

Federated decentralized trusted dAta Marketplace for Embedded finance



D6.1 - Use Cases Specification and Pilot Sites Preparation. I

Title	D6.1 - Use Cases Specification and Pilot Sites Preparation. I
Revision Number	1.0
Task reference	T6.1
Lead Beneficiary	UNP
Responsible	Tiago Teixeira
Partners	AL, BPFI, DAEM, INNOV, INSO, JOT, JRC, KM, MC, MOH, NOVA, NOVO, NRS, SFS,
Deliverable Type	OTHER
Dissemination Level	PU
Due Date	2023-10-31 [Month 10]
Delivered Date	2023-12-23
Internal Reviewers	NOVA INNOV
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
0.5	2023-12-15	MC, SFS, DAEM, JRC, MOH, KM, JOT, INNOV, NOVO, NRS, NOVA, BPFI, INSO, AL, GFT	
0.6	2023-12-15	MC, SFS, DAEM, JRC, MOH, KM, JOT, INNOV, NOVO, NRS, NOVA, BPFI, INSO, AL, GFT	
0.7	2023-12-22	MC, SFS, DAEM, JRC, MOH, KM, JOT, INNOV, NOVO, NRS, NOVA, BPFI, INSO, AL, GFT	Updated base on internal review feedback
1.0	2023-12-23	MC, SFS, DAEM, JRC, MOH, KM, JOT, INNOV, NOVO, NRS, NOVA, BPFI, INSO, AL, CAGS, GFT	Final Version

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

Acronyms	Definition
AI	Artificial Intelligence
AML	Anti Money Laundering
API	Application Programming Interface
ATOS	Atos It Solutions And Services Iberia Sl
BOI	Bank of Ireland
BPFI	Banking and Payments Federation Ireland
BR	Business Requirements
CDS	Copernicus Data Store
CMIP5	Coupled Model Intercomparison Project Phase 5
CNAE	National Classification of Economic Activities
DAEM	Dimos Athinaion Epicheirisi Michanografisis
DIAS	Data and Information Access Services
DSC	Distributed Control System
EF	Embedded Finance
ESG	Environmental, social, and corporate governance
EU	European Union
FAME	Federated decentralized trusted dAta Marketplace for Embedded finance
FDAC	Federated Data Assets Catalogue
FML	Federated Machine Learning
GDP	Gross Domestic Product
GDPR	General Data Protection Regulation
GTFS	General Transit Feed Specification
HRP	Hierarchical risk parity (HRP) algorithm
IBM	International Business Machines
ID	Identity
INNOV	Innov-Acts Limited
INSO	Inneurope Initiative S.L.
ISIN	International Securities Identification Number
IT	Information Technology
JRC	Jrc Capital Management Consultancy and Research Gmbh
JSI	Institut Jozef Stefan
JSON	JavaScript Object Notation
KM	KM Cube Anonymi Etaireia Parochis Ependytikon Ypiresion
KYC	Know Your Customer

MC	MC SHARED SERVICES SA
ML	Machine Learning
MOH	Motor Oil (Hellas) Diilistiria Korinthou A.E.
NDA	Non-Disclosure Agreement
NGEU	Next Generation EU
NOVO	Novomatix Idiotiki Kefalaiouschiki Etaireia
NRS	Norsk Regnesentral
NUIG	National University Of Ireland Galway
OCR	Optical Character Recognition
OEE	Overall Equipment Effectiveness
ROI	Return On Investment
SAP	Service Access Point
TB	Terabyte (1,000 gigabytes)
TBD	To Be Determined
UBI	Ubitech Limited
UC	Use Case
UIA	UNIVERSITETET I AGDER
UNP	Unparallel Innovation
VAR	Value at Risk
WP	Workpackage
XAI	eXplainable Artificial Intelligence

Executive Summary

This deliverable is the first of a total of 2 deliverables, the second being D6.4 which will be delivered in M21, and both are part of Task 6.1 "Use Cases Co-Creation and Sites Preparation".

All the work described here is the beginning of the pilots' process at FAME, this will create the foundations so that the pilots can develop their work during the project.

All information described in the document may undergo some changes as the project progresses. In fact, there is some information that is not finalized and should be in the second version of this deliverable.

To monitor the progress of the pilots and help them during the project, a plan was created that consists of 4 steps: Pilots and Use Cases Description Updates, Co-Creation Requirement Collection, Datasets Available and Systems (e.g. APIs) to connect to FAME.

In the first step, Pilots and Use Cases Description Updates, a template was created for pilots to fill in with information about the Pilot and their Use Cases, namely name, description, motivation, value propositions, etc.

In the second step, Co-Creation Requirement Collection, it was work carried out by WP 2 partners, who through Co-creation workshops collected Business Requirements from the pilots. In total from all Pilots, 121 Business Requirements were collected, with Pilot 3 and Pilot 4 missing (which will be included in the next version of the Business requirements).

In the third step, is related to Datasets, which was work carried out by WP 5 partners, who created an Excel with the aim of collecting the datasets for each pilot.

In the fourth step, Systems (e.g. APIs) to connect to FAME, is divided into 2 stages: the first phase, the pilots identified the systems they already have and analysed the data that could be linked to FAME. In a second phase, the pilots identified the datasets or external services, which could be useful to them, for their development in the project.

Table of Contents

1	Introduction.....	5
1.1	Insights from other Tasks and Deliverables	5
1.2	Structure	5
2	FAME Pilots	7
2.1	Pilots.....	7
2.2	Pilot preparation plan	7
2.2.1	Pilot Sites Preparation – Pilot and Use Cases description	8
2.2.2	Business Requirements - Co-Creation Workshop Collection.....	9
2.2.3	Dataset Status	10
2.2.4	Systems to connect to FAME.....	10
3	Pilot Sites Preparation.....	12
3.1	Pilot #1: FaMLy – A powerful financial recommendation engine for families.....	12
3.1.1	Use Cases	12
3.1.2	Business Requirements	15
3.1.3	Datasets Status	18
3.1.4	Systems to connect to FAME.....	19
3.1.5	Next steps and timeline	19
3.2	Pilot #2: Embedding Finance Services in a Personalized Citizen Wallet	20
3.2.1	Use Cases	20
3.2.2	Business Requirements	23
3.2.3	Datasets Status	25
3.2.4	Systems to connect to FAME.....	27
3.2.5	Next steps and timeline	27
3.3	Pilot #3: Personalized Collaborative Intelligence for Enhancing EmFi Services	28
3.3.1	Use Cases	28
3.3.2	Business Requirements	30
3.3.3	Datasets Status	31
3.3.4	Systems to connect to FAME.....	33
3.3.5	Next steps and timeline	34
3.4	Pilot #4: The EU Funds Application Process Made Easy	35
3.4.1	Use Cases	35
3.4.2	Business Requirements	36
3.4.3	Datasets Status	36
3.4.4	Systems to connect to FAME.....	37
3.4.5	Next steps and timeline	37
3.5	Pilot #5: ESG Scorecard Ranking & Sustainable Portfolio Optimisation.....	38
3.5.1	Use Cases	38

3.5.2	Business Requirements	40
3.5.3	Datasets Status	42
3.5.4	Systems to connect to FAME.....	45
3.5.5	Next steps and timeline	46
3.6	Pilot #6: Embedding Climatic Predictions in Property Insurance Products.....	46
3.6.1	Use Cases	46
3.6.2	Business Requirements	48
3.6.3	Datasets Status	51
3.6.4	Systems to connect to FAME.....	52
3.6.5	Next steps and timeline	53
3.7	Pilot #7: Assessing the Quality and Monetary Value of Data Assets	53
3.7.1	Use Cases	54
3.7.2	Business Requirements	55
3.7.3	Datasets Status	57
3.7.4	Systems to connect to FAME.....	58
3.7.5	Next steps and timeline	59
4	Conclusions.....	60
4.1	Next Steps.....	61

List of Figures

Figure 1 - Plan for the Pilots	8
Figure 2 - Miro boards business requirements example	9
Figure 3 - Excel: Dataset Collection	10
Figure 4 – Systems (e.g. APIs) to connect to FAME	11
Figure 5 - Pilot #2: Proposed Architecture	27
Figure 6 - Proposed Model 1: Use Case #1	34
Figure 7 - Proposed Model 1: Use Case #2	34
Figure 8 - Pilot #7 Internal Architecture	58
Figure 9 - Pilot #7 Internal to FAME components	59

List of Tables

Table 1 - Pilots Summary	7
Table 2 - Pilot Template	8
Table 3 - Use Case Template	8
Table 4 - Pilot 1: Description	12
Table 5 - Pilot 1: Use Case 1	12
Table 6 - Pilot 1: Use Case 2	14
Table 7 - Pilot 1: Business Requirements	15
Table 8 - Datasets Status: Pilot #1 - Use Case #1	18
Table 9 - Datasets Status: Pilot #1 - Use Case #2	18

Table 10 - Pilot 2: Description	20
Table 11 - Pilot 2: Use Case 1	20
Table 12 - Pilot 2: Use Case 2	22
Table 13 - Pilot 2: Business Requirements	24
Table 14 - Datasets Status: Pilot #2 - Use Case #1	25
Table 15 - Datasets Status: Pilot #2 - Use Case #2	26
Table 16 - Pilot 3: Description	28
Table 17 - Pilot 3: Use Case 1	28
Table 18 - Pilot 3: Use Case 2	29
Table 19 - Pilot 3: Business Requirements	31
Table 20 - Datasets Status: Pilot #3 - Use Case #1	31
Table 21 - Datasets Status: Pilot #3 - Use Case #2	32
Table 22 - Pilot 4: Description	35
Table 23 - Pilot 4: Use Case 1	35
Table 24 - Datasets Status: Pilot #4 - Use Case #1	36
Table 25 - Pilot 5: Description	38
Table 26 - Pilot 5: Use Case 1	38
Table 27 - Pilot 5: Use Case 2	39
Table 28 - Pilot 5: Business Requirements	40
Table 29 - Datasets Status: Pilot #5 - Use Case #1	42
Table 30 - Datasets Status: Pilot #5 - Use Case #2	44
Table 31 - Pilot 6: Description	46
Table 32 - Pilot 6: Use Case 1	46
Table 33 - Pilot 6: Use Case 2	47
Table 34 - Pilot 6: Use Case 3	47
Table 35 - Pilot 6: Business Requirements	48
Table 36 - Datasets Status: Pilot #6 - Use Case #1	51
Table 37 - Datasets Status: Pilot #6 - Use Case #2	51
Table 38 - Datasets Status: Pilot #6 - Use Case #3	52
Table 39 - Pilot 7: Description	53
Table 40 - Pilot 7: Use Case 1	54
Table 41 - Pilot 7: Use Case 2	54
Table 42 - Pilot 7: Business Requirements	55
Table 43 - Datasets Status: Pilot #7 - Use Case #1	57
Table 44 - Datasets Status: Pilot #7 - Use Case #2	57

1 Introduction

This deliverable is part of WP 6 - “Integration, Validation and Evaluation of EmFi Use Cases” and the general objectives of the WP 6 are:

- To specify in detailed the pilots of the project and to prepare the various sites/partners for their implementation and validation;
- To implement the use cases based on the FAME marketplace i.e., leveraging FAME data assets and tools via Open APIs;
- To evaluate the UCs from a techno-economic and socio-economic perspective, as well as based on the analysis of stakeholders’ feedback.

This deliverable is the first of a total of two deliverables that are part of Task 6.1 “Use Cases Co-Creation and Sites Preparation” – “This task will undertake all preparatory actions towards the integration, deployment, and evaluation of the project’s UCs at the pilot sites. Specifically, it will co-create the use cases through the organization of proper workshops and other co-creation activities. Likewise, it will mobilize, engage and when required train all relevant stakeholders including data providers, data analysts, and non-tech users. Moreover, it will collect and prepare the data assets for integration in the FAME marketplace, and it will undertake preparatory activities for using the FAME platform and its services.”.

1.1 Insights from other Tasks and Deliverables

This Deliverable (D6.1) relates to Deliverable 2.1 “Requirements Analysis, Specifications and Co-Creation” which is part of the Task 2.1 “Requirements, Specifications and Co-Creation”, with regard to the requirements that were analyzed, which will be necessary for the pilots’ execution plan.

Moreover, this document is also related to other Work Packages and in particular with the WP 5 and the Task 5.1 “Catalogue of AI/ML Techniques for EmFi” - This task will specify, implement, and make available in the marketplace a library of AI/ML techniques for EmFi use cases (focus of ATOS). The library will comprise classical ML techniques and most popular deep learning techniques, which will serve as a basis for the implementation of the analytical components/modules of the following tasks. Hence, it will comprise models that are appropriate for supporting incremental analytics, as well as models that can be explained based on the FAME XAI techniques (focus of IBM). ML assets to be used in the use case will be specified as well (focus of MOH, UNP). In this Task, Pilot Datasets will be collected, which will be important for both WP 5 and WP 6.

All the work carried out in WP 6 is related with this deliverable, D6.1 being the base point for all the work of the pilots throughout the project. With this, Task 6.2, Task 6.3 and Taks 6.4, reaffirm the implementation of each pilot and to finalize all the work, Task 6.5 evaluates the entire pilot process, and the impact of the pilots' work during the project.

1.2 Structure

This document is divided into 4 main chapters which are the follows:

- **Chapter 1 – Introduction:** This chapter introduces the deliverable highlighting its objective and its relation to other deliverables.
- **Chapter 2 – FAME Pilots:** In this chapter is a small summary of all the pilots who are part of the project.

- **Chapter 3 – Pilots Sites Preparation:** In this chapter is a description of the Pilot and the Uses Cases, the Business Requirements, the Dataset Status, the Systems to connect to the FAME, and the Next Steps and timeline.
- **Chapter 4 – Conclusion:** This chapter concludes the deliverable.

2 FAME Pilots

2.1 Pilots

In the table below, you will find a small summary of all the pilots who are part of the project, with the leader, name, and Use Cases.

Table 1 - Pilots Summary

Pilot #	Leader	Name	Use Cases
Pilot #1	MC	FaMLy – A powerful financial recommendation engine for families	<ul style="list-style-type: none"> • UC #1 – Recommendation Engine for Financial Decisions in MCs Business Ecosystem • UC #2 – Consumers’ Incentives and Interfaces
Pilot #2	DAEM	Embedding Finance Services in a Personalized Citizen Wallet	<ul style="list-style-type: none"> • UC #1 – Personalized Payment and Loyalty Programs for Smart Parking • UC #2 – Extended Payments Across Multiple Services
Pilot #3	BPFI	Personalized Collaborative Intelligence for Enhancing EmFi Services	<ul style="list-style-type: none"> • UC#1 – Pay Facts-as-a-Service for Embedded Payments • UC#2 – Anti Money Laundering (AML) as a Service
Pilot #4	INSO	The EU Funds Application Process Made Easy	<ul style="list-style-type: none"> • UC#1 – The EU Funds Application Process Made Easy
Pilot #5	KM	ESG Scorecard Ranking & Sustainable Portfolio Optimisation	<ul style="list-style-type: none"> • UC#1 –ESG On-line Reporting • UC#2 – Portfolio Optimization
Pilot #6	NRS	Embedding Climatic Predictions in Property Insurance Products	<ul style="list-style-type: none"> • UC#1 – Climate Aware Real Estate Pricing • UC#2 – VaR of an Insurers’ Portfolio • UC#3 – Climate-Aware Analysis of Alternative Portfolios
Pilot #7	MOH	Assessing the Quality and Monetary Value of Data Assets	<ul style="list-style-type: none"> • UC#1 – Quality Assessment and Pricing of Industrial Data Assets • UC#2 – Data Analysis and Production of Value-Added assets using FAME Tools

2.2 Pilot preparation plan

During this first period of the project, a plan was defined to harmonize the progress of the pilots. This plan consists of 4 steps, as you can see in the Figure 1, which are:

1. Pilots and Use Cases Description Updates
2. Co-Creation Requirement Collection (WP 2)
3. Datasets Available (WP 5)
4. Systems (e.g. APIs) to connect to FAME

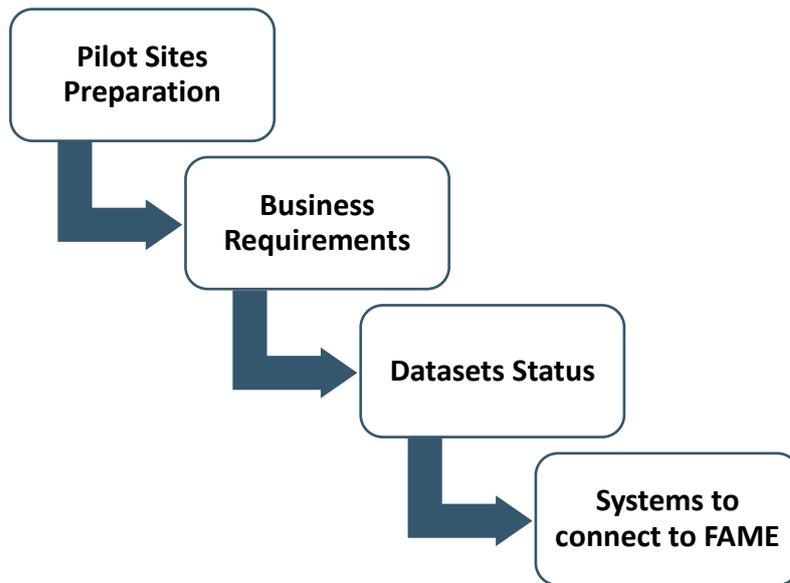


Figure 1 - Plan for the Pilots

All these steps are defined in the following subchapters and will be detailed and specified for each pilot in the following chapter.

2.2.1 Pilot Sites Preparation – Pilot and Use Cases description

So that we could collect information about the pilots and their Use Cases, we created a template with the type of information we needed to be filled out. This template is divided into two sections, the first (Table 2) dedicated exclusively to the pilot, where the pilot's name, a summary, a description and their motivation are requested. The second section is dedicated to the Use case, which asks for the pilot's name, a summary, a description, value proposition(s), documentation, image(s), website, keywords and available data assets. The use cases section must be repeated depending on the number of use cases in the pilot.

Table 2 - Pilot Template

Pilot #
Pilot Name
Summary
Description
Motivation

Table 3 - Use Case Template

Use Case #
Use Case Name
Summary
Description
Value Proposition(s)
Documentation
Image(s)
Website
Keywords

Available Data Assets

2.2.2 Business Requirements - Co-Creation Workshop Collection

To make it possible to survey the business requirements, online collaboration tools were used that will ensure the successful execution of the collaborative and structured process required for the co-creation workshops in a productive and efficient manner. For the Specific Requirements of the Pilot, Co-creation Workshops were used, focusing on the use of User Stories. Each demonstrator participated in a co-creation workshop where both the demonstrator and technical partners collaboratively analysed business requirements and formulated corresponding technical requirements. A total of 38 Generic Requirements, which are intended to be supported by the FAME federated asset space. These requirements encompass both functional and non-functional aspects and originate mainly from the FAME Description of Action, as well as functionalities existing in established marketplaces external to the project. These requirements may be updated or refined as the project progresses.

Regarding the business requirements elicitation process, the consortium decided to utilise the online collaboration platform Miro (www.miro.com). Miro is an online collaborative whiteboard platform. Miro empowers remote, in-office, and hybrid teams to communicate and collaborate across formats, tools, channels, and time-zones without the constraints of physical location, meeting space, and whiteboards.

The figure below represents an example of the Miro board with collected information. The outcomes of these workshops were provided as input to the next series of workshops that were dedicated the technical requirements extraction.

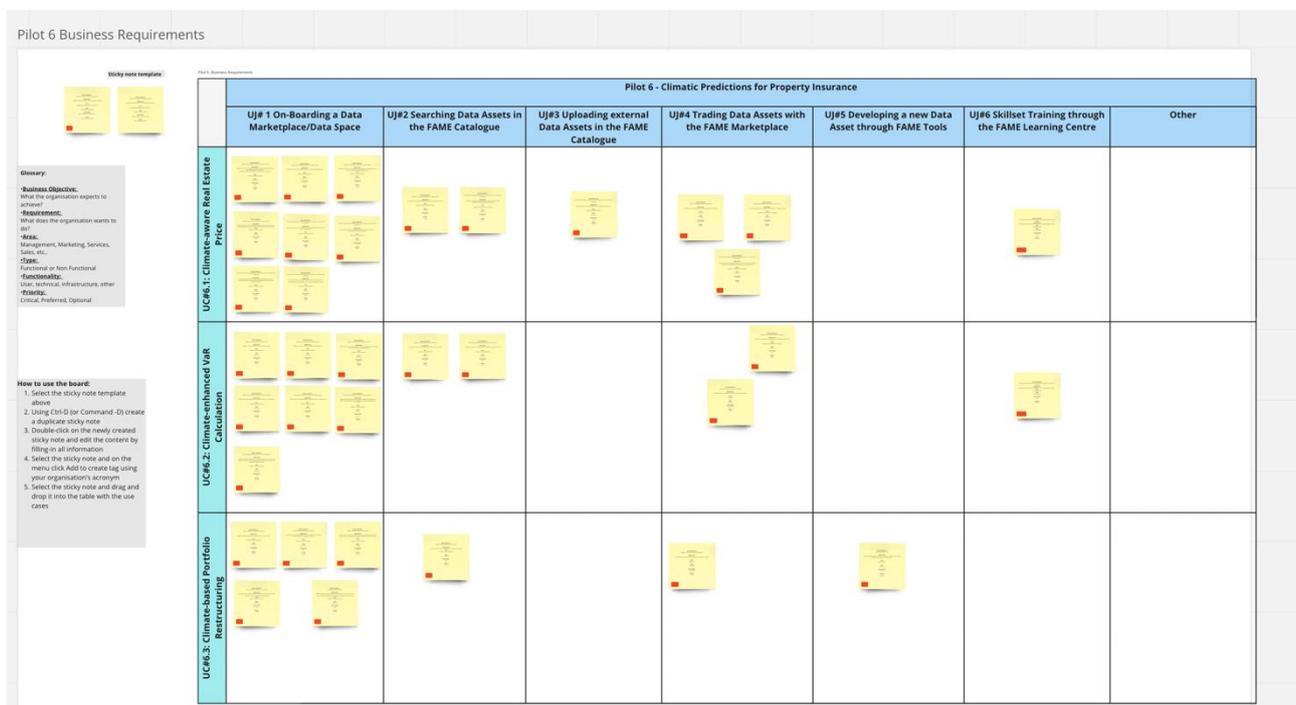


Figure 2 - Miro boards business requirements example

In chapter 3, which is dedicated to each pilot, there is a small summary of each one and in D2.1 there is more detailed information.

2.2.3 Dataset Status

To collect the datasets, WP 5 created an excel to request information from pilots. This work of collecting the datasets came from WP 5, because they needed to know what data the pilots have, due to the algorithms

Pilot ID	Use Case	Pilot lead	Technical Contact	Name	Category	Data Type	Input Data	Expected Output	Type of analytics to implement	Input source	Output storage	Processing location	Confidentiality
P1	UC1	SONAE	Adriana Ramos (adramos@fame.pt)	Recommendation Engine for Financial Decisions in MCI Business Ecosystem	FAMEy – A powerful financial recommendation engine for families	Tabular Text Timeseries	Client behaviour (transactions, digital channels) Client sociodemographics Client queries and reviews ~100 million rows	Ranking	Machine Learning	Inside pilot's premises	Inside pilot's premises	Cloud	Yes, such as client data
P1	UC1	SONAE	Adriana Ramos (adramos@fame.pt)	Consumers' Incentives and Interfaces	FAMEy – A powerful financial recommendation engine for families	Tabular Text Timeseries	client transactions, store, district	Customers data and transactions	IA, Machine Learning	Inside pilot's premises	Inside pilot's premises	Cloud	yes, fiscal num phone number
P2	UC1			Personalized Payment and Loyalty Programs for Smart Parking	Embedding Finance Services in a Personalized Citizen Wallet	Spatial Timeseries / Images	Type - Parking Location Data - User ID - Parking Transaction Data - Parking Duration Data - Incident Location Data - Incident Category Data						
P2	UC2	DAEM	Isabel Bourantas (isabel@daem.pt)	Extended Payments Across Multiple Services	Embedding Finance Services in a Personalized Citizen Wallet	Spatial Timeseries / Images	Type - Parking Location Data - User ID - Parking Transaction Data - Parking Duration Data - Incident Location Data - Incident Category Data	Classification & Regression	Potentially XAI, but we will likely use parametric models	Migrated	Available from FAME or Pilot's premises	Anywhere	No
P3	UC1	DAEM	Isabel Bourantas (isabel@daem.pt)	Pay Facts-as-a-Service for Embedded Payments	Personalized Collaborative Intelligence for Enhancing EnFi-Services								
P3	UC1			Anti Money Laundering (AML) as a Service	Personalized Collaborative Intelligence for Enhancing EnFi-Services								
P4	TBO	INCOMIA	Michèle Vavallo (mvavallo@incomiaconsulting.com)	The EU Funds Application Process Made Easy	personalized optimized model allowing a perfect business model (loan cash flow synergy between both stakeholders (bank and granted companies).	Tabular Text Timeseries	(i) General scheme grants data (amount, prefinancing, payment scheme and installments) (ii) Customers' data (company, cashflow, solvency, debt capital), (iii) Financial products' data (loans, leasing, Renting).	Simulated optimized plan for loan intervention (time and quantity)	simulator algorithm	Inside pilot's premises	Inside pilot's premises	anywhere	yes, fiscal num phone number
P5	UC1	KM	lilipaton@voynote.com and helen@kmj.gr	ESG On-line Reporting	ESG Scorecard Ranking & Sustainable Portfolio Optimisation	JRC: Tabular data, Timeseries and csv data KMC: Tabular data, Timeseries, json objects KMC: (company data) I) Tabular list of a ESG ratings for a sample US	JRC: Price (Market) data for any financial instrument (i) Provide ESG table of recommendations, (ii) Develop a ranking system for stocks and/or bonds based on ESG criteria, alongside fundamental and price-based factors.	Analytical models to estimate data quality		Inside Pilot's premises	Available in FAME's data space & Inside Pilot's premises	Anywhere	NO

Figure 3 - Excel: Dataset Collection

2.2.4 Systems to connect to FAME

All pilots will not develop systems from scratch, because they already have data and operate using that data. Therefore, FAME will not start from scratch, but rather with information that the pilots will share.

The first step the pilots had to take was to identify the systems they already use and that provide them with the data they need. They will identify the data that they will make available in FAME, which will take place during the development of WP 6.

The pilots' second step will be to identify datasets or services external to the pilots, which contain data that will be useful and that they would like to use. As FAME is a federation, the pilots intend for FAME to make this data available so that they can have access.

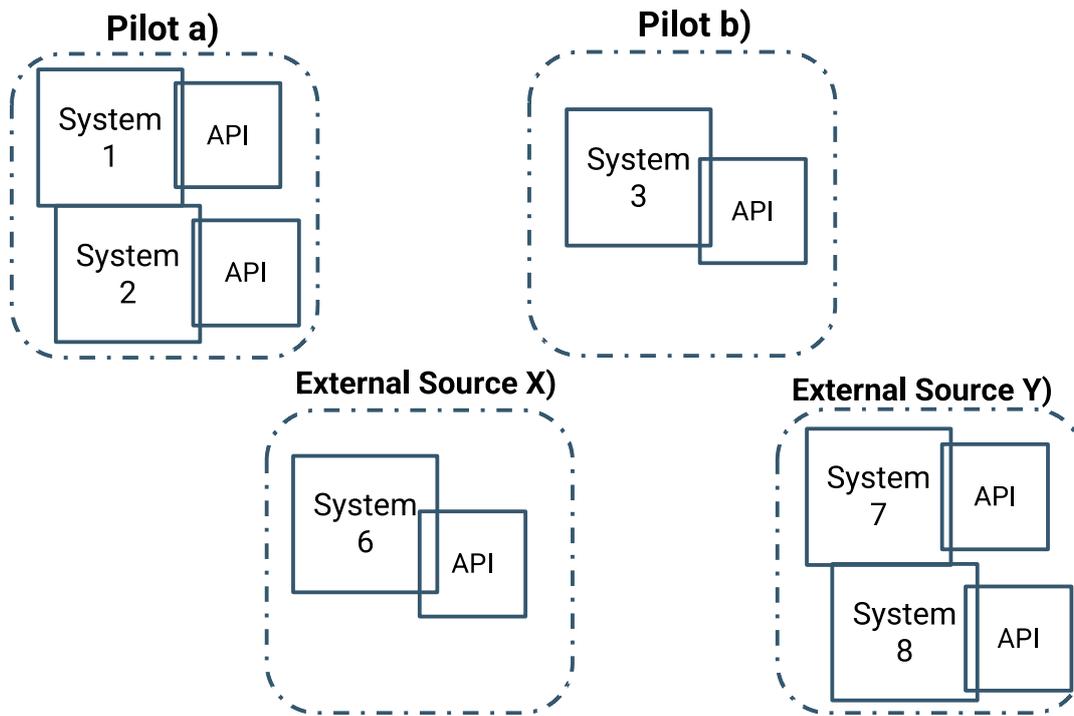


Figure 4 – Systems (e.g. APIs) to connect to FAME

3 Pilot Sites Preparation

All the information described in this chapter was collected in the 1st period of the project, as such, the information may change during the course of the project.

3.1 Pilot #1: FaMLy – A powerful financial recommendation engine for families

Table 4 - Pilot 1: Description

Pilot #1	
Pilot Name	FaMLy – A powerful financial recommendation engine for families
Summary	<p>Using Universo brand, data and expertise, it is aimed to test FAME analytical tools and capabilities by creating:</p> <ul style="list-style-type: none"> - A recommendation engine for financial products and/or other products from Universo partnership's network for families; - Customer profiles and a ML model to address instalment risk. <p>It is also aimed to test FAME marketplace and federation capabilities by sharing and trading in FAME the data assets which will be the outputs from this work.</p>
- Description	<p>MC is a multinational corporation managing a wide portfolio of companies. Universo MC is the group sub-holding for financial services. Started in 2016, transforming the financial management of families and small businesses, innovating through simple financial solutions with positive impact on daily lives.</p> <p>Universo has different financial products under its portfolio, namely, payments solutions, cards, personal credits, insurances, and instalment solutions. Universo is also enriched with a vast partnership network, that promotes better purchasing opportunities for Universo clients.</p> <p>With FAME it is aimed to enlarge client' usage of Universo products by (1) creating a recommendation engine for financial products and/or other products from Universo partnership's network for families and by (2) generating a risk model (based on creation of client profiles) that will enable a wider client access to the Universo instalment solution (Universo Flex).</p>
Motivation	<ul style="list-style-type: none"> - Usage of innovative features of AI, ML and analytics; - Testing data marketplace and federation concept; <p>Product development and innovation.</p>

3.1.1 Use Cases

3.1.1.1 Use Case #1

Table 5 - Pilot 1: Use Case 1

Use Case #1	
Use Case Name	Recommendation Engine for Families
Summary	This use case will aim at developing a robust recommendation engine based on Machine Learning algorithms that will ultimately provide Universo clients with optimal offers and campaigns, by benefiting from

	MC and Universo's partnership network - which encompasses several partners in multiple sectors that address the specific needs of families.
Description	<p>Universo aims at developing a powerful recommendation engine for families, to suggest financial products and other products from Universo partnerships' network by leveraging FAME's analytical capabilities over an extensive pool of data assets originating from Universo's ecosystem. The recommendation engine will provide the user with the offers and campaigns that best suit their needs. Data used for training will include sociodemographical features, as well as transactional and digital behaviour.</p> <p>After training the model, XAI capabilities and other relevant analytical tools will be leveraged to extract valuable insights, guide marketing strategies, and provide transparent explanations to the final client.</p> <p>The model and its outputs will be appraised and negotiated in FAME's marketplace, with the objective of testing the data marketplace concept. Emphasis will be paid on the development of an intuitive way to present the recommendations, as user-friendliness of the application will impact the level of the users' engagement, and their willingness to share data. The recommendations themselves will be integrated into a mobile app, under App Universo, leveraging an already trusted brand and a vast customer base. Engagement will be measured and, to further reinforce data collection and supply, incentives may be provided to end-users. Finally, the incorporation of new partners into a Federation will be validated, as well as retrieving data from other sources, and granting secure access to proprietary assets.</p>
Value Proposition(s)	<ul style="list-style-type: none"> • Generate a powerful recommendation engine for financial and other partnerships under Universo brand • Improve the Customer Experience • Increase sales • Monetize the model and its outputs • Create a Data Federation for future use cases
Documentation	-
Image(s)	-
Website	Universo - A Nova Dimensão Financeira Conta Crédito e Seguros
Keywords	Recommendation engine; Explainable AI; Financial products
Available Data Assets	<p>The data to be utilized in the development of this UC will include information such as:</p> <ul style="list-style-type: none"> • Individual and family profiles, encompassing demographic and socioeconomic data, such as age, income, family size, and geographic location. • Data related to spending history, i.e., detailed information on spending patterns (daily, monthly, or yearly) of families. • Data associated with financial transactions, including transaction types, channels, categories, dates, and locations. • Purchase history, containing information on products or services acquired, merchants and frequency. User feedback, comprising returns and evaluations provided by users on previous recommendations, to adjust and continually enhance the system. <p>All shared data will adhere to compliance and applicable privacy and data security regulation.</p>

3.1.1.2 Use Case #2

Table 6 - Pilot 1: Use Case 2

Use Case #2	
Use Case Name	Consumer interfaces for families
Summary	This use case will generate different customer profiles, based on consumer data, and cross them with instalment risk, leveraging and enlarging the scope of Buy Now Pay Later solutions implementation. It is intended to use analytical tools to support customer profiling and use ML techniques to develop a scoring risk model on instalments availability.
Description	<p>Universo Flex is the solution from Universo that enables the instalment of online shopping. The service materializes in a payment method that merchants can include in their website checkout. After closing his purchase cart and selecting this payment method, the customer is guided through a simplified onboarding experience – fully digital. To this day, it is identified two main onboarding blocks: customer data + bank card data. Customer must provide a document ID photo (capture by the platform) and a valid credit card.</p> <p>Regarding data required for customer onboarding, there are 3 sources that may be used:</p> <ul style="list-style-type: none"> • pre-filled fields as per data shared by merchant (automatic, through API) • captured by OCR, retrieved from customer document ID photo captured in real-time • manually inserted by customer in the interface <p>This data, despite being required for KYC/AML, is also valuable to segment customers in different profiles. The challenge in this use case is to use this customer data generate customer profiles that will feed Universo business.</p> <p>With this first step of profiles generation, it is aimed to develop a risk model to address the enlargement of instalment solution use. The current solution has the limitation of only using credit cards with blocked amounts to reduce internal risk. The objective is to test the impact of opening the risk in some profiles, allowing them to use debit cards and/or not block the purchase full amount. The final goal will be to create a credit risk model, using all possible variables to have the highest accuracy possible. With is aim in mind, it is aimed to use external data sources, included in FDAC, to enrich the training data for the risk model.</p> <p>As a nice-to-have, it is also objective test new approaches to improve customers conversion goals. Some of them will be with a A B test approach – e.g. test screen changes to check impact on customer conversion rate.</p>
Value Proposition(s)	<p>Generate data to define different customer and transactional profiles, that will feed ML models</p> <p>Generate data to create a highly accurate risk model</p>
Documentation	
Image(s)	
Website	Universo Flex Universo - A Nova Dimensão Financeira

Keywords	Customer profiles; Instalment risk
Available Data Assets	<p>The data to share will include information such as:</p> <ul style="list-style-type: none"> • Customer Data: <ul style="list-style-type: none"> ○ Age ○ Gender ○ Location Profession • Payment Data: <ul style="list-style-type: none"> ○ Card Type ○ Amount ○ Timestamp ○ Payment Status • Merchant Data: <ul style="list-style-type: none"> ○ Merchant Activity Type • Others (to be defined) <p>All shared data will adhere to compliance with and applicable privacy and data security regulation</p>

3.1.2 Business Requirements

Pilot 1 aims to develop a powerful recommendation engine for families and customise the user experience of pilot's clients as well as to develop more user-friendly consumer interfaces to provide these recommendations. UC 2 aims to generate different customer profiles, based on consumer data, and cross them with instalment risk, leveraging and enlarging the scope of Buy Now Pay Later solutions implementation. It is intended to use analytical tools to support customer profiling and use ML techniques to develop a scoring risk model on instalment availability.

Table 7 - Pilot 1: Business Requirements

Business Req. ID	Use Case ID	Business Objective	Business Requirement
P1_BR1	UC 1	Generate the best recommendations for existing customers based on historical data	FAME should be able to train a recommender system based on historical data about customers and their behaviour.
P1_BR2	UC 1	Generate the best recommendations for existing customers based on historical data	FAME should be able to explain the recommendations of the recommender system, or any other model employed.
P1_BR3	UC 1	Find relevant data assets for the recommender system and monetize outputs	FAME should have mechanisms to handle returns or disputes (ex: bad data asset quality, not as expected...).
P1_BR4	UC 1	Find relevant data assets for the recommender system and monetize outputs	FAME should allow trading data assets, including those created through FAME tools (ex: I publish data asset X inside my private federation, transform it using FAME analytical tools into Y into Z and trade Z publicly).

P1_BR5	UC 1	Find relevant data assets for the recommender system and monetize output	FAME should support parametrization of transformation pipelines (ex: ML, anonymization, and whichever tools become available) in order to replicate the steps with new data updates.
P1_BR6	UC 1	Find relevant data assets for the recommender system and monetize output	FAME should have searching and filtering capabilities, including unstructured data. Nice-to-have would be intelligent search and not only keyword based as in classical engines. ex: "find me data about consumption patterns in Europe in 2023".
P1_BR7	UC 1	Find relevant data assets for the recommender system and monetize output	FAME should allow discussing, reviewing and inquiring about data assets.
P1_BR8	UC 1	Find relevant data assets for the recommender system and monetize output	FAME should allow publishing new data as it becomes available, possibly in streaming (to be also consumed in streaming by acquires).
P1_BR9	UC 1	Comply with regulation	FAME should have tools that identify, alert, alter or anonymise client sensitive information on publication.
P1_BR10	UC 1	Comply with regulation	FAME should have tools that can generate new data assets containing synthetic data generated from other privately owned data assets.
P1_BR11	UC 1	Ensure the sovereignty of data	FAME should have mechanisms to ensure data usage control policies.
P1_BR12	UC 1	Train finance sector professionals on customers' profiling mechanisms	FAME should provide training materials on how to use the platform and its tools (tutorials, Webinars, How-to videos, Jupyter notebooks...).
P1_BR13	UC 1	Find relevant data assets for the recommender system and monetize outputs	FAME should provide statistics about data asset usage, views, and downloads.
P1_BR14	UC 1	Comply with regulation	FAME should be able to trace each data asset to its origin when possible (ex: column A comes from dataset B and is a sum of column C from dataset D and E, uploaded by X).
P1_BR15	UC 1	Find relevant data assets for the recommender system and monetize outputs	FAME should allow buying fractional parts of data assets, possibly at reduced price (ex: cost per row).
P1_BR16	UC 1	Find relevant data assets for the recommender system and monetize outputs	FAME should allow requesting data assets to other entities in FAME.
P1_BR17	UC 1	Find relevant data assets for the recommender	FAME should provide estimates of data asset value, based on similar data assets.

		system and monetize outputs	
P1_BR18	UC 1	Find relevant data assets for the recommender system and monetize outputs	FAME should recommend related data assets to the one currently being viewed.
P1_BR19	UC 1	Comply with regulation	FAME should have a way to manage and enforce policies at a federation level.
P1_BR20	UC 2	Generate costumer profiles, through FAME tools	FAME should be able to, through FAME analytical tools, analyse costumer data and identify different costumer profiles (eg: Profile costumer by age, gender, location, profession. Card type, ...).
P1_BR21	UC 2	Generate costumer profiles, through FAME tools	FAME analytical tools should be able to cross various dimensions based on customer profile or merchant business type.
P1_BR22	UC 2	Generate an instalment risk model	FAME tools should be able to create a scoring ML model to predict instalment risk.
P1_BR23	UC 2	Generate an instalment risk model	FAME analytical tools should be able to consume data to predict product consumption and friction points.
P1_BR24	UC 2	Generate an instalment risk model	FAME analytical tools should be able to cross product risk with other similar products.
P1_BR25	UC 2	Understand dataset usage	FAME should provide statistics about data asset usage, views, and downloads.
P1_BR26	UC 2	All participants having the ability access FAME data marketplace	FAME should allow uploading data to the marketplace and monitor data usage by other marketplace users.
P1_BR27	UC 2	Search Data regarding similar Credit and instalment services usage	Filter aggregated or individual data that could be useful to our analytics and management team to explore and analyse.
P1_BR28	UC 2	Search Data regarding similar Credit and instalment services usage	FAME should include in the FDAC relevant data to support costumer profiles generation.
P1_BR29	UC 2	Search Data regarding similar Credit and instalment services usage	FAME should include in the FDAC relevant data to support instalment risk ML model development.
P1_BR30	UC 2	Search Data regarding similar Credit and instalment services usage	FAME should allow project participants to interact to share knowledge and data clarification.
P1_BR31	UC 2	Enrich FAME Marketplace	FAME should allow data asset download from the marketplace.
P1_BR32	UC 2	Enrich FAME Marketplace	Allow continuous upload of data to the marketplace.
P1_BR33	UC 2	Comply with regulation	FAME should have tools that identify, alert, alter or anonymise client sensitive information on publication.

3.1.3 Datasets Status

As mentioned in the previous chapter, the collection of datasets was prepared by WP 5. This information is divided into Use Cases for each pilot and this may change during the course of the project.

Table 8 - Datasets Status: Pilot #1 - Use Case #1

Pilot #1 – Use Case #1	
Pilot Leader	MC
Name	Recommendation Engine for Universo
Category	FaMLy – A powerful financial recommendation engine for families
Data Type	<ul style="list-style-type: none"> • Tabular • Text • Timeseries
Input Data	<ul style="list-style-type: none"> • Client behaviour <ul style="list-style-type: none"> ○ transactions (type, location, dates, amounts, etc) ○ Digital interaction (app downloads, logins, etc) ○ Client sociodemographic data (age, address, income, etc)
Expected Output	Marketing proposal ranking
Type of Analytics	Machine Learning
Input Source	Inside pilot’s premises
Output Storage	Inside pilot’s premises
Processing Location	Cloud
Confidential/Private data	Yes
Is the data part of a business process?	Yes
Dataset tentative delivery deadline	2023 Q4
Data Model in use	Relational model with client data
Data Accessibility	No
Historical Data	TBD

Table 9 - Datasets Status: Pilot #1 - Use Case #2

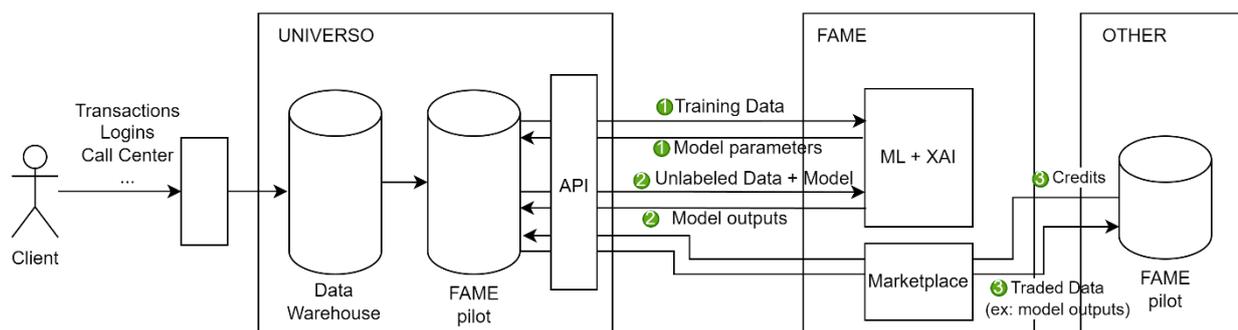
Pilot #1 – Use Case #2	
Pilot Leader	MC
Name	Consumers’ Incentives and Interfaces
Category	FaMLy – A powerful financial recommendation engine for families
Data Type	<ul style="list-style-type: none"> • Tabular • Text • Timeseries
Input Data	<ul style="list-style-type: none"> • Client transactions • Store • District

Expected Output	Customers data and transactions
Type of Analytics	<ul style="list-style-type: none"> • IA • Machine Learning
Input Source	Inside pilot's premises
Output Storage	Inside pilot's premises
Processing Location	Cloud
Confidential/Private data	Yes, fiscal number, client address, phone number, card
Is the data part of a business process?	Yes, purchasing a new product
Dataset tentative delivery deadline	TBD
Data Model in use	Relational model with client data
Data Accessibility	No for now
Historical Data	TBD

3.1.4 Systems to connect to FAME

Data from Universo clients is stored in Universo's data warehouse. A relational database can be made available behind an API for the purpose of serving FAME's pilot 1.

Access control, data asset visibility, trading, model training, inference – and all other usages and configurations pertaining to the pilot's scope – are expected to be handled by FAME's systems.



3.1.4.1 Internal Systems

At the moment of writing, and with the sole purpose of retrieving and sending data, access to Universo's data assets is planned to take place through API, thus constituting the only internal system of relevance.

3.1.4.2 External Data Sources

Access to external data sources is currently not expected for use-case 1 purpose. However, for use-case 2 it is expected to have the ability to access data available in FAME platform, shared by other pilots or from other sources included in the FDAC, if considered to have potential valuable insights for the work to be done.

3.1.5 Next steps and timeline

Short-term next steps under pilot 1 are to share the internal data available with WP 5 partners, to start working on identification of possible dataset limitation to be overcome and start first iterations of

algorithm training, leveraging ML models development. In parallel, it is aimed to research for external data assets to be included in the models' development process.

3.2 Pilot #2: Embedding Finance Services in a Personalized Citizen Wallet

Table 10 - Pilot 2: Description

Pilot #2	
Pilot Name	Embedding Finance Services in a Personalized Citizen Wallet (NOVO, DAEM, UBI)
Summary	Offer more citizen-centred financial services and extend in additional services (e.g. transportation), based on existing data produced from myAthensPass smart parking app.
Description	Since 2019 the City of Athens (DAEM) and its technology partner NOVO have re- branded the popular smart parking app of the city (myAthensPass). The new app enables drivers to buy parking time quickly, easily and conveniently. It also enables motorists to find their exact location, to select their desired parking duration, to extend their parking time remotely, to access information about how much they need to pay, and to parking time in advance. Relevant payment transactions can be carried out via NOVO's app, which serves as a mini citizen wallet. DAEM is also collecting large amounts of data about citizens' parking activity, including their payments. DAEM and NOVO are interested in using the collected data for offering more citizen-centred financial services and for extending the paradigm in additional services (e.g., transport, medical needs).
Motivation	Provide citizen centric services. Extend supported services.

3.2.1 Use Cases

3.2.1.1 Use Case #1

Table 11 - Pilot 2: Use Case 1

Use Case #1	
Use Case Name	Personalized Payment and Loyalty Programs for Smart Parking
Summary	Provide personalized user centric services. Data from user activities (with the consent of user) will be traded through platform and combined with data from external sources will provide model loyalty programs.
Description	In this use case, we aim to create a comprehensive EmFi profile for citizens by utilizing existing parking data (e.g., payment frequency, timeliness, and parking duration and location), along with behavioral insights from other NOVO app services. We will complement this with external data related to public transportation traffic. This robust profile will serve as the foundation for generating a wide array of personalized, citizen-centered offers and recommendations such as "Park & Ride" options. These recommendations will offer citizens incentives such as discounts on bulk parking ticket purchases, reduced fares for using public

transportation, or suggestions for utilizing eco-friendly public transportation options, thus promoting sustainable urban mobility. To make customer profiling and recommendation provision possible, different types of citizen data will be securely traded through our marketplace and shared with DAEM/NOVO, contingent on each citizen's consent to include their data in the marketplace. Furthermore, it's imperative that public transportation traffic data be made accessible to the FAME platform to enable seamless processing.

Value Proposition(s)

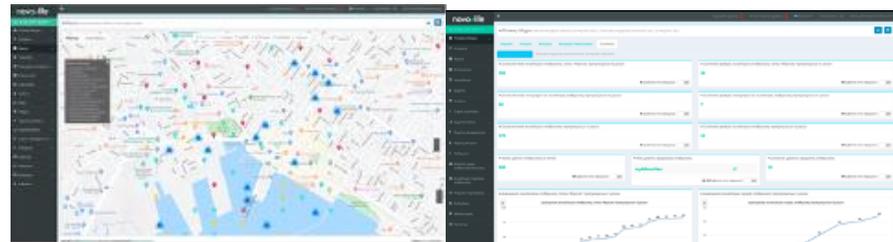
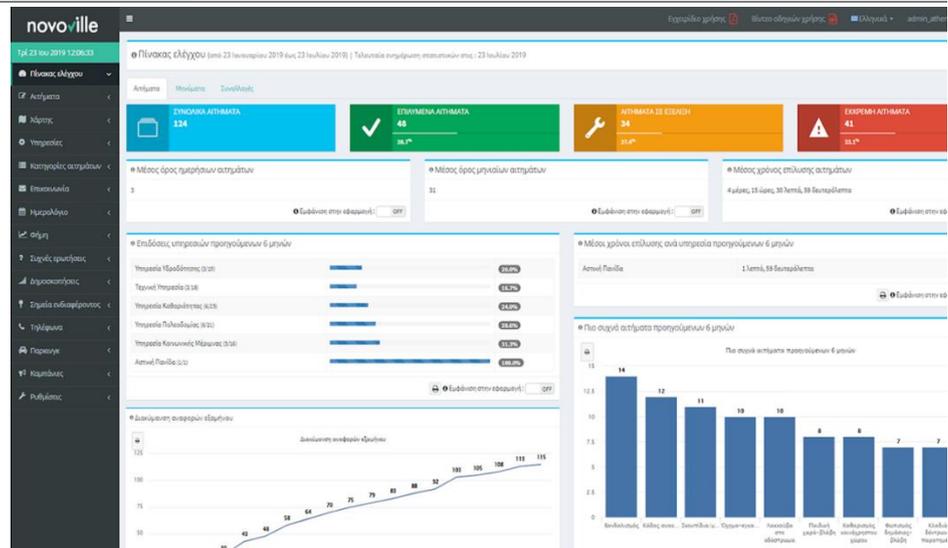
Bullet list with value proposition(s) of the Use Case:

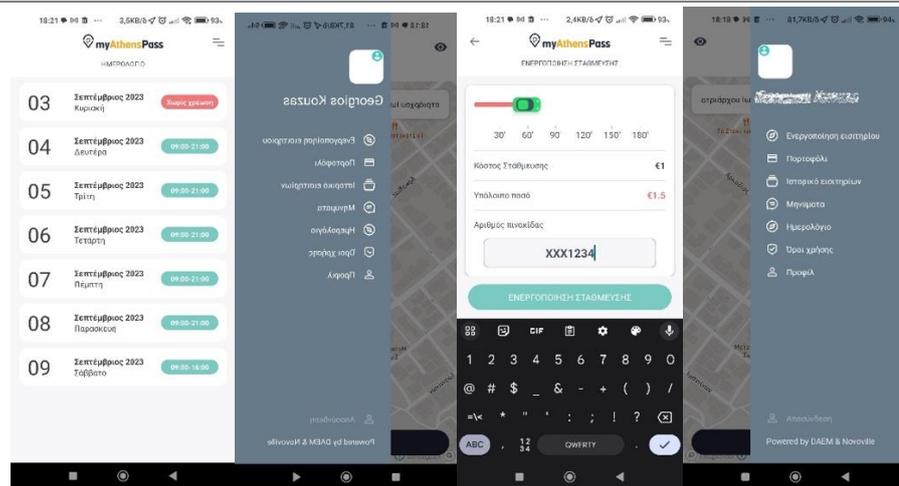
- Improve citizen services (personalized citizen centric services)
- Increase user experience.
- Increase parking areas utilization.
- Increase the green mobility maturity level.

Documentation

https://gft365.sharepoint.com/:b:/r/sites/GST-InnovationUnit-EuropeanProjects-Italy/Shared%20Documents/FAME-PROJECT/020-Project/006%20-%20WP6/Pilot%232/MyAthensPass_App_Novoville.pdf?csf=1&web=1&e=B7bHJP

Image(s)





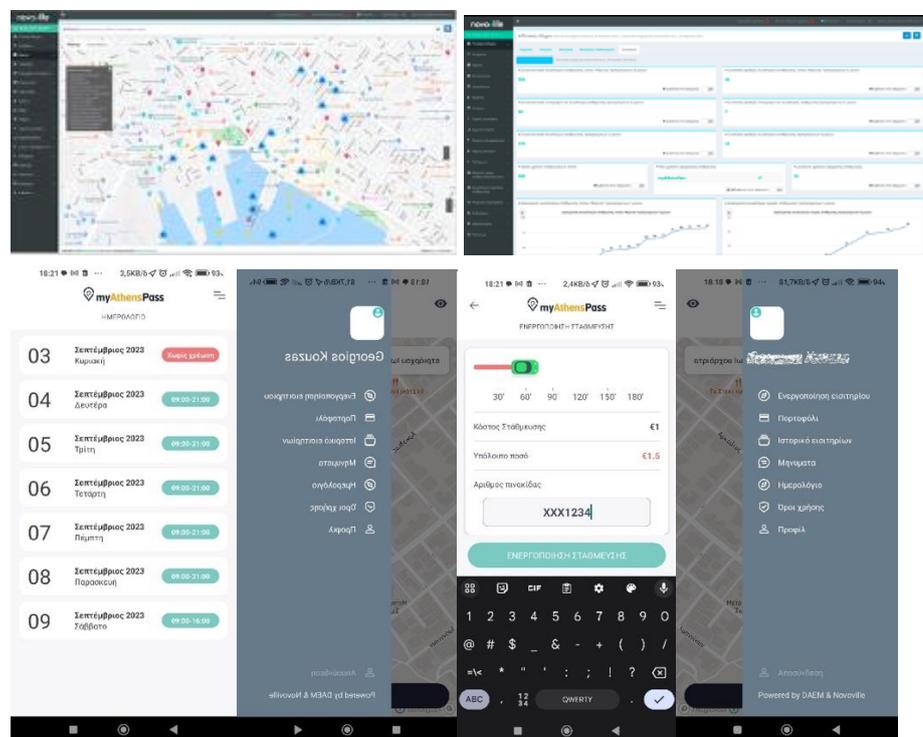
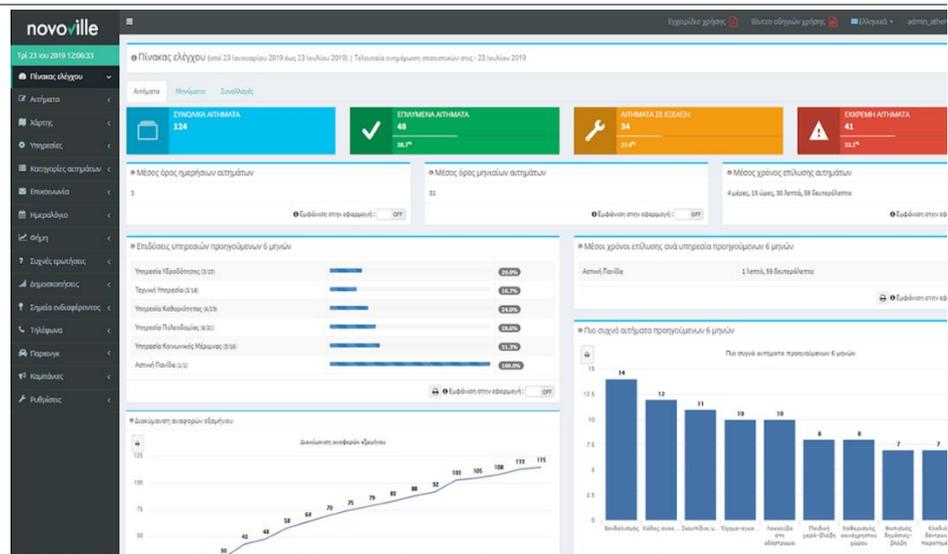
Website	http://parkinathens.gr/
Keywords	Smart Cities, citizen centric, citizen wallet, green mobility
Available Data Assets	<ol style="list-style-type: none"> 1. Parking Data 2. GTFS routing 3. Bus Telematics

3.2.1.2 Use Case #2

Table 12 - Pilot 2: Use Case 2

Use Case #2	
Use Case Name	Extended Payments Across Multiple Services
Summary	Extent payments to other services. Data of user activities or other sources will be traded to support the extension of the financial services
Description	<p>The use case aims to expand NOVO's citizen wallet to encompass a wider array of services, including transportation services and fine payments. This will empower citizens to conveniently fund their citizen wallet using their credit/debit cards and subsequently utilize these credits for various services.</p> <p>Moreover, within their wallet, citizens will have the option to purchase discounted e-tickets for public transportation by leveraging the "Park & Ride" loyalty scheme offered by UC 1. Similar to the previous scenario, there will be an exchange of citizens' behavioral data among the developers, facilitating the provision of cross-service offers.</p>
Value Proposition(s)	<p>Bullet list with value proposition(s) of the Use Case:</p> <ul style="list-style-type: none"> ▪ Expand financial services. ▪ Increase supported services to citizens.
Documentation	https://gft365.sharepoint.com/:b:/r/sites/GST-InnovationUnit-EuropeanProjects-Italy/Shared%20Documents/FAME-PROJECT/020-Project/006%20-%20WP6/Pilot%232/MyAthensPass_App_Novoville.pdf?csf=1&web=1&c=B7bHJP

Image(s)



Website

<http://parkinathens.gr/>

Keywords

Smart Cities, citizen centric, Emended financial services, data sharing

Available Data Assets

1. Parking Data
2. Public Transportation Fare Policy
3. GTFS Routing

3.2.2 Business Requirements

The objective of Pilot 2 is to develop new EmFi services in an urban context, where data from smart services is used to provide additional financial services. To achieve this, the pilot will create a citizen's EmFi profile by leveraging existing parking data, integrating other citizen behavioral data, and using citizen profiles to generate and deliver personalized citizen-centric offers and recommendations. On the other hand, Pilot 2 aims to extend the citizen's card to a wider range of services, such as transport services, fine payments, etc.

Table 13 - Pilot 2: Business Requirements

Business Req. ID	Use Case ID	Business Objective	Business Requirement
P2_BR1	UC 1	Analyse parking data in terms of locality, frequency, and time of parking	FAME should optimize the data analysis part in order to be pinpoint parking habits
P2_BR2	UC 1	Request data from other organisations that have loyalty programs in order to analyse the rewarding mechanisms	FAME should offer an added value to an external entity in order to add their data and offer an authorized and usable onboarding/registration procedure
P2_BR3	UC 1	Acquire knowledge and training on the platform so as to leverage the city's IT personnel competencies	FAME should offer training feature (e.g., webinars, user guide, MOOCs, online training sessions etc)
P2_BR4	UC 1	Request and onboard data from other organizations so as to leverage the services of the city e.g., traffic data	FAME should offer an added value to an external entity in order to add their data and offer an authorized and usable onboarding/registration procedure
P2_BR5	UC 1	Reduce the costs that refer to the management of the parking system of the city	FAME should optimize the data analysis in order to identify solutions that can make the system less costly for the city
P2_BR6	UC 1	Analyse parking and behavioural data in terms of locality in order to design new services e.g. if a citizen owns a shop, then short term parking for customers could be foreseen	FAME should optimize the data analysis part in order to be pinpoint parking and other habits and localities
P2_BR7	UC 1	Ensure the sovereignty of the data being traded	FAME should have mechanisms to ensure data usage control policies
P2_BR8	UC 1	Correlate data sources to offer new services	FAME should offer a view of correlated data and usable visualizations for a city employee
P2_BR9	UC 1	Exploit data on parking payments to design specific offers e.g., discounts	FAME should offer an analysis of parking payments and usable visualizations e.g., time, duration, location etc
P2_BR10	UC 2	Request and onboard data from other organizations so as to leverage the services of the city e.g. traffic data	FAME should offer an added value to an external entity in order to add their data and offer a authorized and usable onboarding/registration procedure
P2_BR11	UC 2	Analyse parking and behavioural data in terms of locality in order to foresee the addition of other municipal services	FAME should optimize the data analysis part in order to be pinpoint parking and other habits and localities
P2_BR12	UC 2	Analyse parking and behavioural data in terms of high/low demand	FAME should optimize the data analysis part in order to be able to

		in order to develop a dynamic pricing scheme	foresee parking demand based on historical data
P2_BR13	UC 2	Offer citizen wallet to citizens	Develop a citizen wallet for citizens to consume services in one app
P2_BR14	UC 2	Develop a dynamic parking pricing scheme	Develop a dynamic pricing mechanism based on parking demand
P2_BR15	UC 2	Ensure the sovereignty of the data being traded	FAME should have mechanisms to ensure data usage control policies

3.2.3 Datasets Status

As mentioned in the previous chapter, the collection of datasets was prepared by WP 5. This information is divided into Use Cases for each pilot and this may change during the course of the project.

Table 14 - Datasets Status: Pilot #2 - Use Case #1

Pilot #2 – Use Case #1	
Pilot Leader	DAEM
Name	Personalized Payment and Loyalty Programs for Smart Parking
Category	Embedding Finance Services in a Personalized Citizen Wallet
Data Type	<ul style="list-style-type: none"> • Spatial Timeseries • Timeseries • Images
Input Data	Type: <ul style="list-style-type: none"> • Parking Location Data • User ID • Parking Transaction Data <ul style="list-style-type: none"> ○ Parking Duration Data ○ Incident Location Data ○ Incident Category Data Size: <ul style="list-style-type: none"> • 1 TB
Expected Output	Classification & Regression
Type of Analytics	Potentially XAI, but we will likely use parametric models
Input Source	Migrated
Output Storage	Available from FAME or Pilot's premises
Processing Location	Anywhere
Confidential/Private data	No
Is the data part of a business process?	No
Dataset tentative delivery deadline	September 2023 (TBD)
Data Model in use	Timeseries data containing date, time, coordinates, numerical data for transactions
Data Accessibility	Not at the moment
Historical Data	Yes

Table 15 - Datasets Status: Pilot #2 - Use Case #2

Pilot #2 – Use Case #2	
Pilot Leader	DAEM
Name	Extended Payments Across Multiple Services
Category	Embedding Finance Services in a Personalized Citizen Wallet
Data Type	<ul style="list-style-type: none"> • Spatial Timeseries • Timeseries Images
Input Data	Type: <ul style="list-style-type: none"> • Parking Location Data • User ID • Parking Transaction Data <ul style="list-style-type: none"> ○ Parking Duration Data ○ Incident Location Data ○ Incident Category Data Size: <ul style="list-style-type: none"> • 1 TB
Expected Output	Classification & Regression
Type of Analytics	Potentially XAI, but we will likely use parametric models
Input Source	Migrated
Output Storage	Available from FAME or Pilot's premises
Processing Location	Anywhere
Confidential/Private data	No
Is the data part of a business process?	No
Dataset tentative delivery deadline	September 2023 (TBD)
Data Model in use	Timeseries data containing date, time, coordinates, numerical data for transactions
Data Accessibility	Not at the moment
Historical Data	Yes

3.2.4 Systems to connect to FAME

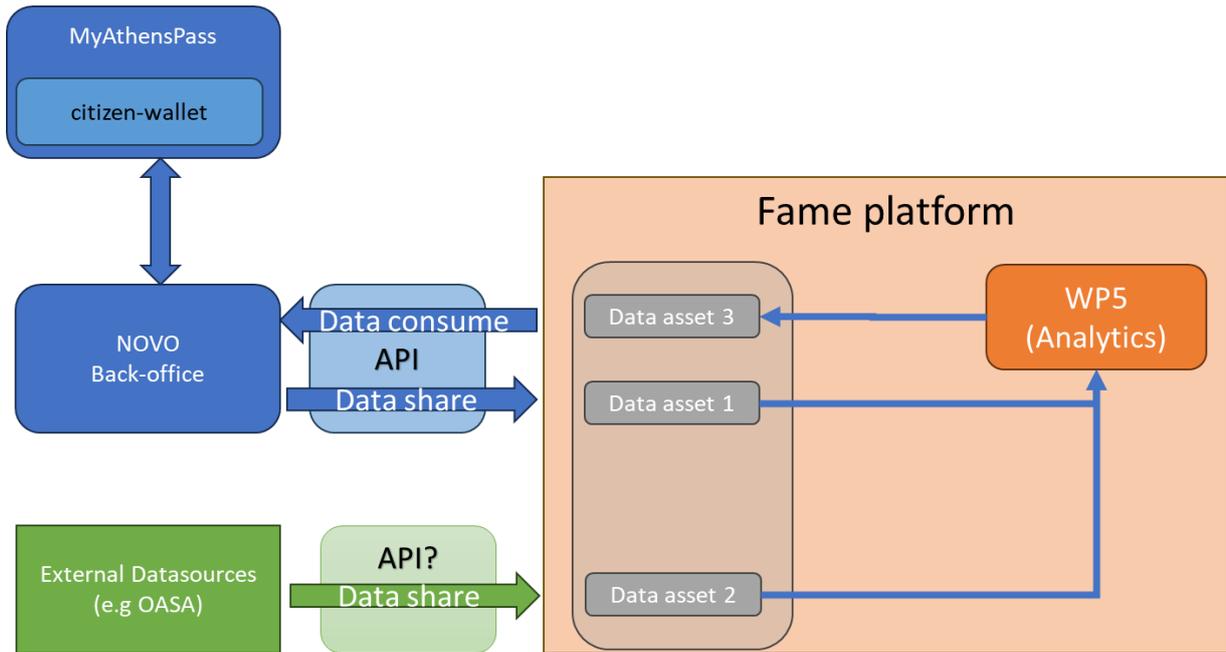


Figure 5 - Pilot #2: Proposed Architecture

3.2.4.1 Internal Systems

- Mobile app myAthensPass. The existing mobile app that is used from citizens in order to buy parking time. This mobile app will be enhanced with a citizen wallet.
- Novo back-office. All transactions from citizens are stored in the back-office system (with an embedded database).
- API for the integration between FAME and back-office database. It will be developed in order to provide data to the FAME platform.

3.2.4.2 External Data Sources

Telematics data from public transportation organization in city of Athens that are provided for free.

3.2.5 Next steps and timeline

- The second round of business requirements elicitation process is currently in progress. Finalize business requirements (till end of January 2024)
- Define analytics (ML, AI) that are applicable to the pilot's use cases.
 - Start cooperation with WP 5 (Mid of Dec 2023).
 - Train models and test (2024)
- Consolidate Pilot 2 architecture and integration with FAME.
 - Define and implement connectors with FAME (in parallel with WP 3- WP 4 implementation)
 - Develop and integrate digital wallet (2024).

3.3 Pilot #3: Personalized Collaborative Intelligence for Enhancing EmFi Services

Table 16 - Pilot 3: Description

Pilot #3	
Pilot Name	Personalized Collaborative Intelligence for Enhancing Embedded Finance (EmFi) Services
Responsible	BPFI, NUIG, JRC
Target	Asset managers, Asset Management Firms, Private Equity Firms, Wealth Management Departments of Banks & Financial Organizations, Financial and Investment stakeholders at large, FAME marketplace members and data exchangers at large.
Stakeholders	<p>Stakeholders' Roles:</p> <ul style="list-style-type: none"> (i) Data providers: JRC based on their Subscriptions in finance data marketplaces (e.g., Bloomberg, Quantal, Intrinio); (ii) NUIG as providers of Synthetic Data models for personal profiling and portfolio risk assessment and BPFI portfolio profile validation and weights from other reference BPFI funds for reference & opportunities. (iii) Data-Driven Application Developers: NUIG & JRC will leverage data assets for the development of the JRC risk assessment and Anti-Money laundering applications. (iv) FAME marketplace members: Data assets for anti-money laundering FAME, the FAME reference stack and FAME Reference Architecture, the FAME methodology and AI algorithms of JRC portfolio Risk Assessment and AML, the NUIG educational material and the market sentiment in relation to BPFI ecosystem.

3.3.1 Use Cases

3.3.1.1 Use Case #1

Table 17 - Pilot 3: Use Case 1

Use Case #1	
Use Case Name	Risk Return Profiling According to Robo Free Black–Litterman model
Summary	<p>This use case will produce new data assets that will be sold as “facts” to embedded finance services providers through the FAME marketplace. The FAME analytical tools (XAI, Energy Efficient Analytics, FML) will be used over datasets from multiple payment providers (incl. BOI, BPFI), as well as alternative data sets (e.g., news, blogs, social media). EmFi providers will be able to locate these insights, pay for them and use them to develop their own services. The FAME marketplace will ensure that data from the various services are consolidated and accessible (FDAC) and fair priced through blockchain smart contracts. This use case will produce new data assets that will be sold as “facts” to embedded finance services providers through the FAME marketplace. The FAME analytical tools (XAI, Energy Efficient Analytics, FML) will be used over datasets from multiple payment providers (incl. BOI, BPFI), as well as alternative data sets (e.g., news, blogs, social media). EmFi providers will be able to locate these insights, pay for them and use them to develop their own services. The FAME marketplace will ensure that data from the various services are</p>

	consolidated and accessible (FDAC) and fair priced through blockchain smart contracts.
Goal	<p>According to discussion with JRC representatives we arrived at the decision to define Model 1 in Risk- Return Profiling Framework consistent with robo free Black-Litterman Model. A robo-advisor (also sometimes spelled as roboadvisor) is a digital platform that provides automated, algorithm-driven financial planning and investment services with little to no human supervision. A typical robo-advisor asks questions about your financial situation and future goals through an online survey. It then uses the data to offer advice and automatically invest for you. Other common designations for robo-advisors include "automated investment advisor," "automated investment management," and "digital advice platforms." The best robo-advisors offer easy account setup, robust goal planning, account services, and portfolio management. Additionally, they offer security features, comprehensive education, and low fees.</p> <p>In finance, the Black–Litterman model is a mathematical model for portfolio allocation developed in 1990 at Goldman Sachs by Fischer Black and Robert Litterman, and published in 1992. It seeks to overcome problems that institutional investors have encountered in applying modern portfolio theory in practice. The model starts with an asset allocation based on the equilibrium assumption (assets will perform in the future as they have in the past) and then modifies that allocation by taking into account the opinion of the investor regarding future asset performance.</p>
Constraints	Embedded Finance (EF) is the process of adoption of financial services or products into non-finance company's business model. In words, EF is when non-financial companies or organizations include financial services as part of their services. In this way, financial operations such as loans, insurance, debit cards, savings and investment instruments are integrated into a non-finance company, organization, or institution operations. Therefore, when products, services or processes of non-financial companies are financialized it is called EF. It is important to note that the process of EF is done through partnership of a non-finance company with technology providers instead of banks or other traditional financial institution.
Actors	Everybody who executes a use case (uploads, downloads, exchanges data and methods), everybody who derives outputs from our developed tools, interested parties which follow our educational content and trainings for JRC investments
Dependency with other scenarios	Pilot 5, potentially pilot 4 on Client Scoring for AML.

3.3.1.2 Use Case #2

Table 18 - Pilot 3: Use Case 2

Use Case #2	
Use Case Name	Anti Money Laundering (AML) as a Service
Summary	Anti Money Laundering (AML) as a Service: This use case will implement an AML service based on the identification of potential links

	<p>to fraudulent payments or other criminal activities. The service will be made available to EmFi application developers and services providers in order to facilitate their AML checks. Stakeholders' Roles: (i) Data providers: JRC and BPFi will provide the models for payments data in the FAME marketplace, while NUIG will leverage FAME tools to develop synthetic data sets and value added analytical insights for facts-as-a-service and AML; (ii) Data-Driven Application Developers: NUIG will have the role of application developer as well, through using data assets in simple EmFi applications like credit risk scoring and recommendations for customers. Various pricing/trading models will be tested for the data providers/developers' interactions. Data Assets: Multi-million (>1 M) Customer payments records in Europe, including customer-to-customer relationships and millions of customer account data to experiment and generate decentralised payments data sets; Millions of profiles and events from alternative data sources for embedded payments.</p>
Goal	<p>This use case will implement an AML service based on the identification of potential links to fraudulent payments or other criminal activities. The service will be made available to Embedded Finance (EmFi) application developers and services providers in order to facilitate their AML checks.</p> <ol style="list-style-type: none"> 1) JRC portfolio comparison tool. 2) JRC portfolio optimization tool.
Constraints	<p>EF provides a tremendous opportunity for finance industry to provide solutions to real- world problems. For example, in the era of Open Banking, payment providers have significant opportunities to gain value-added information about their services and their customers based on the aggregation, consolidation and sharing of data across institutions and organizations. The consolidation of analytics across financial organizations enables them to gain timely and accurate insights for processes like onboarding, services personalization, and credit risk scoring.</p>
Constraints	<p>EF provides a tremendous opportunity for finance industry to provide solutions to real- world problems. For example, in the era of Open Banking, payment providers have significant opportunities to gain value-added information about their services and their customers based on the aggregation, consolidation and sharing of data across institutions and organizations. The consolidation of analytics across financial organizations enables them to gain timely and accurate insights for processes like onboarding, services personalization, and credit risk scoring.</p>
Actors	<p>Everybody who executes a use case (uploads, downloads, exchanges data and methods), everybody who derives outputs from our developed tools, interested parties which follow our educational content and trainings for JRC investments.</p>
Dependency with other scenarios	<p>Pilot 5, potentially pilot 4 on Client Scoring for AML.</p>

3.3.2 Business Requirements

These Business Requirements are still in Draft, the final ones will be in Deliverable 2.1.

Table 19 - Pilot 3: Business Requirements

Non-functional requirements Use Case #1	<ul style="list-style-type: none"> • Profile client data on JRC-type scoring. • EU global Transactions data through FAME marketplace • Allow market data to be inserted beyond price and synthetic data sets. • Investment algorithms over the client transactions to be: <ul style="list-style-type: none"> (i) data mining-coded. (ii) run algorithms and compare with outputs. (iii) allow data output as new input (structured client portfolio scoring).
Non-functional requirements Use Case #2	<ul style="list-style-type: none"> • Profile client data on JRC-type scoring. • EU global Transactions data through FAME marketplace. • Allow market data to be inserted beyond price and synthetic data sets. • Investment algorithms over the client transactions to be: <ul style="list-style-type: none"> (iv) data mining-coded. (v) run algorithms and compare with outputs.

3.3.3 Datasets Status

As mentioned in the previous chapter, the collection of datasets was prepared by WP 5. This information is divided into Use Cases for each pilot and this may change during the course of the project.

Table 20 - Datasets Status: Pilot #3 - Use Case #1

Pilot #3 – Use Case #1	
Pilot Leader	
Name	Pay Facts-as-a-Service for Embedded Payments
Category	Personalized Collaborative Intelligence for Enhancing EmFi Services
Data Main Characteristics	<ul style="list-style-type: none"> • Available Tabular data (Modelled). • Recorded Timeseries (Historical and/or Synthetic). • Data format Csv, Json data (Price, JRC score, metadata). • Txt for the formulas of JRC weightings (Modelled).
Input Data Main Types	<ol style="list-style-type: none"> 1) Price data (in txt and Json) formats to be discussed and provided by JRC as examples and pending approval for formatting in order to expand the formats according to a data model. 2) Customer profiles for indicative scenarios (UC 3.1 and UC 3.2), including risk client profile and JRC client profile to be provided by both partners (synthetic and indicative scenarios). 3) Profiles being prepared by JRC and NUIG partners until end of year. 4) JRC scorecard data (UC 3.1 and UC 3.2), this list includes JRC scorecard data from approx. 50 EU companies in tabular format – Status Uploaded,

	<p>pending approval for formatting.</p> <p>5) Metadata, such as (example) News data for potential fraud and money laundering extraction (UC 3.1) and other metadata (UC 3.2) from companies - news on ESG scores – Indicative datasets provided.</p>
Data Model in use	Timeseries containing Datetime and numerical values.
Data Accessibility	Planning to Generate Synthtetic data at the moment. Indicative JSON and txt Data available on SharePoint repository for evaluation.
Data Assets	Market Data, JRC Scores, Synthetic portfolio data based and constructed from combinations of anonymized portfolio data, risk profiles and JRC preference profiles of synthetic clients and synthetic portfolios, Market Metadata, Newsfeeds.
Historical Data	Have to check the volume and how it can be provided. Only indicative data for back-testing will be needed.

Table 21 - Datasets Status: Pilot #3 - Use Case #2

Pilot #3 – Use Case #2	
Pilot Leader	
Name	Anti Money Laundering (AML) as a Service
Category	Personalized Collaborative Intelligence for Enhancing EmFi Services.
Data Main Characteristics	<ul style="list-style-type: none"> • Available Tabular data (Modelled). • Recorded Timeseries (Historical and/or Synthetic) • Data format Csv, Json data (Price, JRC score, metadata). • Txt for the formulas of JRC weightings (Modelled).
Input Data Main Types	<p>1) Price data (in txt and json) formats to be discussed and provided by JRC as examples and pending approval for formatting in order to expand the formats according to a data model.</p> <p>2) Customer profiles for indicative scenarios (UC 3.1 and UC 3.2), including risk client profile and JRC client profile to be provided by both partners (synthetic and indicative scenarios).</p> <p>3) Profiles being prepared by JRC and NUIG partners until end of year.</p> <p>4) JRC scorecard data (UC 3.1 and UC 3.2), this list includes JRC scorecard data from approx 50 EU companies in tabular format – Status Uploaded, pending approval for formatting.</p> <p>5) Metadata, such as (example) News data for potential fraud and money laundering extraction (UC 3.1) and other metadata (UC 3.2) from companies - news on ESG scores – Indicative datasets provided.</p>

Data Model in use	Timeseries data containing Datetime and numerical sensor values.
Data Accessibility	Not at the moment. Indicative JSON and txt Data available on SharePoint repository for evaluation.
Data Assets	Market Data, ESG Scores, Synthetic portfolio data based and constructed from combinations of anonymized portfolio data, risk profiles and ESG preference profiles of synthetic clients and synthetic portfolios, Market Metadata, Newsfeeds.
Historical Data	Have to check the volume and how it can be provided. Only indicative data for back-testing will be needed. Full stream of Bloomberg Data is too big and not fully needed.

3.3.4 Systems to connect to FAME

3.3.4.1 Internal Systems

- Use Case #1

Proposed Model 1- Risk Return Profiling According to Robo Free Black– Litterman model

According to discussion with JRC representative we arrived at the decision to define Model 1 in Risk-Return Profiling Framework consistent with robo free Black-Litterman Model.

A robo-advisor (also sometimes spelled as roboadvisor) is a digital platform that provides automated, algorithm-driven financial planning and investment services with little to no human supervision. A typical robo-advisor asks questions about your financial situation and future goals through an online survey. It then uses the data to offer advice and automatically invest for you. Other common designations for robo-advisors include "automated investment advisor", "automated investment management," and "digital advice platforms." The best robo-advisors offer easy account setup, robust goal planning, account services, and portfolio management. Additionally, they offer security features, comprehensive education, and low fees.

In finance, the Black–Litterman model is a mathematical model for portfolio allocation developed in 1990 at Goldman Sachs by Fischer Black and Robert Litterman, and published in 1992. It seeks to overcome problems that institutional investors have encountered in applying modern portfolio theory in practice. The model starts with an asset allocation based on the equilibrium assumption (assets will perform in the future as they have in the past) and then modifies that allocation by taking into account the opinion of the investor regarding future asset performance.

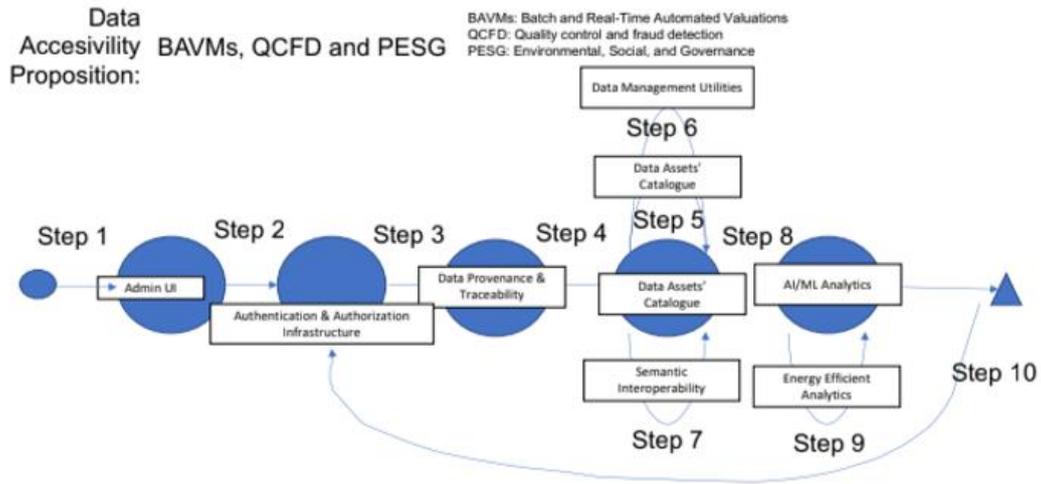


Figure 6 - Proposed Model 1: Use Case #1

- Use Case #2

This use case will implement an AML service based on the identification of potential links to fraudulent payments or other criminal activities. The service will be made available to EmFi application developers and services providers in order to facilitate their AML checks.

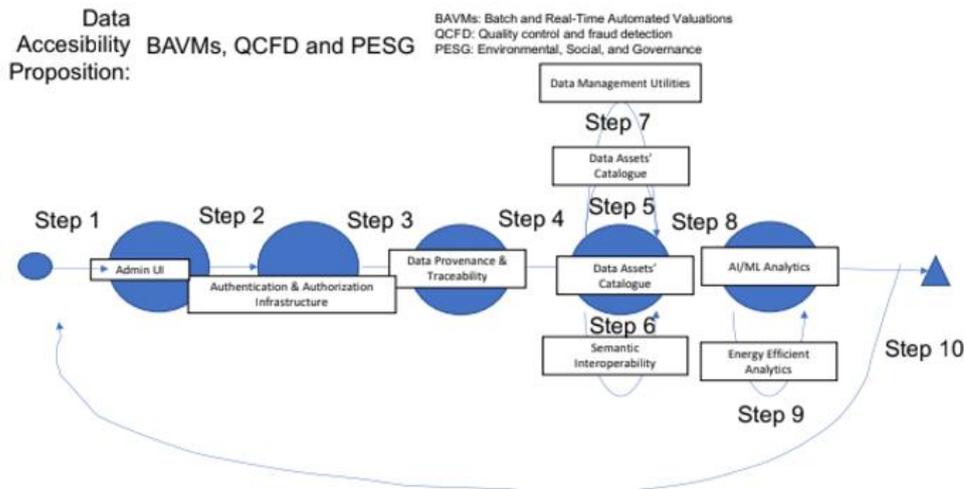


Figure 7 - Proposed Model 1: Use Case #2

3.3.4.2 External Data Sources

3.3.5 Next steps and timeline

(information about next steps to be made by the pilot, and a timeline for the activities)

3.4 Pilot #4: The EU Funds Application Process Made Easy

Table 22 - Pilot 4: Description

Pilot #4	
Pilot Name	The EU Funds Application Process Made Easy (INSO)
Summary	Parametrizing the financial supporting schemes between financial institutions and clients to facilitate the effectiveness of all the EU Funds innovation projects implementation.
Description	This pilot will simplify the process for applying for NGEU funds for companies, SMEs and self-employed in Spain through a smart platform for information and identification of grant programs and personalised advice on the processing and submission of projects that they offer to their customers. Santander Bank offers financial products that allow SMEs to guarantee, anticipate or complement the financing that they can obtain through European Funds, reducing the time gap between the grant concession and the first payment (ensuring liquidity to kick-off the project). In this cross-selling activity, Santander Bank will leverage FAME data assets and functionalities to profile grants applicants towards providing them with recommendations about optimal financial products. Companies' profiles and interests registered in the grant's platform will be matched to 4 financial products (Loans for Project Cofinancing/Prefinancing, Leasing, Renting, Financial Guarantee), towards producing accurate suggestions. Specifically, recommendation techniques that mix and match information from the users of the platform to business offerings that suit their needs and financial profiles will be employed, parametrising the financial payment scheme of the grants. The algorithm recommendations will be structured as data assets of FAME and will be available in the marketplace to be transferred to other financial entities internationally (e.g., entities with similar profiles).
Motivation	The advent of Next Generation EU (NGEU) funds has resulted in an increase in public grants and therefore a unique opportunity for self-employed, SMEs and companies to finance the twin (digital, green) transformation their businesses. Nevertheless, identifying and processing these grants is no mean feat and, on many occasions, put the brakes on investment.

3.4.1 Use Cases

3.4.1.1 Use Case #1

Table 23 - Pilot 4: Use Case 1

Use Case #1	
Use Case Name	The EU Funds Application Process Made Easy
Summary	Parametrizing the financial supporting schemes between financial institutions and clients to facilitate the effectiveness of all the EU Funds innovation projects implementation.
Description	When a company is granted with the funds, there are steps where they need a financial support from banks to be able to carry on with the process due to the fact that the public fund is granted in different moment

of the project (depending to the Grant policy). Therefore, the Bank has a good business opportunity, but all the process should be harmonized and automatized delivering a methodology with which assuring the most effective approach. The objective of this Pilot is achieving this optimized model allowing a perfect business model synergy between both stakeholders (bank and granted company).

We have access to this data through our platform and we can involve in this experiment others banks outside the FAME consortium.

In this sense, this pilot proposes sharing 20 companies data which have been granted with 3 public funds and 4 financial products. With this data we will have the opportunity to set the methodology which allow the bank pre-financing, complementary/additional financing model and create a win-win business opportunity for both.

According to the outcome pilot experience affordability, the result/product will be placed inside the FAME Marketplace for future exploitation agreement.

Value Proposition(s)	Bullet list with value proposition(s) of the Use Case: <ul style="list-style-type: none"> • Assuring granted company cash flow for its project, • Create a business opportunity for banks, • Improve the granting funding parameters (time, feasibility, bureaucracy, etc).
Documentation	<ul style="list-style-type: none"> • Granted company data anonymized. • Resulting product
Image(s)	-
Website	-
Keywords	-
Available Data Assets	Granted company data coming from other banks, will be anonymized. This data specifies the company profile which got the grant with categories like, Size, CNAE, Region, Fund description, type, total grant and € granted, financial scheme execution linked to % amount of the fund reimbursed in the calendar, bank product associated to each grant, etc.).

3.4.2 Business Requirements

This is a working process and updates will be made available in the 2nd version of the del.

3.4.3 Datasets Status

As mentioned in the previous chapter, the collection of datasets was prepared by WP 5. This information is divided into Use Cases for each pilot and this may change during the course of the project.

Table 24 - Datasets Status: Pilot #4 - Use Case #1

Pilot #4 – Use Case #1	
Pilot Leader	INSOMNIA
Name	The EU Funds Application Process Made Easy

Category	Personalized optimized model allowing a perfect business model (loan cash flow) synergy between both stakeholders (bank and granted company).
Data Type	<ul style="list-style-type: none"> • Tabular • Text • Timeseries
Input Data	<ul style="list-style-type: none"> (i) General scheme grants data (amount, prefinancing, payment scheme and instalments), (ii) Customers' data (company, cashflow, solvency, debt capacity), (iii) Financial products data (Loans, Leasing, Renting, Financial).
Expected Output	Simulated optimized plan for loan intervention (time and quantity)
Type of Analytics	Simulator algorithym
Input Source	Inside pilot's premises
Output Storage	Inside pilot's premises
Processing Location	Anywhere
Confidential/Private data	Yes, fiscal number, client address, phone number, card, etc.
Is the data part of a business process?	Yes, loan request
Dataset tentative delivery deadline	Upon request
Data Model in use	Relational model with client data
Data Accessibility	Not at the moment
Historical Data	Yes

3.4.4 Systems to connect to FAME

This is a working process and updates will be made available in the 2nd version of the del.

3.4.4.1 Internal Systems

3.4.4.2 External Data Sources

3.4.5 Next steps and timeline

This is a working process and updates will be made available in the 2nd version of the del.

3.5 Pilot #5: ESG Scorecard Ranking & Sustainable Portfolio Optimisation

Table 25 - Pilot 5: Description

Pilot #5	
Pilot Name	ESG Scorecard Ranking & Sustainable Portfolio Optimisation
Summary	<p>The two Fintech firms of the consortium (KM, JRC) will access data and tools from the FAME marketplace to develop a toolset for ESG investments:</p> <ol style="list-style-type: none"> 1) ESG ranking 2) ESG online Reporting 3) ESG recommendations (monthly-quarterly) 4) ESG portfolio comparison tool 5) ESG portfolio optimization tool 6) Relevant ESG educational material and training
Description	<p>(KM, JRC) provide pre-processed market and profile data to FAME and will access data and tools from the FAME marketplace to develop an online reporting tool that will provide weighted portfolio sustainability scores for ESG investments, ESG on-line reporting and recommendations, and also an ESG portfolio comparison and optimization tool which will also serve as a reference point, with focus on the particular needs and sensitivities of investors, stakeholders, FAME marketplace users-visitors- data traders.</p> <p>The ESG tool will be able to access and use standardized and semantically interoperable data from FAME's FDAC catalogue, which will be shared in a trusted way that ensures the sovereignty of the data.</p>
Motivation	<p>Despite the rising interest in ESG investing and the strong support from national and international financial and governmental authorities,</p> <ol style="list-style-type: none"> 1) there is still a lack of data and analytical tools for assessing, benchmarking, and comparing ESG portfolios. This is a setback to constructing "alpha" portfolios that can lead to excess returns above the average benchmarks, especially with an Environmental-Social sensitivity. 2) Current ESG portfolios around the globe use "secret in-house recipe" weightings for the asset selection with no transparent way to third interested parties of the selection criteria and weights, and most importantly no "globally accepted" standard by big players in the investment sector. FAME has the opportunity to propose its "transparent recipe". 3) Investors, and especially those with small-mid portfolios, are only offered "pre-determined" ESG funds with few options of aligning to various levels of ESG sensitivity and according to their needs and personal preference, with no customisation and no ability for optimisation metrics comparison. <p>Comparison of different ESG portfolios and also clear ESG recommendations based on unique client needs and client profiles, are still not properly addressed, since there is lack of such solutions to non-institutional investors.</p>

3.5.1 Use Cases

3.5.1.1 Use Case #1

Table 26 - Pilot 5: Use Case 1

Use Case #1	
Use Case Name	UC 1 –ESG On-line Reporting
Summary	<ol style="list-style-type: none"> 1) Online ESG-Reporting Tool from the “universe” of preselected financial Assets, 2) Table of ESG recommendations made public (monthly-semester) according to the FAME “optimisation recipe”. <p>Relevant explanatory trainings on the methodologies above as well as ESG introductory courses for FAME marketplace stakeholders.</p>
Description	<p>The Fintech experts (KM, JRC) will access data and tools from the FAME marketplace, will gather, organize, and cleanse financial data from various sources (time series, ESG scores, prices, tickers etc.), expand them with curated metadata, and combine them with a distinct set of synthetic portfolios, in order to develop a toolset that will provide:</p> <ol style="list-style-type: none"> 1) weighted portfolio sustainability scores for ESG investments 2) An online ESG reporting tool that will accumulate ESG scores, and in a universe of pre-selected financial Assets (stocks, bonds, ETFs, etc) and will provide a monthly-semester recommendations, made available to the marketplace. <p>ESG recommendations Table (monthly-quarterly).</p>
Value Proposition(s)	<ol style="list-style-type: none"> 1) ESG reporting tool available to FAME marketplace users-stakeholders, 2) ESG recommendations Table (monthly-quarterly) available to FAME marketplace users-stakeholders. <p>Selection according to the novel and transparent FAME recipe that can also be exchanged and compared to other best-in-class approaches around the globe.</p>
Documentation	<p>Documentation will be provided on:</p> <p>ESG selection criteria, ESG recommendations Table, ESG online reporting tool usage, including relevant trainings</p>
Image(s)	-
Website	-
Keywords	-
Available Data Assets	Market Data, ESG Scores, Synthetic portfolio data based and constructed from combinations of anonymized portfolio data, risk profiles and ESG preference profiles of synthetic clients and synthetic portfolios, Market Metadata, Newsfeeds.

3.5.1.2 Use Case #2

Table 27 - Pilot 5: Use Case 2

Use Case #2	
Use Case Name	UC 2 – Sustainable Portfolio Optimization
Summary	<ol style="list-style-type: none"> 1) ESG portfolio comparison tool 2) ESG portfolio optimization tool <p>Relevant ESG portfolio analysis educational material and training.</p>
Description	<p>The Fintech experts (KM, JRC) will first define the “universe” of preselected financial Assets, their market data and ESG data, the synthetic client profiles (risk and ESG sensitivity), and will:</p>

	<ol style="list-style-type: none"> 1) Build accordingly the proposed ESG portfolio based on the above requirements with specific weights. 2) Will then compare the above, as well as compare the ESG “weights recipe” with any other ESG portfolio from world renowned Financial institutions, if this is also be inserted in the same data format, as reference in the FAME marketplace/platform for comparison and knowledge exchange. <p>Will provide a tool that based on the above parametric input will provide portfolio optimization according to ESG sensitivity, risk profile and other similar parameters of each client/portfolio.</p>
Value Proposition(s)	<ol style="list-style-type: none"> 1) ESG comparison and optimisation tool outputs available to FAME marketplace users-stakeholders. 2) The unique FAME ESG “weighting recipe” can be the reference point available to FAME marketplace users-stakeholders, and a basis for further comparisons, as well as a key asset to be exploited. 3) Selection according to the novel and transparent FAME recipe that can also be exchanged and compared to other best-in-class approaches around the globe. <p>The calibration of the tool will be based on various methodologies, which range from regression analysis to AI/ML/XAI models using FAME trusted and energy efficiency tools.</p>
Documentation	Documentation will be provided on ESG selection criteria, ESG portfolio comparison and optimisation methodology, and relevant tools developed, usage, including relevant trainings and how-to guides.
Image(s)	-
Website	-
Keywords	-
Available Data Assets	Market Data, ESG Scores, Synthetic portfolio data based and constructed from combinations of anonymized portfolio data, risk profiles and ESG preference profiles of synthetic clients and synthetic portfolios, Market Metadata, Newsfeeds.

3.5.2 Business Requirements

The objective of Pilot 5 is to develop an online reporting tool that will provide weighted portfolio sustainability scores for ESG investments. This Pilot will develop ESG reporting tools that will combine multiple ESG metrics and aggregate different dimensionalities of data to provide synthetic measures that classify assets using a multi-criteria model. It will leverage ESG reporting tools developed to perform portfolio optimizations towards the transition to sustainable finance.

Table 28 - Pilot 5: Business Requirements

Business Req. ID	Use Case ID	Business Objective	Business Requirement
P5_BR1	UC 1	Develop a ranking system for stocks and/or bonds based on ESG criteria,	Rank Stocks based on their ESG scoring. Publish it globally through FAME marketplace

		alongside fundamental and price-based factors	
P5_BR2	UC 1	ESG Tutorials	Provide educational content explaining its ESG approach
P5_BR3	UC 1	ESG Ranking, Weightings and Recommendation Tables Tutorials	Introduce detailed ESG "recipe" of weights and the mechanisms that govern our recommendations
P5_BR4	UC 1	ESG Reporting	Introduce detailed ESG reporting and impact analysis, allowing users to select based on E,S,G criteria
P5_BR5	UC 1	Provide ESG table of recommendations	Allow market data to be inserted beyond price and ESG data from Bloomberg etc
P5_BR6	UC 1	Provide ESG analytics on a set of listed companies from market feeds (only indicative for testing)	Allow market data to be inserted beyond price and ESG data from Bloomberg etc
P5_BR7	UC 1	ESG Table of Recommendations (quarterly-monthly) according to the proprietary FAME recipe of ESG weights	Allow the calculated table to be exchanged through FAME Marketplace
P5_BR8	UC 2	ESG-focused Model Portfolios	Create monthly ESG-powered model portfolios, offering clients a range of investment options
P5_BR9	UC 2	ESG scoring Calculator Tool	1)Customer Questionnaire during KYC, on how 'ESG' sensitive you feel, how much E,S,G; 2)Get recommendations: i) the global FAME ESG recommendation for Stock Allocation ii) Your "personalized" one (if time permits)
P5_BR10	UC 2	Optimizing Investment Portfolio with ESG	Optimize client portfolios by considering ESG criteria and client-specific ESG preferences, resulting in more sustainable and responsible investments
P5_BR11	UC 2	Use AI to extract sentiment and portfolio impact of our recommendation text for portfolio optimisation	Create monthly AI-powered text analysis on our ESG portfolio recommendations
P5_BR12	UC 2	Whitepaper: Hierarchical risk parity (HRP) algorithm	Provide educational content explaining approach Hierarchical Risk Parity (HRP) algorithm
P5_BR13	UC 2	Portfolio Optimization Tutorials with custom ESG weights	Provide educational content explaining Portfolio Optimization methods

P5_BR14	UC 2	Comparison of proprietary FAME "recipe" of ESG weights with the weights of other renowned investment organisations	Compare a list of 10 weights of ESG parameters to best-in-class paradigms around the world
P5_BR15	UC 2	Back-testing Analysis for ESG portfolio performance	Conduct back-testing on our ESG ranking system, ensuring that the rankings support investment rationale and the profitability of portfolios
P5_BR16	UC 2	Insert 3-4 indicative portfolios as testing sample from renowned global investment sources (yahoo, Bloomberg, banks, investment houses etc.)	Allow a portfolio of assets with their ISIN number or Ticker and with weights to be constructed. Insert examples from data feeds if possible
P5_BR17	UC 2	Provide indicative E,S,G weights of 2-3 comparative investment institutions.	Allow weights of E,S,G in various "recipes" to be held
P5_BR18	UC 2	Provide indicative portfolios of "synthetic" fictional customers for testing and for FAME marketplace visitors to consult	Allow construction of a set of 4-5 portfolios of 5-10 assets each
P5_BR19	UC 2	Compare ESG weights of other established investment institutions that registered in FAME marketplace with the recommendation weights of FAME	Allow for 10 custom weights of E+S+G for every Asset in a portfolio
P5_BR20	UC 2	Explanation of ESG Reporting and Recommendations	Introduce detailed ESG reporting and impact analysis, allowing clients to track the E,S,G impacts of their investments
P5_BR21	UC 2	ESG Portfolio Recommendations published through FAME to global investors	FAME marketplace can provide ESG portfolio Recommendations with custom E,S,G weights

3.5.3 Datasets Status

As mentioned in the previous chapter, the collection of datasets was prepared by WP 5. This information is divided into Use Cases for each pilot and this may change during the course of the project. Pilot#5 will provide 4 categories of data (Market, Profile, ESG Scorecard, Metadata).

Some of them are already uploaded in the project SharePoint repository indicatively.

Table 29 - Datasets Status: Pilot #5 - Use Case #1

Pilot #5 – Use Case #1	
Pilot Leader	KM
Name	ESG On-line Reporting
Category	ESG Scorecard Ranking & Sustainable Portfolio Optimisation
Data Type	<ul style="list-style-type: none"> • Tabular data • Timeseries (Historical)

	<ul style="list-style-type: none"> • Csv, Json data (price, ESG score, metadata) • Txt for the formulas of ESG weightings
Input Data	<p>1) Price data (UC 1 and UC 2) stock data, Forex data, daily data, (in txt and json) formats already provided by KM & JRC as examples and pending approval for formatting in order to expand them further</p> <p>2) Customer profiles for indicative scenarios (UC 1 and UC 2), including risk profile and ESG profile to be provided by both partners (synthetic and indicative scenarios). Profiles being prepared by Fintech partners until end of year.</p> <p>3) ESG scorecard data (UC 1 and UC 2), this list includes ESG scorecard data from approx. 50 EU companies in tabular format – Status Uploaded, pending approval for formatting.</p> <p>4) Metadata, such as (example) News data for potential sentiment extraction and other metadata (UC 2) from companies - news on ESG scores – Indicative datasets provided, more to be uploaded</p>
Expected Output	<ul style="list-style-type: none"> • Provide ESG table of recommendations • Develop a ranking system for stocks and/or bonds based on ESG criteria, alongside fundamental and price-based factors
Type of Analytics	Analytical models to estimate data quality
Input Source	Inside pilot's premises
Output Storage	Available in FAME's dataspace & Inside Pilot's premises
Processing Location	Anywhere. TBD can be in Fintech partner premises for pre-processing and at FAME dataspace for post-processing
Confidential/Private data	<ul style="list-style-type: none"> • ESG Scores – No • Market Data – No, but proprietary streams from Bloomberg used • Client Data – Yes (use synthetic portfolios instead) • HRP and Unicorn Proprietary algorithm data used – Confidential • Weighting Formulas used and derived - No
Is the data part of a business process?	No
Dataset tentative delivery deadline	Indicative Data uploaded in September 2023. Upon approval, Data Delivery in January 2024.
Data Model in use	Timeseries data containing Datetime and numerical sensor values.
Data Accessibility	Not at the moment. Indicative JSON and txt Data available on SharePoint repository for evaluation.
Historical Data	Have to check the volume and how it can be provided. Only indicative data for back-testing will be needed. Full stream of Bloomberg Data is too big and not fully needed.

Table 30 - Datasets Status: Pilot #5 - Use Case #2

Pilot #5 – Use Case #2	
Pilot Leader	KM
Name	Portfolio Optimization
Category	ESG Scorecard Ranking & Sustainable Portfolio Optimisation
Data Type	<ul style="list-style-type: none"> • Tabular data • Timeseries (Historical) • Csv, Json data (price, ESG score, metadata) • Txt for the formulas of ESG weightings • Csv data on Client profile • Csv data on client preferences for optimization parameters
Input Data	<p>1) Price data (UC 1 and UC 2) stock data, Forex data, daily data, (in txt and Json) formats already provided by KM & JRC as examples and pending approval for formatting in order to expand them further.</p> <p>2) Customer profiles for indicative scenarios (UC 1 and UC 2), including risk profile and ESG profile to be provided by both partners (synthetic and indicative scenarios). Profiles being prepared by Fintech partners.</p> <p>3) ESG scorecard data (UC 1 and UC 2), this list includes ESG scorecard data from approx. 50 EU companies in tabular format – Status Uploaded, pending approval for formatting.</p> <p>4) Metadata, such as (example) News data for potential sentiment extraction and other metadata (UC 2) from companies - news on ESG scores – Indicative datasets provided, more to be uploaded.</p>
Expected Output	<ul style="list-style-type: none"> • Provide ESG table of recommendations • Develop a ranking system for stocks and/or bonds based on ESG criteria, alongside fundamental and price-based factors • Develop AI optimization algorithms based on formulas of weightings • Derive indicative ESG portfolio comparison and optimization metrics, based on slider selection of both client needs and ESG preference as well as risk profile • Compare with market sentiment and metadata, in order to further exploit INFINITECH outputs
Type of Analytics	Analytical models to estimate data quality
Input Source	Inside pilot's premises
Output Storage	Available in FAME's dataspace & Inside Pilot's premises.

Processing Location	Anywhere. TBD can be in Fintech partner premises for pre-processing and at FAME dataspace for post-processing.
Confidential/Private data	ESG Scores – No Market Data – No, but proprietary streams from Bloomberg used Client Data – Yes (use synthetic portfolios instead) HRP and Unicorn Proprietary algorithm data used – Confidential Weighting Formulas used and derived - No
Is the data part of a business process?	No
Dataset tentative delivery deadline	Indicative Data uploaded in September 2023. Upon approval, Data Delivery in January 2024.
Data Model in use	Timeseries data containing Datetime and numerical sensor values.
Data Accessibility	Not at the moment. Indicative JSON and txt Data available on SharePoint repository for evaluation.
Historical Data	Have to check the volume and how it can be provided. Only indicative data for back-testing will be needed. Full stream of Bloomberg Data is too big and not fully needed.

3.5.4 Systems to connect to FAME

3.5.4.1 Internal Systems

The 2 Fintech partners will leverage their knowledge and experience with selected outputs from their proprietary portfolio optimization tools, namely:

- HRP Risk Parity Tree from JRC, and
- UNICORN by KM,

Which are protected as background knowledge by corporate confidentiality but will be used for the purpose of the Pilot as sources of initial optimization methods. Furthermore, the formulas of ESG optimization will be based on selected outputs of these proprietary tools.

Furthermore, the internal pilot systems will be selectively used for gathering and cleansing -pre-processing of ESG data as derived from external sources (example Bloomberg). The pre-processed data will include:

- Synthetic client data, but constructed based on anonymized data of combinations of real persons (protected by privacy and GDPR)
- Pre-processed, cleansed Metadata (example newsfeeds)
- The Asset universe (set of stocks, bonds, etfs etc.) to be used as basis for the pilot portfolio optimization.

3.5.4.2 External Data Sources

Financial datafeeds both free, and in the case of proprietary subscription (Bloomberg) to the extent that the license allows, in order to fulfill the pilot needs:

- Bloomberg, Quantl, Yahoo Finance, Trading View et al.

- Stock, Forex, Daily, Historical, ESG scorecards, ESG analytical scores, Asset names, tickers, etc.

3.5.5 Next steps and timeline

First round of indicative data is uploaded. Upon approval and format agreement, the next round of data uploaded will be in January 2024. Algorithms for ESG optimization and tabular selection of Assets are currently being developed theoretically so that they will be coded in the second half of the project.

3.6 Pilot #6: Embedding Climatic Predictions in Property Insurance Products

Table 31 - Pilot 6: Description

Pilot #6	
Pilot Name	Embedding Climatic Predictions in Property Insurance Products
Summary	There is a need for novel data-driven insurance services that incorporate climatic parameters in the estimation of insured liabilities, while facilitating climate aware investment decisions. This pilot will show how the FAME platform can assist these analyses by combining climate information, statistical post processing and user-specific information.
Description	The pilot will use data assets and tools from the FAME marketplace to develop climate aware property insurance UCs. Leveraging raw datasets about property prices and climatic parameters estimation (e.g., Copernicus/DIAS datasets) based on downscaled models from federating sources, the pilot will produce and integrate in FAME novel value-added data assets, including: (i) Data models and datasets for risk analytics in property insurance; (ii) Models for the valuation of fixed assets enhanced with climatic parameters; (iii) Value-at-Risk (VaR) Calculation methods for real estate portfolios of assets, enhanced with climatic parameters; (iv) Local Level Climatic Predictions based on downscaled prediction models; (v) External data sources (e.g., energy prices, carbon prices, Real GDP estimation, inflation estimates, government bond yields, central bank rates, real-estate prices projections, etc.
Motivation	Understand the effect on climate change will become an increasingly important component for corporations to take into account, including financial organizations. However, climate data is currently difficult for non-specialistics to access and make use of. This pilot will show how FAME can help bridge that gap.

3.6.1 Use Cases

3.6.1.1 Use Case #1

Table 32 - Pilot 6: Use Case 1

Use Case #1	
Use Case Name	Climate Aware Real Estate Pricing
Summary	Using FAME, we will show how historical and projected climate information can be accessed that will allow an insurance company to model the changes in real estate prices expected by climate change.

Description	This use case will project valuations of real-estate assets based on local-level predictions. It will use the climatic-enhanced models for property pricing (e.g., enhanced indemnity basis method and the enhanced Reinstatement Value Basis approach) to produce the climate aware valuations of a list of assets. The differences produced as part of the asset repricing will be compared to known information about damages and compensations associated with hazardous events that impacted real estate asset.
Value Proposition(s)	<ul style="list-style-type: none"> • Inform pricing of real estate insurance products • Inform risk analysis related to real estate valuations
Documentation	<ul style="list-style-type: none"> • Documentation on statistical post processing and downscaling of climate projections to be provided. <p>Example code showing how analysis can then be performed.</p>
Image(s)	-
Website	-
Keywords	-
Available Data Assets	Statistically downscaled climate projections

3.6.1.2 Use Case #2

Table 33 - Pilot 6: Use Case 2

Use Case #2	
Use Case Name	VaR of an Insurers' Portfolio
Summary	This use case will show how the FAME platform can be used to understand the manner by which an insurer's Value at Risk (VaR) will change on account of climate change.
Description	This use case will calculate the climate aware VaR of entire portfolios of assets possessed by insurers, including bonds, stocks and real-estate assets. It will integrate climate risk assessments in the VaR calculation, based on the enhanced VaR Calculators (e.g., Historic Method, Variance-Covariance Methods, Monte Carlo Simulations). The climate enhanced VaR calculations will be compared to conventional calculations, as well as to what happened in past hazards that resulted in lower prices for assets.
Value Proposition(s)	<ul style="list-style-type: none"> • Improved risk management • Better understanding of the potential for leverage
Documentation	<ul style="list-style-type: none"> • Documentation on statistical post processing and downscaling of climate projections to be provided <p>Example code showing how analysis can then be performed</p>
Image(s)	-
Website	-
Keywords	-
Available Data Assets	Statistically downscaled climate projections

3.6.1.3 Use Case #3

Table 34 - Pilot 6: Use Case 3

Use Case #3	
Use Case Name	Climate-Aware Analysis of Alternative Portfolios
Summary	Each asset in a portfolio is expected to have an exposure to climate change. Taking this exposure into account will allow for alternative portfolios to be considered which may outperform current portfolios.
Description	This use case will analyse different portfolios of assets in the light of climate change. It will offer a tool for insurance companies to identify how to restructure their portfolio in order to be less susceptible to climate change.
Value Proposition(s)	Higher financial returns by incorporating climate change risk into considerations of portfolios.
Documentation	<ul style="list-style-type: none"> Documentation on statistical post processing and downscaling of climate projections to be provided Example code showing how analysis can then be performed
Image(s)	-
Website	-
Keywords	-
Available Data Assets	Statistically downscaled climate projections

3.6.2 Business Requirements

The goal of Pilot 6 is to develop new climate-conscious property insurance products. It will leverage FAME's value-added datasets to project real estate asset valuations based on local-level forecasts, as well as to calculate climate-aware VaR for entire portfolios of assets held by insurers, including bonds, stocks, and real estate. This will analyse different asset portfolios in light of climate change.

Table 35 - Pilot 6: Business Requirements

Business Req. ID	Use Case ID	Business Objective	Business Requirement
P6_BR1	UC 1	Supply featurised climate projections for a specific location	FAME must be able to train a statistical downscaling model that relates coarse grid climate projections to finer grid ERA5 Land data
P6_BR2	UC 1	Supply featurised climate projections for a specific location	FAME must connect to CDS to download historical reanalysis data (ERA5 Land)
P6_BR3	UC 1	Supply featurised climate projections for a specific location	FAME must connect to CDS (Copernicus data store) to download climate project data (CMIP5) and historical reanalysis
P6_BR4	UC 1	Supply featurised climate projections for a specific location	FAME must be able to create ML model to downscale climate projections
P6_BR5	UC 1	Supply featurised climate projections for a specific location	An Analysts can find that the climate risk feature for a specific location exists

P6_BR6	UC 1	Forecast the effect climate change will have on property prices for a specific asset and location	An analysts can find that the model exists which allows climate risk to be calculated for a user-supplied location and asset
P6_BR7	UC 1	Forecast the effect climate change will have on property prices for a specific asset and location	FAME should offer value to an organization that uploads property price data
P6_BR8	UC 1	Ensure the sovereignty of the data being traded	FAME should have mechanisms to ensure data usage control policies
P6_BR9	UC 1	Supply featurized climate projections for a specific location	An analyst can purchase the featurized climate projection for a specific location
P6_BR10	UC 2	Train insurance sector professionals how to perform climate-aware Real Estate Pricing	Development of financial courses, Webinars, How-to videos, Jupyter notebooks
P6_BR11	UC 1	Supply featurised climate projections for a specific location	FAME must be able to take the downscaling model and future climate projections to supply a projection of changes in climate project features at any given location
P6_BR12	UC 1	Forecasts the effect climate change will have on property prices for a specific asset and location	Historical property price information will need to be uploaded that has both prices and asset characteristics
P6_BR13	UC 1	Forecast the effect climate change will have on property prices for a specific asset and location	A model will need to be trained that links the downscaled climate features to fluctuations in historical property prices
P6_BR14	UC 1	Forecast the effect climate change will have on property prices for a specific asset and location	An analyst can purchase a projected effect on the real estate price for a supplied location and the featurized climate projection for a specific location
P6_BR15	UC 1	Forecast the effect climate change will have on property prices for a specific asset and location	A model will need to be trained that links the downscaled climate features to fluctuations in historical property prices that takes asset characteristics and location into account
P6_BR16	UC 1	Forecast the effect climate change will have on property prices for a specific asset and location	A new asset and location will need to be provided, for which the model and downscaled climate features will be used to forecast property price changes
P6_BR17	UC 2	Supply Seasonal Forecasts of Climate Risk Features	An analyst can purchase the seasonal forecast of a catalogue of climate risk features
P6_BR18	UC 2	Supply Seasonal Forecasts of Climate Risk Features	FAME must connect to CDS to download seasonal forecasts
P6_BR19	UC 2	Supply Seasonal Forecasts of Climate Risk Factors	FAME must connect to CDS to download ERA5 Reanalysis Data

P6_BR20	UC 2	Supply Seasonal Forecasts of Climate Risk Factors	FAME must bias correct seasonal forecasts relative to the ERA5 data, to then derive the corrected forecasts of climate risk features
P6_BR21	UC 2	Supply Seasonal Forecasts of Climate Risk Factors	An analyst can find that a seasonal forecast of climate risk factors exists
P6_BR22	UC 2	Determine Risk Exposure of an Asset to Climate Risk Features	An analyst can find that a climate-aware VAR calculation for a given asset exists
P6_BR23	UC 1	Train insurance/finance sector professionals how to climate-aware assess their portfolios	Development of financial courses, Webinars, How-to videos, Jupyter notebooks
P6_BR24	UC 1	Ensure the sovereignty of the data being traded	FAME should have mechanisms to ensure data usage control policies
P6_BR25	UC 2	Determine Risk Exposure of an Asset to Climate Risk Features	FAME must be able to download historical assets prices for e.g. a given equity
P6_BR26	UC 2	Determine Risk Exposure of an Asset to Climate Risk Features	FAME must be able to run a quantile regression model that can determine the extent to which historical asset price volatility was correlated with climate risk features
P6_BR27	UC 2	Determine Risk Exposure of an Asset to Climate Risk Features	FAME must be able to load up seasonal forecasts of the climate risk features and feed them through the asset-specific risk model to obtain a climate aware VAR estimate for the next six months
P6_BR28	UC 3	Assess the climate risk exposure of a portfolio	FAME must be able to download historical prices for a portfolio of assets
P6_BR29	UC 3	Assess the climate risk exposure of a portfolio	For each asset in the portfolio, FAME must be able to model the "risk premium" and "risk" related to climate features from (UC 1)
P6_BR30	UC 3	Assess the climate risk exposure of a portfolio	For each asset in the portfolio, FAME must be able to model the portfolio level excess return over a supplied index and risk from this portfolio related to climate risk
P6_BR31	UC 3	Assess the climate risk exposure of a portfolio	An analyst can find that this tool exists and can upload their portfolio
P6_BR32	UC 2	Ensure the sovereignty of the data being traded	FAME should have mechanisms to ensure data usage control policies
P6_BR33	UC 3	Propose climate-safe indices	FAME should allow external users access to our tools to enable them to propose new stock indices which are "climate proof"
P6_BR34	UC 3	Assess the climate risk exposure of a portfolio	For a given portfolio, FAME must be able to indicate how excess returns and risk will change on account of changes in climate features

P6_BR35	UC 3	Propose Restructuring	portfolio	FAME must be able to determine which assets to drop from an existing portfolio in order to improve performance based on changes in projected climate risk features
----------------	------	--------------------------	-----------	--

3.6.3 Datasets Status

As mentioned in the previous chapter, the collection of datasets was prepared by WP 5. This information is divided into Use Cases for each pilot and this may change during the course of the project.

Table 36 - Datasets Status: Pilot #6 - Use Case #1

Pilot #6 – Use Case #1	
Pilot Leader	NRS
Name	Climate Aware Real Estate Pricing
Category	Embedding Climatic Predictions in Property Insurance Products
Data Type	Spatial Timeseries
Input Data	1 TB
Expected Output	Regression
Type of Analytics	Potentially XAI, but we will likely use parametric models.
Input Source	Migrated
Output Storage	Available from FAME
Processing Location	Anywhere
Confidential/Private data	No
Is the data part of a business process?	No
Dataset tentative delivery deadline	No
Data Model in use	Jan 2024
Data Accessibility	Netcdf data for spatial timeseries
Historical Data	Not at the moment

Table 37 - Datasets Status: Pilot #6 - Use Case #2

Pilot #6 – Use Case #2	
Pilot Leader	NRS
Name	VaR of an Insurers' Portfolio
Category	Embedding Climatic Predictions in Property Insurance Products.
Data Type	Spatial Timeseries
Input Data	1 TB
Expected Output	Regression
Type of Analytics	Migrated
Input Source	Available from FAME
Output Storage	Anywhere
Processing Location	No

Confidential/Private data	No
Is the data part of a business process?	No
Dataset tentative delivery deadline	Jan 2024
Data Model in use	Netcdf data for spatial timeseries
Data Accessibility	Not at the moment
Historical Data	Yes

Table 38 - Datasets Status: Pilot #6 - Use Case #3

Pilot #6 – Use Case #3	
Pilot Leader	NRS
Name	Climate-Aware Analysis of Alternative Portfolios
Category	Embedding Climatic Predictions in Property Insurance Products.
Data Type	Spatial Timeseries
Input Data	1 TB
Expected Output	Regression
Type of Analytics	Migrated
Input Source	Available from FAME
Output Storage	Anywhere
Processing Location	No
Confidential/Private data	No
Is the data part of a business process?	No
Dataset tentative delivery deadline	Jan 2024
Data Model in use	Netcdf data for spatial timeseries
Data Accessibility	Not at the moment
Historical Data	Yes

3.6.4 Systems to connect to FAME

3.6.4.1 Internal Systems

NRS will use its extensive experience with climate projections to create a curated dataset of climate projections, regionalization methods and statistical bias correction techniques to upload an ensemble climate projection dataset.

In addition, NRS will upload an observational dataset created by meshing together multiple observational products. Finally, NRS will provide the core code and logic necessary to combine these two databases in order to—on the fly—create a downscaled climate projection for any location requested by a user of FAME.

3.6.4.2 External Data Sources

The core datasets used by NRS will come from the Copernicus climate datastore.

3.6.5 Next steps and timeline

The first exemplars of the data will be uploaded by the end of 2023. Subsequent refinements to the methodology are expected throughout 2024. Example scripts that show how the use cases can be accomplished using the uploaded data and the FAME platform more generally will be provided by the end of 2024.

3.7 Pilot #7: Assessing the Quality and Monetary Value of Data Assets

Table 39 - Pilot 7: Description

Pilot #7	
Pilot Name	Assessing the Quality and Monetary Value of Data Assets (MOH, INNOV, JSI, UIA)
Summary	This pilot will assess the quality of different types of MOH's data assets, while using the quality assessment and the type of each data asset for pricing and trading purposes inside the FAME marketplace. Information on the value of the data assets will be accordingly used to estimate premiums for potential cyber-insurance services.
Description	This pilot encompasses a holistic evaluation of MOH's existing and potential data assets derived from the various sensors implemented in different parts of its production chain. Leveraging the FAME marketplace, the pilot aims to assess, price, and if judged beneficial, trade these data assets while analysing their potential for enhancing cyber-insurance services. The pilot encompasses a meticulous quality analysis, including examining the data assets' volume, completeness, locality and context, and other relevant attributes. The ultimate objective is to unearth the hidden value in the data assets to pave the way for strategic investments and robust ML systems that can foresee potential challenges in assets' behaviours.
Motivation	For nearly a decade MOH is heavily investing in the digitization of its production based on the introduction of sensors and cyber-physical systems in its oil refineries and on the implementation of novel use cases. The company manages a great amount of raw data from various sensors (e.g., pressure, thermometers, vibration, axial displacement), which is analysed hourly to identify abnormalities, to predict compressors' behaviour for the next 24 hours and to identify potential problems ahead of time. The company's experience from the deployments has unveiled the importance of the quality of the data assets, which is critical for implementing effective ML systems that can accurately predict assets' behaviours. Leveraging the FAME marketplace, MOH is interesting in extracting valuable insights on the quality and potential monetary value of its data assets, to drive the exploitation and use of these assets in processes like cyber-insurance contracts.

3.7.1 Use Cases

3.7.1.1 Use Case #1

Table 40 - Pilot 7: Use Case 1

Use Case #1	
Use Case Name	Quality Assessment and Pricing of Industrial Data Assets
Summary	Different data assets will be assessed against their quality, including: (i) Raw sensor datasets; (ii) Pre-processed, filtered and labelled datasets; (iii) Predictive models (ML models) and algorithms. The different assets will be audited against different characteristics, including their volume, completeness, locality and context, variety of data sources, use in industrial applications etc. A quality score will be assigned to each of the assets to allow their comparative evaluation and to foster the implementation of trading and pricing schemes inside the marketplace. The pricing schemes will be confronted to the actual work allocated to advance the quality of the assets, including cleansing, pre-processing, filtering and analytics. This will provide MOH with invaluable insights on the potential ROI (Return on Investment) of tedious processes like data labelling. Likewise, an insured value/premium for the various data assets will be specified to be used for future cyber-insurance calculation.
Description	In this use case, MOH aims to carry out a comprehensive assessment of different industrial data assets, ranking them based on their quality through various established metrics, such as volume and completeness. This process involves devising a pricing mechanism for these assets, correlating with the work put into enhancing their quality. It seeks to establish a harmonious trading ecosystem within the FAME marketplace where assets can be evaluated and traded based on a defined quality score. Simultaneously, it lays the groundwork for establishing an insured value/premium for the cyber insurance of these data assets, founded on their calculated monetary value.
Value Proposition(s)	Bullet list with value proposition(s) of the Use Case: <ul style="list-style-type: none"> • Facilitates a systematic approach to assess the quality of different data assets. • Unlocks the potential ROI on data pre-processing activities. • Creates a system for defining insured values for cyber-insurance services based on data asset quality.
Documentation	N/A
Image(s)	N/A
Website	N/A
Keywords	data duality, IoT, data profiling, industrial sensor data
Available Data Assets	Raw data from ~150 sensors monitoring 5 machine groups

3.7.1.2 Use Case #2

Table 41 - Pilot 7: Use Case 2

Use Case #2	
Use Case Name	Data Analysis and Production of Value-Added assets using FAME Tools

Summary	The project's analytics tools, including XAI and energy efficient analytics, will be used to develop new ML models on top of existing data. The new assets will be made available in the marketplace as trusted and energy efficient assets. They will be priced, traded, and compared to other assets produced outside FAME to showcase the value of trust and energy efficiency.
Description	This use case revolves around utilizing FAME project's advanced analytics tools to craft new, trustworthy, and energy-efficient ML models from existing data. The fresh set of data assets thus created will be introduced into the FAME marketplace, emphasizing their trusted nature and energy efficiency. It envisages creating a competitive edge for assets fostered within FAME, highlighting the added value in terms of trust and energy efficiency compared to assets originating from outside FAME.
Value Proposition(s)	Bullet list with value proposition(s) of the Use Case: <ul style="list-style-type: none"> • Facilitates the creation of trusted secondary data assets. • Allows the monetization of confidential data without exposing them.
Documentation	N/A
Image(s)	N/A
Website	N/A
Keywords	XAI, predictive maintenance, analytics
Available Data Assets	Raw data from ~150 sensors monitoring 5 machine groups

3.7.2 Business Requirements

The objective of Pilot 7 is to assess the quality of your different types of data assets, while also utilizing the quality assessment and type of each data asset for pricing and trading purposes within the FAME market. To make this possible, the Pilot will audit its assets for different characteristics, including their volume, completeness, location and context, variety of data sources, use in industrial applications, etc., while developing different pricing schemes for trading on the FAME market. Ultimately, the pilot aims to develop new ML models based on existing XAI-based data and energy efficiency analytics that will be traded in the market.

Table 42 - Pilot 7: Business Requirements

Business Req. ID	Use Case ID	Business Objective	Business Requirement
P7_BR1	UC 1	Quality Assessment of (IoT) Data Assets	Perform quantitative quality assessment of data assets considering various quality dimensions such as data accuracy, completeness, consistency, timeliness, and validity
P7_BR2	UC 1	Indicative Pricing of Data Assets (ML Models, Labelled data) used for Maintenance	FAME should be able to assign a quality score to each asset for comparison and pricing purposes

P7_BR3	UC 1	Indicative Pricing of Data Assets (ML Models, Labelled data) used for Maintenance	Estimate the value of each data asset based on factors such as usage, impact on business processes, accuracy, and maintenance costs. The valuation should be flexible, allowing for adjustments based on specific business requirements and market conditions
P7_BR4	UC 1	Ensure the sovereignty of the data being traded	FAME should have mechanisms to ensure data usage control policies
P7_BR5	UC 1	Re-purpose/reuse/sell Data Assets for Industrial Maintenance	Develop analytical models providing operational insights on equipment used in industrial environments
P7_BR6	UC 1	Re-purpose/reuse/sell Data Assets for Industrial Maintenance	Trade of industrial data assets (i.e., sensor data, AI/ML models) through FAME
P7_BR7	UC 1	Improving Overall Equipment Effectiveness (OEE) through Predictive Maintenance	Obtain more precise predictions in the scope of preventive/predictive maintenance by developing more accurate predictive models than the existing ones. These models should ensure zero downtime and increase overall equipment effectiveness (OEE).
P7_BR8	UC 1	Increase in the utilization of the available data assets for predictive maintenance purposes	FAME should provide tools that allow the curation of multi-type data produced from different sources
P7_BR9	UC 1	Streamline decision-making by assessing the quality and value of data assets	Identify underutilized resources and opportunities for further investment
P7_BR10	UC 1	Train industrial workers on how to assess and understand data produced by IoT devices and sensors	Develop relevant training materials
P7_BR11	UC 2	Increase Trustworthiness of AI models used in IoT	An Analyst can search for an XAI solution based on criteria such type of the underlying ML model and data.
P7_BR12	UC 2	Ensure the sovereignty of the data being traded	FAME should have mechanisms to ensure data usage control policies
P7_BR13	UC 2	Stakeholder Interaction	FAME should allow access to the primary data assets to develop secondary data assets and applications
P7_BR14	UC 2	Increase Trustworthiness of AI models used in IoT	Develop XAI techniques for timeseries forecasting models
P7_BR15	UC 2	Increase Trustworthiness of AI models	FAME should provide explanations of results of AI (ML) models in proper form
P7_BR16	UC 2	Increase acceptance of novel AI-based systems by industrial workers	Train industrial workers on how to assess, use, and interpret the outcomes of AI/XAI Systems related to machinery health
P7_BR17	UC 2	Increase Trustworthiness of AI models in order to	Provide explanations of results of AI (ML) models in proper form and users should be aware of those explanations and be able to use

use AI models in proper those explanations in order to update their
manner (AI Governance) business and organizational processes

3.7.3 Datasets Status

As mentioned in the previous chapter, the collection of datasets was prepared by WP 5. This information is divided into Use Cases for each pilot and this may change during the course of the project.

Table 43 - Datasets Status: Pilot #7 - Use Case #1

Pilot #7 – Use Case #1	
Pilot Leader	MOH
Name	Quality Assessment and Pricing of Industrial Data Assets
Category	Assessing the Quality and Monetary Value of Data Assets
Data Type	Timeseries
Input Data	<ul style="list-style-type: none"> • Type <ul style="list-style-type: none"> ○ Sensor data ○ Machine group ID ○ Maintenance data • Size <ul style="list-style-type: none"> ○ 8 TB
Expected Output	Data assessment
Type of Analytics	Analytical models to estimate data quality
Input Source	Into pilot's premises (Data warehouse/SAP BTS)
Output Storage	Into pilot's premises (Data warehouse/SAP BTS)
Processing Location	MOH, INNOV, JSI (with NDA)
Confidential/Private data	Confidential data (Have to check again with chief information security officer about sharing policies).
Is the data part of a business process?	No
Dataset tentative delivery deadline	December 2023 (TBD)
Data Model in use	Timeseries data containing Datetime and numerical sensor values of several machines
Data Accessibility	Sample IoT data from different types of sensors and machines
Historical Data	Yes

Table 44 - Datasets Status: Pilot #7 - Use Case #2

Pilot #7 – Use Case #2	
Pilot Leader	MOH
Name	Data Analysis and Production of Value-Added assets using FAME Tools
Category	Assessing the Quality and Monetary Value of Data Assets

Data Type	Timeseries
Input Data	<ul style="list-style-type: none"> • Type <ul style="list-style-type: none"> ○ Sensor data ○ Machine group ID ○ Maintenance data • Size <p>8 TB</p>
Expected Output	Classification (alerts) & Regression (forecasting).
Type of Analytics	XAI models for explaining timeseries forecasting models (Prophet).
Input Source	Into pilot’s premises (Data warehouse/SAP BTS)
Output Storage	Into pilot’s premises (Data warehouse/SAP BTS)
Processing Location	MOH, INNOV, JSI (with NDA)
Confidential/Private data	Confidential data (Have to check again with chief information security officer about sharing policies).
Is the data part of a business process?	No
Dataset tentative delivery deadline	September 2023 (TBD)
Data Model in use	Timeseries data containing Datetime and numerical sensor values of several machines.
Data Accessibility	Not at the moment
Historical Data	Yes, but have to check the volume and how it can be provided

3.7.4 Systems to connect to FAME

The architectural framework of Pilot #7 is tailored for efficient data processing and advanced predictive analytics in a refinery setting. Data collection begins with Refinery Machines outfitted with Sensors, capturing real-time operational data. This data is relayed to the Distributed Control System (DSC), which performs initial data aggregation and annotation, including sensor identification and type, at an increased frequency for manageability.

Subsequently, data flows to the Historian, a time-series database that logs data with corresponding timestamps and granularity. This database is pivotal for advanced analytics and visualization. Furthermore, the pilot integrates an SAP component for enterprise resource planning, emphasizing preventive and predictive maintenance functionalities.

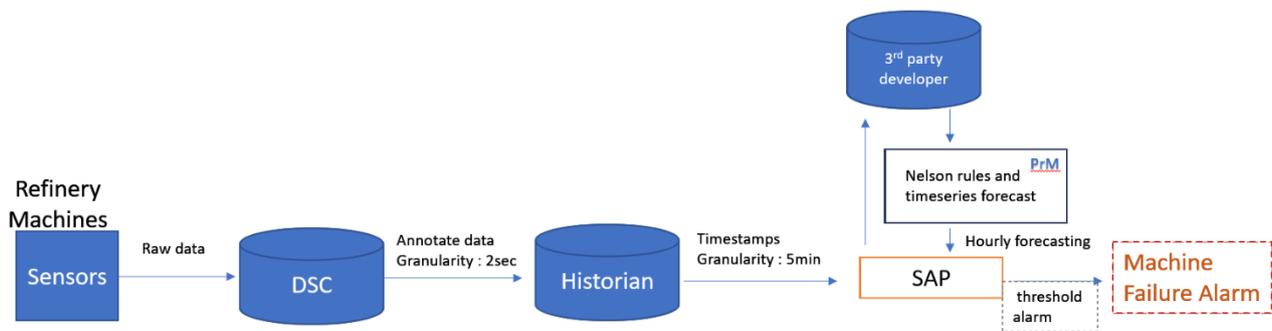


Figure 8 - Pilot #7 Internal Architecture

3.7.4.1 Internal Systems

To align with MOH's security protocols, the pilot introduces an API layer to securely manage data asset interactions with FAME. This layer, illustrated in Figure 7 (green components), provides a controlled environment for data exchange, ensuring compliance with privacy and critical infrastructure protection standards. The components within this layer include:

- **Datastore:** Receives and stores new data batches uploaded by MOH.
- **Data Quality Assessment Service:** Evaluates the incoming data's integrity and quality.
- **Analytics Service:** Generates insights from the assessed data, contributing to informed decision-making within FAME.

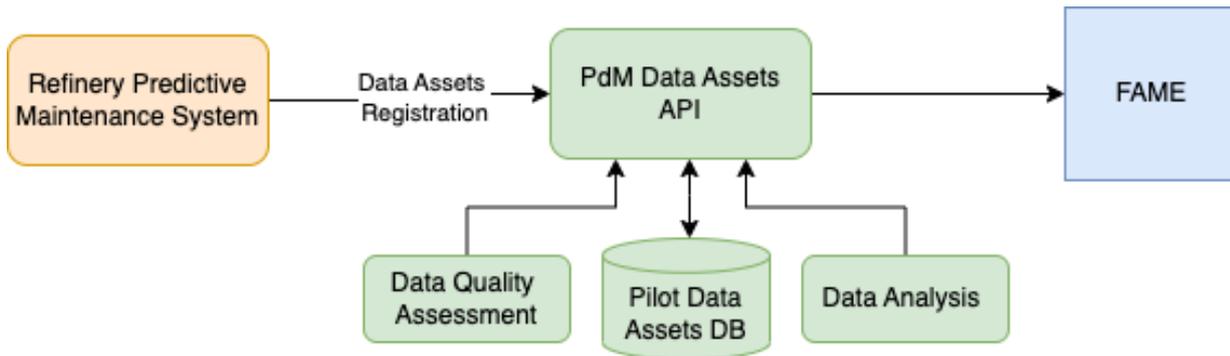


Figure 9 - Pilot #7 Internal to FAME components

3.7.4.2 External Data Sources

Due to the security and privacy imperatives, the processes in Figure 8 are external to FAME. These processes encompass sensor data, predictive forecasts from existing models, and alarm or failure notifications. Nonetheless, selected historical, annotated, and lower-frequency sensor data will be accessible to FAME through the PdM Data Assets API, which will operate in batch mode to ensure controlled data flow.

3.7.5 Next steps and timeline

By M18, Pilot #7 aims to establish the necessary API endpoints and develop the PdM Data Assets API and its associated database to facilitate data integration with FAME. Concurrently, the pilot will integrate the Data Quality Assessment service, enhancing data with quality annotations. These annotations will assess the reliability of datasets, such as sensor data linked to specific machinery, ultimately supporting more astute transactions within the marketplace.

4 Conclusions

This deliverable is the first of a total of 2 deliverables, the second being D6.4 which will be delivered in M21. These are part of Task 6.1 "Use Cases Co-Creation and Sites Preparation" and aim to "undertake all preparatory actions towards the integration, deployment, and evaluation of the project's UCs at the pilot sites. Specifically, it will co-create the use cases through the organization of proper workshops and other co-creation activities. Likewise, it will mobilize, engage and when required train all relevant stakeholders including data providers, data analysts, and non-tech users. Furthermore, it will collect and prepare the data assets for integration into the FAME marketplace, and it will undertake preparatory activities for using the FAME platform and its services."

All the work described here is the beginning of the pilots' process at FAME, this will create the foundations so that the pilots can develop their work during the project.

Due to some setbacks that occurred at the beginning of the project, namely the restructuring of partners in the case of pilot 3 and 4, the information is not complete, but it will be resolved during the course of the project.

In chapter 2 of the document, a brief overview of the pilot is given, namely, its leader, name and use cases. During the first period of the project, a plan was outlined to help the pilots progress. This plan consists of 4 steps, which are:

- Pilots and Use Cases Description Updates: At this stage of the plan, a template was created to collect information from the pilot and its Use Cases.
- Co-Creation Requirement Collection (WP 2): This phase was prepared by WP 2 partners, who collected information about the business requirements of each pilot.
- Datasets Available (WP 5): This phase, like the previous one, was also carried out by the WP 5 partners, who created an Excel with several topics so that they could collect the datasets for each pilot.
- Systems (e.g. APIs) to connect to FAME: At this stage, firstly, it was necessary to identify the systems that the pilots already have in use and that can share data with FAME and secondly, to identify datasets or external services that the pilots find necessary, or that are useful for their development in the project.

In chapter 3, each phase of the plan that was described in the previous chapter is discussed in depth. Each pilot will have a first section, dedicated to the description of their pilot and their use cases, using the template that was made available to the pilot. In the business requirements section, each pilot has a table containing the business objectives and the business requirements for each business requirement ID. In the case of Pilot 1, he has a total of 33 Business Requirements, Pilot 2 has a total of 15 BR, Pilot 3 and Pilot 4, due to the situation of restructuring pilots need to do this phase, Pilot 5 has a total of 21 BR, Pilot 6 has a total of 35 and finally, Pilot 7 has a total of 17 BR. In total, there are 121 Business Requirements. In the Datasets section it is divided into Use Cases for each pilot, and each table has as topics the name, category, data type, input data, expected output, type of analytics, input source, output storage, processing location, confidential/private data, is the data part of a business process, dataset tentative delivery deadline, data model in use, data accessibility and historical data. In the systems to connect to FAME section, each pilot identified the internal systems and the external data sources that they require. Finally, each pilot describes the next steps that will follow.

4.1 Next Steps

All information described in this deliverable may change during the course of the project and information that is currently missing will be completed and available in D6.2, as previously mentioned.

One of the next steps will quantify and qualify the kpis, so that we know the values and what they intend to achieve. Another important factor is defining who the stakeholder engagement is, that is, who will use FAME and benefit from the FAME tool.