Tailored IoT & BigData Sandboxes and Testbeds for Smart, Autonomous and Personalized Services in the European Finance and Insurance Services Ecosystem

D7.20 - Pilots' Evaluation and Stakeholders' Feedback - I

Title	D7.20 - Pilots' Evaluation and Stakeholders' Feedback - I
Revision Number	3.0
Task reference	T7.8
Lead Beneficiary	ABILAB
Responsible	Gabriele Gamberi
Partners	ABILAB AGRO AKTIF ATOS BANKIA BOC BOI BOS BOUN BPFI CP CTAG DYN ENG GEN GFT ISPRINT JRC JSI LIB NBG PI PRIVE RRD SIA SILO UNP WEA
Deliverable Type	Report
Dissemination Level	PU
Due Date	2021-06-30 [M21]
Delivered Date	2021-07-16 08:07:36
Internal Reviewers	CTAG RRD
Quality Assurance	INNOV
Acceptance	Accepted by Coordinator
EC Project Officer	Pierre-Paul Sondag
Programme	HORIZON 2020 - ICT-11-2018
	This project has received funding from the European Union's horizon 2020 research and innovation programme under Grant Agreement no 856632

Contributing Partners

Partner	Role[1]	Author(s)[2]	Section(s)
ABILAB	Lead Beneficiary	Gabriele Gamberi	0. 1. 2. 3. 3.1. 3.2. 3.3. 3.4. 3.5. 3.6. 3.7. 3.8. 3.9. 3.10. 3.11. 3.12. 3.13. 3.14. 3.15. 3.16.
ABILAB	Contributors		0. 1. 2. 3. 3.15. 3.16. 4. 5.
ABILAB BANKIA	Contributors		3.1.
ABILAB JRC	Contributors		3.2.
ABILAB BOI	Contributors		3.3.
ABILAB PRIVE	Contributors		3.4.
ABILAB BOC	Contributors		3.5.
ABILAB NBG	Contributors		3.6.
ABILAB CXB	Contributors		3.7.
ABILAB BOS	Contributors		3.8.
ABILAB AKTIF	Contributors		3.9.
ABILAB PI	Contributors		3.10.
ABILAB ATOS	Contributors		3.11.
ABILAB SILO	Contributors		3.12.
ABILAB WEA	Contributors		3.13.
ABILAB GEN	Contributors		3.14.
CTAG	Reviewer		all sections
RRD	Reviewer		all sections
INNOV	Quality Assurance		all sections

1. Lead Beneficiary, Contributor, Internal Reviewer, Quality Assurance
2. Can be left void

Revison History

Version	Date	Partners	Description
0.1	2021-06-18		TOC and preliminary contents

Version	Date	Partners	Description
0.9	2021-06-30		Contributions from Pilots
1.0	2021-06-30		Version for Internal Review
2.0	2021-07-10		Version for QA
2.9	2021-07-13		Version after QA
3.0	2021-07-16		Version for Submission

Executive Summary

One of the main objectives of the INFINITECH project is to introduce, validate and evaluate advanced BigData and AI-based Digital Finance services in real-life pilot settings. This deliverable (D7.20 – Pilots' Evaluation and Stakeholders' Feedback – I) is devoted to set up the basis of the framework that will be used to evaluate the fifteen pilots that make up the whole project. As a matter of fact, this deliverable represents the inception of a sequentially driven strategy, a.k.a. Evaluation Framework, made up of two main phases: the first phase, is meant to find across-the-board Key Performance Indicators (KPIs) as to obtain a standardized way of evaluating the pilots; the second phase is focused on opening the way to a full-fledged periodic evaluation of the pilots' progress (which will be subject to future refinements as to improve its efficacy in its profiling activities, as well as to reduce the pilots' burden into providing periodic feedback).

For this reason, within the deliverable it is not yet presented any outcome of the evaluation as the framework is currently in a preliminary stage. It is therefore illustrated the status of the Pilots' KPIs, as to embody an initial snapshot from which to base the evaluation, as well as the methodology that will be used to carry out the future monitoring process (which belongs to the second phase of the framework).

Such outputs are based on the continuous interaction with the pilots, as well as the WP7 task leaders, from which ABI Lab obtained an understanding of the aspects of the diverse use-cases, such as figuring out the needs over the KPIs, finding a proper terminology to encompass all pilots' use-cases, defining the number of requested KPIs per category, standardizing their format, who are the pilots that already achieved their KPIs, etc. These aspects have all been discussed in the deliverable.

Table of Contents

Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
AML	Anti Money Laundering
API	Application Programming Interface
AWS	Amazon Web Services
BERT	Bidirectional Encoder Representations from Transformers
BFM	Business Financial Management
BOC	Bank of Cyprus
BOI	Bank of Ireland
BOS	Bank of Slovenia
DL	Deep Learning
ETF	Electronic Toll Fraud
ETSI	European Telecommunications Standards Institute
FI	Financial Innovation (INFINITECH beneficiary)
FIU	Financial Investigations Unit
GDPR	General Data Protection Regulation
HPC	High Performance Computing
IBM	International Business Machines
IT	Information Technology
IoT	Internet of Things
KOM	Kick Off Meeting
KPI	Key Performance Indicator
KYB	Know Your Business
KYC	Know Your Customer
ML	Machine Language
MRVF	Min Return gap Vol Factor
MVP	Minimum Viable Product|Platform
N/A	Not Available / Not Applicable
NLU	Natural Language Understanding
OCR	Optical Character Recognition
PDF	Portable Document Format (for Adobe Acrobat Reader)
SME	Smalland Medium-Sized Enterprises
SOC	Security Operations Center
SRF	Sharpe Ratio Factor

Abbreviation	Definition
UI	User Interface
VaR	Value at risk
WP	Work Package
WP7	Work Package 7 dealing with Pilots

1 Introduction

1.1 Objective of the Deliverable

One of the main objectives of the Task 7.8, from which the deliverable at hand is being conceived, is to collect stakeholders' feedback for all pilots, as well as evaluating all pilot systems. Such evaluation will be multi-facet, including technical evaluation, techno-economic evaluation, evaluation of usability, evaluation of customer satisfaction and more. A set of common standards will be devised and applied to manage the evaluation process in a unified way. To this end, each pilot will be evaluated against pilot specific KPIs, as illustrated in the pilot descriptions.

The approach followed for the evaluation is based on a shared-responsibility logic, offering a methodological landscape that will allow the pilots to adopt their own peculiar strategy, as well as the tools, as to better tailor them to their specific needs. Indeed, it is paramount that the Task 7.8 does not regulate, nor impose constraints, nor other binding rules to the project's pilots by imposing a rigid model. On the contrary, the Task focuses on an agile approach to better support the innovation within the INFINITECH project. Such methodological background, which is the driver of all the task's activities, has been agreed upon with the pilots at the very inception of the task.

The purpose of this deliverable is to introduce and describe the approach and the methodology used to start the pilots' evaluation framework. Furthermore, the deliverable will illustrate a first draft of the KPI structure along with their status reporting, as to depict the current pilots' situation within the INFINITECH project.

1.2 Insights from other Tasks and Deliverables

The Task 7.8 takes input from the whole WP7 where the definition of all use cases takes place. Such inputs include, among a wide array of aspects, the technologies stored by the WP7 task leaders within dedicated spreadsheets, make use of the cluster meetings as to collect pilots' experiences and challenges, as well as understanding of the use cases implemented within the project.

Out of the synergy gained within the scope of WP7, a two-dimensional model has been devised. Such model will consider both operational and business-related aspects, enabling the task to evaluate the potential of the platform in supporting innovative initiatives.

1.3 Structure

The deliverable at hand is structured as follows:

- Section 2 outlines the methodology employed for the evaluation process, as well as the following sub-sections:
  - Sub-section 2.1: lists the Task's achievements.
  - Sub-section 2.2: introduces the Evaluation Framework.
  - Sub-section 2.3: describes the monitoring process (surveys).
- Section 3 collects the KPIs identified by each pilot, with the goal to show an overall picture of what the pilots will be measuring in the future, as well as to display a straightforward summary of the current pilots' status.
- Section 4 contains the conclusions, namely Task's insights and plan for future actions.

2 Methodology

The INFINITECH project is known to be a very complex environment, in which there are a broad spectrum of differentiated use cases which embrace a wide range of business types. Above all, we are working on the edge of innovation.

In the light of this scenario, it is essential to come up with a solid strategy with which to define a methodology that measures the effectiveness and the value that INFINITECH can give to the Pilots.

To this end, we the Task 7.8 started from a path of exchange and discussion with the various pilots, and we met several times during the periodic Cluster Meetings as to share our experiences. Such experiences were expressed in terms of either the business side (related to the user, as well as the technology meant for the business objectives) and the technical side with the pilots' developers.

From all these interactions, it turned out that the topic of measuring is considered extremely valuable for pilots, but it was also very laborious to fully standardize a measurement platform for all possible use cases developed within the INFINITECH project.

In fact, the existing methodologies useful for identifying KPIs are applicable as long as the measurements refer to the same techniques used or to the same context. In this specific case, the INFINITECH pilots are extremely varied, with different objectives and often with different application areas and technologies used. All of this forced us to rely on a focused reasoning aimed at finding potential points of overlap (technical or business related) which would allow us to identify KPIs that are not only measurable but also equally significant for all pilots.

Accordingly, it has been agreed with the various pilots to be based on a two-dimensional model:

- The first one that considers the technical/operational efficiency with the so-called **Operational-driven KPIs**, which measures "how much" in terms of capability the technology, the platform and the technical tools achieve the results (capable of supporting the business).
- The second dimension that concerns the **Business-driven KPIs**, that is the feedback that the previously mentioned technical capabilities give back to the business.

Moreover, it has been established a shared-responsibility logic for the evaluation process, offering a methodological approach that allows the pilots to adopt their own way to customize the landscape to their specific needs. This is meant not to impose constraints (method and tools) in the way every Pilot measure their own KPIs, as well as other aspects within their scope. Therefore, the pilot themselves are accountable for the use of the appropriate tools to make their own inner assessments over the KPIs.

2.1 Overall Achievements

As ABILab took the leadership for Task 7.8, they started devising the Pilots' Evaluation Framework. To this end, ABILab illustrated the overall process to the pilots during the Kick off Meeting (KOM), providing at the same time instructions about how to fill in the KPIs within the foreseen periodic surveys that will be assessing all pilot systems.

ABILab actively, and periodically, established an interaction with the pilots' representatives in order to initiate the "KPIs gathering" process, whose main objective is to obtain the identified KPIs from each pilot in order to commonly standardize its format for future assessments. In this regard, each pilot provided two (or more) operational-driven KPIs (which measure technical aspects), as well as two business-driven KPIs (which measure, as the name implies, business-oriented aspects).

Indeed, all pilot systems provided the required feedback for the first phase of the Evaluation Process (which will be thoroughly described in the next section). Hence, the collected KPIs represents what each pilot will be measuring overtime and make up the first draft for the KPIs structure, as they might potentially be subject to minor refinements in the near future.

Following such collection, it has been found an adequate level of standardizations for the KPIs and the survey has been adjusted accordingly in order to start the periodical assessments of the pilots.

The KPIs reported down below (Figure 1) are just an example of the ones gathered from the pilots, collected in a dedicated spreadsheet.

4	Personalized Portfolio Management ("Why Private Banking cannot be for everyone?")	1) Flexible portfolio construction based on personal risk profiling for retail clients 2) Accuracy of individual portfolio construction based on individual customer preferences 3) Accuracy of portfolio reporting after portfolio construction	1) Improved Advisor Productivity 2) Hyper-personalisation of portfolio construction 3) Increase Customer satisfaction with better risk-adjusted portfolios (Advisor / End-user) 4) Remove barrier to entry for professional wealth management
5b	Business Financial Management (BFM) tools delivering a Smart Business Advise	1) Transaction Categorization Accuracy 2) Smart Virtual advisor Reponse time	1) Customer Engagement 2) Customer Satisfication
6	Personalied Closed-Loop Investment Portfolio Management for Retail Customers	1) Increase efficiency by allocating resources properly 2) Increase effectiveness through prioritization based on expected Customer Investment	1)Set appropriate targets based on existing Customer portfolio and potentials 2) Make more targeted proposals to Customers and increase sales and CSat
7	Operation Whitetail – Avoiding Financial Crime	* False positive rate * Percentage of number of frauds detected	* Increased automation in fraud detection processes (operational cost saving) * User (SOC & eCrime employees) satisfaction

Figure 1 - Few entries of the KPIs structure's first draft

The analysis made over the collected KPIs highlighted even more the strong differences between the various pilots and the complexity in finding a standardized KPI to measure different phenomenon, as shown in Table 18.

For instance, by taking a loot at the 4th and 7th entries in the above figure, we undesirably spot strong differences with respect to the business driven KPIs (listed in the 3° column). Instead, it might seem plausible to come up with a common operational KPI (2° column), by choosing the recurrent term « accuracy ». However, considering that there are currently 15 Pilots which diversify within the INFINITECH project, such a solution was impracticable.

Thus, the only viable solution that the Consortium identified was to establish with the partners macro categories of KPIs that can be used as common guidelines that fall in two categories:

- ***Increase of efficiency***, that is the measure of how much the pilot has improved their efficiency overtime in relation to its core task (from an operational point of view – thanks to the improvement of the tools, tweaks within its code/parameters, etc.). Notably, it is an indicator which is the ratio between the "efficiency level at an early state of the project" T_0 and the "efficiency level at the latest state of the project" T_i;
- ***User satisfaction***, that is how much the product's quality has improved from the business side. This is measured as a change in the users' positive opinion compared to the initial situation. Such indicator, for the Pilots where its applicable, will also include the external stakeholder feedbacks.

This reasoning that has been made over the KPIs, is entirely technology-agnostic and such solution allows to encompass all pilot's use-cases. Indeed, the KPIs are interpreted/evaluated in evolutionary/ comparative terms, regardless of the technology being used by the pilot. In such a way, it will be feasible to include not only AI application (DL/ML algorithms), but also Blockchain, BigData/IoT solutions, and more.

The following (Figure 2) is an overview of the current Task 7.8 situation with respect to the steps involved in the Pilots' Evaluation Framework.

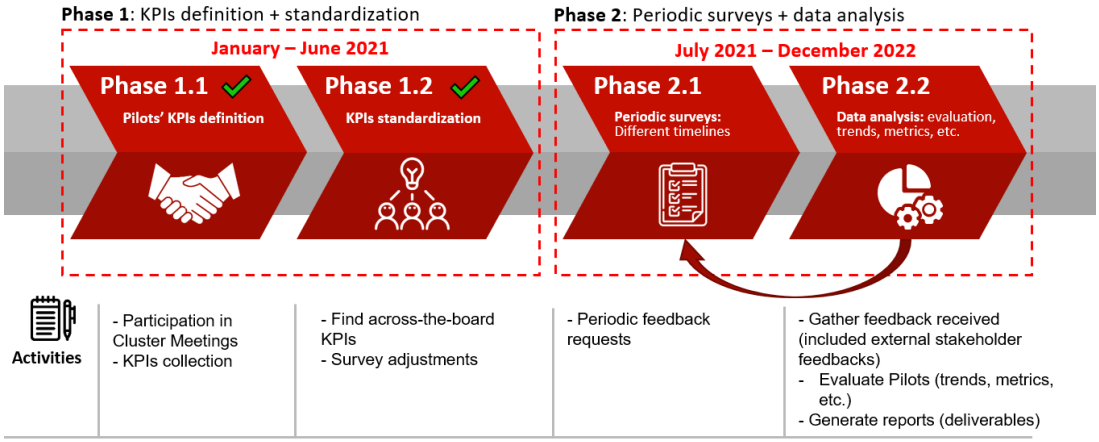

Figure 2 – Evaluation Framework (status and planned activities)

As a matter of fact, ABI Lab have recently finished up the Phase 1.2 that required our team to ask for a validation for the identified across-the-board KPIs and make the proper adjustments to the surveys to simplify the process of collecting the pilots' feedback.

The next phase would be to undertake the Phase 2 which is basically meant to start with the periodic measurements and monitoring process. To this end, ABI Lab will deliver two periodic surveys which are better described the Monitoring Process section.

Moreover, also the stakeholders' feedback and the techno-economic assessment will be addressed within the scope of the Task. Whereas the first aspect is partially reported in the monitoring process section (which will be subject to future refinement), the latter is foreseen to be implemented in the final deliverable D7.22 (Pilots' Evaluation and Stakeholders' Feedback – III [M39, ABILAB]).

2.2 Evaluation Schema

The following is a high-level description of the phases comprising the Evaluation Framework, which has the goal of assessing the progress of each pilot overtime:

- **Phase 1**: First interaction with the pilots followed by the KPIs standardization. To be more specific:
  - Define Pilots' KPIs through the first round of survey: each pilot specifies two operational-driven KPIs, as well as two business-driven KPIs.
  - Select common KPIs (standardized): among the KPIs received, a couple of across-the-board KPIs are selected and added to the survey designed in the second phase of the Framework. These represents the common ground for the pilots' evaluation.
- **Phase 2**: Delivery of two periodic surveys: definition of trends, periodic monitoring, identification of critical points. Notably:
  - **Collect Pilot's KPIs measurements**. In such way, it will be possible to periodically assess and evaluate pilot's use cases, whose results will then be reported in T7.8 related deliverables. Therefore, it will be possible to draw a picture of how the entire project is evolving, as well as potential critical points that need to be addressed.
  - **Gather information about the tools and business cases**: such information might give an insight of what are the currently used tools by the developers, whether such tools needed adjustments, as well as whether there is any current business practicality in a real-world scenario.

2.3 Monitoring Process

The monitoring process all boils down to the delivery of two periodic surveys (each of which has its own periodicity) to every pilot within the project, with the aim of profiling them in terms of both technical and business perspectives.

During the execution of such a profiling activity, ABI Lab allows the pilots to come back with a feedback in a timespan of a couple of weeks. This is meant not to overburden the pilot's development, as well as to allow the pilot systems to better collect (with ease) the required measurements. Therefore, the surveys are developed with the purpose of being brief and tendentially effortless.

One the one hand, we have a survey focused on obtaining a screenshot of the currently employed technologies within each pilot, to gain at the same time a gradual view of the pilot's development; on the other hand, we have a statistical-oriented spreadsheet within which the pilots will report their measurements with regard to various technical and business aspects, as well as the status of the standardized KPIs.

Notably, for the first survey, we have the following information:

- **Pilot**: pilot number.
- **Name**: pilot name.
- **Used Tools**: list of tools used for the pilot development.
    - Provided by INFINITECH: list of tools provided by the INFINITECH Project.
    - Open-source: list of open-source tools (tools obtained from the internet – publicly available).
    - Proprietary: self-made/tailored tools (tools developed by the Pilot and/or by third parties).
- **"What next?"**: this field is useful to get an idea of what the pilot will be used for (e.g., is it going to be used as a product/service or is it a demonstrator/pure research? etc.).
- **"Is there a business case?"**: this field is meant to gather information from the pilot as to understand whether they already have conducted a market analysis.

As for the second survey, we have the following self-explanatory information:

- **Estimation of the percentage of completion** (i.e., 36%): percentage of currently achieved milestones against the total number of future milestones (namely, at what point the Pilot is located on the timelime they have set).
- **Deviation from the Timeline** (i.e., +2 months): the time difference between the planned baseline against the actual schedule (e.g., +1 month delay, -1 month ahead with the schedule).
- **Deviation from Expected Effort** (i.e., +15%): difference between the expected effort against the actual effort spent. That is, the percentage of the additional effort spent during the execution of the project, over what was planned/approved since the beginning: This is calculated with the following formula: (*Actual effort* – *Baseline effort*) * 100 / (*Baseline effort*).
- **Number of critical problems reported**: criticalities (bigger picture) encountered that have implied adjustments to the project.
- **Project Adjustments**: number of changes made to the project in terms of contents, timetable (rescheduling activities), number of non-planned/extra activities, etc.
- **Core Model Accuracy**: accuracy Level of the core AI model in making predictions/ classifications.
- **Goodness of Fit**: measure of the quality of observation data.
- **Core Task Efficiency** (common KPI).
- **User Satisfaction** (common KPI).

Given the actual simultaneity of the survey and the deliverable, the Pilot Specific KPIs are not currently included within the survey. However, these will be included in the next surveys (which will be constantly refined) as to guarantee a proper and continuous pilots monitoring. To this regard, ABI Lab planned to refine the survey in the next future to better reflect the goal of the framework, as well as to simplify the process of retrieving the feedback from pilot systems. Such refinements will also include the possibility to ask each pilot to express their position to compare the Pilots results with respect to the state of the art, state of practice and other systems available in the market. Such evaluation will be the starting point to enrich the overall evaluation framework with the summary of the main comments.

With respect to the monitoring process, the idea is to progressively track the status of the KPIs overtime and consequently deriving its increase/decrease in different points in time. This is achieved by getting "screenshots" (i.e., *T*0, *T*1, *T*2, etc.) of each pilot-provided-KPI, as well as the standardized ones, and to extract its improvement indicator by reasoning in evolutionary/differential terms.

Example: Operational-driven KPI (Increase of Efficiency) associated to the core task of a pilot's use-case: "How many documents had extraction errors and couldn´t be processed."

- Status at time *T*0: 50 documents out 100 (measure of efficiency) – NOW
- Status at time *T*1: 20 documents out of 100 (measure of efficiency) – December 2021
- Status at time *T*2: 15 documents out of 100 (measure of efficiency) – May 2022

By giving a glance to the acquired statuses reported above, it is possible to build an indicator which is the ratio between the efficiency level at an early state of the project (*T*0) and the efficiency level at the latest state of the project (*T*2). In other terms, we want to measure the variation in performance over time, which is the difference of how much efficient the pilot is today (Ti) to perform a task and how much efficient the pilot was in a previous situation (*T*0), still related to the same task.

However, to better understand the context within which such KPIs measurements are obtained, in the next rounds of sharing with the Pilots, the Task 7.8 will take care of asking the pilots to provide more details on their measurement methodology (which is foreseen to be included in further versions of the deliverable).

As previously mentioned in the introduction, the monitoring process will also need to address the process of gathering the stakeholders' feedback of all the pilots. However, even though such process has not started yet, the main idea remains the following: as the pilots begins to increase their maturity level and maintain a constant interaction the stakeholders, from which it is possible to picture an advancement in their activities, the Task will serve as a hub. That is, by relying on a decentralized approach, through which the Pilots themselves will take care of periodically gathering the stakeholders' feedback from the planned workshops, ABI Lab will then be responsible to gather such outcomes and present them finally in an aggregated fashion as to have a high-level vision of the pilots' progress. Furthermore, the task does not preclude the possibility to rely on the already established channels between the Pilots and the Cluster Leaders, with the aim of carrying on such approach as efficiently as possible. Further details of such plan will surely be included in the next versions of the deliverable as to properly achieve a multi-facet evaluation from the stakeholders, as well as to better define the evaluation process in a unified way.

3 Drafting KPI structure

In this section the Pilots provide their first draft for the KPI structure, showing an overall picture of what each one of them will be measuring in the future.

Every Pilot has its own dedicated chapter within which they will have to describe the following aspects:

1. **Description of the KPIs**
2. **Main Actors Involved**
3. **Outline of the Current Status**
4. **Main Challenges**

To avoid redundancy and facilitate the reader in the understanding of the tables reported in every sub-section 3 (Outline of the Current Status), we report below their legend:

Type: Operational-driven KPI (O), Business-driven KPI (B)

Measurement Mode: how the KPI will be measured

Status: represents the current status of the Pilot. Put an X under the column you relate to (the columns are not mutually exclusive):

- **(ID) = Identified**: the KPI has been identified
- **(M) = Measured**: the KPI has been measured
- **(NR) = Needs refinements**: the KPI will be subject to future refinements
- **(CM) = Continuously monitored**: the KPI is in advanced/mature state by which we are currently under monitoring

Initial KPI Measurement: measurement prior to the start of the project (if it exist)

Current KPI Measurement: current value of the KPI

Target Level: specify the level towards which to make the KPI tend (and eventually achieve)

Achieved: insert yes or no as to state whether the target level has been currently achieved

3.1 Pilot #1 (BANKIA)

3.1.1 Description of the KPIs

The following are the identified KPIs:

- **Number of documents accepted for processing by the pilot application**: there are documents that the OCR is not able to process due to their poor quality, so we are working to improve the OCR solution to increase the number of accepted documents.
- **Effectiveness of the extraction of low level concepts** e.g. prices, copy concepts, quantity, etc. It will be measured by standard algorithmic measures of precision and recall.
- **Effectiveness of the recognition of high level concepts** formed by the association of low level concepts (e.g. an item to be audited is composed of price, copy concept, quantity, etc. this information is scattered in the document and a specific neural network has been designed to identify them). It will be measured by standard algorithmic measures of precision and recall.
- **Effectiveness of the business rules validity** when applied, the criteria to accept or not the invoice according to different high-level concepts.
- **Effectiveness of the auditing in other intangible features** like detecting invoices with excessive copies billed, so we can recommend a reduction of the billed copies.

The identified core task over which we will be measuring our “Increase of Efficiency” over time.

> **Identified core Task**: *"how many invoices have been validated according to the business rules"*
>
> **Status** at time **T0**: ND - not calculated yet (measure of efficiency) – NOW

For the present Pilot, the core task that summarizes from one side the business impact of the pilot, and on the other hand condenses the impact of the technical performance is the KPI regarding effectiveness of the business rules.

It should be noted that a quite acceptable effectiveness has already been attained from the first developments.

> **Identified core Task**: *"how many invoices had extraction errors and could not been processed"*
>
> **Status** at time **T0**: 60% (measure of efficiency) – NOW
>
> **Identified core Task**: *"how many expedited copies and paper sheets were identified and how much paper and costs could be saved"*
>
> **Status** at time **T0**: ND (measure of efficiency) – NOW

- **How do you plan to measure the KPIs**?

As is typical in projects involving machine learning algorithms, we have to monitor not only the increase in efficiency (for which we expect a slow increase) but also monitor that a decrease in efficiency does not take place. This is consistent with the characteristics of machine learning algorithms, since one of the main goals of the algorithms is to include new unseen cases. Testing with new documents and testing the robustness of the solution to different regional or idiosyncratic varieties of the invoices will be a key parameter to monitor.

- **With which frequency do you plan/foresee to monitor the KPIs**?

They will be measured every time a new slot of invoices (either defined by a time interval or by a geographic provenance) are received, every month or every three months coincidentally with the business normal operations cycle. To measure the KPIs is relatively costly, since it requires human verification of quite a wide number of parameters and judgement.

- **Why are such indicators relevant to your Pilot**?

> Because these indicators allow us the effectiveness of the solution and the progress we are making.

3.1.2 Key Actors Involved

The relevant actors involved in this process (see Figure 3):

Clients of the bank: they can choose the notary to formalize the lending contract.

Notaries: they formalize the lending contract, make the invoices with the associated costs and send to the agencies to get the payment.

Bank Lending Backoffice: They check the invoices to know if they are correct to authorize the payment.

Bank Budget Management Department: they pay the invoices of the lending operations and supervise the suppliers.

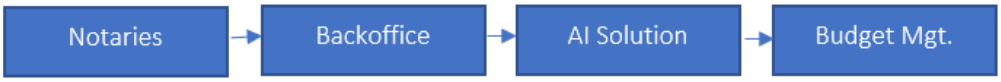

Figure 3 - Pilot #1 Relevant Actors

3.1.3 Outline of the Current Status

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Core Task Efficiency: Effectiveness of the business rules (common-KPI)	**B**	Estimated savings (processing time, number of processed invoices, estimated economic savings).	X				n.d.	n.d.	n.d.	NO
User Satisfaction (common-KPI)	**B**		X				n.d.	n.d.	n.d.	NO
No Processable: how many documents had extraction errors and couldn´t be processed (%)	**O**	Precision & recall over human validated examples	X	X			n.d.	60%	30%	NO
Effectiveness of the extraction of low level concepts	**O**	Precision & recall over human validated examples	X	X			n.d.	60%	90%	NO
Effectiveness of the recognition of high level concepts	**O**	Precision & recall over human validated examples	X	X			n.d.	50%	90%	NO

Table 1 - Pilot #1 KPIs status

3.1.4 Main Challenges

- Increase the number of processed invoices by the solution.
- Increase the number of validated invoices by the solution.
- Improve the accuracy, precision and recall in the extraction and graphical association of the invoice's concepts.
- Achieve a completely explainable notaries sustainability score.

3.2 Pilot #2 (JRC)

3.2.1 Description of the KPIs

The KPIs regarding Pilot #2 could be divided in two main categories, the operational and the business ones. However, meeting an operational goal could also lead to meeting a business objective as they are probably interconnected.

From a business perspective, Pilot's main goal is to provide risk assessments of financial portfolios in (near) real time, enabling quicker reactions and giving the possibility to make adjustments in portfolio composition if needed. To this end, the provided solution should be able to leverage the latest market data and based on the input trading positions to deliver valid risk metrics. Furthermore, the Pilot should increase the user (e.g. trader, risk manager) satisfaction through a friendly user interface providing features such as sentiment analysis on financial news and pre-trade analysis.

As for the operational KPIs, in order for the Pilot to provide reliable risk estimations, the utilised risk models should be able to accurately predict portfolio VaR/ES, under a given probability, within a time period such as a day. Thus, this indicator could be measured through back-testing on the historical performance of the underlying models. For instance, in case of daily-VaR and confidence probability α = 99%, the VaR model should not produce more than 1 VaR violation (i.e. the actual return r of the portfolio is less than the predicted VaR) within 100 trading days. In addition, the pilot should deliver the aforementioned risk assessments in (near) real-time in order to control the performance of a given portfolio closely. To achieve this, the risk predictions, based on the latest data available, should be recalculated frequently. Finally, the pre-trade analysis feature should be able to yield results, according to the user inputs, in a timely manner.

Identified core Task 1: *"how big were the financial portfolios for which the risk should be calculated in real time?"*

> **Status** at time **T0**: 4 FX instruments per portfolio (measure of efficiency) – NOW
>
> **Status** at time **T1**: xx FX instruments per portfolio (measure of efficiency) – December 2021
>
> **Status** at time **T2**: xx FX instruments per portfolio (measure of efficiency) – May 2022

Identified core Task 2: *"What is the level of user satisfaction (score 1-5, worst to best scaling)?"*

> **Status** at time **T0**: 3.9/5 Based on stakeholders workshop– NOW
>
> **Status** at time **T1**: xx 1st JRC survey results – December 2021
>
> **Status** at time **T2**: xx 2st JRC survey results – May 2022

• **How do you plan to measure the KPIs**? (describe the operational activities)

> As mentioned above, the two main operational aspects of Pilot #2 are related firstly to the accuracy of the risk measurements and secondly to the latency in recalculating the portfolio risk as market data is updated.
>
> The KPI regarding the risk estimation accuracy is measured in terms of VaR/ES violations. To be more concrete, the probability of a violation must be equal to the α coverage rate (Eq. 1).
>
> $P(r < R_\alpha) = 1 - \alpha$ (1)
>
> Where r, R, α the observed portfolio returns, the risk metric value and the confidence probability of the prediction respectively.
>
> The second indicator, related to the real-time delivery of the employed risk models, is measured in relation to the time required by the developed system to recalculate the portfolio risk. This task should be completed in less than 1 minute. Moreover, the pre-trade analysis feature should return risk measurements, as the user enters a new portfolio composition, in less than 5 seconds.
>
> With respect to the measurement of the user satisfaction (Business-driven KPI), we will implement an evaluation based on the following KPIs:

- intuitivity/understanding
- usability
- effectiveness

> The evaluation will be based on subjective feedback taken from a survey that will involve potential end users and their user experience. For each KPI a scale/range from 1 (worst) to 5 (best) will be used. The overall level of user satisfaction is determined by the arithmetic average of the levels obtained on the above mentioned three KPIs.

- **With which frequency do you plan/foresee to monitor the KPIs**?

> In terms of the accuracy of the risk models, each of the embedded models has been validated in an extensive volume of historical data in order to prove that they are able to meet the corresponding KPI.
>
> As for the KPIs with respect to the real time data management, Pilot #2 is constantly monitored for being able to provide risk assessments within the required time constraints. It is also noted that, as the pilot is still under development, a dedicated environment has been developed for end-to-end simulation of the data pipeline that allows monitoring of these KPIs.

- **Why such indicators are relevant to your Pilot**?

> The first KPI, the number of VaR/ES violations, is considered, between the scientific community and regulators, to be one of the most important indicators for the validity of a VaR/ES model[1][2]. Additionally, selected indicators that measure the pilot's ability to perform real-time risk assessment were determined in relation to stakeholder's feedback. As current practice is that portfolio risk is recalculated once a day in batch mode, switching to a platform that allows updates to risk metrics every minute and takes into account intra-day trading positions is a great performance improvement.

3.2.2 Key Actors Involved

The following types of stakeholders are involved in this use case.

- *Traders*: traders can use the platform to assess the risk of a trade in real time using VaR and ES. (individually and at portfolio level). Furthermore, they can conduct a pre-trade analysis to assess the risk of a potential additional trade.
- *Risk managers*: they can use the platform to evaluate and monitor the risk of a given portfolio without having to resort to end-of-day data. In this way, they can compare the real time risk levels with pre-defined levels based on specific rules and investment policies.
- *Financial institutions* (as a whole): they would be equipped with a flexible and efficient real-time risk analysis system that can be useful at different levels within the company, to meet actual and future requirements in terms of risk assessment.

3.2.3 Outline of the Current Status

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Core Task Efficiency (common-KPI)	**O**	Number of assets per FX portfolio for which the risk is calculated in real time	X	X		X	0	4	10	NO

User Satisfaction (common-KPI)	**B**	Arithmetic average based on the evaluation of three sub-KPIs (intuitivity, usability, effectiveness)	X				N/A	N/A	>4	NO
Fewer VaR violations compared to existing models [P(r	**O**	Percentage of daily VaR/ES violations per year in confidence level 99% (mean value between utilized portfolios)	X	X		X	2.5%	<1%	<1%	YES
VaR estimation of multi-asset portfolios in less than 5 seconds	**O**	Seconds	X	X		X	3 sec (4 FX assets per portfolio)	1 sec (4 FX assets per portfolio)	1 sec (10 FX assets per portfolio)	NO
Real Time VaR availability instead of daily (batch) VaR estimation	**B**	Seconds	X	X		X	300 secs (4 assets per portfolio)	60 secs (4 assets per portfolio)	60 sec (10 assets per portfolio)	NO

Table 2 - Pilot #2 KPIs status

3.2.4 Main Challenges

The two main challenges of the Pilot are related initially to the risk models accuracy and then to optimize these models to provide their assessments in real-time. As this is a blueprint pilot of INFINITECH, these objectives have been largely addressed. However, in the coming months the Pilot will scale to more instruments per portfolio to evaluate the real-time functionality over a larger amount of data. Moreover, it is planned to integrate the "sentiment analysis on financial news" feature in the near future. This will act as an extra risk indicator enhancing the user satisfaction.

3.3 Pilot #3 (BOI)

The flexibility of Pilot #3 solution is a requirement from the Bank of Ireland Financial Investigations Unit (FIU) and other Financial Institutions operating in the Irish market (i.e., the Banking Payment Federation of Ireland member organizations). The objective of the defined KPIs is to understand the way to measure the common capabilities that can be reused and quantify any additional requirements which would be needed for other organizations.

3.3.1 Description of the KPIs

Here the Pilot #3 provides a description in the form of a narrative related to KPIs definitions as follow:

- *Data Sharing (Data Sets-Related) KPIs*

***P3-KPI-D1**: The use of Synthetic Data sets as proof of realistic development on Pilot #3 applications. 2 x Data sets will come from customers, 2 x financial applications and 2 x Financial Services in a Bank. A total of 6 Data Sets to Enhance transparency and consumer data ownership.*

T0 June 2021 Status: 2 Data Sets Samples (Indicator) – Availability

T1 Dec 2021 Status: 6 Data Sets Samples (Indicator) – Num. of Available Data Sets

T2 June 2022 Status: 6 Synthetic Data Sets (Indicator) – Num. of Created Data Sets

T3 Dec 2022 Status: 6+ Synthetic Data Sets (Indicator) – Num. of Used Data Sets

Measurement: This KPI will be measured by enabling the use of the sample datasets during the project, firstly after the first working implementation and secondly during the project production simulation test. A direct measuring can be done also using the number of records/entries per each sample data set.

Frequency: every month

***P3-KPI-D2**: The use of data sets related to Customer, financial transactions and Financial Services demonstrating at least 1 application involving the following customer on-boarding, pricing of credit and traffic analytics hub automation services.*

T0 June 2021 Status: 0 Data Sets Available (Indicator) – Availability

T1 Dec 2021 Status: 6 Data Sets Samples (Indicator) – Num. of Available Data Sets

T2 June 2022 Status: 6 Synthetic Data Sets (Indicator) – Num. of Created Data Sets

T3 Dec 2022 Status: 6+ Synthetic Data Sets (Indicator) – Num. of Used Data Sets

Measurement: This KPI will be measured by providing the prototype implementation and enabling the demonstration and use of the data sets during the project, firstly after the first working implementation and secondly during the project production simulation test.

Frequency: every six months

***P3-KPI-D3**: The adoption and/or implementation of 1-2 applications or methods ensuring data security, data protection of the available data sets with the objective of improving customer trust through transparency.*

T0 June 2021 Status: 0 Data Sets Available (Indicator) – Availability

T1 Dec 2021 Status: 6 Data Sets Samples (Indicator) – Num. of Available Data Sets

T2 June 2022 Status: 6 Synthetic Data Sets (Indicator) – Num. of Created Data Sets

T3 Dec 2022 Status: 6+ Synthetic Data Sets (Indicator) – Num. of Used Data Sets

Measurement: This KPI will be measured by providing the prototype implementation and the demonstration and use of the applications and/or methods during the project, firstly after the first working implementation and secondly during the project production simulation test starting in June 2022.

Frequency: every three months

***P3-KPI-D4**: The definition and demonstration of a reference use case where Sharing financial data to enable KYC/ AML and Credit Scoring processes as main purpose.*

T0 June 2021 Status: 0 Data Sets Available (Indicator) – Availability

T1 Dec 2021 Status: 6 Data Sets Samples (Indicator) – Num. of Available Data Sets

T2 June 2022 Status: 6 Synthetic Data Sets (Indicator) – Num. of Created Data Sets

T3 Dec 2022 Status: 6+ Synthetic Data Sets (Indicator) – Num. of Used Data Sets

Measurement: This KPI will be measured by providing the prototype implementation providing the demonstration and use of the reference use case during the project, firstly after the first working implementation and secondly during the project production simulation test starting in June 2022.

Frequency: every three months

***P3-KPI-D5**: The use of common data model(s) for reducing complexity and less time-consuming process and reduction in paperwork/documentation.*

T0 June 2021 Status: 0 Data Sets Available (Indicator) – Availability

T1 Dec 2021 Status: 6 Data Sets Samples (Indicator) – Num. of Available Data Sets

T2 June 2022 Status: 6 Synthetic Data Sets (Indicator) – Num. of Created Data Sets

T3 Dec 2022 Status: 6+ Synthetic Data Sets (Indicator) – Num. of Used Data Sets

Measurement: This KPI will be measured by providing the demonstration and use of the common data model during the project, firstly using the proposed BOI Data model after the first working implementation and secondly the INFINITECH Data Model test starting in June 2022.

Frequency: every three months

• *Know Your Customer (Technology/Platform/Framework-Related) KPIs*

***P3-KPI-KYC1**: Define two to three frameworks and forms to enable the collaboration of Data Ownership and multi-tenancy of the different stakeholders involved*

T0 June 2021 Status: 0 Tool/Framework Available (Indicator) – Availability

T1 Dec 2021 Status: 1 Tool/Framework Available (Indicator) – Num. of tools

T2 June 2022 Status: 1-2 Tool/Framework Available (Indicator) – Num. of Created Tools

T3 Dec 2022 Status: 2+ Tool/Framework Available (Indicator) – Num. of Used Tools

Measurement: This KPI will be measured by the organization of a KYC/KYB for banking and Insurance Sector workshop triggering an analysis and the survey first collection of Feedback over INFINITECH Pilot #3 project results.

Frequency: every six months

***P3-KPI-KYC2**: Define at least one procedure and engagement activity that allows Central Bank / CCPC / DOF Improve competition and operational resilience.*

T0 June 2021 Status: 0 Procedure/Activity Available (Indicator) – Availability

T1 Dec 2021 Status: 1 Procedure/Activity Available (Indicator) – Num. of procedure(s)

T2 June 2022 Status: 1+ Procedure/Activity (Indicator) – Num. of Activities

T3 Dec 2022 Status: 1+ Procedure/Activity (Indicator) – Num. of Activities

Measurement: This KPI will be measured by defining the procedure for aggregating the confidence scores from the NLU models for each red-flag indicator contained in the typology and ranking the highest scoring typologies.

Frequency: every six months

***P3-KPI-KYC3**: Involvement of at least 50 end user / consumers of Pilot #3 applications for understanding and educating about the opportunities and benefits of using financial data for KYC/KYB Services like Credit Scoring, Profile building, etc. Services that will promote the participation and at the same time the generation of more financial services.*

T0 June 2021 Status: 0 end user / consumers (Indicator) – Num. of Participants

T1 Dec 2021 Status: 0-10 end user / consumers (Indicator) – Num. of Participants

T2 June 2022 Status: 11-30 end user / consumers (Indicator) – Num. of Participants

T3 Dec 2022 Status: 31-50 end user / consumers (Indicator) – Num. of Participants

Measurement: This KPI will be measured by enabling a KYC/KYB for banking and Insurance Sector workshop triggering an analysis and the survey first collection of Feedback over INFINITECH Pilot #3 project results.

Frequency: every six months

***P3-KPI-KYC4**: Provide 1 x Data Platform to improve sharing and Increased empowerment of consumer.*

T0 June 2021 Status: 0 Data Platform Available (Indicator) – Availability

T1 Dec 2021 Status: 1 Data Platform Design Available (Indicator) – Data Platform Design

T2 June 2022 Status: 1 Data platform Deploy Ready (Indicator) – Data Platform Deploy

T3 Dec 2022 Status: 1 Data Platform Use and Tested (Indicator) – Data Test and Use

Measurement: This KPI will be measured by the Identification & generation of five human trafficking centric red-flag typologies which are integrated within the Bank of Ireland FIU Know Your Customer processes and/or systems. This process will be tested against a training set of data which has been manually scored & ranked to evaluate the accuracy & veracity of the typologies.

Frequency: every six months

P3-KPI-KYC5*: A comprehensive analysis for integrating solutions Sharing of Standardized information with customer consent and Increased KYC/KYB product offering awareness.*

T0 June 2021 Status: 0 Tool/Framework Available (Indicator) – Availability

T1 Dec 2021 Status: 1 Study/Analysis Plan Available (Indicator) – Study Analysis Plan

T2 June 2022 Status: 1 Study/Analysis Plan Tool (Indicator) – Analysis Framework

T3 Dec 2022 Status: 1 Study/Analysis Results (Indicator) – Analysis Results

Measurement: This KPI will be measured by having ready a comprehensive report that study the different integration and interoperability solutions for data sharing in the context of KYC/KYB using standardized information with customer consent tools.

Frequency: every six months

• *Application / Solution (Application-related) KPIs*

***P3-KPI-AS1**: Enable 2 x events that enable the collaboration of Data Ownership and multi-tenancy of the different stakeholders involved*

T0 June 2021 Status: 0 Events Organised (Indicator) – Num. of Events

T1 Dec 2021 Status: 1 Event Organised (Indicator) – Num. of Events

T2 June 2022 Status: 1+ Events Organised (Indicator) – Num. of Events

T3 Dec 2022 Status: 2 Events Organised (Indicator) – Num. of Events

Measurement: This KPI will be measured by organizing two banking and Insurance Sector stakeholders and organizations event/workshop triggering the awareness and collaboration of their participants either in the form of validation or testing and providing Feedback over INFINITECH Pilot #3 approach.

Frequency: every six months

***P3-KPI-AS2**: Involvement and participation of 50-100 consumers as stakeholders during application processes, including KYC that helps to identify and verify customer identity*

T0 June 2021 Status: 0 Consumers/Stakeholders (Indicator) – Num. of Participants

T1 Dec 2021 Status: 0-25 Consumers/Stakeholders (Indicator) – Num. of Participants

T2 June 2022 Status: 26-75 Consumers/Stakeholders (Indicator) – Num. of Participants

T3 Dec 2022 Status: 76-100 Consumers/Stakeholders. (Indicator) – Num. of Participants

Measurement: This KPI will be measured through a combination of Bank of Ireland FIU team end-user participants and the Financial Institutions operating in the Irish market (Banking Payment Federation of Ireland member organization). The objective is to understand the level of acceptance and awareness about the mechanism to help identify and verify customer identity.

Frequency: every six months

***P3-KPI-AS3**: 1 x application/solution for improving and enabling Intelligent Analysis for KYC Transactions Data and Data Services for Traffik Analysis Hub (NGOs Name)*

T0 June 2021 Status: 0 Application/Solution Design (Indicator) – Availability

T1 Dec 2021 Status: 1 Application/Solution Available (Indicator) – Application Design

T2 June 2022 Status: 1 Data platform Deploy Ready (Indicator) – Application Deploy

T3 Dec 2022 Status: 1 Data Platform Use and Tested (Indicator) – Application Use

Measurement: This KPI will be measured by enabling the FIU teams or other organizations interested to have a repeatable & reliable application for identifying human trafficking related services for KYC transactions in the context of Data Services for Traffik Analysis Hub, this solution will enable them to gain Intelligent Analysis for human trafficking provided for the Traffik Analysis Hub platform.

Frequency: every six months

***P3-KPI-AS4**: 1 x Traffik Analysis Hub application that works towards increasing the capacity to identify unlawful activity in relation to financial activities.*

T0 June 2021 Status: 0 Application/Solution Design (Indicator) – Availability

T1 Dec 2021 Status: 1 Application/Solution Available (Indicator) – Application Design

T2 June 2022 Status: 1 Data platform Deploy Ready (Indicator) – Application Deploy

> T3 Dec 2022 Status: 1 Data Platform Use and Tested (Indicator) – Application Use
>
> *Measurement:* This KPI will be measured by enabling the FIU teams to have a repeatable & reliable process for identifying human trafficking related red-flag typologies to use in their KYC processes, this solution will enable them to gain an AI generated notification of human trafficking red-flag typologies from the incident reports & survivor stories provided by the Traffik Analysis Hub platform.
>
> *Frequency:* every six months
>
> ***P3-KPI-AS5**: Define efficient process driving up NIM (Net Interest Margin) and Improved Banking Reputation. Identifying & Generating 1 x 5 human trafficking centric red-flag typologies which are integrated within the Bank of Ireland FIU Know Your Customer processes and/or systems*
>
> T0 June 2021 Status: 0 Red-Flag Typologies Design (Indicator) – Availability
>
> T1 Dec 2021 Status: 2 Red-Flag Typologies Available (Indicator) – Typology Design
>
> T2 June 2022 Status: 3 Red-Flag Typologies Deploy (Indicator) – Typology Deploy
>
> T3 Dec 2022 Status: 5 Red-Flag Typologies Use (Indicator) – Typology Use
>
> *Measurement:* This KPI will be measured by aggregating the confidence scores from the NLU models for each red-flag indicator contained in the typology and ranking the highest scoring typologies. This process will be tested against a training set of data which has been manually scored & ranked to evaluate the accuracy & veracity of the typologies.
>
> *Frequency:* every six months

3.3.2 Key Actors Involved

Financial Investigations Unit (FIU) Investigator at Bank of Ireland (BOI). This actor needs access to domain (human trafficking) specific red-flag typologies and technology to integrate/consume these red-flag typologies in internal transaction monitoring solutions for the purpose of identifying suspicious customer & transactional data points.

Banking and Payments Federation of Ireland (BPFI) Organization member(s) and other participant(s). This actor facilitates the provisioning of information and engagement alike the evaluation from the Irish-based ecosystem in relation to banking and insurance organizations and enable the participation in events.

Human Trafficking Subject Matter Expert Analyst (Traffik Analysis Hub). This actor needs access to big data specific to the domain of human trafficking and liaises with the FIU Investigations team to implement NLU models for identifying red-flag indicators in the survivor stories & incident data.

3.3.3 Outline of the Current Status

The KPIs defined in Pilot #3 will be measured through a combination of Bank of Ireland FIU team end-user surveys and an evaluation of the customization effort required for additional Financial Institutions operating in the Irish market (as part of the Banking Payment Federation of Ireland (BPFI) member organization). The objective of the KPIs is to understand the common capabilities that can be reused and quantify any additional mandatory requirements which would be needed for other organizations.

Pilot #3 KPIs will be evaluated once for each organization during the project production simulation test.

The proposed Pilot #3 solution needs to be compatible to the core requirements of multiple Financial Institutions to facilitate industry adoption and promote active collaboration within the FIU teams in the Irish market.

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Use of different data sets from financial and social media	**P3-KPI-D1** **O**	Enabling the use of the sample datasets during the project	X				Not Yet	Not Yet	6+ Synthetic Data Sets	Pending
	P3-KPI-D2 **O**	Providing the prototype implementation and enabling the demonstration and use of the data sets during the project.	X				Not Yet	Not Yet	6+ Synthetic Data Sets	Pending
Use of Single Data Model for data processing	**P3-KPI-D3** **O**	Providing the prototype implementation and the demonstration and use of the applications and/or methods during the project.	X				Not Yet	Not Yet	6+ Synthetic Data Sets	Pending
	P3-KPI-D4 **O**	Providing the prototype implementation providing the demonstration and use of the reference use case during the project.	X				Not Yet	Not Yet	6+ Synthetic Data Sets	Pending
	P3-KPI-D5 **O**	Providing the demonstration and use of the common data model during the project, using the proposed BOI Data model.	X				Not Yet	Not Yet	6+ Synthetic Data Sets	Pending
Core Task Efficiency (common-KPI)	**P3-KPI-KYC1** **O**	Organization of a KYC/KYB for banking and Insurance Sector workshop triggering an analysis and the survey first collection of Feedback	X				Not Yet	Not Yet	2+ Tools/ Framework Available	Pending
	P3-KPI-KYC2 **O**	Defining the procedure for aggregating the confidence scores from the NLU models for each red-flag indicator contained in the typology and ranking the highest scoring typologies.	X				Not Yet	Not Yet	1 Procedure Available	Pending

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
	P3-KPI-KYC3 **O**	Enabling a KYC/KYB for banking and Insurance Sector workshop triggering an analysis and the survey first collection of Feedback.	X				Not Yet	Not Yet	50 End user / Consumers	Pending
Profile identification accuracy based on input data	**P3-KPI-KYC4** **B**	Identification & generation of five human trafficking centric red-flag typologies which are integrated within the Bank of Ireland FIU Know Your Customer processes and/or systems.	X				Not Yet	Not Yet	1 Data Platform Available Used and tested	Pending
	P3-KPI-KYC5 **B**	Having ready a comprehensive report that study the different integration and interoperability solutions for data sharing in the context of KYC/KYB.	X				Not Yet	Not Yet	1 Study Analysis Report Available	Pending
User Satisfaction (common-KPI)	**P3-KPI-AS1** **B**	Organizing two banking and Insurance Sector stakeholders and organizations event/workshop triggering the awareness and collaboration.	X				Not Yet	Not Yet	2 Events Organised	Pending
	P3-KPI-AS2 **B**	Combination of Bank of Ireland FIU team end-user participants and the Financial Institutions operating in the Irish market (Banking Payment Federation of Ireland member organization).	X				Not Yet	Not Yet	100 Consumers / Stakeholders	Pending
Reduction of false positives based on enriched input data	**P3-KPI-AS3** **B**	Enabling the FIU teams or other organizations interested to have a repeatable & reliable application for identifying human trafficking related services.	X				Not Yet	Not Yet	1 Data Platform Available Used and tested	Pending

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
	P3-KPI-AS4 **B**	Enabling the FIU teams to have a repeatable & reliable process for identifying human trafficking related red-flag typologies to use in their KYC processes.	X				Not Yet	Not Yet	1 Data Platform Available Used and tested	Pending
	P3-KPI-AS5 **B**	Aggregating the confidence scores from the NLU models for each red-flag indicator contained in the typology and ranking the highest scoring typologies.	X				Not Yet	Not Yet	5 Red-Flag Typologies Used	Pending

Table 3 - Pilot #3 KPIs status

3.3.4 Main Challenges

There are several challenges being managed by the Pilot #3 team. Of note are:

1. **Pilot Partners Constitution Challenge:** One of the Pilot partners had to modify their role in the deliverables. Their technical department were no longer able to commit their development resources. This resulted in on boarding a new partner joining the team as to provide new development resources. Onboarding this new partner is taking some time.
   Mitigation of challenge: If IBM cannot join the partnership directly, they will be able to contribute through the Traffic Analysis HUB platform that is being used in Pilot #3.
2. **Customer Datasets Construction:** The initial intention was to use real customer transactional data, but the banking partner outlined the difficulties this would bring when conforming with GDPR obligations.
   Mitigation of challenge: Pilot #3 will transform real data sets into synthetic data sets, respecting the requirements for privacy/anonymization and competitive considerations; these data sets will be representative of real-world financial institution transactions.
3. **Technologies Identification:** Identifying applicable technologies to be leveraged to deliver an integrated solution which supports the FIU requirements for Know Your Customer scenarios related to financial transactions and its impact on Human Trafficking activities.
   Mitigation of challenge: Consortium members will work towards integrating INFINITECH-ready developed technologies and may need to develop other custom technology elements to bridge any gaps in the available technology.
4. **Measuring Pilot Impact:** It will be difficult to measure the positive impact the Pilot #3 technology is having on the use case of “Stop the Traffic”. The nature of how real-world cases are identified does not lend itself to clear and repeatable step changes in the reduction of false-positive alerts.
   Mitigation of challenge: Leverage AI approaches to facilitate the identification of human trafficking typologies, represented as red flag indicators, and support the FIU investigators in discovering these typologies in existing FI data.

3.4 Pilot #4 (PRIVE)

3.4.1 Description of the KPIs

Operational-driven KPIs:

The first operational-driven KPI is to construct a flexible portfolio based on personal risk profiling for retail clients. The second KPI is to achieve accuracy of individual portfolio construction based on individual customer preferences. Finally, accuracy of portfolio reporting after portfolio construction and execution of investment decisions is pursued as a third operational-driven KPI.

Business-driven KPIs:

The first business-driven KPI is to achieve Advisor Productivity Improvement. Secondly, hyper-personalisation of portfolio construction is pursued. The third KPI is to increase customer satisfaction with better risk-adjusted portfolios (Advisor / End-user). The fourth KPI is to remove entry barriers of professional wealth management solutions for retail customers (mass-affluent) and their advisors.

Identified core Task: "*How many fitness factors, that enable efficiency measurement, have been created?*"

Status at time **T0** (Figure 4): Current fitness factors: Min Return gap Vol Factor (MRVF), Sharpe Ratio Factor (SRF), Product Diversification Factor (PDF), Preferred Currency Factor (PCF)

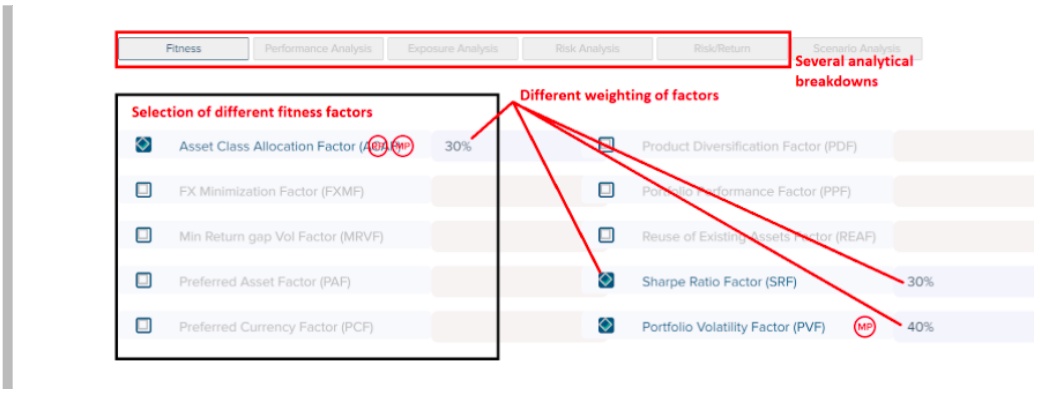

Figure 4 - Current measure of efficiency (NOW)

Status at time **T1**: Creation of 2 further fitness factors: Sustainability, ETF

Sustainability fitness factor (Figure 5)**:**

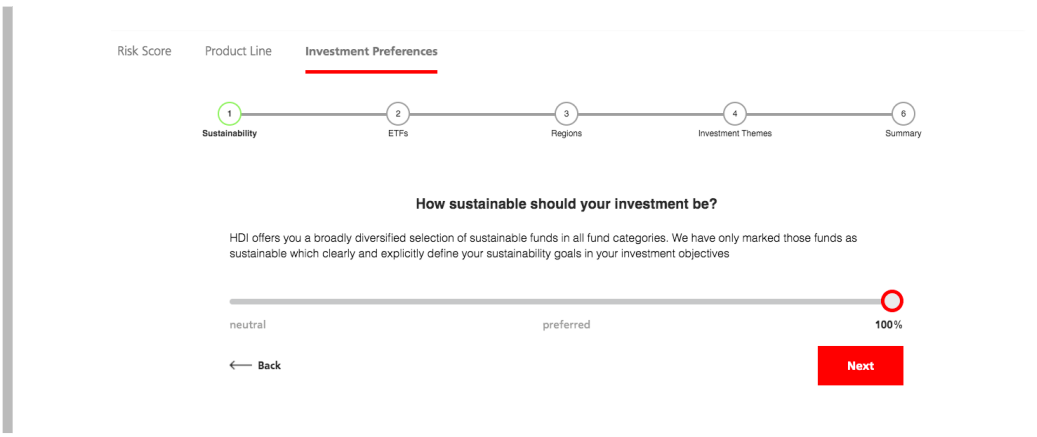

Figure 5 - Investment preferences – Sustainability (measure of efficiency) – December 2021

ETF (reduced implicit product cost) fitness factor (Figure 6)**:**

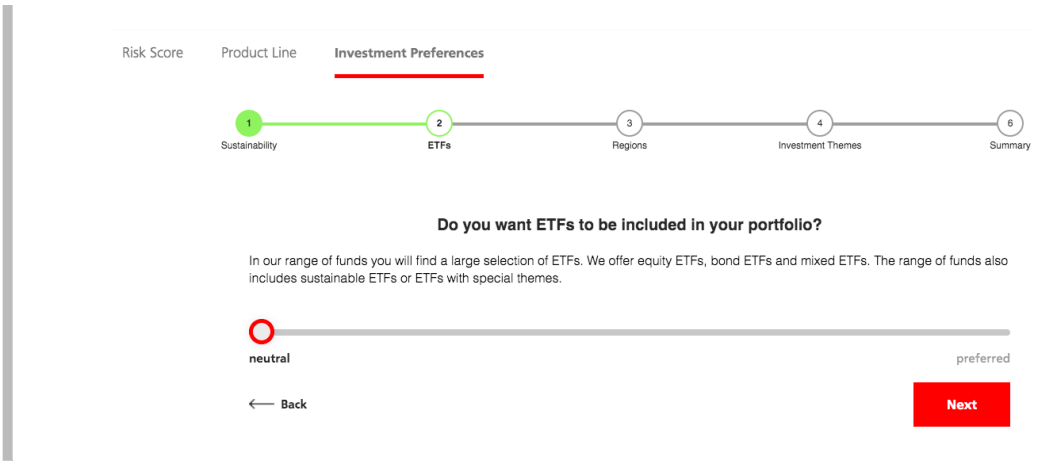

Figure 6 - Investment Preferences – ETFs (measure of efficiency) – December 2021

Status at time T2: Creation of 2 new fitness factors: News Article Sentiment Analysis Fitness Factor, Regional Fitness Factor. Continuous improvement of the previously developed and created fitness factors. Final fitness factor number: depending on customer requirements.

Regional fitness factor (Figure 7)**:**

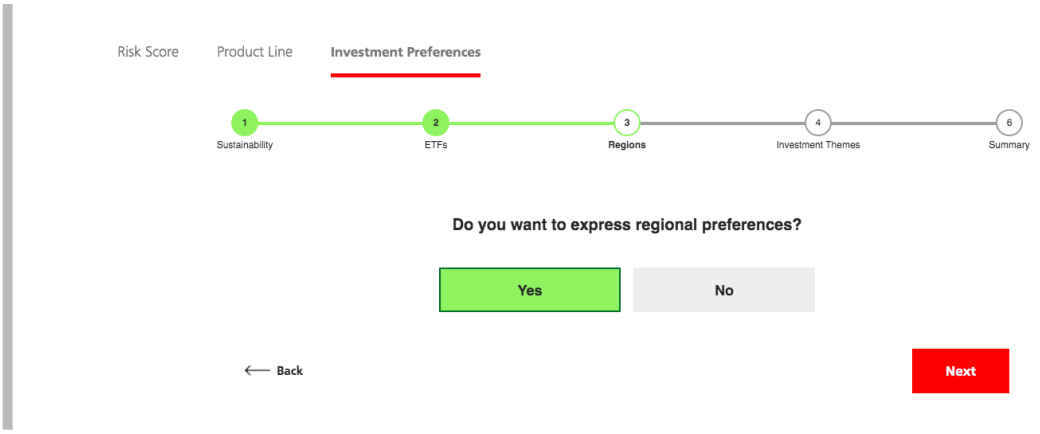

Figure 7 - Investment Preferences – Regions (measure of efficiency) – May 2022

News Article Sentiment Analysis Fitness Factor (in cooperation with the project partner Report Brain): to be designed and integrated.

- **How do you plan to measure the KPIs**? (Describe the operational activities)

Previously, personalized investment proposals were generated in a 2-4-hour time frame depending on various business approaches. The planned measurement approach is to continuously monitor the overall time that will be needed to create an investment proposal. The final goal of Pilot #4 is to develop AI GO portfolio optimization tool that will create such personalized investment proposals in less than 10 minutes. Thus, a time frame, within which investment proposals can be generated, will be taken as a variable to measure the identified KPIs. For information on other KPIs measurement see the table below (Table 4).

- **With which frequency do you plan/foresee to monitor the KPIs**?

Pilot #4 plans to monitor the KPIs on a yearly basis, depending on customer onboarding procedures.

- **Why such indicators are relevant to your Pilot**?

This continuous monitoring and overview will create an opportunity to adapt the AI GO portfolio optimization tool to clients' needs and requirements. Moreover, Pilot #4 has the goal of enabling portfolio optimization solutions not only for high-volume portfolios but also for the investors with

even smaller investment amounts. Mainly, the goal is to provide such a service to the masses as well. Ensuring that the personalized investment proposals are generated within a shorter time frame (less than 10 minutes), will enable Pilot #4 to satisfy the single investors needs with both "high and lower volume" investment space. Measurement of the portfolio time generation is a clear and transparent indicator of the successful improvement and implementation of the technology. This would act as a proof of concept which would then be used to acquire new customers. End-clients could be thus serviced better and at lower cost.

3.4.2 Key Actors Involved

Here (Figure 8) the Pilot provides a **description of the relevant actors/stakeholders involved**:

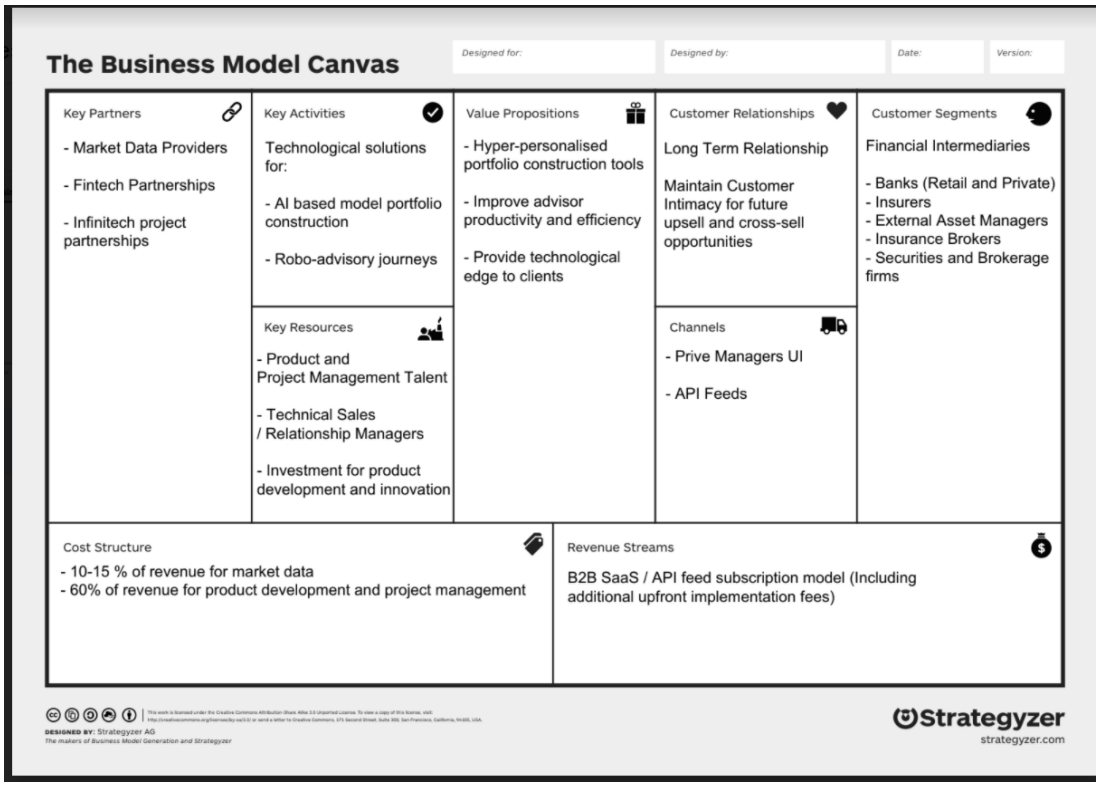

Figure 8 - Pilot #4 Relevant actors/stakeholders involved

As the goal of Pilot #4 is to combine or establish mainly Digital Workflows within highly regulated Securities Business, and possibly to create fully digitised Wealth Management / Advisory Journeys, the business model can be described as a B2B2C offering ("Robo like investments"). As the target market sets its focus on all financial services intermediaries who provide advisory and wealth management services, involved relevant actors/stakeholders are as follows:

- INFINITECH Partners as potential clients. PRIVE could enable interested parties of the financial sector to concentrate on customer advisory and an automated offering for “Private Banking” like services to the retail space.
- FinTech Partners who want to offer professional portfolio optimization services to end-clients.
- Banks and their Advisors for their end-clients.
- Asset Managers & their Retail Customers (“direct sales B2C”).
- Securities and Brokerage firms.
- Financial Intermediaries with a fully digitized user journey for portfolio construction. As previously identified, financial intermediaries and other relevant actors will have the opportunity to use a leading technology that will enable a fully scalable digitized advisory and wealth management journey for financial institutions and market participants.
- Market Data Providers.

3.4.3 Outline of the Current Status

In the following table (Table 4) the Pilot provides the information needed to outline the status of the KPIs:

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Core Task Efficiency (common-KPI)	**O**	How quickly a personalized and optimized portfolio can be generated?	X	X			Normal Advisory Journey. Typically, 2-4 hours	Currently, less than 1-3 minutes	10-20 seconds	NO
User Satisfaction (common-KPI)	**B**	UI Journey user friendly experience score	X	X			Portfolio score risk-return is developed, as well as individual factor score. This allows us to understand the quality of the results. The goal was to create a first prototype allowing the user to check on understandability & accuracy of the tool (whether it is easy for the user to interact with the tool).	Further and ongoing measurement of the portfolio return score. First user prototype of the UI Journey (as described) is created.	Most accurate version of the defined measurement. Achieved.	YES
Flexible portfolio construction based on personal risk profiling for retail clients	**O**	Number of fitness factors to choose from	X			X	4 fitness factors	5+ fitness factors	+10 fitness factors. Depending on the customer requirements.	NO
Accuracy of individual portfolio construction based on individual customer preferences	**O**	Tools for the user to understand the quality of the results	X	X			Portfolio score risk-return tool is developed. The goal: quality & transparency of the results.	Improvement of the tool showing the accuracy of the results.	Most accurate version of the defined measurement. Achieved.	YES
Accuracy of portfolio reporting after portfolio construction and execution of investment decisions	**O**	Portfolio health score to monitor the quality of the portfolio over time	X	X			The health score API as an ongoing monitoring tool	The health score API	Completed. API is available	YES
Improved **Advisor Productivity**	**B**	Time frame needed for the portfolio generation Same to Core Task Efficiency	X	X			Normal Advisory Journey. Initially, 2-6 hours	Currently, 5-10 minutes to generate the portfolio	5 minutes	NO

Hyper-personalization of portfolio construction	**B**	Generic investment theme method already implemented in the part of the fitness function topic	X			X	Covered by some of the fitness factors	Will be covered by some of the fitness factors	n.d.	NO
Increased Customer satisfaction with better risk-adjusted portfolios (Advisor / End-user)	**B**	to be defined once at the later stage of the project once the development stage is completed and prototype is available	X				To be defined once the prototype is delivered	To be defined once the prototype is delivered	To be defined once the prototype is delivered	NO
Remove barrier to entry for professional wealth management solutions for retail customers (mass-affluent) and their advisors	**B**	The number of investors with lower investment portfolio horizons The growth in the number of mass affluent customers	X			X	n.d.	n.d.	n.d.	NO

Table 4 - Pilot #4 KPIs status

3.4.4 Main Challenges

1. One major challenge is associated with the potential availability and accessibility of the market data. Mainly, no general availability of open-source market data hub for the required historical data used in the optimization processes that are used by the Pilot#4.
2. Another challenge is that there are no unified data providers, as a data hub service is required to import and transform data from different providers within a unified framework.
3. Finally, data cost including index license cost and the complexity of contracts by data providers is a challenge for the scalability of the commercial model and implementation.

3.5 Pilot #5b (BOC)

3.5.1 Description of the KPIs

Pilot #5b, namely Business Financial Management (BFM) tools delivering a Smart Business Advise, offers an innovative Business Financial Management (BFM) toolkit to SME customers of Bank of Cyprus. The KPIs chosen to measure the efficiency of the platform consist of both operational and business KPIs.

Regarding the operational KPIs, the following two have been identified to accurately measure the increase in efficiency of the proposed mechanism:

- **Transaction Categorization Rate**

Measures the percentage of transactions categorized by the transaction categorization engine, with the ideal target being above 90%. The accurate categorization is considered the foundation of all other microservices found within the developed BFM platform, thus considered a core task activity to retain credibility and increase the value of the business insights provided.

When the hybrid transaction categorization model was tested with the data provided by Bank of Cyprus, only 6% of total transactions were left uncategorized. However, once new SMEs transaction data are ingested at an operational stage, the results would decrease as the rule-based aspect of the model is biased due to the development process implemented. Thus, the transaction categorization accuracy will be revised on future times once new data are available. This means the injection of new SME transactions batch, which is preferred to decrease biased outcomes, or future transactions of the same SME customers currently utilized.

Status at time **T0** (*Now*): 96% transactions categorized out of initial 3,5 million SME transactions

Status at time **T1** (*December 2021*): >90% transactions left uncategorized out of an extra 2 million transactions

Status at time **T2** (*July 2021*): >90% transactions left uncategorized out of an extra 4 million transactions

- **Smart Virtual Advisor Response Time:**

The quick response time of the Smart Virtual Advisor is also considered vital for the pilot's success, as a reduced reaction over 1 sec would cause the users to lose attention or undermining the state-of-the-art aspects of the pilot. Thus, a KPI measuring the response time has been identified, which is measured by the time required to produce the expected outcomes of the BFM tool under 0,7 secs.

This KPIs has still not been measured, as it requires the introduction of newly developed data analytics components included in the BFM, setting new data streams and in general a higher overall Technology Readiness Level (TRL).

As far as the business KPIs are concerned, the following has been chosen to measure the impact of the developed BFM toolkit and the user's engagement. Both of the KPIs are still not measurable since an operational phase is required and are estimated to be introduced by Q2 2022.

- **Overall Efficiency**

As both Transaction Categorization Rate and Smart Virtual Response Time are vital for the pilot's success, this KPI combines the two pilot-specific operational KPIs to indicate the overall efficiency. It is noted that this KPI is subject to change, integrating further efficiency metrics related to newly introduced pilot data analytics microservices.

Status at **T0** (Now): 50%, sourcing from the currently achieved Transaction Categorization Rate, with the Response Time not being yet measurable

Status at **T1** (December 21): 80%, where a decrease to Categorization Rate is expected due to new transaction data ingestion and an operational Smart Virtual Advisor with measurable response time

Status at **T2** (July 2021): 100%, with both operational KPIs exceeding their target values

- **Customer Satisfaction**

One of the core KPIs, measuring the SME's overall content or discontent of the BFM microservices provided. The users will be asked to measure their satisfaction on a 2-choice option, namely Like and Don't like. The anticipated target is 80% like ratings.

- **Customer Engagement**

Measured by the increase in customer logins and time spent online. Deciding whether to track logins on a daily or weekly basis, as well as the ideal target spent online, will be done at a future point.

3.5.2 Key Actors Involved

- ***Bank of Cyprus (BOC)***: The bank is providing all data currently being used, setting the data analytics requirements and envisioning how the completed BFM toolkit should operate to cover the high expectations of their SME customers. Moreover, the bank is assisting in setting the required data streams as well as exploring and connecting external data sources utilized by the pilot. The Smart Business Advisor, currently being developed in the project's testbed, will also be migrated to the bank's own testbed, running also on the AWS environment.
- ***University of Piraeus Research Centre (UPRC)***: UPRC is responsible for all the development of the analytical components and the AI models, while also serving as a technical proxy to all technical needs and pilot development. Moreover, UPRC assists the bank in the design of the featured BFM microservices ensuring the provision of value adding tools to SMEs.
- ***SME customers of BOC***: As previously mentioned, SME customers of the Bank are the end users of the developed BFM toolkit, with tending their financial needs and providing accurate personalized insights being the main focus of the pilot.

3.5.3 Outline of the Current Status

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Overall Efficiency	**O**	Combined efficiency (mean) of the pilot's two operational KPIs	X	X	X		50%	50%	100%	No
Transaction Categorization Rate	**O**	>90% transactions categorized	X	X		X	88%	96%	> 90%	yes
Smart Virtual advisor Response time	**O**	Response time < 0.7 secs	X				n.d.	n.d.	n.d.	n.d.
Customer Engagement	**B**	Increase in customer logins & time spent logged in	X				n.d.	n.d.	n.d.	n.d.
Customer Satisfaction	**B**	Actionable insights will be presented with a Like / No Like option. An average value of 80% like is anticipated.	X		X		n.d.	n.d.	n.d.	n.d.

Table 5 - Pilot #5b KPIs status

3.5.4 Main Challenges

Main challenges are not strictly related to the definition or measurement of the KPIS. However, as further development is required to enable most KPI measurements, a key challenge faced in the pilot is migrating all development progress to the Bank's own testbed from the project's shared testbed provided by GFT. To this direction, a significant effort is required by the bank related to revisioning Terms and Conditions, enabling the required bank resources, setting the cloud environment as drafted in the scope of WP6 and migrating all development process. Of course, the collaboration of involved partners from different WP, as well as the main actor's engagement will assist overcoming potential barriers and having an early MVP, capable of measuring the KPIs described above.

3.6 Pilot #6 (NBG)

3.6.1 Description of the KPIs

Pilot #6 focuses on providing personalized investment recommendations for the retail customers of the bank. National Bank of Greece (NBG) will leverage large customer datasets and large volumes of customer-related alternative data sources (e.g., social media, news feeds, on-line information) to make the process of providing investment recommendations to retail customer **more targeted, automated, effective, as well as context-aware (i.e., tailored to state of the market)**.

Here the Pilot **provide a description** that answers the **following questions:**

- What are the identified KPIs?

> The KPIs we identified as the more appropriate measures in order to monitor the progress of our pilot can be found below:
>
> **1) Increase efficiency by allocating resources properly**
>
> **2) Increase effectiveness through prioritization based on expected Customer Investment**
>
> **3) Set appropriate targets based on existing Customer portfolio and potentials**
>
> **4) Proceed with more targeted proposals to Customers and increase sales and CSat**

- *Specify what is the core task* you have *identified over which you will be measuring its "Increase of Efficiency"* over time.

> Until now, RM (Relationship Manager) was able to make relevant propositions to the client without taking a deeper analysis of the investment profile and the most appropriate financial instruments that will maximize each individual customer's profit and investment performance. All the above, without taking into account neither an analytics methodology nor an algorithm.
>
> The main goal of Pilot #6 is to create a mechanism in order for the RM to propose more personalized investments to the client. The above can increase RM's efficiency as, from now on, it will have the opportunity through a recommendation engine, to suggest not only a more suitable instrument to the client, but also has the option to propose more than one instrument to the client.
>
> Apart from the day-to-day benefits that support both efficiency and effectiveness, mainly in the area of Operations, the final target is to lead Customers to fruitful investments and increase Customer satisfaction.
>
> The KPI that will measure the efficiency will be calculated at the end of the implementation in order to be more precise.

- **How do you plan to measure the KPIs**? (describe the operational activities)

> Our goal is to capture the % increase in the acceptance of the propositions for each branch and each RM through a KPI capable of measuring the above.
>
> All KPIs are planned to be measured either utilizing internal information, especially in regards to resources allocation and efficiency, while for Customer Satisfaction, we may consider performing Customer Experience Surveys so as to capture the sentiment through the particular experience. Given that this affects also other Departments , it is not yet finalized, but will be after the mechanism is fully operational and of course after implemented for a sufficient time period.

- **With which frequency do you plan/foresee to monitor the KPIs**?

> There is no specific time frame that the KPI should be measured.

- **Why such indicators are relevant to your Pilot**?

> The above-mentioned indicators fulfil the main requirement for more targeted, automated, effective investment recommendations to retail customer. Both operational and business indicators intend to measure the results of utilizing Pilot #6 developments as part of the business goals that a bank intends to achieve, minimizing the operational costs and increase customer satisfaction.

3.6.2 Key Actors Involved

The first stakeholder, in any case, is the client. The feedback received from the clientele, is the first measure of efficiency in addition to the number of proposals that lead to final acceptance. Furthermore, the capacity of the recommendation engine the pilot develops, it will be derived through mechanisms that will capture both the efficiency and the effectiveness of the platform.

3.6.3 Outline of the Current Status

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Core Task Efficiency (common-KPI)	**O**	% Increase of acceptance of the clientele propositions	X				n.d.	n.d.	n.d.	NO
User Satisfaction (common-KPI)	**B**	Questionnaires sent to targeted clientele segments	X				n.d.	n.d.	n.d.	NO
Increase efficiency by allocating resources properly	**O**	Measure number of RMs used for Personalized Investments	X				n.d.	n.d.	n.d.	NO
Increase effectiveness through prioritization based on expected Customer Investment	**O**	Measure and/or project the % net profit for recommendations provided utilizing Pilot's technologies	X				n.d.	n.d.	n.d.	NO
Set appropriate targets based on existing Customer portfolio and potentials	**B**	Calculate % increase of portfolio improvement for Customer portfolio and potentials	X				n.d.	n.d.	n.d.	NO

Make more targeted proposals to Customers and increase sales and CSat	**B**	Measure % difference of effectiveness for a) Customers' targeted proposals b) Sales Volumes increase c) Customer Satisfaction	X				n.d.	n.d.	n.d.	NO

Table 6 - Pilot #6 KPIs status

Note: The KPIs will be measured near the end of Pilot's implementation where will be more precise.

3.6.4 Main Challenges

The challenges of the Pilot are mainly related to the effectiveness of the targeted proposals that will be provided based on the relative algorithms implemented and improve the user satisfaction. In the next months, Pilot will focus on making the models and algorithms used for personalized investment proposals to be more targeted and build a better user interface for RMs to provide better propositions to clients and achieve higher level of satisfaction.

3.7 Pilot #7 (CXB)

3.7.1 Description of the KPIs

Here the Pilot **provide a description** in the form of a narrative that answers the **following questions:**

- **What are the identified KPIs?**

  - False positive rate
  - Percentage of number of frauds detected
  - Increased automation in fraud detection processes (operational cost saving)
  - User (SOC & eCrime employees) satisfaction

  Increase the number of frauds and the efficiency in “express loans” fraud detection. Now it is a completely manual process and the time spent for processing the data and detecting fraud is very dependent on the case. However, it is estimated by the fraud detection team as an average of 2-3 hours per case to identify and verify each fraud.

  Increased Automation (automated processes)

  Status at time T0: Completely manual (0%) NOW

  Status at time T1: Preliminary Analysis process (30%) July 2021

  Status at time T2: Automated Analysis process (50%) December 2021

  Status at time T3: Completely Automated process (100%) May 2022

  Increased Automation (time efficiency/reduction)

  Status at time T0: 0%* NOW

  Status at time T1: 20%* July 2021

  Status at time T2: 30%* December 2021

  Status at time T3: 50%* May 2022

  * [1- (Time needed to automated fraud detection / Time to manually detect fraud)] x 100

  Automated Fraud Detection Quality

  Status at time T0: 0%* NOW

  Status at time T1: 30%* July 2021

  Status at time T2: 80-90%* December 2021

  Status at time T3: >=100%* May 2022

  * (Number of automatically detected frauds / Number of manually detected frauds) x 100

- **How do you plan to measure the KPIs**?
  - False positive rate & percentage of detected frauds:
    - Identification of verified frauds.
    - Analysis of data.
    - Compare results of the analysis towards verified frauds.
    - Investigate and verify the false positives.
    - Compute KPIs (false positive rate and percentage of detected frauds)
  - Increased automation in fraud detection processes:
    - Calculate the total amount of time spent in manual fraud investigation.
    - Calculate the total amount of time needed in automatic fraud detection with the INFINITECH way.
  - User (SOC & eCrime employees) satisfaction:
    - Share a user satisfaction survey among the solution potential users, after testing a demo of the INFINITECH approach.
- **With which frequency do you plan/foresee to monitor the KPIs**?
  - False positive rate & percentage of detected frauds: Monthly/Every 3 months.
  - Increased automation in fraud detection processes: Once, after solution deployment.
  - User (SOC & eCrime employees) satisfaction: Once, after solution testing.
- **Why such indicators are relevant to your Pilot**?

  On the one hand, these indicators provide the usage feasibility of the solution in the long term. First, the quality of the obtained results, and especially “false positive rate” should be manageable to manage it completely automatically and avoid dedicated resources to validate a large number of cases. Second, the “percentage of detected frauds” KPI will indicate if the solution is able to automatically detect at least the same amount of frauds that are detected manually. And third, “User satisfaction” KPI will indicate the willingness and support of the target users to change the process and use the proposed solution.On the other hand, “increased automation in fraud detection processes” KPI will indicate the amount of resources, time and costs saved.

3.7.2 Key Actors Involved

Digital Security & IT team (CXB): Coordination of CXB efforts, pilot objectives, KPIs and requirements definition. Facilitating the data acquisition, pre-processing and extraction out of CXB premises when necessary. It is also managing the deployment of internal resources in CXB and guiding the evaluation of the results.

Tech Proxy (FTS): In charge of the technical coordination of the pilot, encompassing the activities of pilot architecture definition and alignment with the rest of INFINITECH ecosystem. It leads the technical development of the pilot, from the data analysis to the technical

Data owners and providers -eCrime & SOC employees- (CXB): Data providers and fraud investigators of eCrime and Security Operations Center from CXB, who are in charge of analysing cases of potential frauds. Providers of the field knowledge and pilot results evaluation.

Data analysts (FTS and FBK): Data analysts experts, who will be in charge of processing the data. They will research algorithms and fraud models that can automatically identify fraud cases. They contribute to scientific and technical tasks on data analysis, fraud prevention model generation and pilot evaluation.

3.7.3 Outline of the Current Status

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Core Task Efficiency (common-KPI)	**O**	Weighted sum of the level of achievement on KPIs (False Positive Rate, Percentage of detected frauds and Increased automation).	X				0%	n.d.	100%	NO
User Satisfaction (common-KPI)	**B**	User satisfaction survey	X				n.d.	n.d.	4.5 (out of 5 on user satisfaction feedback	NO
False positive rate	**O**	False positive ratio: FPR = FP / FP + TN FP (False Positives), TN (True Negatives)	X				50%	n.d.	0	NO
Percentage of detected frauds	**O**	%frauds detected: Number of automatically detected frauds / Number of manually detected frauds	X	X	X		60%	n.d.	=>100%	NO
Increased automation in fraud detection processes (time / operational cost saving)	**B**	[1- (Time needed to automated fraud detection / Time to manually detect fraud)] *100	X				2.5 hours/ case	n.d.	0.5 hours	NO

Table 7 - Pilot #7 KPIs status

3.7.4 Main Challenges

Improve the detection of Financial Fraud using refreshed customer data and AI based analysis of customers' transaction behavior.

Pilot #7 focuses on enhancing the current state of the art on machine learning tools and models for fraud detection and building them especially for tackling financial fraud, which is a use case with very specific characteristics (i.e., unsupervised learning on highly unbalanced datasets with very low fraud rate). More in detail, the problem of fraud detection and scoring for instance loans is rather specific and we do not have a tool for solving it. Generalistic tools like DataRobot (https://www.datarobot.com/) or Dataiku (https://www.dataiku.com/) do not provide enough suitable support on data preparation or they lack specific models to focus on the problem. The INFINITECH solution will foster both ends of the data chain from preparation to the specific detection and scoring models.

- Provide a solution that is able to automatically detect "express loans" fraud cases with a very low false positive rate.
- Provide a user-friendly solution that is able to be used by fraud detection teams, reduces its investigation time and means a smooth transition to them.

3.8 Pilot #8 (BOS)

3.8.1 Description of the KPIs

- **What are the identified KPIs?**

Conventional supervisory planning is time consuming and effort prone due to the requirement to collect data from various sources, complex process of data search and data analysis especially when taking in consideration big data. Several tasks for the purpose of supervision are currently managed manually, therefore the efforts of the Pilot #8 and functionalities developed within the Platform for AML Supervision are focused on the improvement of the supervisory process.

Main KPIs for Pilot #8 are:

- Decrease possibility of human error ≥ 60%.
- Increase the quality in the data analysis (discovery of patterns with additional data sources) ≥ 30%.
- Decrease time and effort for supervisory planning ≥ 50% ManDay (60MD → 30MD).
- Increase the quality of supervisory planning ≥ 50%.

Identified core Task: "Improved automatization in process of risk calculation and data ingestion"*:*

- **Status** at time **T0 (NOW)**: ≥ 20% of the process automated (1st version of the Risk Assessment Tool already implemented).
- **Status** at time **T1 (September 2021):** ≥ 40% of the process automated (Risk Assessment methodology update and implementation of the Distribution Channel for automated data gathering).
- **Status** at time **T2 (September 2022):** ≥ 60% of the process automated (final integration of the Risk assessment tool with the big data analysis results within the Screening tool).

Identified core Task: "Improved data analysis process"*:*

- **Status** at time **T0 (NOW)**: ≥ 0%.
- **Status** at time **T1 (September 2021):** ≥ 15% of the process improved (1st version of the ML components).
- **Status** at time **T2 (March 2022):** ≥ 25% of the process improved (ML components on production data).

- **Status** at time **T3 (September 2022):** ≥ 30% of the process improved due to final integration of the Risk assessment tool with the big data analysis results within the Screening tool.

Identified core Task: "Improved supervisory planning (time and effort)":

- **Status** at time **T0 (NOW)**: 60MD Pilot #8 tools under development.
- **Status** at time **T1 (December 2021):** 40MD improved supervisory planning (analysis of FI data based on the improved Risk Assessment methodology, data delivered through Distribution Channel).
- **Status** at time **T3 (September 2022):** 30MD improved supervisory planning (integration of the Risk assessment tool with the big data analysis results within the Screening tool.

Identified core Task: "Improved supervisory planning (quality)" :

- **Status** at time T0 (NOW): ≥ 0%.
- **Status** at time **T1 (September 2022):** ≥ 50% improved supervisory planning (integration of the Risk assessment tool with the big data analysis results within the Screening tool).
- **How do you plan to measure the KPIs?**

The progress and achievement of the KPIs will be monitored through quantitative and qualitative methods at the level of the milestones set in the Pilot #8 timeline (for example at the end of the Risk Assessment methodology update - how does the improved methodology improve the risk calculation process, what the KPI achieved?). Measurement methodology for specific Identified Core Tasks are provided in the Table 8 - Pilot #8 KPIs status.

- **With which frequency do you plan/foresee to monitor the KPIs**?

Depend on the PAMLS functionality under development, minimum on the quarterly level.

- **Why such indicators are relevant to your Pilot**?

As already mentioned, Pilot #8 is focused on the supervisory needs, therefore the partner BOS and JSI are developing a Platform for AML Supervision (PAMLS) that pursues digitalisation and automation of existing supervisory procedures in a way to enhance effectiveness, reduce costs and improve supervisory capabilities.

3.8.2 Key Actors Involved

Bank of Slovenia (BOS) as an end user and pilot leader provides the content, business overview & requirements and validation of Pilot #8. Key actors involved in Pilot #8 are Payment systems and settlement systems (data provider), IT & information security (technical and security requirements for the testbed), Legal and Compliance (support and overview of regulatory requirements, data protection and data confidentiality standards), Banking Supervision (testing of functionalities, usability and of ML/AI methods and end user).

Josef Stefan Institute (JSI) as technical leader provides all technical development for Pilot #8 – development of PAMLS- including Machine learning components, data visualization and business services, integrated as PAMLS. Key actor involved in Pilot #8 form JSI is Department of Artificial Intelligence (developer of the platform, developer of user requirements, developer of different AI /ML methods, provider of anonymization service).

- ***BOS -*** *Pilot Leader: Central bank, supervisory authority*
    - * End user - > set the user requirements
    - * Data provider (the right to process data)
    - * Testing the usability of ML / AI methods Location of the testbed
- ***JSI -*** *Tech Proxy: Research institute, technical partner*
    - * Developer of the platform
    - * Developer of the user requirements

- * Developer of different AI /ML methods
- * Provider of anonymization service

3.8.3 Outline of the Current Status

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Core Task Efficiency (common-KPI)	**O**	Functionalities tested on production data and comparison of current process vs. PAMLS process Number of hours spent for risk assessment. Number of data analyzed. Time spent for data analysis.			X		n.d.	n.d.	n.d.	NO
User Satisfaction (common-KPI)	**B**	Questionnaire to AML analysts to evaluate the user experience within PAMLs (per functionality). Number of manual tasks for: - data gathering, - data quality, - risk assessment, - reassessment.			X		n.d.	n.d.	n.d.	NO
Decrease possibility of human error ≥ 60%	**O**	Evaluation of current manual process vs. PAMLS automated process and controls.	X			X	0%	20%	60%	NO
Increase the quality in the data analysis (discovery of patterns with additional data sources)	**O**	Evaluation of existing data sources and number cases identified vs. number of cases identified via PAMLS big data analysis through data sources used Number of data sources for data analysis. Number of cases identified.	X			X	0%	O%	30%	NO
Decrease time and effort for supervisory planning (measurable in man/day)	**B**	Evaluation of time spent in current supervisory process vs. time spent for the supervisory process with support of PAMLS Man Day spent for risk assessment: - data gathering, - data quality, - risk assessment, - reassessment.	X			X	60 MD	n.d.	30MD	NO

Increase the quality of supervisory planning	**B**	Evaluation of the number of factors and data sources taken into account for supervisory planning now vs. the number of factors and data sources taken into account for supervisory planning using PLAMS tools. Number of data sources for risk assessment. Number of risk factors evaluated.	X			X	0%	0%	50%	NO

Table 8 - Pilot #8 KPIs status

3.8.4 Main Challenges

During the course of the Pilot #8 development following challenges were identified:

- ***AI regulation/governance:*** There are several documents addressing common AI diversity and complexity, however there is still no consistent regulatory framework. To address this issue in Pilot #8 will include the so-called "human in the loop" principle. While the analytical tool suggests the alerts about the potential risks and ML/TF typologies, the analyst in the supervisory department will decide on the output and the need for any supervisory actions.
- ***Data quality issues:*** Initial ML components development is running on the historical transaction data and the quality and completeness of the data was raised as a big challenge. However, the data quality is important for the PAMLS functionalities, more effort and time was appointed to the data preparation. However, the ML component development was on hold due to that reason.
- ***Regulatory and Compliance issues:*** BOS has to apply high regulatory and compliance standards also those arising from data privacy laws and confidentiality standards. Therefore, the data with personal and confidential information has to be pseudo-anonymized prior to entering the PAMLS. The level of the anonymization can affect the level of the collected information and consequently valuable analysis. Data was for this reason enriched prior pseudo-anonymization and pseudo-anonymization tool had to be redefined.

3.9 Pilot #9 (AKTIF)

3.9.1 Description of the KPIs

Here the Pilot **provide a description** in the form of a narrative that answers the **following questions:**

- **What are the identified KPIs?**

Pilot #9's main objective is analyzing blockchain transactions for fraudulent activities. Since blockchain transaction graphs are massive with billions of transactions, efficient computation is very important. Therefore, KPIs of the Pilot are identified according to efficiency.

Operational KPIs of Pilot #9 are related to size of the processed data and running time of the system. First KPI is the size of the transaction graph extracted from blockchain data. This KPI indicates the number of blockchain transactions that can be processed. Second KPI is loading times of partitioned transaction graphs from disk. Third KPI is related to scalability. This KPI checks the running time of graph algorithms as the size of blockchain increases. The last KPI is response time of queries.

In addition to operational KPIs, two business driven KPIs are identified. First one is increased automation in fraudulent activity tracing processes. This KPI indicates the operational cost saving. Second one is user satisfaction with the program's interface.

Identified Core Task1: *"How many blockchain transactions and addresses have been processed using 16 node cloud based HPC cluster?"*

Identified Core Task2: *"What is the best parallel transaction graph construction time using 16 node cloud based HPC cluster?"*

Identified Core Task3: *"What is the best parallel computation time for a trace of a very small subgraph (< 20 edges) of a blacklisted address using 16 node cloud based HPC cluster?"*

Identified Core Task4: *"What is the best parallel time for shortest path blacklisted node trace forest construction using 16 node cloud based HPC cluster?"*

Status at Time	Core Task 1	Core Task 2	Core Task 3	Core Task 4	Month
T0	Ethereum: 633M transactions, 69 M addresses	188 secs	few seconds	N/A	June 2020
T1	Ethereum: 766 M transactions, 78 M addresses	219 secs	1 sec	33 secs	May 2021
T2	Bitcoin: 625 M transactions, 800 M addresses	1910 secs	1 sec	240 secs	June 2021

Table 9 - Pilot #9 Core tasks evolution

- **How do you plan to measure the KPIs**? (Operational activities)
  - Since blockchain data continuously grows, periodically newly generated Bitcoin and Ethereum blockchain datasets are extracted and merged with existing datasets.
  - Tests involving graph construction and trace queries are run on an HPC cluster and timings obtained.
- **With which frequency do you plan/foresee to monitor the KPIs**?

Operational tests are carried out every three months after newly generated blockchain datasets are appended to existing datasets.

- **Why such indicators are relevant to your Pilot**?

Sustainability of the developed graph analysis system depends on the ability to handle continuously growing blockchain data - that is its scalability. When blockchain data size grows, it may not fit into single node computer memories. Therefore, the system is developed as parallel and distributed software. But due to communication costs in parallel programs, it is important that parallel software running times decrease or if growing, grow slowly. Therefore, our indicators which mainly measure time and sizes of transactions are relevant.

3.9.2 Key Actors Involved

Boğaziçi University (BOUN): is developing (i) "Blockchain Transaction Dataset Preparation Component" and (ii). "Scalable Transaction Graph Analysis Component".

Aktif Bank (AKTIF): is developing "User Interface for Blockchain Transaction Reports and Visualization Component". Aktif Bank business units are going to be the key actor for evaluating the pilot.

3.9.3 Outline of the Current Status

Core Task Efficiency = 0.15*A + 0.15*B + 0.15*C + 0.15*D + 0.4*E

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved

Core Task Efficiency (common-KPI)	**O**	Ability to do fraudulent analysis on Bitcoin and Ethereum blockchain data within similar (or slowly growing) time frames provided the number of nodes in the HPC cluster will be increased	X	X			n.d.	60%	n.d.	NO
User Satisfaction (common-KPI)	**B**	User satisfaction survey	X				n.d.	n.d.	n.d.	NO
Sizes of the transaction graph extracted from blockchain data, partitioned and stored on disk **(A)**	**O**	Number of Transactions and addresses.	X	X			n.d.	Bitcoin: 625 M transactions and 800 M addresses Ethereum: 766 M transactions and 78M addresses	n.d.	YES
Loading and construction times of partitioned transaction graphs **(B)**	**O**	Time to load and construct graph	X	X			n.d.	Bitcoin: 1910 sec	n.d.	YES
Scalability: how running times of the graph algorithms grow as the number of nodes and/or the size of blockchain increases **(C)**	**O**	Parallel Running Time / Sequential Running time (as the transaction size is increased)	X	X			n.d.	speed-up obtained for fixed sized dataset	n.d.	YES
Response time of queries **(D)**	**O**	Blacklisted address to user address trace query time.	X	X			n.d.	few seconds	n.d.	YES

Increased automation in fraudulent activity tracing processes (operational cost saving) **(E)**	**B**	[1- (Time needed to trace fraudulent transactions using transaction graph analysis system / Time needed to manually trace fraudulent transactions on public blockchain explorers)] *100	X				n.d.	n.d.	n.d.	NO

Table 10 - Pilot #9 KPIs status

3.9.4 Main Challenges

Public blockchain data is huge. Processing huge blockchain data requires HPC cluster and massive storage. Hence, cloud computational and storage costs can be challenging especially when new blockchains like Ethereum2 with high throughputs will be deployed in the future.

Another challenge is blacklisted address compilation. We are dependent on surfing and extracting blacklisted addresses from publicly available Internet sources. This requires a lot of human effort.

3.10 Pilot #10 (PI)

3.10.1 Description of the KPIs

Meaning and relevance of KPIs - Measuring efficiency upgrades introduced by Pilot#10 in operations

In order to measure the efficiency and effectiveness of the Fraud detection system developed in Pilot #10, four significant KPIs have been identified, considering them as the most relevant to evaluate if the Pilot reached its pre-set objectives.

Over the years several approaches have been developed and tested in order to improve the effectiveness of the rule-based detection methodologies, but current trends suggest that promising results could be obtained by adopting analytics at scale based on an agile data foundation and solid Machine Learning (ML) technologies. By measuring the *number of frauds detected* which previously passed under the radar of a traditional rule-based fraud detection system, we clearly aim at demonstrating by facts that the approach to fraud detection followed in Pilot #10 is a winning approach.

Another area of great potential improvement in fraud detection is related to the high number of *false positives* which are really common with traditional rule-based systems and costly to manage, wasting hours of investigations performed by highly-skilled analysts. That's why we focused this area and expect to demonstrate that technologies developed in Pilot #10 will be able to get a relatively low number of false positives, considering average performance[3] of current technologies available on the market.

The increasing number of financial and payment transactions driven by several phenomena including the exponential growth of sales in e-commerce seen in recent years (just to name one) requires greater and greater efforts for investigating suspect fraudulent transactions; Pilot #10 introduced

innovative technical measures for increasing automation and therefore operational efficiency for the personnel involved in analysis and investigation processes. That's why we wanted to measure the level of automation in operation (*increased automation in fraud detection processes*) and we expect a sensible increase of the overall detection process led by INFINITECH-ALIDA technology.

When talking about *a more efficient and effective transaction investigation by means of data visualization and analysis tools* we aim at measuring the impressive leaps forward brought by a total change of paradigm in transaction data analysis: integrating the mere and difficult analysis of huge-sized tabular data with graphical representations of not evident correlations among features, of hidden patterns and trends, enables just-in-time fraud mitigation, allowing analysts to detect criminal phenomena just when they are occurring and not retrospectively when the "valuables" are lost forever.

Most important, as a core task for representing in synthesis the results of the entire innovation framework introduced by Pilot #10, we would like to measure the *overall decrease of operational costs* we expect from the introduction of INFINITECH technologies.

How we collect and evaluate KPIs

Number of frauds detected ever recognized before > 0

The system will be tested against a synthetic dataset which includes a majority of regular transactions and a number of fraudulent transactions which have not been previously detected by rule-based fraud-detection systems. The number of fraudulent transactions detected will be recorded. The measurement will be executed once per each synthetic dataset that will be made available during project lifespan.

False positives rate <= 20%

The system will be tested against a synthetic dataset which includes a majority of regular transactions and a number of fraudulent transactions. The number of false positives notified by the system will be recorded. The measurement will be executed once per each synthetic dataset that will be made available during the project lifespan.

Increased automation in fraud detection processes

A typical fraud analysis process will be represented including the most demanding steps. The number of manual process steps will be measured before and after introducing INFINITECH-ALIDA to demonstrate increased automation. The representation will include an esteem made by process owners (a value scaling from 0 to 10) of the operational load for analysts associated with each step. A new process will be represented considering analysts who make use of the Data Science and Machine Learning (DSML) tool adopted within such use-case to automation features. Each step will be weighted by process owners (a value scaling from 0 to 10). Summing up the weights with old and new processes will demonstrate the efficiency brought by automation. The measurement will be executed once per each synthetic dataset that will be made available during the project lifespan.

More efficient and effective transaction investigation by means of data visualization and analysis tools

Average time for establishing a fraud case will be requested to process owners. Average time for establishing a fraud case with visualization and analysis tool integrated within the ML-based Fraud detection system proposed, will be measured when testing them against a synthetic dataset which includes a majority of regular transactions and a number of fraudulent transactions. The measurement will be executed once per each synthetic dataset that will be made available during the project lifespan.

Overall decrease of operational costs (core measurement)

Average cost of fraud management activities per single fraud will be requested to process owners. Reference average cost will be retrieved from literature if available also. The cost of operations with the ML-based Fraud detection system in place will be estimated and compared to the average in order to measure the expected savings. The measurement will be executed once per each synthetic dataset that will be made available during the project lifespan.

User satisfaction (core measurement)

End users involved in the pilot will be surveyed at the end of piloting phase therefore collecting their subjective evaluations on three important aspects:

- (usability) Pls. evaluate ML-based Fraud detection system in terms of ease of use (choose a grade from 0 - 5)
- (effectiveness) Pls. evaluate ML-based Fraud detection system in terms of effectiveness (choose a grade from 0 - 5)
- (efficiency) Pls. evaluate ML-based Fraud detection system in terms of transaction analysis' speed (choose a grade from 0 - 5)

3.10.2 Key Actors Involved

The pilot will involve POSTE IT playing the role of *End-Users/Process Owners* and being able to evaluate the performance of the ML-based Fraud detection system as a real-time fraud-detection engine by identifying user requirements and KPIs, providing transactional data, developing and running the Real-Time Transaction Generator component and by using the Pseudonymization Tool provided by *JSI Tech Partner* while *ENGINEERING as a Tech Proxy* will go for technical coordination of the piloting phase, for developing the pilot, running the service infrastructure, collecting measurements data, fine-tuning the system.

3.10.3 Outline of the Current Status

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Core Task Efficiency Overall decrease of operational costs (common-KPI)	**O**	Calculate difference Δ between average cost of operations and estimated cost of operations with the ML-based Fraud detection system implemented	X				n.d.	n.d.	Δ > 0	NO
User Satisfaction (common-KPI)	**B**	Collect end-users' feedbacks by surveys and calculate the average M_1 of all scores assigned by users	X				n.d.	n.d.	M_1 >= 4 M_1 ranging from 0 to 5	NO
Number of frauds detected >= 5	**O**	Collect number of frauds *F* detected with the ML-based Fraud detection system implemented that were not recognized by rule-base engine	X				n.d.	n.d.	F >= 5	NO
False positives rate <= 20%	**O**	Collect number F_p of false positives wrongly identified by the ML-based Fraud detection system implemented	X				n.d.	n.d.	F_p <= 20%	NO

Increased automation in fraud detection processes	**B**	Calculate difference Δ_2 between overall cost (weight) of operations with and without the ML-based Fraud detection system implemented $\Delta_2 = \sum_{i=1}^{n} W_i - \sum_{j=1}^{m} W_j$ Wi & Wj = weights of process steps with or without ML-based Fraud detection system implemented	X				n.d.	n.d.	$\Delta_2 > 0$	NO
Availability of graphical tools for more efficient and effective transaction analysis	**B**	Collect end-users' feedbacks by surveys and calculate the average M_2 of all scores assigned by users	X				n.d.	n.d.	$M_2 >= 8$ M_2 ranging from 0 to 10	NO

Table 11 - Pilot #10 KPIs status

KPI Plan

KPI Name, status & formula	When?
Overall decrease of operational costs (core measurement)	
Status at time T0: Completely manual (0%)	NOW
Status at time T1: Preliminary Analysis process (30%)	July 2021
Status at time T2: Automated Analysis process (50%)	December 2021
Status at time T3: Completely Automated process (100%)	May 2022
User satisfaction (core measurement)	
Status at time T0: Completely manual (0%)	NOW
Status at time T1: Preliminary Analysis process (30%)	July 2021
Status at time T2: Automated Analysis process (50%)	December 2021
Status at time T3: Completely Automated process (100%)	May 2022
Number of frauds detected ever recognized before > 0	
Status at time T0: Completely manual (0%)	NOW
Status at time T1: Preliminary Analysis process (30%)	July 2021
Status at time T2: Automated Analysis process (50%)	December 2021
Status at time T3: Completely Automated process (100%)	May 2022
False positives rate <= 20%	
Status at time T0: Completely manual (0%)	NOW
Status at time T1: Preliminary Analysis process (30%)	July 2021
Status at time T2: Automated Analysis process (50%)	December 2021
vStatus at time T3: Completely Automated process (100%)	May 2022
Increased automation in fraud detection processes	
Status at time T0: Completely manual (0%)	NOW
Status at time T1: Preliminary Analysis process (30%)	July 2021
Status at time T2: Automated Analysis process (50%)	December 2021
Status at time T3: Completely Automated process (100%)	May 2022

More efficient and effective transaction investigation by means of data visualization and analysis tools	
Status at time T0: Completely manual (0%)	NOW
Status at time T1: Preliminary Analysis process (30%)	July 2021
Status at time T2: Automated Analysis process (50%)	December 2021
Status at time T3: Completely Automated process (100%)	May 2022
Overall decrease of operational costs (core measurement)	
Status at time T0: Completely manual (0%)	NOW
Status at time T1: Preliminary Analysis process (30%)	July 2021
Status at time T2: Automated Analysis process (50%)	December 2021
Status at time T3: Completely Automated process (100%)	May 2022
User satisfaction (core measurement)	
Status at time T0: Completely manual (0%)	NOW
Status at time T1: Preliminary Analysis process (30%)	July 2021
Status at time T2: Automated Analysis process (50%)	December 2021
Status at time T3: Completely Automated process (100%)	May 2022

Table 12 - Pilot #10 KPIs Plan

3.10.4 Main Challenges

Frauds on financial services are an ever-increasing phenomena and cybercrime generates multi-million revenues, therefore even a small improvement in fraud detection rates would generate significant savings. It is well-known that the fraud phenomenon is endemic and could never be totally eradicated, but only mitigated more or less effectively. In this framework, wide areas for improvement still currently exist: one need only partially addressed is banally to improve the fraud detection rate, decrease the number of false-positives – analyzing them requires considerable resources – and at the same time reducing the number of false-negatives which impacts negatively organizations first by the costs of the undetected frauds but also generating a misleading sense of security in users: if the system does not detect frauds, this doesn't mean there aren't. Improve the efficiency and effectiveness of a fraud-detection system, allowing the implementation of mechanisms for automated or semi-automated decisions, ensuring high-impact business results and a significant reduction of CAPEX & OPEX.

Pilot #10 arises from the need to overcome the limitations of the rule-based systems to block potentially fraudulent transactions. In order to overcome the limitations of the rule-based approach, we adopted machine learning techniques which are able to address many of these limitations, and more effectively identify risky transactions. A novel AI-based fraud detection system - built over a Data Science and Machine Learning – has been developed for the pre-processing of transaction data and model training in a batch layer (to periodically retrain the predictive model with new data) while in a stream layer, the real time fraud detection is handled based on new input transaction data. The developed architecture makes this solution a valuable tool for supporting fraud-analysts, for automating the fraud detection processes and for facing new threats led by unpredictable, ever seen before, fraud schemas.

3.11 Pilot #11 (ATOS)

3.11.1 Description of the KPIs

Operational Driven KPIs

These KPIs refer to technical indicators, provided by (and supported from) the different components of the involved frameworks: the Smart Fleet (data gathering) that captures and builds first data aggregations; and the EASIER (AI framework) used to develop the AI powered driving profile model.

Volume (number) of routes captured/evaluated

Whitin Pilot's #11, a "route" comprises all technical data captured from a given vehicle within a fixed time interval, e.g. from the very moment its engine is started till it is stopped, so a new route would be created every time a vehicle starts. The vehicle's technical data includes speed, acceleration, heading, location, fuel consumption, gases emissions, and status, among other datasets captured by its On Board Unit (OBU). From the vehicle's perspective, the route is considered as the basic data unit to train and test the Driving Profiling AI model. The more captured, identified and classified routes, the better testing and training scenario for the Driving Profiling model we get.

Task: *"capturing vehicles' routes by aggregating vehicles' technical data (from simulated and real environments) and processing them (classify each route according defined clusters)"*

Status at time **T0**: 60 routes identified and classified from the Simulation environment (measure of efficiency) – NOW

Status at time **T1**: 30k routes combining Simulated routes and live routes (of the participating vehicles equipped by CTAG) (measure of efficiency) – December 2021

Status at time **T2**: over 60k routes identified and classified from (mainly) live vehicles and real drivers (participant vehicles equipped by CTAG) (measure of efficiency) – May 2022

Number of Driving profiles identified and defined

Based on the captured routes and the correlations obtained with the reported weather conditions and related traffic alerts, we will obtain a set of initial clusters that will define, as we analyse the characteristics of each cluster, the later Driving Profiles. A (very) short number of identified drivers' profiles would provide a poor classification (e.g., a simple good/bad profile may not be enough to design customized services) whilst a (too) wide set of profiles could not group enough the drivers to get valuable feedback. We aim to get about 6 different profiles.

Task: *"identify and define the clusters obtained by the captured data processing by using AI methodologies"*

Status at time **T0**: 4 initial clusters obtained from simulated routes (measure of efficiency) – NOW

Status at time **T1**: 5-6 preliminary clusters (pre-driving profiles) including data from participant vehicles equipped by CTAG (measure of efficiency) – December 2021

Status at time **T2**: > 5 stable driving profiles defined from clusters attributes and supported by the insurance company (DYN) (measure of efficiency) – May 2022

Business Driven KPIs

These KPIs reflects the exploitation, from the insurance business' perspective, of the outcomes provided by the technical implementation of Pilot #11 and refers to the customized products and services envisioned by the car insurance companies.

Number of new products/services offered by the Insurance Company

The driving profiling AI model will be used by the insurance company to develop insurance tools to define and develop novel customized services for both, the insured clients and the insurance company itself or even 3rd interested companies.

Task: *"design and develop final apps/services customized to insured clients' behaviours and/or to insurance companies needs"*

Status at time **T0**: Identified 2 basic services (Pay as you Drive and Fraud Detection). 0 final services/apps/products (offered to the insured client/insurance company) developed. (measure of efficiency) – NOW

Status at time **T1**: at least 2 final services/apps/products defined (not yet implemented) – December 2021

Status at time **T2**: > 2 final services/apps/products (offered to the insured client/insurance company) developed. – May 2022

Accuracy of Driver's (insured client) classification

This KPI evaluates the Driving Profiling model accuracy by performing with tagged datasets from insured vehicles. This KPI will be applied to each of the identified driving profiles.

Task: *"compare the driving profiling model outcomes using testing (and tagged) routes from the live data captured (CTAG), simulated data, and/or given by the insurance company (DYN) from previously profiled insured clients"*

Status at time **T0**: no accuracy has been tested (no available driving profile tool) (measure of efficiency) – NOW

Status at time **T1**: accuracy > 60% (for initial AI models and per each profile) by using smartfleet testing datasets – December 2021

Status at time **T2**: final Driving Profiling accuracy > 75% (per each profile) by testing with insurance company (DYN) datasets (measure of efficiency) – May 2022

Pilot's efficiency (overall) measurement

The pilot's efficiency will be managed by averaging the "Accuracy of Driver's classification" KPI applied to each of the identified driving profiles, using previously tagged routes.

Identified Core Task: *"average the accuracies obtained for each defined driving profile"*

Status at time **T0**: no accuracy has been tested (no available driving profile tool) (measure of efficiency) – NOW

Status at time **T1**: efficiency > 60% (for initial AI models) by using smartfleet testing datasets – December 2021

Status at time **T2**: final efficiency (average Driving Profiling accuracy) > 70% by testing with insurance company (DYN) datasets (measure of efficiency) – May 2022

3.11.2 Key Actors Involved

According to the Pilot's diagram, we identified the relevant actors as follows:

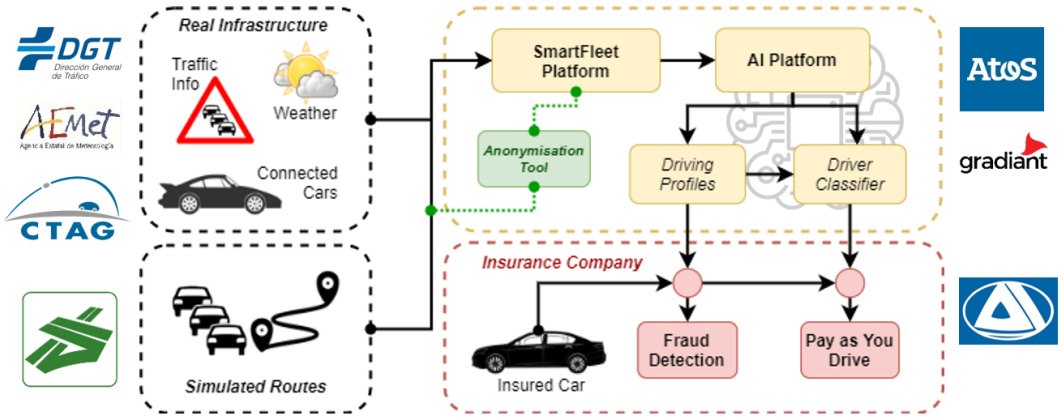

Figure 9 - Pilot #11 Relevant Actors

Data providers, remarked in black, data providers supply the required raw datasets to implement the final services. We can identify two main data providers according to the nature of the raw data:

- Real datasets providers: here are included the connected cars (main data source), infrastructure provided by CTAG, and the context information, composed by the weather info (supported by ATOS and provided by the AEMET) and the Traffic information, in turn including traffic alerts (supported by CTAG and extracted from the DGT) and roads information (extracted from Open Street Map)
- Simulated routes: provided by the SUMO based simulation scenarios supported by ATOS

Framework providers. Two main actors provide the different frameworks (remarked in yellow) that build the technical core of the solution:

- ATOS provides its Smart Fleet framework to collect and homogenize all the raw data according to selected standards; and the EASIER solution to develop and serve the AI models
- Gradiant supports the anonymization component

Insurance stakeholder, in red dotted box, exploits the AI models defining and developing the final applications to be finally offered.

- Dynamis (DYN) plays this role in Pilot #11, helping to identify the different driving profiles from the obtained clusters, to define the final services offered whilst providing data from their own insured clients to evaluate accuracy of the final models and services acceptance.

3.11.3 Outline of the Current Status

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Core Task Efficiency (common-KPI)	**O/B**	average the accuracies obtained for each defined driving profile	X		X		n.d.	n.d.	> 70%	NO
User Satisfaction (common-KPI)	**B**	compare Driving Profiling model outcomes with current classification of the provided routes (from insurance company and/or testing environment)	X		X		n.d.	n.d.	> 70%	NO

Volume (number) **of routes captured/ evaluated**) (Set of attributes captured; real vehicles measurements; traffic alerts measured)	**O**	list all the aggregated routes for testing/ training the AI model	X	X			0	60	>10K	NO
Number of Driving profiles identified and defined	**O**	identify and detail the clusters derived from the AI analysis	X				0	4 clusters (pre-driving profiles)	>5	NO
Number of new products/services offered by the Insurance Company	**B**	list all novel services/apps developed exploiting Pay as You Drive and/or Fraud detection	X	X	X		0	0	>=2	NO
Accuracy of Driver's (insured clients) **classification**	**B**	compare Driving Profiling model outcomes with current classification of the provided routes (from insurance company and/or testing environment)	X		X		n.d.	n.d.	> 75% per each profile	NO

Table 13 - Pilot #11 KPIs status

3.11.4 Main Challenges

The main Pilot's #11 target is to develop an accurate driving profiling AI powered model that, by exploiting IoT capabilities to gather real time datasets from connected cars plus related context sources, allows car insurance companies to redesign their current business models by customizing the clients' primes according to their real and updated risks. At the same time, this information will improve the way the different responsibilities are assigned in case of a traffic incident, helping to detect and avoid possible frauds. In this sense, and to get the best results in a potential final scenario, we have detected several challenges to be addressed, during the pilot's life and in a foreseeable future:

- The first challenge deals with the main data sources, the cars, and pursues to **involve as many connected cars as possible**. To make this easier, we develop an open framework based on standards from automotive communication protocols such as OBD2, CANBus, etc. and standards from IoT infrastructures, e.g. ETSI NGSI or FIWARE.
- In terms of AI, we have to **Identify the most relevant datasets** to properly define and differentiate diverse driving profiles; **select the most proper ML/DL techniques** among all new existing methodologies; to finally **develop accurate AI powered driving profiling models**.
- From the business perspective, the pilot should identify those **services or applications** that better exploit the new calculated risks and lead to a **better acceptance** by final users.
- Finally, and related to the trustworthy AI, we need to evaluate the **impact of data anonymization** in the derived Driving Profiling Models, so these final models can be offered according to AI ethics.

3.12 Pilot #12 (SILO)

3.12.1 Description of the KPIs

Pilot #12 aspires to demonstrate that continuous risk assessment of healthcare insurance premiums is possible and useful to health insurance professionals.

“Possible” refers to being able to provide the infrastructure and secure the acceptance of the clients. The former is to be proven by the setup of the pilot testbed for training models and the employing of Healthentia, an eClinical platform by Innovation Sprint for the data collection at the client side and the dashboard provision at the healthcare professional side. The latter is twofold: clients should on the one hand accept sharing data with their insurance companies in the hope of lowering their premium, while on the other hand they should be able and willing to use the data collection system in the long run. Willingness to share the data has already been addressed via an online survey answered by 216 individuals, of which 181 responses are complete. The complete responses are being analysed and the results will be reported in one of the forthcoming WP7 deliverables. It is the ability and willingness of the clients to use the data collection system that is enumerated via the following pilot-specific KPIs addressing the “possible” aspect:

1. At least 75% of the clients complete more than 85% of all triggered questionnaires forwarded to them in the first 2 months.
2. At least 75% of the clients wear their sensors to provide the automatic measurements for at least 6 out of 7 days in each of the first 2 months.
3. At least 75% of the clients are satisfied with the mobile app after 2 months of use.
4. Upon completing the 2nd month of their usage of the mobile app, at least 50% of the clients are willing to keep using it in the long run.

All these KPIs can be continuously monitored on a weekly basis, starting on the 6th week of the pilot (mid October 2021). The first two KPIs are evaluated by the Pilot #12 tech proxy monitoring the progress from within the Healthentia platform. KPIs 3 and 4 are evaluated by RRD employing user satisfaction surveys at the end of each client’s 2nd month of usage.

The average of the client percentages satisfying the conditions in KPIs 1 and 2 will provide the efficiency of the pilot (efficiency common KPI).

The average of the percentages satisfying the conditions in KPIs 3 and 4 will provide part of the user satisfaction of the pilot (user satisfaction common KPI). The other part will address the “useful” aspect of the goal of the pilot, i.e. the opinion of the health insurance end user on the effectiveness of the provided dashboard for their continuous risk assessment. This is evaluated by the fifth pilot-specific KPI:

1. The health insurance professionals can confidently determine insurance premiums (increase, decrease or stay) monthly for at least 50% of the clients.

This KPI will be evaluated by the insurance partner on Pilot #12, Dynamis, monthly involving only the active clients, as they are identified by KPIs 1 and 2. The tools at the disposal of the health insurance professional will change throughout the pilot:

- During a first period, the decision will be based on the collected data only since no trained ML model exists.
- During a second period of the pilot, the expert decision will be based on the data and the model decisions on the health outlook of each client. Not enough such decisions are yet collected though, to allow their accumulation into a risk score.
- During a third period of the pilot, the expert decision will be based on the data and the model decisions accumulated into a risk score for each client.

The pilot partners are interested to establish if the increased involvement of the ML model in the decision process will be helpful to the health insurance professionals as it matures throughout the pilot.

The user satisfaction common KPI will be evaluated as the weighted sum of the percentages in KPIs 3, 4 and 5. KPIs 3 and 4 refer to client satisfaction and will each carry a weight of 0.25, while KPI 5 refers to health insurance professional satisfaction and will carry a weight of 0.5.

3.12.2 Key Actors Involved

There are three relevant Pilot #12 actors:

- **Insurance professionals**: They will be aided by the provided dashboard that depicts the general population data and the suggestions by the predictive models in deciding about clients' premium adaptation.
- **General population** (prospective insurance company clients): Every adult eligible to sign some sort of health insurance contract with an insurance company can be in this category. We are excluding children from our study to avoid unnecessary ethical implications. The individual must consent to the use of their personal health and behavioral data in the decision process for insurance premiums adaptation.
- **ML engineers**: They will be utilizing the pilot testbed to train and test models of different underlying algorithms and complexity, as well as performing tests with different levels of anonymization.

3.12.3 Outline of the Current Status

The following table summarizes the KPIs of Pilot #12 and their status. Please note that the initial KPI measurement refers to the continuous data collection phase of the pilot, being carried out with a limited number of people from the pilot organizations. During this pilot phase there have been no user satisfaction surveys, thus none of the business KPIs have been evaluated. There are no current KPI measurements (see Table 14); they will start being available after the 6th week of the pilot validation phase with participants from the general population (mid of October 2021).

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Core Task Efficiency (common-KPI)	**O**	Average of the percentages in the two operational pilot-specific KPIs	X	X		X	44.6%	n.d.	75%	NO
User Satisfaction (common-KPI)	**B**	Weighted average of the percentages in the three business pilot-specific KPIs	X	X		X	n.d.	n.d.	60%	NO
At least 75% of the clients complete more than 85% of all triggered questionnaires forwarded to them in the past 2 months	**O**	Objectively determined by the data collected at Healthentia	X	X		X	46.4%	n.d.	75%	NO

At least 75% of the clients wear their sensors to provide the automatic measurements for at least 6 out of 7 days in each of the past 2 months	**O**	Objectively determined by the data collected at Healthentia	X	X		X	42.9%	n.d.	75%	NO
At least 75% of the clients are satisfied with the mobile app after 2 months of use	**B**	Subjectively reported by the clients in a survey	X	X		X	n.d.	n.d.	75%	NO
Upon completing the 2nd month of their usage of the mobile app, at least 50% of the clients are willing to keep using it in the long run	**B**	Subjectively reported by the clients in a survey	X	X		X	n.d.	n.d.	50%	NO
The health insurance professionals can confidently determine insurance premiums (increase, decrease or stay) monthly for at least 50% of the clients	**B**	Subjectively reported by the health insurance professional	X	X		X	n.d.	n.d.	50%	NO

Table 14 - Pilot #12 KPIs status

3.12.4 Main Challenges

The main challenge we face in the pilot is the lack of a health insurance partner that can help us train our models by annotating the data and provide the expert assessment for KPI 5. Dynamis is active in the insurance sector, but not specifically in the domain of health insurance, and will have to play this role in KPI 5. The annotations for model training are going to be based on the clients' self-assessment of their health.

3.13 Pilot #13 (WEA)

3.13.1 Description of the KPIs

Operational Driven KPIs

These KPIs are what we use to measure the technical and scientific process and how the results obtained bring us closer to the objectives of the pilot. They are based on the data collection, volume, density and validity of the data for the purposes of the pilot.

Percentage of SMEs relevant information found in external sources used

Based on the data sources found and the density of information available for each of the companies analysed. To determine whether a company is found or not, minimum information fields such as name, address, activity, tax identification number, etc. are used.

This indicator is fundamental because it is the first and most relevant indicator to validate that the data sources are adjusted and provide value.

Identified core Task: *"Identify and obtain relevant information in each of the areas of the pilot's data map so that the developed algorithms can work correctly and develop the risk models."*

Status at time **T0**: 50% of the information found from open sources from a giving SMEs data set – NOW

Status at time **T1**: 60% of the information found from open sources from a giving SMEs data set – September 2021

Status at time **T2**: 85% of the information found from open sources from a giving SMEs data set – April 2022

Density of information in each area

The data map of the pilot organises the collection of information from the different sources into 6 different areas that serve to service the different insurance products that an SME consumes, ranging from property to cyber risk, general liability, etc.Therefore, this KPI allows us to evaluate the information obtained and how dense it will be to service each of the functionalities of the pilot such as information improvement, customer knowledge and process automation.

Identified core Task: *"Identify and obtain relevant information in each of the areas of the pilot's data map so that the developed algorithms can work correctly and develop the risk models*."

Status at time **T0**: 30% of the different areas with relevant information – NOW

Status at time **T1**: 30% of the different areas with relevant information – September 2021

Status at time **T2**: 30% of the different areas with relevant information – March 2022

Business Driven KPIs

Percentage of fields found for automated company insurance issuance

In order to evaluate business indicators, it is essential to understand how insurers can increase the efficiency of their underwriting and quoting processes by reducing data entry times and improving the user experience for both the insurer and the end customer through an online interface. Therefore we need to know the % of information fields that we are able to obtain in an automated way.

Identified core Task: *"Automatic collection of as many fields as possible for the underwriting of a company product, e.g. general liability.."*

Status at time **T0**: 30% Number of fields to be completed in the case of issuance of an SME liability product– NOW

Status at time **T1**: 50% Number of fields to be completed in the case of issuance of an SME liability product– October 2021

Status at time **T2**: 90% Number of fields to be completed in the case of issuance of an SME liability product– April 2022

Level of use

This KPI is set to measure the level of usage by an insurer's stakeholders to see if the service and information provided is of value and is the first step in scaling the solution within the insurer. It is established as the number of users who continue to use the service on a recurring basis once it has been implemented.

Identified core Task: *"Number of users who use the system in relation to the total number of potential users of an insurer***.**"

Status at time **T0**: 10% of the potential users have adopted the service/platform/tool– NOW

Status at time **T1**: 20% of the potential users have adopted the service/platform/tool – December 2021

Status at time **T2**: 60% of the potential users have adopted the service/platform/tool – May 2022

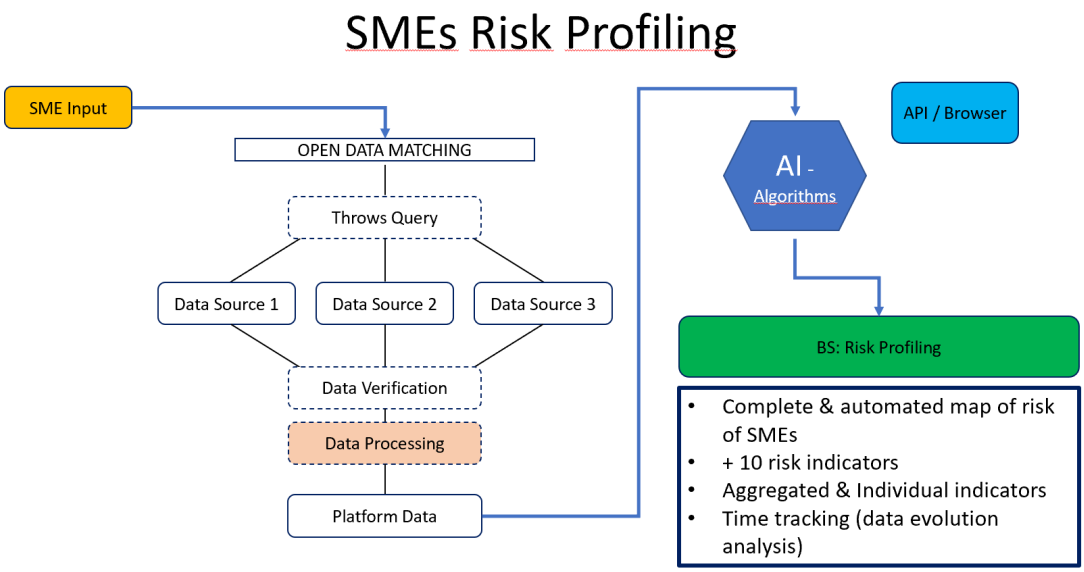

Figure 10 - Pilot #13 SME Risk Profiling diagram

3.13.2 Key Actors Involved

The key actors involved in this pilot project are companies offering commercial insurance cover or policies to companies, in particular SMEs, and their staff, who come from various departments: the actuary, the product developer and the underwriter, who are involved in risk assessment, pricing and evaluation of product conditions, will benefit from the increased availability of data on company risks. These risks affect several insurance products such as commercial and industrial multi-risk, fire, cyber security, general and professional liability. In addition, the sales manager or agent has a larger risk database of the target market(s) and can therefore better identify target regions to diversify and balance the insurance portfolio.

Brokers and agents can use the improved risk database and information for their advisory services (e.g. product development) and a broader approach to achieve individual risk transfer solutions for insurance companies in the international reinsurance market.

Finally, it should be noted that also the automation of processes benefits the end policyholder by improving the process and giving the policyholder the possibility to obtain the insurance price directly from the internet.

Partners for this project are Wenalyze for the platform to obtain the data and ML algorithms and LeanXcale to manage the information. There are no data providers because the info is got form open data sources

3.13.3 Outline of the Current Status

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Core Task Efficiency (common-KPI)	**O**	Average of the number of targets founds and density of information	X		X		40%	50%	72,5%	
User Satisfaction (common-KPI: Number of errors corrected in data base)	**B**	Comparing original and new found information in the data base	X		X		n.d.	n.d.	n.d.	

Percentage of SMEs relevant information found in external sources used	**O**	number of targets with information found versus no information found	X	X		X	50%	60%	85%	No
Density of the information found in each of the SMEs	**O**	Information found in each of the areas into which the data map is divided.	X	X			30%	40%	60%	No
Percentage of fields found for automated company insurance issuance	**B**	Percentage of fields that can be filled in or obtained automatically from the sources without human intervention	X		X		30%	50%	90%	No
Level of usage (from employees or system integrators)	**B**	Number of users who use the system in relation to the total number of potential users of an insurer.	X		X		10%	20%	60%	No

Table 15 - Pilot #13 KPIs status

3.13.4 Main Challenges

The main objective of Pilot #13 is to improve the collection of information that insurers obtain from their SME clients to improve risk assessment, increase the automation of processes through the collection of underwriting information and improve the quality of the data that these financial institutions have on their business clients in order to better understand them.

The challenges we face are as follows:

Firstly, that of obtaining information. Insurance companies do not usually have data available for use due to legacy problems, information silos and the low quality of the information available to them. Therefore, we have to look for open sources that allow us to overcome this problem. We resort to both structured and unstructured open sources whose connectivity is not always well defined. The integration of the sources in the pilot is the initial challenge.

Regarding the design of the algorithms, the challenge is to be able to select the SME variables obtained from the open sources that best explain the risks and the greater aversion or predisposition to them. To this end, modelling will be carried out with data sets obtained from insurance companies that will allow us to select these variables. Also, with a view to automating and cleaning the database, it is important to know the level of accuracy, veracity and timeliness of the information contained in these variables.

In terms of business, the challenge lies in disseminating the solution in a suitable way and impacting decision-makers in the insurance companies and transmitting the usefulness of the solution and its impact on improving the business models of the insurance companies and how this will not only contribute to improving results but also to increasing the competitiveness of the company.

3.14 Pilot #14 (GEN)

3.14.1 Description of the KPIs

Pilot #14 aims to provide a full operational monitoring tool for agricultural insurers enabling the monitoring of land parcels in terms of weather events and crop growth. The AgI toolbox is designed to support the efficiency of policy handling in agricultural insurance.

In order to measure efficiency and usability of the product, operational-driven and business-driven KPIs were identified comparing the savings in time and costs to the traditional processing of policies. Besides the KPIs valid for all pilots for measuring efficiency and user satisfaction (KPI1+2), the pilot specific KPIs are designed to monitor opportunities of the AgI toolbox improving the daily tasks an insurance company needs to conduct for agro insurance portfolio management. Agricultural insurers are enabled to increase the status of automatization of insurance processes especially in underwriting, loss adjustment and claims handling in crop insurance.

The status of automatization can be measured either by time saving in underwriting a policy or claims handling (KPI4), as well as the number of insurance tasks and process steps the AgI toolbox can take over (KPI3). Information (current situation, weather alerts, forecasts) about natural perils is provided tailored to the insurer's portfolio. For each feature (Damage Assessment Tool, Risk Mapping, Early Warnings/Weather Risk probability) at least one KPI was designed to monitor the development progress and usability of the service (KPI6-9). The more calamities are covered by the AgI toolbox, the better the insurer can rely on the provided data and analysis for its portfolio monitoring process including single risk cover, multi-peril crop insurance or parametric insurance products (KPI5).

Finally, the mitigation of "fraud attempts or false claims", is measured by the number of incidents - false claims that the Agri-Insurance actor will identify through INFINITECH AgI toolbox (KPI10).

The user target group are actors within an agro insurance company. In order to measure the operational and business performance of the tool, a close cooperation and feedback loop with the pilot customers will be established to gather their experience. Based on the different services provided in the AgI toolbox different professions within the company are asked to fill in a questionnaire of quantitative and qualitative indicators per crop season, but also in some cases in a number of iterations (after significant calamities, launch of new products etc.), starting in autumn 2021.

In addition to the requested evaluation by the users or rather the insurance companies, the project team of Pilot #14 will monitor operational KPIs constantly regarding the measurement of accuracy of predictions for weather alerts and yield damages in comparison with observed claim data gathered from pilot customers and other sources.

3.14.2 Key Actors Involved

The key actors involved in this pilot are companies offering Agricultural Insurance coverage/policies and their personnel, coming from a diverse set of departments; the *actuary*, *product developer and underwriter*, being involved in risk evaluation, pricing and product terms and conditions assessment, will benefit from the increased weather and pest and disease risk data availability and the underwriting automation. The *loss adjuster & claims manager* are mainly concerned with the event prediction and damage estimation before and after the event occurrence. While the loss adjuster is responsible for the field-specific assessments, the claims manager has a broader view (responsible for the whole portfolio) and is also highly interested in event damage mitigation and fraud detection. Both will be supported by remote earth observation (EO) data analysis, as well as by automation of damage reporting and contractual payout. Furthermore, the sales manager or agent has an increased risk data basis of the target market(s) and can therefore better identify target regions in order to diversify and balance the insurance portfolio.

Brokers & Consultants can use the improved risk data basis and information for their advisory services (e.g. product development) and a more comprehensive approach to achieve individual risk transfer solutions for insurance companies on the international reinsurance market.

3.14.3 Outline of the Current Status

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved

KPI1: Premium Calculation Efficiency (common-KPI)	**O**	Time required to calculate premium for a new coverage/policy Vs time required through traditional calculation process	X				n.d.	n.d.	n.d.	NO
KPI2: User Satisfaction (common-KPI)	**B**	Pilot insurance companies are asked to evaluate (on a scale between 1-10 from very bad to very good) their usage satisfaction after testing the AgI tool at certain intervals	X				n.d.	n.d.	n.d.	NO
KPI3: % of automation achieved on AgI company level (quantitative criterium)	**O**	Pilot insurance companies are asked to evaluate % of policy processing tasks in underwriting and claims handling that can be conducted automatically via the tool	X				n.d.	n.d.	n.d.	NO
KPI4: AgI Process chain-time decrease (%) (quantitative criterium)	**O**	Pilot insurance companies are asked to evaluate Time saving (duration per policy) in underwriting and claims handling by using the tool compared to before	X				n.d.	n.d.	n.d.	NO
KPI5: Number of calamities covered (drought, fire, ice, flood) (quantitative criterium)	**O**	Number of weather risks covered by the available datasets and analyses of the tool	X				n.d.	n.d.	n.d.	NO
KPI6: Increase accuracy of Risk Mapping and premium calculation for the Insurance Company (qualitative criterium)	**B**	Pilot insurance companies are asked to evaluate (on a scale between 1-10 from very bad to very good) their risk mapping and premium calculation capacities (risk data availability, spatial and temporal resolution) before and during using the AgI tool. In addition, G&Co as a consultant will evaluate (on a scale between 1-10 from very bad to very good) the usefulness for product development.	X				n.d.	n.d.	n.d.	NO

KPI7: Improved alerting (Early Warnings/ Weather Risk probability) send to Ins. companies and damage mitigation measures deployment (quantitative criterium)	**B**	(1) Pilot insurance companies are asked to evaluate % of weather alerts in comparison to experienced insurance claims; (2) % of alerts compared to actual occurrence of the forecasted events (3) Pilot insurance companies are asked to evaluate Number of alerts forwarded to the insurer's clients for mitigation actions	X				n.d.	n.d.	n.d.	(1-3) NO
KPI8: Improved damage assessment (quantitative criterium)	**O**	% of damage predictions after event within -20%/+20% range of loss adjuster estimations	X				n.d.	n.d.	n.d.	NO
KPI9: O&A Cost reduction (savings as a result of the automated process incl. Damage Assessment) (quantitative criterium)	**B**	Pilot insurance companies are asked to evaluate % of savings for loss adjustment costs per event	X				n.d.	n.d.	n.d.	NO
KPI10: Fraud attempts or false claims mitigation number of false claims detected using EO data vs false claims detected through on-site assessment (quantitative criterium)	**B**	% of clients with insurance claim but no event detection.	X				n.d.	n.d.	n.d.	NO

Table 16 - Pilot #14 KPIs status

*All KPI's will be calculated and compared on the basis of Agri-Insurance operations with and without the use of INFINITECH AgI toolbox. At present there are no specific average sectoral values (given the extensive heterogeneity of actors in terms of size, network and capacity) to compare our findings with. *

3.14.4 Main Challenges

Pilot #14 identified challenges include:

- **Data integration:** insurance company's data needs to be compiled in such a way that it can be imported into the AgI toolbox.
- **System integration:** how does the AgI toolbox and its functionalities fit the current workflow in an insurance company & reluctance to change from Agriculture Insurance employees.
- **Pilot user time resources:** the pilot users are often very busy with their everyday work and need to allocate a certain amount of time to spend in the feedback process.
- **Regulatory inefficiencies:** how does national legislation/regulation hinder the adoption of AgI toolbox (e.g obligatory in-field visits to perform damage assessment).

3.15 Pilot #15 (ABILAB)

3.15.1 Description of the KPIs

The main objective of Pilot #15 is to develop, integrate and deploy a data-intensive system to extract key concepts from banks internal documents for the definition of common business glossaries against a reference taxonomy (ABI Lab taxonomy), and support the standardization of the classes associated to the documentation analyzed.

In order to achieve this objective, ABI Lab is developing a ML engine able to recognize and make inferences about banking concepts, as described or cited by free texts. These inferences are semantic ones for metadata creation, where “accept” or “reject” correspond to reasoning about the text.

The architectural design is based on BERT (Bidirectional Encoder Representations from Transformers), the transformer-based architecture that reaches the state-of-the-art for Text classification, Sequence labelling and Question Answering.

BERT support Embeddings for the whole sentence in input as well as, simultaneously, for all the occurring words. It allows us to mining natural language sentences to solve Natural Language Inference tasks.

For the scope of the pilot we:

- Adapted BERT to the banking language.
- Injected the ABI Lab's taxonomy evidence in BERT.
- Designed an ABI Lab taxonomy classifier.

To illustrate how effectively the pilot objectives associated with the model performance are achieved (performance of the algorithm and impact on the business**),** KPIs monitoring and evaluation will be crucial throughout the project. For this purpose, ABI Lab grouped different metrics in two main domains: technological/operational performance and business performance, which are strictly interconnected as they both lead to evaluate the model’s efficiency.

Technological/Operational Metrics:

- Classification (Recall, Precision)
- Impact: % of saved manual analysis (automation of the workflow)

Business Metrics

- Usability / Satisfaction (the speed and simplicity to which a user can accomplish tasks)
- Availability of real time responses (compare the results obtained with the as is situation)

The chosen metrics represents ad-hoc KPIs for modelling the semantic accuracy from a business perspective. The core task identified for ABI Lab’s model to measure the performance variation over time (Increase of efficiency), is to **tag correctly the documents analyzed** (automatic metadata extraction from documents, automatically associating ABI Lab taxonomy processes to portions of bank documents). To achieve this goal, the F1 score metric will be used, namely a combination of precision and recall, to obtain accuracy in the pilot binary classification evaluation.

Identified core Task: *“how accurate is the capability to* automatically associate ABI Lab taxonomy processes to portions of bank document”

> **Status** at time **T0**: F1 score less than 10% (measure of efficiency) – Prior to the pilot
>
> **Status** at time **T1**: F1 score above 70% (measure of efficiency) – December 2021
>
> **Status** at time **T2**: F1 score above 75% (measure of efficiency) – May 2022

Regular assessment and semester reporting starting from the status (T0) in June 2021, will be implemented throughout the project. As far as concerns technical/operational KPIs, they will be provided by the system by default, while for the business KPIs we foresee the involvement of the banks end users at later stage with a questionnaire to measure the degree of usability/satisfaction and the availability of real time response.

3.15.2 Key Actors Involved

Starting from the Pilot's 15 reference architecture (Figure 11), we identified the relevant actors as follows:

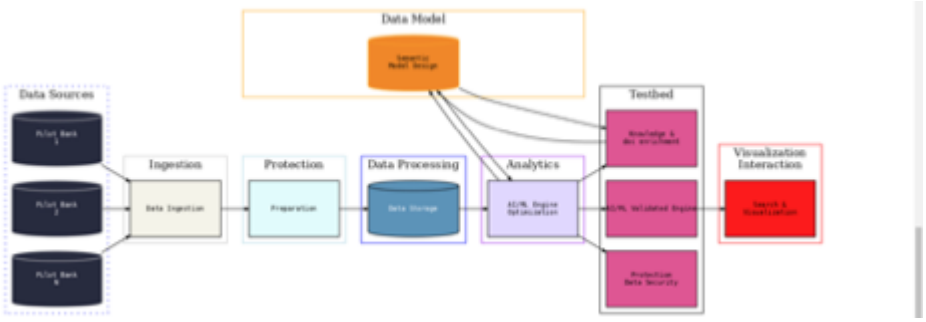

Figure 11 - Pilot #15 Reference Architecture

Data providers: to supply the required raw datasets (Banks involved in the project)

Researcher partner: to test the most relevant NLU models, verifying their ability to support the analysis of internal document within a bank to design the semantic model

Technical partners: GFT and HPE

End users: Data Governance Officers and Enterprise Architecture from the banks involved in the project.

3.15.3 Outline of the Current Status

KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
Core Task Efficiency (common-KPI)	**O**	F1 score, monitored through technical measurement process	X	X		X	9,09%	74,04%	Above 75%	NO
User Satisfaction (common-KPI)	**B**	Interviews with end users	X				n.d.	n.d.	n.d.	NO
% Automation of workflow	**O**	Interviews with end users	X				n.d.	n.d.	n.d.	NO
Usability	**B**	Interviews with end users	X				n.d.	n.d.	n.d.	NO

Table 17 - Pilot #15 KPIs status

3.15.4 Main Challenges

The main Pilot's #15 target is to develop an accurate semantic model to detect key concepts related to the internal regulatory process and to map the tagged concepts against common standardized taxonomies (e.g., data, processes, business applications, etc.). This information would allow the standardization of the classes associated to the documentation analyzed as well as the creation of a unified reference glossary and continuous improvement of reference taxonomies and architectural maps.

In order to achieve the target, and to get the best results, several challenges had and will need to be addressed during the pilot's cycle:

> · Governance: the involvement of several stakeholders needed a clear and shared governance.
>
> · Data Protection and Data Governance, in order to be compliant with trustworthy AI.
>
> · Identify and select the most proper ML techniques among all new existing methodologies to develop an accurate semantic model.

3.16 Summary Tables

Here ABILab took care of gathering all the feedback provided by the Pilots and fill in the tables accordingly (Table 18-19). This is meant to achieve a straightforward overview of all the Pilots in terms of their status with the KPIs, related measurements and goals.

The following is a legend to enable the reader to understand every aspect of the table reported down below:

Type: Operational-driven KPI (O), Business-driven KPI (B)

Measurement Mode: how the KPI will be measured

Status: represents the current status of the Pilot. Put an X under the column you relate to (the columns are not mutually exclusive):

- **(ID) = Identified**: the KPI has been identified
- **(M) = Measured**: the KPI has been measured
- **(NR) = Need refinements**: the KPI will be subject to future refinements
- **(CM) = Continuously monitored**: the KPI is in advanced/mature state by which we are currently under monitoring

Initial KPI Measurement: measurement prior to the start of the project (if it exist)

Current KPI Measurement: current value of the KPI

Target Level: indicates the level towards which to make the KPI tend (and eventually achieve)

Achieved: yes/no statement to indicate whether the target level has been currently achieved

#Pilot	KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
1	**No Processable**: how many documents had extraction errors and couldn´t be processed (%)	O	Precision & recall over human validated examples	X	X			n.d.	60%	30%	NO

#Pilot	KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
	Effectiveness of the extraction of low level concepts	**O**	Precision & recall over human validated examples	X	X			n.d.	60%	90%	NO
	Effectiveness of the recognition of high level concepts	**O**	Precision & recall over human validated examples	X	X			n.d.	50%	90%	NO
	Core Task Efficiency: Effectiveness of the business rules (common-KPI)	**B**	Estimated savings (processing time, number of processed invoices, estimated economic savings).	X				n.d.	n.d.	n.d.	NO
	User Satisfaction (common-KPI)	**B**		X				n.d.	n.d.	n.d.	NO
2	**Fewer VaR violations** compared to existing models [P(r < VaR_a) = 1-a]	**O**	Percentage of daily VaR/ES violations per year in confidence level 99% (mean value between utilized portfolios)	X	X		X	2.5%	<1%	<1%	YES
	VaR estimation of multi-asset portfolios in less than 5 seconds	**O**	Seconds	X	X		X	3 secs (4 FX assets per portfolio)	1 sec (4 FX assets per portfolio)	1 sec (10 FX assets per portfolio)	NO
	Real Time VaR availability instead of daily (batch) VaR estimation	**B**	Seconds	X	X		X	300 secs (4 assets per portfolio)	60 secs (4 assets per portfolio)	60 sec (10 assets per portfolio)	NO
	User satisfaction	**B**	Arithmetic average based on the evaluation of three sub-KPIs (intuitivity, usability, effectiveness)	X				n.d.	n.d.	>4	NO
3	**Use of different data sets** from financial and Social media	**O**	Enabling the use of the sample datasets during the project	X				Not Yet	Not Yet	6+ Synthetic Data Sets	Pending
		O	Providing the prototype implementation and enabling the demonstration and use of the data sets during the project.	X				Not Yet	Not Yet	6+ Synthetic Data Sets	Pending

#Pilot	KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
		O	Providing the prototype implementation and the demonstration and use of the applications and/ or methods during the project.	X				Not Yet	Not Yet	6+ Synthetic Data Sets	Pending
	Use of Single Data Model for data processing	**O**	Providing the prototype implementation and the demonstration and use of the applications and/ or methods during the project.	X				Not Yet	Not Yet	6+ Synthetic Data Sets	Pending
		O	Providing the prototype implementation providing the demonstration and use of the reference use case during the project.	X				Not Yet	Not Yet	6+ Synthetic Data Sets	Pending
		O	Providing the demonstration and use of the common data model during the project, using the proposed BOI Data model.	X				Not Yet	Not Yet	6+ Synthetic Data Sets	Pending
	Profile identification accuracy based on input data	**B**	Identification & generation of five human trafficking centric red-flag typologies which are integrated within the Bank of Ireland FIU Know Your Customer processes and/or systems.	X				Not Yet	Not Yet	1 Data Platform Available Used and tested	Pending
		B	Having ready a comprehensive report that study the different integration and interoperability solutions for data sharing in the context of KYC/ KYB.	X				Not Yet	Not Yet	1 Study Analysis Report Available	Pending

#Pilot	KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
	Reduction of false positives based on enriched input data	**B**	Enabling the FIU teams or other organizations interested to have a repeatable & reliable application for identifying human trafficking related services.	X				Not Yet	Not Yet	1 Data Platform Available Used and tested	Pending
		B	Enabling the FIU teams to have a repeatable & reliable process for identifying human trafficking related red-flag typologies to use in their KYC processes.	X				Not Yet	Not Yet	1 Data Platform Available Used and tested	Pending
		B	Aggregating the confidence scores from the NLU models for each red-flag indicator contained in the typology and ranking the highest scoring typologies.	X				Not Yet	Not Yet	5 Red-Flag Typologies Used	Pending
4	**Flexible portfolio construction** based on personal risk profiling for retail clients	**O**	Number of fitness factors to choose from	X			X	4 fitness factors	5+ fitness factors	+10 fitness factors. Depending on the customer requirements.	NO
	Accuracy of individual portfolio construction based on individual customer preferences	**O**	Tools for the user to understand the quality of the results	X	X			Portfolio score risk-return tool is developed. The goal: quality & transparency of the results.	Improvement of the tool showing the accuracy of the results.	Most accurate version of the defined measurement. Achieved.	YES
	Accuracy of portfolio reporting after portfolio construction and execution of investment decisions	**O**	Portfolio health score to monitor the quality of the portfolio over time	X	X			The health score API as an ongoing monitoring tool	The health score API	Completed. API is available	YES
	Improved **Advisor Productivity**	**B**	Time frame needed for the portfolio generation Same to Core Task Efficiency	X	X			Normal Advisory Journey. Initially, 2-6 hours	Currently, 5-10 minutes to generate the portfolio	5 minutes	NO

#Pilot	KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
	Hyper-personalization of portfolio construction	**B**	Generic investment theme method already implemented in the part of the fitness function topic	X			X	Covered by some of the fitness factors	Will be covered by some of the fitness factors	n.d.	NO
	Increase Customer satisfaction with better risk-adjusted portfolios (Advisor / End-user)	**B**	to be defined once at the later stage of the project once the development stage is completed and prototype is available	X				To be defined once the prototype is delivered	To be defined once the prototype is delivered	To be defined once the prototype is delivered	NO
	Remove barrier to entry for professional wealth management solutions for retail customers (mass-affluent) and their advisors	**B**	The number of investors with lower investment portfolio horizons The growth in the number of mass affluent customers	X			X	n.d.	n.d.	n.d.	NO
5b	**Transaction Categorization Rate**	**O**	>90% transactions categorized	X	X		X	88%	96%	> 90%	YES
	Smart Virtual advisor Response time	**O**	Response time < 0.7 secs	X				n.d.	n.d.	n.d.	NO
	Customer Engagement	**B**	Increase in customer logins & time spent logged in	X				n.d.	n.d.	n.d.	NO
	Customer Satisfaction	**B**	Actionable insights will be presented with a Like / No Like option. An average value of 80% like is anticipated.	X		X		n.d.	n.d.	n.d.	NO
6	**Increase efficiency** by allocating resources properly	**O**	Measure number of RMs used for Personalized Investments	X				n.d.	n.d.	n.d.	NO
	Increase effectiveness through prioritization based on expected Customer Investment	**O**	Measure and/or project the % net profit for recommendations provided utilizing Pilot's technologies	X				n.d.	n.d.	n.d.	NO
	Set appropriate targets based on existing Customer portfolio and potentials	**B**	**Calculate % increase of portfolio improvement for** Customer portfolio and potentials	X				n.d.	n.d.	n.d.	NO

#Pilot	KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
	Make more targeted proposals to Customers and increase sales and CSat	**B**	Measure % difference of effectiveness for a) Customers' targeted proposals b) Sales Volumes increase c) Customer Satisfaction	X				n.d.	n.d.	n.d.	NO
7	**False positive rate**	**O**	False positive ratio: FPR = FP / FP + TN FP (False Positives), TN (True Negatives)	X				50%	n.d.	0	NO
	Percentage of detected frauds	**O**	%frauds detected: Number of automatically detected frauds / Number of manually detected frauds	X	X	X		60%	n.d.	=>100%	NO
	Increased automation in fraud detection processes (time / operational cost saving)	**B**	[1- (Time needed to automated fraud detection / Time to manually detect fraud)] *100	X				2.5 hours/ case	n.d.	0.5 hours	NO
	User Satisfaction	**B**	User satisfaction survey	X				n.d.	n.d.	4.5 (out of 5 on user satisfaction feedback	NO
8	**Decrease possibility of human error ≥ 60%**	**O**	Evaluation of current manual process vs. PAMLS automated process and controls	X			X	0%	20%	60%	NO
	Increase the quality in the data analysis (discovery of patterns with additional data sources)	**O**	Evaluation of existing data sources and number cases identified vs. number of cases identified via PAMLS big data analysis through data sources used Number of data sources for data analysis. Number of cases identified.	X			X	0%	0%	30%	NO

#Pilot	KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
	Decrease time and effort for supervisory planning (measurable in man/day)	**B**	Evaluation of time spent in current supervisory process vs. time spent for the supervisory process with support of PAMLS. Man Day spent for risk assessment: - data gathering, - data quality, - risk assessment, - reassessment.	X			X	60 MD	n.d.	30 MD	NO
	User Satisfaction	**B**	Questionnaire to AML analysts to evaluate the user experience within PAMLs (per functionality) Number of manual tasks for: - data gathering, - data quality, - risk assessment, - reassessment.			X		n.d.	n.d.	n.d.	NO
	Increase the quality of supervisory planning	**B**	Evaluation of the number of factors and data sources taken into account for supervisory planning now vs. the number of factors and data sources taken into account for supervisory planning using PLAMS tools. Number of data sources for risk assessment. Number of risk factors evaluated.	X			X	0%	0%	50%	NO
9	**Sizes of the transaction graph** extracted from blockchain data, partitioned and stored on disk	**O**	Number of Transactions and addresses.	X	X			n.d.	Bitcoin: 625 M transactions and 800 M addresses Ethereum: 766 M transactions and 78M addresses	n.d.	YES

#Pilot	KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
	Loading times of partitioned transaction graphs	**O**	Time to load and construct graph	X	X			n.d.	Bitcoin: 1910 sec	n.d.	YES
	Scalability: how running times of the graph algorithms grow as the number of nodes and/or the size of blockchain increases	**O**	Parallel Running Time / Sequential Running time (as the transaction size is increased)	X	X			n.d.	speed-up obtained for fixed sized dataset	n.d.	YES
	Response time of queries	**O**	Blacklisted address to user address trace query time.	X	X			n.d.	few seconds	n.d.	YES
	Increased automation in fraudulent activity tracing processes (operational cost saving)	**B**	[1- (Time needed to trace fraudulent transactions using transaction graph analysis system / Time needed to manually trace fraudulent transactions on public blockchain explorers)] *100	X				n.d.	n.d.	n.d.	NO
	User satisfaction	**B**	User satisfaction survey	X				n.d.	n.d.	n.d.	NO
10	**Number of frauds detected >= 5**	**O**	Collect number of frauds *F* detected with the ML-based Fraud detection system implemented that were not recognised by rule-base engine	X				n.d.	n.d.	F >= 5	NO
	False positives rate <= 20%	**O**	Collect number F_p of false positives wrongly identified by the the ML-based Fraud detection system implemented	X				n.d.	n.d.	F_p <= 20%	NO

#Pilot	KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
	Increased automation in fraud detection processes (operational cost saving)	**B**	Calculate difference Δ_2 between overall cost (weight) of operations with and without the ML-based Fraud detection system implemented $\Delta_2 = \sum_{i=1}^{n} W_i - \sum_{j=1}^{m} W_j$ Wi & Wj = weights of process steps with or without ML-based Fraud detection system implemented	X				n.d.	n.d.	$\Delta_2 > 0$	NO
	Availability of graphical tools for more efficient and effective transaction analysis	**B**	Collect end-users' feedbacks by surveys and calculate the average M_2 of all scores assigned by users	X				n.d.	n.d.	M_2 >= 8 M_2 ranging from 0 to 10	NO
11	**Volume** (number) **of routes captured/ evaluated**) (Set of attributes captured; real vehicles measurements; traffic alerts measured)	**O**	List all the aggregated routes for testing/ training the AI model	X	X			0	60	>10K	NO
	Number of Driving profiles identified and defined	**O**	Identify and detail the clusters derived from the AI analysis	X				0	4 clusters (pre-driving profiles)	>5	NO
	Number of new products/services offered by the Insurance Company	**B**	List all novel services/apps developed exploiting Pay as You Drive and/or Fraud detection	X	X	X		0	0	>=2	NO
	Accuracy of Driver's (insured clients) **classification**	**B**	Compare Driving Profiling model outcomes with current classification of the provided routes (from insurance company and/or testing environment)	X		X		n.d.	n.d.	> 75% per each profile	NO

#Pilot	KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
12	At least 75% of the clients complete more than 85% of all triggered questionnaires forwarded to them in the past 2 months	**O**	Objectively determined by the data collected at Healthentia	X	X		X	46.4%	n.d.	75%	NO
	At least 75% of the clients wear their sensors to provide the automatic measurements for at least 6 out of 7 days in each of the past 2 months	**O**	Objectively determined by the data collected at Healthentia	X	X		X	42.9%	n.d.	75%	NO
	At least 75% of the clients are satisfied with the mobile app after 2 months of use	**B**	Subjectively reported by the clients in a survey	X	X		X	n.d.	n.d.	75%	NO
	Upon completing the 2nd month of their usage of the mobile app, at least 50% of the clients are willing to keep using it in the long run	**B**	Subjectively reported by the clients in a survey	X	X		X	n.d.	n.d.	50%	NO
	The health insurance professionals can confidently determine insurance premiums (increase, decrease or stay) on a monthly basis for at least 50% of the clients	**B**	Subjectively reported by the health insurance professional	X	X		X	n.d.	n.d.	50%	NO
13	**Percentage of SMEs relevant information found** in external sources used >50%	**O**	number of targets with information found versus no information found	X	X		X	50%	60%	85%	NO
	Density of the information found in each of the SMEs	**O**	Information found in each of the areas into which the data map is divided.	X	X			30%	40%	60%	NO
	User satisfaction (based on level of accuracy)	**B**	Percentage of fields that can be filled in or obtained automatically from the sources without human intervention	X		X		30%	50%	90%	NO

#Pilot	KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
	Level of usage (from employees or system integrators)	**B**	Number of users who use the system in relation to the total number of potential users of an insurer.	X		X		10%	20%	60%	NO
14	**% of automation workflow achieved on AgI company level** (quantitative criterium)	**O**	Pilot insurance companies are asked to evaluate % of policy processing tasks in underwriting and claims handling that can be conducted automatically via the tool	X				n.d.	n.d.	n.d.	NO
	AgI Process cycle-time decrease (%) (quantitative criterium)	**O**	Pilot insurance companies are asked to evaluate Time saving (duration per policy) in underwriting and claims handling by using the tool compared to before	X				n.d.	n.d.	n.d.	NO
	Number of calamities covered (drought, fire, ice, flood) (quantitative criterium)	**O**	Number of weather risks covered by the available datasets and analyses of the tool	X				n.d.	n.d.	n.d.	NO
	Increase accuracy of Risk Mapping and premium calculation for the Insurance Company (qualitative criterium)	**B**	Pilot insurance companies are asked to evaluate (on a scale between 1-10 from very bad to very good) their risk mapping and premium calculation capacities (risk data availability, spatial and temporal resolution) before and during using the AgI tool. In addition, G&Co as a consultant will evaluate (on a scale between 1-10 from very bad to very good) the usefulness for product development.	X				n.d.	n.d.	n.d.	NO

#Pilot	KPI Denomination	Type	Measurement Mode	ID	M	NR	CM	Initial KPI measurement	Current KPI measurement	Target Level	Achieved
	Improved alerting (Early Warnings/ Weather Risk probability) send to Ins. companies and damage mitigation measures deployment (quantitative criterium)	**B**	(1) Pilot insurance companies are asked to evaluate % of weather alerts in comparison to experienced insurance claims ; (2) % of alerts compared to actual occurrence of the forecasted events (3) Pilot insurance companies are asked to evaluate Number of alerts forwarded to the insurer's clients for mitigation actions	X				n.d.	n.d.	n.d.	(1-3) NO
	KPI8: Improved damage assessment (quantitative criterium)	**O**	% of damage predictions after event within -20%/+20% range of loss adjuster estimations	X				n.d.	n.d.	n.d.	NO
	O&A Cost reduction (savings as a result of the automated process incl. Damage Assessment) (quantitative criterium)	**B**	Pilot insurance companies are asked to evaluate % of savings for loss adjustment costs per event	X				n.d.	n.d.	n.d.	NO
	Fraud attempts or false claims mitigation number of false claims detected using EO data vs false claims detected through on-site assessment (quantitative criterium)	**B**	% of clients with insurance claim but no event detection.	X				n.d.	n.d.	n.d.	NO
15	**Semantic Accuracy**	**O**	F1 score, monitored through technical measurement process	X	X		X	9,09%	74,04%	Above 75%	NO
	% Automation of workflow	**O**	Interviews with end users	X				n.d.	n.d.	n.d.	NO
	Usability	**B**	Interviews with end users	X				n.d.	n.d.	n.d.	NO
	User Satisfaction	**B**	Interviews with end users	X				n.d.	n.d.	n.d.	NO

Table 18 - Overall Pilot-Specific KPIs status

The following are the Pilots' Common KPIs measurements:

Pilot#	Common KPIs (denomination)	Measurement	Target Level	Achieved
1	Core Task Efficiency	n.d.	n.d.	NO
	User Satisfaction	n.d.	n.d.	NO
2	Core Task Efficiency	4	8	NO
	User Satisfaction	n.d.	> 4	NO
3	Core Task Efficiency	Not Yet	2+ Tools/Framework Available; 1 Procedure Available; 50 End user / Consumers	NO
	User Satisfaction	Not Yet	100 Consumers / Stakeholders	NO
4	Core Task Efficiency	Currently, less than 1-3 minutes	10-20 seconds	NO
	User Satisfaction	Further and ongoing measurement of the portfolio return score. First user prototype of the UI Journey (as described) is created.	Most accurate version of the defined measurement. Achieved.	YES
5b	Core Task Efficiency	50%	100%	NO
	User Satisfaction	n.d.	n.d.	NO
6	Core Task Efficiency	n.d.	n.d.	NO
	User Satisfaction	n.d.	n.d.	NO
7	Core Task Efficiency	0%	100%	NO
	User Satisfaction	n.d.	4.5 (out of 5 on user satisfaction feedback)	NO
8	Core Task Efficiency	n.d.	n.d.	NO
	User Satisfaction	n.d.	n.d.	NO
9	Core Task Efficiency	60%	n.d.	NO
	User Satisfaction	n.d.	n.d.	NO
10	Core Task Efficiency	n.d.	$\Delta > 0$	NO
	User Satisfaction	n.d.	$M_1 >= 4$ M_1 ranging from 0 to 5	NO
11	Core Task Efficiency	n.d.	> 70%	NO
	User Satisfaction	n.d.	> 70%	NO
12	Core Task Efficiency	44.6%	75%	NO
	User Satisfaction	n.d.	60%	NO
13	Core Task Efficiency	50%	72,5%	NO
	User Satisfaction	n.d.	n.d.	NO
14	Core Task Efficiency	n.d.	n.d.	NO
	User Satisfaction	n.d.	n.d.	NO
15	Core Task Efficiency	74,04%	> 75%	NO
	User Satisfaction	n.d.	n.d.	NO

Table 19 - Overall Common KPIs status

4 Conclusions

The deliverable is the first one of a sequence made up of three deliverables meant to progressively evaluate the pilots. Notably, these are: Pilots' Evaluation and Stakeholders' Feedback I (D7.20 - current deliverable); Pilots' Evaluation and Stakeholders' Feedback II (D7.21 – month 30 of the project) within which it will be taken into consideration the continuous pilots' monitoring with respect to their own KPIs advancements and the overall picture of the stakeholders' feedback; Pilots' Evaluation and Stakeholders' Feedback III (D7.22 – month 39 of the project) in which the final evaluation, comprised of the techno-economic assessment as well as diverse operational/business aspects, will be reported.

As part of this deliverable, we have illustrated the various steps involved in the devised Evaluation Framework, which currently lead to obtain an overall picture of the Pilots KPI status. The document also presents a description of the monitoring process that will be carried out to conduct future Pilots' assessments.

With respect to the Task 7.8's future actions, that will be reflected in the next deliverables, ABI Lab foresee to start with the continuous monitoring process (namely, the Phase 2 of the presented Evaluation Framework) as to achieve three main goals. First, to monitor the evolution of the pilots' KPIs measurements. This will enable ABI Lab to define analytical means which aims at evaluating pilot's use cases, as well as finding potential critical points needed to be taken care of. The second goal would be to gather information about the tools employed by the developers, as well as business cases, roughly biannually. This is meant to understand further potential setbacks and adjustments, as well as to figure out whether there is any development with regards to their perspectives for their use in a real-world scenario. Finally, even the process of gathering stakeholders' feedbacks will be addressed as to achieve a multi-faced and high-level vision of the pilots' progress.

5 References

[1] Supervision, B.: Basel committee on banking supervision. Principles for Sound Liquidity Risk Management and Supervision (September 2008) (2011)

[2] Christoersen, P.F.: Evaluating interval forecasts. International economic review pp. 841{862 (1998)

[3] Wedge R., Kanter J. M., Veeramachaneni K., Rubio S. M., and Perez S. I., (2018). Solving the false positives problem in fraud prediction using automated feature engineering. In Proceedings of the ECML-PKDD 2018 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, *http://www.ecmlpkdd2018.org/wp-content/uploads/2018/09/567.pdf* (http://www.ecmlpkdd2018.org/wp-content/uploads/2018/09/567.pdf)