs.eu/specifications/>
59. Gaia-X Trust Framework, <https://docs.gaia-x.eu/policy-rules-committee/trust-framework/22.10/>
60. Second Phase of the XFSC Specification, <https://www.gxfs.eu/second-phase-of-the-xfsc-specification/>

61. International Data Spaces Associations, <https://internationaldataspaces.org/>
62. International Data Spaces, <https://www.fraunhofer.de/en/research/lighthouse-projects-fraunhofer-initiatives/international-data-spaces.html>
63. 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>
64. 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](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)
65. 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](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)
66. 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](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)
67. 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](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)
68. 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](https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/perspectives-of-the-reference-architecture-model/4_perspectives/4_1_security_perspective)
69. 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](https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/perspectives-of-the-reference-architecture-model/4_perspectives/4_2_certification_perspective)
70. 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](https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/perspectives-of-the-reference-architecture-model/4_perspectives/4_3_governance_perspective)
71. [https://internationaldataspaces.org/?page\\_id=8679](https://internationaldataspaces.org/?page_id=8679)
72. Dataspaces Protocol 2024-1, <https://docs.internationaldataspaces.org/ids-knowledgebase/v/dataspaces-protocol>
73. DSSC, <https://dssc.eu/>
74. Data Spaces Blueprint v1.0, <https://dssc.eu/space/BVE/357073006/Data+Spaces+Blueprint+v1.0>
75. Building Block Overview, <https://dssc.eu/space/BVE/357073899/Building+Block+Overview>
76. FAME, D5.1 - Trusted and Explainable AI Techniques I (2024).
77. H. B. McMahan et al. (2017), “Communication-Efficient Learning of Deep Networks from Decentralized Data”. In *Artificial intelligence and statistics*. PMLR.
78. M. Abadi et al. (2016), “Deep Learning with Differential Privacy”. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308-318.
79. FAME, D7.4 - Training Programs and Learning Centre (2024).
80. C4 Architecture Model, <https://c4model.com/>
81. Unified Modelling Language (UML), <https://www.uml.org/>
82. Kruchten, P. B. (1995). The 4+1 view model of architecture. *IEEE software*, 12(6), pp. 42-50.
83. C4 model for system architecture design, <https://icepanel.io/blog/2022-10-03-c4-model-for-system-architecture-design>
84. C4 Model for Software Architecture, <https://www.infoq.com/articles/C4-architecture-model/>
85. FAME, D4.3 - Operational Models, Business Models and Governance I (2024).

86. TypeScript, <https://www.typescriptlang.org/>
87. Node.js, <https://nodejs.org/>
88. NestJS, <https://nestjs.com/>
89. MySQL, <https://www.mysql.com/>
90. MongoDB, <https://www.mongodb.com/>
91. Solidity, <https://soliditylang.org/>
92. Hyperledger Besu, <https://www.hyperledger.org/projects/besu>
93. Sphereon SSI SDK, <https://sphereon.com/sphereon-ssi-sdk/>
94. OpenID for Verifiable Credentials, <https://openid.net/sg/openid4vc/>
95. OpenID for Verifiable Presentations, [https://openid.net/specs/openid-4-verifiable-presentations-1\\_0.html](https://openid.net/specs/openid-4-verifiable-presentations-1_0.html)
96. OpenID for Verifiable Credential Issuance, [https://openid.net/specs/openid-4-verifiable-credential-issuance-1\\_0.html](https://openid.net/specs/openid-4-verifiable-credential-issuance-1_0.html)
97. Auth0, <https://auth0.com/>
98. Decentralized Identifiers (DIDs) v1.0, <https://www.w3.org/TR/did-core/>
99. Verifiable Credentials Data Model v2.0, <https://www.w3.org/TR/vc-data-model-2.0/>
100. TypeScript, <https://www.typescriptlang.org/>
101. React Native, <https://reactnative.dev/>
102. Sphereon, <https://sphereon.com/>
103. Spring Boot, <https://spring.io/projects/spring-boot>
104. PostgreSQL, <https://www.postgresql.org/>
105. Casbin, <https://casbin.org/>
106. Joinup Licensing Assistant, <https://joinup.ec.europa.eu/collection/eupl/solution/joinup-licensing-assistant/jla-find-and-compare-software-licenses>
107. Hyperledger Besu Ethereum client, <https://besu.hyperledger.org/>
108. ExpressJS, <https://expressjs.com/>
109. EU Vocabularies, <https://op.europa.eu/en/web/eu-vocabularies/dcat-ap>
110. Natural Language Processing with Python's NLTK Package, <https://realpython.com/nltk-nlp-python/>
111. Elasticsearch Python Client, <https://www.elastic.co/guide/en/elasticsearch/client/python-api/current/index.html>
112. TensorFlow, <https://www.tensorflow.org/>
113. PyTorch, <https://pytorch.org/>
114. SciKit Learn, <https://scikit-learn.org/stable/>
115. Jimenez-Peris, R., Burgos-Sancho, D., Ballesteros, F., Patiño-Martinez, M., & Valduriez, P. (2022). Elastic scalable transaction processing in LeanXcale. Information Systems, 108, 102043.
116. JDBC Drivers, <https://www.geeksforgeeks.org/jdbc-drivers/>
117. Udemy, <https://www.udemy.com/>
118. STAR Project, <https://star-ai.eu>
119. Angular, <https://angular.dev/>
120. HTML5, <https://html.com/html5/>
121. SASS, <https://www.npmjs.com/package/sass>
122. TypeScript, <https://www.npmjs.com/package/typescript>
123. Angular Material, <https://v16.material.angular.io/>
124. FAME, D2.3 - Integrated FAME Data Marketplace I (2024).