


Infinitech

D7.2 – Report on Pilot Sites Preparation – II

Title	D7.2 – Report on Pilot Sites Preparation - II
Revision Number	3.0
Task reference	T7.1
Lead Beneficiary	ATOS
Responsible	Jose Gato Luis/Ignacio Elicegui Maestro
Partners	ATOS, GFT, IBM, FTS, HPE, SILO, ENG, ISPRINT, SIA, LIB, NBG, AKTIF, BOS, PI, BPI, DYN, GEN, JRC, PRIVE, CP, WEA, RB, LXS, UNP, RRD, FBK, NUIG, NOVA, BOUN, JSI, ORT, CTAG, GRAD, ABILAB, BOI, AGRO, BANKIA, BOC, UPRC
Deliverable Type	Report
Dissemination Level	PU
Due Date	2021-09-30 [M24]
Delivered Date	2021-10-29 [M25]
Internal Reviewers	LXS CCA
Quality Assurance	CCA
Acceptance	Coordinator Accepted
EC Project Officer	Beatrice Plazzotta
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 Acronym	Role¹	Author(s)²
ATOS	Lead Beneficiary	Jose Gato Luis, Ignacio Elicegui Maestro
FTS	WP2 Leader	Juergen Neises
LXS	WP3 Leader	Pavlos Kranas
NUIG	WP4 Leader	Martin Serrano
UBI	WP5 Leader	Dimitrios Miltiadou
HPE	WP6 Leader	Alessandro Mamelli
UPRC	WP8 Leader	Dimosthenis Kyriazis
GFT	WP9 Partner	Vittorio Monferrino
GFT	Cluster Leader	Vittorio Monferrino, Maurizio Megliola
CrowdPolicy	Cluster Leader	Marinos Xynarianos, Angeliki Kitsiou
WeAnalyze	Cluster Leader	Carlos Albo (WeAnalyze)
Cluster 1	Pilot Leaders and Technical Proxies	Barbara Cacciamani, Marco Rotoloni (ABILAB) Javier Sancher (Bankia), Antoni Munar (GFT) Petra Ristau (JRC), John Soldatos (Innov), Nikos Kapsoulis (INNOV), Despoina Kyriazis (INNOV)
Cluster 2	Pilot Leaders and Technical Proxies	Richard Walsh (NUIG), Martin Serrano (INSIGHT) Pablo Carballo (PRIVE) John Kaldis (RB) Silvio Walser (BOC) Dimitris Kotios (UPRC) Manolis Syllignakis, Nikos Droukas, George Kanellis (NBG)
Cluster 3	Pilot Leaders and Technical Proxies	Maja Skrjanc (JSI) Sabina Podkriznik (BOS)Can Ozturan (BOUN)Susanna Bonura (ENG) Massimiliano Aschi (PI)
Cluster 4	Pilot Leaders and Technical Proxies	Ignacio Elicegui Maestro (ATOS) Andreas Politis (DYN) Aristodemos Pnevmatikakis (ISPRINT)
Cluster 5	Pilot Leaders and Technical Proxies	Carlos Albo (WeAnalyze) Gregory Mygdakos (AGRO) Lukas Linden (GEN)

¹ Lead Beneficiary, Contributor, Internal Reviewer, Quality Assurance

² Can be left void

Revision History

Version	Date	Partner(s)	Description
0.1	2021-09-10	ATOS	ToC Version
0.2	2021-09-28	ATOS, WP leaders	WP section structure
0.3	2021-10-06	ATOS, Cluster leaders	Pilots' contributios
0.4	2021-10-11	ATOS, Cluster leaders	Pilots' contributions
0.5	2021-10-18	ATOS, Cluster leaders, WP leaders	Pilot's contributions; WPs' contributions
1.0	2021-10-21	ATOS, WP leaders	WPs' contributions and Conclusions
2.0	2021-10-28	ATOS, CCA, LXS	Internal reviews, Quality assurance
3.0	2021-10-29	ATOS	Final (ready to deliver)

Executive Summary

This deliverable aims to update the first version of the report on pilots, D7.1, which focused on the following aspects of pilots: readiness; development; and validation of different services and components. From that first version, a new set of deliverables has been released, adding more information and progress on each pilot.

For this second report, the readiness will be updated, particularly regarding:

- Relationships of this WP7 with the rest of the project tasks, in terms of the different deliverables released, and how these are progressing, including news and readiness information from pilots.
- A first overview and actions taken (or to be taken) by pilots about the main suggestions coming from the last review.
- A first look on the new reference business model and competitiveness coming from WP9.
- Summary of the general progresses accomplished by each pilot since the last report (understanding by last report, the set of clusters' deliverables released on M18: D7.3, D7.6, D7.9, D7.12, D7.15)

In parallel with this deliverable, the D7.18 Pilot Sites Synergies and Collaboration – I is working on analysing the different services from pilots that potentially could be horizontally exploited by other pilots. Also, the KPIs monitoring task team is working internally to collect the information from each pilot, within an iterative process. This information will be valuable in documenting the creation and status of the synergies.

Finally, by the end of this year (M27), a deeper view of all the use cases and pilots, with a more business-oriented approach will be reflected in the second round of pilots reporting: D7.4, D7.7, D7.10, D7.13 and D7.16.

Table of Contents

1	Introduction.....	10
1.1	Objective of the Deliverable.....	10
1.2	Insights from other Tasks and Deliverables.....	10
1.3	Structure of the Document.....	10
2	Holistic pilots’ readiness.....	11
2.1	Readiness about specifications and requirements (WP2).....	11
2.2	Readiness regarding data management and governance (WP3).....	12
2.3	Readiness regarding semantic interoperability and blockchain (WP4).....	14
2.4	Readiness regarding data analytics tools and AI (WP5).....	16
2.5	Readiness regarding testbeds and sandboxes implementation (WP6).....	17
2.6	Synergies and pilots’ monitoring (WP7 tasks).....	18
2.7	Readiness regarding the INFINITECH Marketplace (WP8).....	19
2.8	Common reference model for business services and competitiveness (from WP9).....	20
3	Quick update from pilots.....	21
3.1	Pilot #1 Invoices Processing Platform for a more Sustainable Banking Industry.....	21
3.2	Pilot #2 Real-time risk assessment in Investment Banking.....	21
3.3	Pilot #3 Collaborative Customer-centric Data Analytics for Financial Services.....	22
3.4	Pilot #4 Personalised Portfolio Management.....	23
3.5	Pilot #5b Business Financial Management (BFM) tools delivering Smart Business Advice.....	24
3.6	Pilot #6 Personalized Closed-Loop Investment Portfolio Management for Retail Customers.....	25
3.7	Pilot #7 Operation Whitetail - Avoiding Financial Crime.....	25
3.8	Pilot #8 Platform for AML supervision.....	27
3.9	Pilot #9 Analysing Blockchain Transaction Graphs for Fraudulent Activities.....	28
3.10	Pilot #10 Real-time cybersecurity analytics on financial transactions’ data.....	29
3.11	Pilot #11. Personalized insurance products based on IoT connected vehicles.....	31
3.12	Pilot #12 Real world data for novel health insurance products.....	32
3.13	Pilot #13 Alternative and automated insurance risk selection and insurance product recommendation for SME’s.....	34
3.14	Pilot #14 Big Data and IoT for the Agricultural Insurance Industry.....	35
3.15	Pilot #15 Inter-Banking Open Pilot.....	37
3.16	Pilot#16: Data Analytics Platform to detect payments anomalies linked to money laundering events.....	38
4	Conclusions.....	40

List of Figures

Figure 1: WP7 cross-relations with INFINITECH WPs	11
Figure 2: Pilot #8 summarised roadmap	28
Figure 3: Pilot #12 summarised roadmap	34

List of Tables

Table 1: Pilot#13 meetings and events with stakeholders.....	35
--	----

Abbreviations

Abbreviation	Definition
4AML	Fourth Money Laundering
AgI	Agricultural Insurance
AI	Artificial Intelligence
AIGO	AI Based Portfolio Optimization Process
AML	Anti-Money Laundering
API	Application Programming Interface
AWS	Amazon Web Services
AWS EC2	Amazon Web Services Elastic Compute Cloud
B2B	Business to Business
BD	Database
BDA	Big Data Analytics
BDVA	Big Data Value Association
BFM	Business Financial Management
BIAS	Bias
BOC	Business Operations and aCtions
CAN	Controller Area network
CFT	Countering Terrorist Financing
CXB	CaixaBank
DNS	Domain Name Service
DUOS	Digital User Onboarding Service
DPO	Data Protection orchestrator
DWH	Data Warehouse
EO	Earth Observation
ERC20	Ethereum Requests for Comment
ETL	Extract, Transform and Load
FI	Financial Institutions
FIBO	Financial Business Ontology
FIGI	Financial Instrument Global Identifier
FX	FOREX
GA	General Assembly
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HDFS	Hadoop Distributed File System
HPC	High Performance Computing

D7.2 - Report on Pilot Sites Preparation - II

ID	Identity
IDM	Identity Management
IoT	Internet of Things
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
KYC	Know Your Customer
LKIF	Legal Knowledge Interchange Format
MB	Message Broker
MC	Monte Carlo
MiFiD	Markets in Financial Instruments Directive
ML/DL	Machine Learning/Deep Learning
ML/FT	Money Laundering/Financing Terrorism
MPI	Message Passing Interface
MS	Microsoft
NLP	Natural Language Processing
NN	Neural Network
PAMLS	Platform for anti-money laundering Supervision
PAYD	Pay As You Drive
PCTU	Top-ups phone credit
PHYD	Pay How You Drive
PoC	Proof of Concept
PSD	Payments Service Directive
RA	Reference Architecture
ReM	Resource Management
REST	Representational State Transfer
RoM	Role Management
RWD	Real World Data
SaaS	Service as a Service
SEPA	Single Euro Payments Area
SHAP	SHapley Additive exPlanations
SHARP	Smart, Holistic, Autonomy, Personalized and Regulatory Compliant
SME	Small Medium Enterprise
SMWCA	Sending Money to people Who do not have a Current Account
SOTA	State of The Art
TRY	Turkish Lira
UAT	User Acceptance Testing
URL	Uniform Resource Locator

V2I	Vehicle to Infrastructure
VaR	Value at Risk
VDIH	Virtualized Digital Innovation Hub
WP	Work Package
XAI	Explainable AI

1 Introduction

1.1 Objective of the Deliverable

This document presents a holistic update of each pilot's readiness and progresses since the last general report in M14, covering the last 11 months. As a 'version II' deliverable, this one will not repeat some of the points that were presented in D7.1, which confirms the basis and starting point of the pilots. These are:

- General overview and status of the pilot.
- Different components and services (technical development status)
- Description and objectives for a first PoC ready by the release of the current deliverable
- Testbed/sandbox availability

During the last 10/11 months, since version I of this deliverable, pilots have evolved to address many different aspects. Many of them have been already covered in the cluster's reporting deliverables: D7.3, D7.6, D7.9, D7.12, D7.15 (M18). Therefore, this new version updates readiness from a general approach to an aspect-by-aspect approach, by enumerating the relevant outcomes and achievements and by indexing the different deliverables and contents related to each pilot's readiness in the last 10/11 months.

The deliverable aims to aid an understanding of the overall progresses of pilots and WP7, serving as an index to go deeper into different pilots' achievements. During this period, tasks T7.7 and T7.8 have started work on synergies and KPIs. Therefore, a preamble of this work is included, presenting the next incoming deliverables and outcomes from these tasks. Finally, the deliverable will make a first introduction to actions related to the main suggestions coming from the Review Report.

This text is intended to be short and straightforward, to help the reader to find the information that he/she is looking for, avoiding unnecessary repetition. The deliverable does not try to paraphrase D7.1 to re-summarize the main objectives of each pilot.

1.2 Insights from other Tasks and Deliverables

WP7 is a horizontal work package, collaborating with the different INFINITECH objectives, tasks, and work packages. For that reason, one of the main outcomes of this deliverable is a cross-index of all these contributions and the rest of the INFINITECH activities, showing, in this way, the existing correlation between the project's packages and its pilots. Chapter 2 will include all these collaborations, structured by WP.

1.3 Structure of the Document

As per the Introduction, the objective of this document is to introduce a) the progress of each of the WP7 internal tasks plus the interactions with the other WPs; and b) the different achievements of its 16 pilots, during the last period of 10/11 months. Within this framework, the document is divided as follows.

Chapter 2: this will include a holistic view of the readiness of each pilot, pointing to the deliverables and collaborations happening in the last 10/11 months. In addition, it presents a quick overview of the WP9 Business and Competitiveness reference model. Also, some brief comments and actions relating to the review report are mentioned.

Chapter 3: each pilot's main areas of progress are summarized, focusing more on the business side, and not so much on the development status (which will be included in the next pilot's reporting).

Chapter 4: Final conclusions.

2 Holistic pilots' readiness

This chapter aims to provide a holistic update of readiness from pilots. WP7, as a horizontal WP, is collaborating in and supporting most of the different tasks and deliverables of the project. WP7 managed pilots, their KPIs and potential synergies (including how they consume outputs from, and contribute inputs to, the different INFINITECH work packages), are as shown in Figure 1.

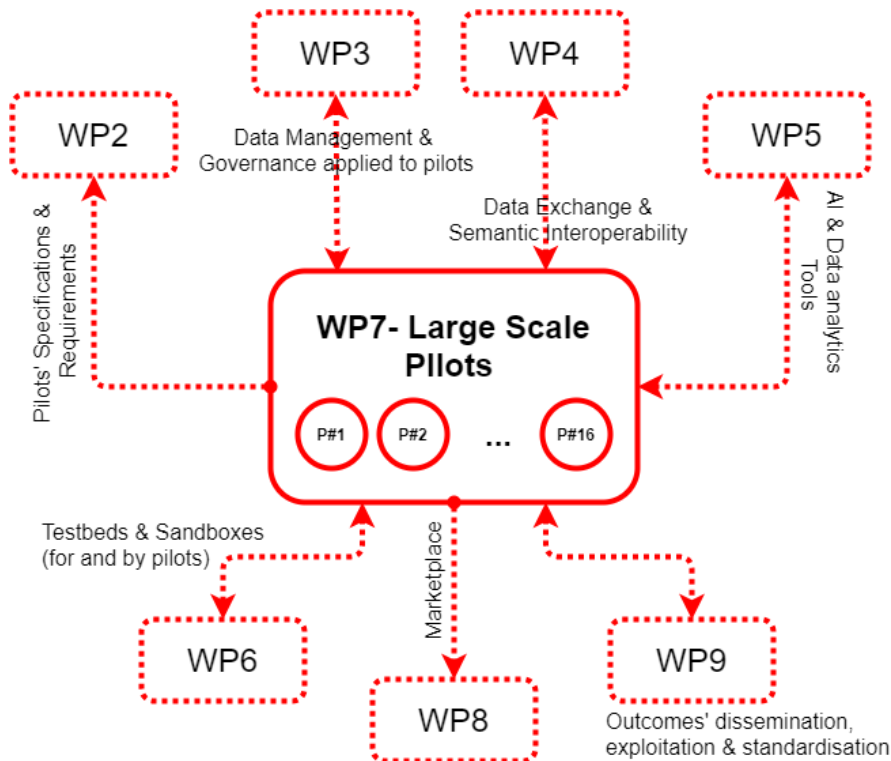


Figure 1: WP7 cross-relations with INFINITECH WPs

The chapter serves as an index of multiple deliverables, assisting the reader to go to a deep dive into each pilot's readiness in different aspects. In that way, the deliverable can be concise, and at the same time, precise and broad, helping readers to have a better understanding of each pilot.

2.1 Readiness about specifications and requirements (WP2)

WP2 provided the foundation for the subsequent work packages in the project. Starting from business, services, technology, technical and operational measures, data and APIs as well as the core ontology of INFINITECH aligning the building blocks, the pilots' objectives and the INFINITECH way to tackle them was elicited and specified in close collaboration with the pilots.

Finally, all the insights obtained in these tasks were bundled and resulted in the INFINITECH Reference Architecture providing a holistic view on the INFINITECH building blocks and an overview of the pilot's workflows linked to the BDVA reference architecture.

Based on the business objectives mapped to technology, the INFINITECH Reference Architecture provides a comprehensive set of building blocks, which can be utilized in the technical work packages and by the pilots themselves. Pilots' incomes and outcomes related to this second report can be found in the following sets of deliverables (second versions of specifications, requirements, architecture, APIs and technologies):

- **D2.2 & D2.18 User Stories and Stakeholders' Requirements – II:** In task T2.1 the user stories from each pilot were developed and refined. This served to document Requirements, User Stories, Pains, Gains, Jobs to be Done; provided summaries for the git repository; and achieved a clear focus on business objectives. In these deliverables, the refined and final version of that documentation can be found.
- **D2.4 & D2.19 Reference Scenarios and Use Cases – Version II:** This task T2.2 focused on mapping the user stories to business services and underlying technical building blocks. Again, the business view was addressed. In addition, the technical services providing the data analytics to the business were identified and clustered along the BDVA reference architecture. The latest version of the pilots' user business drivers for application of BigData, IoT and AI is found in these deliverables.
- **D2.6 Specifications of INFINITECH Technologies – II:** Detailed specifications of 40 component characteristics, including I/O, functionalities and specifications about the implementation technologies are provided in this deliverable. This includes detailed specifications of 60 components identified by the pilots, including information about their Marketplace, containerized components, along with the specification, if any, of their endpoints or REST API URLs.
- **D2.8 & D2.21 Security and Regulatory Compliance Specifications – Version II:** Task T2.4 analysed the relevance of GDPR, PSD II, 4AML and MiFiD II plus ISO27K series for the INFINITECH pilots. In these deliverables the technological and organizational impact of relevant standards and regulations that shall be considered in the INFINITECH testbeds and sandboxes was defined. Technical and organisational measures were specified per pilot. Moreover, a general guideline was developed based on the current state of discussions on AI & Compliance applied to the pilots. Recommendations per pilot are found in these deliverables.
- **D2.10 & D2.23 Initial Specification of Testbeds, Data Assets and APIs – II:** These deliverables outline the latest version of each pilot's testbed specification. This includes detailed description of the relevant data assets and APIs.
- **D2.12 Data Models Specification – II:** In task T2.6 the State of the Art of ontologies available in the Financial Sector was elicited enabling the creation of an INFINITECH core Ontology. This ontology gives an overview over the pilots' terms in data sets that include the pilots joining with the amendment. This way a common vocabulary is ensured across the pilots within INFINITECH.
- **D2.14 INFINITECH Reference Architecture – II:** This is the intermediate version of the INFINITECH Reference Architecture. In addition to the reference architecture itself and the INFINITECH building blocks, this deliverable describes how each pilot implements the Reference Architecture.

2.2 Readiness regarding data management and governance (WP3)

WP3 is related to both the data management and data governance activities of the INFINITECH, and as such, it can be considered as the union of these two major blocks. It is important to highlight that the work that has been carried out in the scope of WP3 is not targeting specific pilot needs, rather it is pilot agnostic.

Work that concerns the data management activities can be further divided into work that relates to the processing of static data, and work that relates to streaming processing. No pilot is contributing to any of the data management activities and therefore, the approach was to provide the fundamental pillars that consist of the data management platform, without focusing on specific pilots, but to implement innovative features that can break through the current technological barriers. These barriers are general enough and can be found in different domains, but they are also crucial for the finance and insurance sectors. The outcomes of the work that has been carried out regarding the data management activities are documented in the following list of deliverables:

- **D3.2 Hybrid Transactional/Analytics Processing for Finance and Insurance Applications – II:** This report describes the INFINISTORE, a novel relational datastore capable of providing both operational and analytical processing of the same data and being able to scale out effectively. This solves the needs from pilots for high rates of data ingestion while ensuring data consistency in terms of database transactions. Moreover, it prevents the pilots from using ETLs for migrating data from the operational database to a data warehouse, which would result in performing data analytics over outdated and obsolete datasets. The outcomes described in this report allow for real-time business intelligence. At the moment, pilots 2, 5b, 6, 12 and 13 are using this component, with the plan for the majority of pilots to be engaged at the forthcoming period of the project.
- **D3.5 Integrated (Polyglot) Persistence – II:** This report describes the polyglot capabilities that can be considered as an extension of the INFINISTORE. The polyglot query processing enables a seamless way of accessing and combining datasets that are stored in different underlying datastores, using a unique connection mechanism and query language. At this moment, this is being used by pilot 13.
- **D3.7 Data Streaming and Data at Rest Queries Integration – II:** The work that is being reported in this report is related to work on the streaming engine of INFINITECH and how it enables the combination of *static* data with data *at-rest*, implementing the INFINISTORE data connectors with the core streaming engine. Having that, it allows the streaming operators to take advantage of the unique characteristics and innovations developed and incorporated into the INFINISTORE. This removes the technological barriers of current (pre-INFINITECH) approaches that trade data consistency for efficiency. It is currently being experimented with under the scope of pilot 2 and 8, with the plan for all pilots that require streaming processing to be engaged in the next period of the project.
- **D3.10 Automatic Parallelization of Data Streams and Intelligent Pipelining – II:** This report focuses on the automation of the parallelization of streaming operators on one hand, and the intelligent data pipelines on the other. It solves the need for streaming processing deployments to be adaptive in diverse workloads that can be found in the finance sector, without downtimes, while it also provides the needs for transparent data migration between different sources. The automatic scalability of the streaming operators is being validated via pilots 2 and 6, while the intelligent data pipelines via its integration with the semantic interoperability engine, that is being validated with pilot 2 and 3.

The activities regarding the data governance can be further divided into two groups: a group related to the data mechanisms themselves, and a group related to the regulatory compliance tools. Some of these activities are closer to the pilots and therefore, in contrast to the data management layer, the pilots were closely involved in the implementation of the outcomes. The work that has been carried out is documented in the following list of deliverables:

- **D3.13 Data Governance Framework and Tools – II:** This document reports the different approaches that can be applied and validated using the INFITECH pilots. More precisely it firstly contains information regarding the Digital User Onboarding Service (DUOS) developed by ATOS for user enrolment and authentication currently used by pilot 4. Then it demonstrates the implementation of the data pseudonymization mechanisms developed by JSI to pseudonymize enriched transactional data used in the context of pilot 8. What is more, it reports the use of the data anonymization tool, in particular, the prototype implementation of the location data anonymization mechanisms, to anonymize location data reported by connected cars in real time for the needs of pilot 11. Finally, it reports the use the data anonymization tool to anonymize data collected in the Healthentia platform of pilot 12.
- **D3.16 Regulatory Compliance Tools – II:** This document provides preliminary definition of the solutions for regulatory compliance for i) pilot 4, which will implement a possible enhancement for customer authentication by adopting DUOS solution (Digital User Onboarding System - a solution for dealing with

virtual identities in a mobile device) on explicit B2B customer request; ii) pilot 6, which will use anonymization in order to comply with regulations; iii) pilot 7, which provides solutions for suspicious behaviour in Financial Crime Fraud with the final objective of stealing the bank customers' identity and money using next generation technical solutions like Machine Learning, together with advanced modelling; iv) pilot 8, which develops a Platform for anti-money laundering Supervision (PAMLS) using pseudo-anonymization prior to data delivery to PAMLS in order to comply with applicable data protection rules; v) pilot 11, for OAuth 2.0 based authorization mechanisms to access the platform and the use of "ask for consent" mechanisms for the data treatment as well as pseudo-anonymization and anonymization; vi) pilot 12, for a user's signed consent for data collection and processing, applying an access control framework, making use of the general regulatory compliance tool based on DPO (Data Protection orchestrator) that orchestrates the calls to different security or privacy tools, interacting with Data Collector and Anonymization component; vii) pilot 13, for standard access security measures delivered by AWS; and viii) pilot 14, for a client's signed consent to handle GDPR regulation and personal data.

2.3 Readiness regarding semantic interoperability and blockchain (WP4)

INFINITECH WP4 (Interoperable Data Exchange and Semantic Interoperability) is focused on developing the tools for data interoperability and trustful data exchange in the project. WP4 has six tasks providing design, implementation, and continuity to relevant project activities. WP4 has worked towards delivering software assets for reducing data fragmentation via tools and demonstrators that addressed community requirements and the identified need to address lack of semantic interoperability tools and methods across diverse financial domains (banking and insurance). WP4 refers to the same data entities, with similar properties but with different semantics, providing in this way the flexibility to re-use the data but in different contexts. This is a best practice to address the limitations in data models for sharing datasets across different stakeholders and for enabling connected applications that span multiple systems across financial domains.

INFINITECH WP4 has organized activities in two clusters: cluster one focuses on technologies for secure data sharing over blockchain infrastructures; cluster two focuses on the use of semantic technologies and more particularly standard financial data models for trusted data exchange and interoperability purposes. The two clusters follow a planned sprint progress in periods of 3 months per sprint, where the implementation of demonstrators and prototypes were also synchronized with delivery of the software assets. In this part of WP4, two solutions for interoperability and data exchange were developed; one is provided as a centralized online graph data model solution that is based on the development of an interoperability (ontology based) database/registry supporting linking of diverse systems and datasets based on shared semantics, as well as semantically-interoperable analytics; a second one is provided as a decentralized solution based on the deployment of hyper-ledger fabric (an open source distributed ledger technology adopted for the project) and the use of permissioned blockchain concepts for decentralized trust. Both solutions become part of the INFINITECH solutions in order to facilitate financial/insurance processes that involve cross-organization data exchange in the defined pilots. All WP4's tasks are active and are continuously providing results and outcomes in the form of deliverables and/or demonstrators as follows:

All achieved results of WP4 regarding requirements definition have been used to advance the design and implementation of data exchange and interoperability data services; other continuing activities are part of the continuity of WP4's defined objectives.

- **D4.2 Semantic Models and Ontologies – II:** The purpose of the deliverable D4.2 – Semantic Models and Ontologies – II is to describe the overall approach used for specifying the models and ontologies for semantic interoperability of diverse applications in the finance and insurance sectors. The document provides the necessary theoretical foundation for the design and implementation of the INFINITECH Interoperability Framework. Furthermore, a methodology for semantic models and ontologies engineering is also presented that defines the overall strategy used to design and specify semantic models. In particular, the semantic models are organized hierarchically according to the domain and the specific application and linked to reference ontologies such as FIBO/FIGI, LKIF, FinReg, etc. The provided models aim to establish the cornerstone for semantic interoperability within INFINITECH while enabling the annotation and linking of diverse data streams.
- **D4.5 Semantic Stream Analytics Engine – II:** This deliverable introduces the INFINITECH Semantics Streams Analytics Engine (SeSA-ME) and the related tools for enabling semantic data exchange, which is based on the development of an interoperability (ontology-based) database/registry supporting linking of diverse systems and datasets based on shared semantics, as well as semantically interoperable analytics. The SeSA-ME system includes tools along with a visual SPARQL query editor providing Swagger APIs for verification and visualization tools for novice users while supporting full access and control over the data mashups for expert users. Tied with the development of the SeSA-ME platform is the development and deployment of the INFINITECH Graph Data Model which enables the support for both the design and deployment of stream-based web applications in a very simple and intuitive way and the analytics services using stream-based applications and services.
- **D4.16 Visualization Front-End for Aggregated Information – II:** This software deliverable reports the current status of the visualization front-end tools and demonstrators that were designed to illustrate all the main functionalities of the visualization front-end for aggregated information tool(s). The main technical requirements and the design architecture of the first version of the tool to be tested and validated within Pilot #10 are presented. Then the adopted design process for the user interaction that is tailored to the needs of the task is elaborated. Following this design process, a comprehensive description of the high-fidelity mock-ups is provided and encapsulates the design choices in terms of color schemes, layouts, typography, iconography, spacing and navigation visuals, as well as the overall atmosphere of the tool. Moreover, for each mock-up presented, the expected user interaction with the corresponding mock-up is defined.

Regarding the design aspects, various best practices, the activities performed, and demonstrators developed for the INFINITECH pilots with regards to the Blockchain technologies and tokenization processes, the following WP4 deliverables containing the various activities are documented as follows:

- **D4.8 Permissioned Blockchain for Finance and Insurance – II:** This deliverable presents the first version of the INFINITECH blockchain network specifications and the design specifications of the blockchain applications. The outcomes of this deliverable provide the design specifications of both the blockchain network and the blockchain application. Deliverable D4.8 constitutes a living document, and it will be continuously updated based on the analysis of the feedback that will be collected from the pilots of the project and the stakeholders of the platform. The purpose of this deliverable is a) to document the analysis of the blockchain technology and define its role within the INFINITECH RA; b) to document the design of a new blockchain capability that can be leveraged within the context of the project; c) to deliver the design specifications of the blockchain applications which are tailored to the needs of the financial and insurance sector as well as the specifications of the INFINITECH blockchain network that will be utilized; and finally d) to document the baseline technologies that will be utilized in pilots.

- **D4.11 Blockchain Tokenization Functionalities and Smart Contracts – II:** This deliverable summarizes the work realized in relation to the “Tokenization and Smart Contracts Finance and Insurance Services”. The aim of this deliverable is “to enhance the permissioned blockchain of the project with tokenization functionalities, as a means of enabling digital assets trading” (from INFINITECH DoA). The starting point for this deliverable is the ERC 20 standard for tokens, chosen for its popularity and familiarity among the partners in the project and the fact that it is based on the account model approach for tokens. This deliverable provides the details about the implementation of ERC 20 functions (with the addition of one function used for the minting of new tokens) as smart contracts in Fabric. Furthermore, we provide a description of future steps that will be carried out in collaboration with other tasks in the project to leverage this initial work. This deliverable also constitutes the first version of the blockchain tokenization and smart contracts template for the use of pilots. Subsequent versions will include elaboration of the foreseen directions to pursue.
- **D4.14 Encrypted Data Querying Algorithms and Personal Data Market – II:** This deliverable includes the current state of the art; highlights shortcomings of the current approaches; surveys interesting existing models; and especially articulates the details of a newly proposed framework and its components. The deliverable starts with an account of the current rise in the role of data, in particular personal data, and the importance of its secure storage, sharing, and manipulation, and with the analysis of two models that address these issues and are part of the design process for INFINITECH encrypted data querying algorithms and the personal data market. In particular, we argue that the current siloed approaches are responsible for the lack of innovation in some sectors, but also that any new paradigm needs to take account of the risks in terms of possible abuses (i.e. privacy violation, discrimination, etc.) in the usage of personal data.

The INFINITECH WP4 reported and presented the progress and results in the Period I project review. It was relevant to consider the use of the interoperability and blockchain tools in the pilots that are not yet prepared for data sharing and data exchange processes. The development of the software assets in WP4 indicated a high level of maturity at the asset level, and additional efforts are underway to integrate these assets in the INFINITECH marketplace, which together with software data assets and information services can be included. A new focus for development towards a more business-oriented approach was also discussed, post project review. More information about the progress on these topics can be found in the corresponding documents but is also accessible via the WP4 INFINITECH documents repository.

2.4 Readiness regarding data analytics tools and AI (WP5)

WP5 is related to the machine learning and deep learning activities of INFINITECH, with the design, implementation and delivery of the tools and technologies that can be leveraged by the finance and insurance sectors during these operations. The list of outcomes spans from data collection, data harmonization and data preparation, to the parallelization of incremental analytics algorithms, the implementation of a declarative and configurable analytics framework, and to the compilation of the INFINITECH library of ML/DL algorithms for applications to the finance/insurance sectors and the discovery and utilization of these algorithms through Open APIs via an API Gateway.

Within the context of WP5, there can be found activities regarding the advanced analytical capabilities of the data management. They are generally applicable and not pilot-specific. These activities concern the work that has been carried out for the incremental and real-time declarative analytics. On the other hand, there are activities related to the preparation and execution of the ML/DL processes that rely on the data management layer. For the first category of activities related with the advanced analytical capabilities, the following list of deliverables document the work that has been carried out:

- **D5.2 Library of Parallelized Incremental Analytics – II:** This report documents the design and implementation of the extensions of the INFINISTORE regarding the provision of incremental scans and analytics. The work that has been carried out is currently being validated in experiments with different architectural design approaches for scenarios related to pilot 2 and pilot 6. These scenarios require the combination of streaming and analytical processing for real-time identification of events of interest.
- **D5.5 Framework for Declarative and Configurable Analytics – II:** This report documents the extension of the INFINISTORE regarding the provision of Online Aggregates. This allows for real-time analytical processing of live data stored in the operational datastore. It is currently being used by pilot 2 and 5b.

Moreover, the various aspects of the activities and work performed on the INFINITECH pilots with regard to the preparation and execution of the ML/DL processes are documented in the series of WP5 deliverables as follows:

- **D5.7 Library of ML/DL Algorithms – I** presents the details of the ML/DL related tasks as planned, established and currently under development by most of the pilots. It provides an overview of the guiding machine learning frameworks that each pilot is planning to test and develop.
- **D5.8 Library of ML/DL Algorithms – II** describes the progress of the ML/DL related tasks of each INFINITECH pilot, detailing the state-of-the-art and innovative algorithms developed and evaluated, some initial results obtained in the tasks, as well as the planned contribution to the INFINITECH library of ML/DL algorithms in terms of microservices or learning frameworks.
- **D5.13 Datasets for Algorithms Training & Evaluation – I** presents an overview of the data collection process performed in the context of each pilot, documenting the details of the datasets, both real and synthetic, that will be collected by the pilots in order to be harmonized, anonymised and ingested into the underlying storage.

During the review, the need was pointed out to go further into XAI and BIAS in the pilots who are developing services related to AI. This is now a new point to be developed in the pilots, starting from this deliverable. In Chapter 0, those pilots, which are related to AI, will give a first view of these topics, how they are going to be considered, and the next steps. More information about the progress in these topics will be presented in the next pilot's reporting of deliverables. The companies, also, will introduce any different policies they have about AI, Ethics and Trustworthiness.

2.5 Readiness regarding testbeds and sandboxes implementation (WP6)

In INFINITECH, WP6 (Tailored Sandboxes and Testbeds for Experimentation and Validation) is mainly focused on providing tools and techniques for creating tailored sandboxes over the digital finance/insurance testbeds of the project and on providing a mechanism for integrated management of testbeds' datasets based on a continuous integration approach: accordingly, WP6's ultimate goal is to establish all the project's foreseen testbeds (on private data centres/cloud providers' accounts for the incumbent partners' organizations, and on a shared data centre for the FinTech/InsuranceTech partners' organizations) for experimentation and validation, including all relevant sandboxes.

In the last period of the project all these aspects have been described in detail in different WP6 deliverables, where the INFINITECH pilots have provided relevant contributions (with full alignment to the WP7 activities), such as:

- **D6.2 (Testbeds Status and Upgrades – II)** contains a report on the readiness and compliance of the infrastructure (hardware & software) deployments of each pilot, compared to the Reference (Blueprint) Testbed provided and deployed in WP6. In such contexts, the different pilots have provided the as-is

current deployment for each pilot and also the specifications of the planned (to-be) actions or modifications that will be performed to match the guidelines of the Reference (Blueprint) Testbed, as well as reports regarding any required upgrades.

- **D6.5 (Tools and Techniques for Tailored Sandboxes and Management of Datasets – II)** contains a report about the proposed tools and techniques integrated and deployed into the INFINITECH blueprint reference testbed setup, designed and implemented using two of the target INFINITECH infrastructures. In this context, two pilots (namely Pilot #2: Real-time risk assessment in Investment Banking and Pilot #5b: Business Financial Management tools delivering Smart Business Advice) have provided the description of the actual and concrete implementation of the “INFINITECH way” Blueprint guidelines, applied to them, the so-called “INFINITECH blueprint pilots”.
- **D6.8 (Sandboxes in Incumbent Testbeds – II)** contains a report about the development of the hardware and software infrastructure of the incumbent testbeds and the blueprint for the development of sandboxes of incumbent organizations, which results in a number of sandboxes configured according to the Continuous Integration / DevOps approach. In such contexts, the involved pilots have provided the related information for each of them.
- **D6.11 (Sandboxes for FinTech/InsuranceTech Innovators – II)** contains a report about the creation of the NOVA testbed and the construction of the required architectural stack needed to manage the FinTech/InsuranceTech Innovators’ sandboxes, and how it is evolved into an actual set of application-and-orchestration-specific artifacts and how they are consumed in order to get a fully working sandbox for each Pilot. In such contexts, the involved pilots have provided the related information for each of them.

2.6 Synergies and pilots’ monitoring (WP7 tasks)

Task 7.7 Pilots’ Synergies and Collaboration

Task 7.7 aims to identify the synergies and collaboration between pilots by analysing User Stories, Technologies and Components; the roadmap is split into three steps:

- Identify User Stories Pilots’ synergies, analysing the user stories details (e.g., goals, reasons) of the pilots of the same cluster.
- Identify Technologies synergies used by pilots of the same cluster, analysing technologies (e.g., needs, components) used/intended to be used by Pilots.
- Identify the transversal synergies of User Stories and Technologies between Pilots in different clusters.

The current result of this task was the creation of the two different lists, one for User Story synergies between Pilots and the other for the synergies of Technologies used by the Pilots.

The work performed had as its main source the results of task T2.1 (Pilots’ User Stories), and since we had some changes in the partners of INFINITECH consortium (Pilots), there is the need to update the list of the User Stories, with information from the new Pilots.

After the collection of User Stories from new pilots, it will be included in the analysis of synergies and potential collaborations and then the results will be presented to pilots, to receive some feedback and validation of the identified synergies.

Task 7.8 Pilots’ Evaluation and Stakeholders’ Feedback

Task 7.8 reached a point where collecting the stakeholders’ feedback is paramount to performing a multi-faceted evaluation of the pilot system. To this end, we produced a new unified survey which the Pilots will be required

to compile to track their improvements in terms of technical aspects (e.g., KPIs) and, most importantly, the business-side (whose metrics are primarily collected out of a Pilot's hosted workshop – e.g., stakeholders' feedback and other business dimensions).

A meeting with Pilots was organized before the GA to evaluate the approach to collecting such required metrics. The main results were the following:

- Need to find the right balance between the complexity of carrying out the survey and the need to collect additional relevant information
- Ensure that each Task's approach works harmoniously with others, with special regard to the business dimensions (e.g., the ones from the INFINITECH Innovation Pentagon), so as to increase the synergies in place and to enforce the business approach
- Emphasize the importance of a continuous dialogue among the Stakeholders and Partners

Additional meetings will soon follow to achieve an agreement on the aggregation model. Once the updated model is reshaped and validated, it will be brought back to the Pilots, starting with the Pilots' monitoring process.

2.7 Readiness regarding the INFINITECH Marketplace (WP8)

WP8 focuses on the design, implementation and delivery of a scalable and adaptable multi-sided marketplace facilitating the utilization of novel Big Data and AI technologies in Fintech and Insurance sectors by providing easy and trusted access to valuable assets such as algorithms, datasets, frameworks, webinars, lectures, end-to-end solutions, and validated results provided/developed by Project partners.

The Marketplace supports the exploitation of these assets either internally among INFINITECH's ecosystem or to external stakeholders. The internal exploitation enables collaboration and knowledge transfer, as the sandboxes described in section 2.5 could also consume offerings and assets hosted on the Market platform. Additionally, the latter facilitates the "external" exploitation of the Project's results to relevant stakeholders such as scientific researchers, financial service providers, AI vendors, and fintech towards the Project's sustainability and broader impact. To this end, the Marketplace will deliver these assets through various monetization and marketing paths, including open access to datasets and AI/ML models, membership fees for specific assets that will not be open to the public, and on a pay-per-use basis for specific solutions. Moreover, the Marketplace aims at creating and attracting various stakeholders while also being open and contributing to BigData and AI communities, amplifying the scientific, societal, and monetary impact on the Finance and Insurance sector.

Accordingly, WP8 performs all the required activities related to the Market Platform's specification, design, and implementation, as well as the establishment of the Virtual Digital Innovation Hub (VDIH). The following list of deliverables documents the work that has been carried out related to the specifications and requirements of the Platform and VDIH capabilities during the period reported in this deliverable:

- D8.2 (Market Platform and VDIH Specifications – II): constitutes the follow-up on the specifications of the INFINITECH market platform as defined in D8.1. In this deliverable, various updates have been introduced based on the feedback on the initial implementation of the market platform, including a completely new design of the presentation layer, a list of the assets to be offered by the market platform, and additional use cases to be supported.

Moreover, the various aspects of the activities and work performed on the INFINITECH Marketplace with regard to the preparation and deployment of the various assets are documented in the series of WP8 deliverables as follows:

- D8.3 (BigData and AI Solutions Marketplace – I): pays emphasis on the accommodated assets, providing a quick overview of the already deployed INFINITECH marketplace, its updated architecture,

functionalities, and asset accommodation capabilities, while presenting in-depth the various APIs developed for users, assets, and their respective descriptions and documentation.

2.8 Common reference model for business services and competitiveness (from WP9)

During the second part of the project, INFINITECH is shifting to a more business-driven view, aiming to identify and highlight the innovativeness and novelty of the different pilots. The INFINITECH approach to Business Modelling and Innovation Management is a continuous validation process of the key aspects of the innovation which leads to adaptive strategies with the objective to maximize opportunities and minimize the risks (threats) connected to innovation development. It is a methodology which along with technological advancements will form the INFINITECH way to propose to the market new services and products. The INFINITECH methodology leverages dynamic and recurrent assessments of the innovation level of different dimensions, thus proposing specific and guided actions intended to raise such levels, increasing as a consequence the innovation of the pilot itself. Business Modelling and Validation will have other purposes which will be fundamental and not ancillary: to define a common language to focus on a business scenario, to suggest the business strategy (what to do and what NOT to do), to highlight the strengths of the solutions and to communicate to customers a sound value proposition. It is (should be) understandable within different functions in an organization, from stakeholders to investors. D9.15 “Business Modelling and Innovation Management” will present and remark on the business innovation of each pilot, along with the INFINITECH reference validation model and its new methodology.

3 Quick update from pilots

Each pilot provides here a quick update based on their main achievements from M14 to M25 and includes concerns from the review report. Each update will also remark on special events, stakeholders' meetings, and market analysis, but will keep the focus on the business innovations, plus relevant explainable AI and roadmap for the next year. Deeper details about all items reported here will be included in the next round of clusters' deliverables (D7.4, D7.7, D7.10, D7.13 and D7.16) in M27.

3.1 Pilot #1 Invoices Processing Platform for a more Sustainable Banking Industry

Pilot#1 will not proceed in its developments and will remain as it is. BANKIA has been acquired by CXB (CaixaBank) and, due to structural changes of the Company, CXB has not the capabilities to continue working consistently on the pilot, having however expressed their interest in it beyond the project. Nevertheless, the work performed in Pilot#1 will be reused for Pilot#15 as they both utilize OCR and AI to extract information from bank documents.

3.2 Pilot #2 Real-time risk assessment in Investment Banking

Pilot #2 has defined two use cases. More specifically the *first use case* implements a real time risk assessment and monitoring procedure in a typical trading process, where there is no real time risk assessment ongoing, but rather a risk assessment based on history (that means for example based on previous day's values). The *second use case* is based on what-if-scenarios and will focuses on pre-trade risk assessment. In addition, the pilot will implement a sentiment-based decision support indicator derived from financial and economic news data and social media channels.

Business Innovation

The pilot introduces a novel and more reliable VaR model which is based on a deep neural architecture instead of simple econometric models. The provided risk estimations are updated in real time, as the developed solution is able to handle and process the most recent available data. Furthermore, the Pilot will offer sentiment analysis in financial news providing an extra risk indicator.

Stakeholders' meetings and events

The types of stakeholders involved in this use case are Traders, Risk managers and financial institutions (as a whole).

Our main objective focuses on stakeholder's feedback. The first step consists in identifying and attracting potential customers and early adopters. The next step is driving their interest (i.e., to discuss the topic of our pilot, inviting them to workshops, following INFINITECH on social networks, etc.) and gather their feedback. Finally, to introduce them to the business/technical development loop.

A 1st workshop has been already taken place on 16 March 2021 and a 2nd one is planned in the near future.

Trusted AI: Explainable, ethical, and unbiased predictions

The Pilot has developed and leverages the DeepVaR algorithm to estimate VaR, which is based on Recurrent Neural Networks, and could be considered as a black box inside a glass box. The glass box aspect arises because the Pilot provides full (transparent) visibility of VaR calculations based on classical approaches (Monte Carlo (MC), Historical Simulation, and Variance-Covariance). This functionality enables the user to compare the outputs of the different models (i.e., statistical models VS DeepVaR) and evaluate their historical performance. Moreover, historical and future predictions of each VaR model are available as ergonomic visualizations at a dedicated

dashboard to ease interpretation of the black box aspect of models' outputs. Besides, this use case follows the "human-in-the-loop" approach with the developed models acting as virtual assistants to the user and not taking any automated decisions. As far as model bias is concerned, the DeepVaR algorithm initially estimates the future (log) returns of each asset in the utilized portfolio and then follows an MC approach to calculate VaR. Thus, bias concerns are applicable indirectly during returns estimation, which is avoided by re-training the DeepVaR model every 15 minutes on the latest market data.

Next steps and timeline

The next steps in the development of the pilot include:

- **Sentiment Analysis Model (Q4 2021):** In the scope of this phase, the pilot will focus on sentiment analysis in financial news such as newspapers, blogs, websites and forums. This field will have the effect of providing an additional risk indicator. On the business side, our objective is to identify and involve stakeholders in discussions of the topic and be more specific about our pilot (stakeholder's Feedback).
- **Scale up to more FX assets and incorporate more than one FX portfolio (Q1 2022):** We will scale-up to more FX assets and portfolios. We also plan to organize a 2nd workshop. In addition, our focus is to keep attracting and involving stakeholders and early adopters.

3.3 Pilot #3 Collaborative Customer-centric Data Analytics for Financial Services

The Pilot is focused on addressing a specific set of KYC (Know Your Customer) challenges relating to the movement of illicit profits generated by criminal entities involved in human trafficking through the legitimate financial system. Human Trafficking is an international business/economy which revolves around the ability of criminals to exploit victims for the purpose of generating money; the proceeds of these activities are moved & washed through the financial systems in many different ways but tend to avoid detection for a number of reasons which range from limited resources, inability to share data, insufficient understanding of the issue, lack of sophisticated financial monitoring systems, etc.

The financial system is the key lever to impacting the profitability of human trafficking but this necessitates technology which is intelligent in the domain of human trafficking, specifically in understanding & identifying the typologies of financial transaction red-flags which can be used for real-time notification of suspicious transactional behaviour patterns. There are internationally published & recognised red-flag transaction indices which are used by the financial sector for monitoring the transactions of their customers for human trafficking related transaction patterns; however these red-flag indices tend not to be specific to human trafficking and typically do not include compound red-flags or typologies which are often present in human trafficking incidents.

Business Innovation

The project will research & develop the ability to use industry standard AI/ML approaches to identify/infer the red-flags transactions which are mentioned in human trafficking incident reports, to provide a weighting of each facet of mentioned red-flag transactions based on importance and/or frequency and assemble these red-flag transactions into a directory of human trafficking specific typologies.

The output from this pilot will be an NLU-driven process which will generate dynamic typologies based on negative news domain data ingested through the Traffik Analysis Hub data ingestion processes. These HT specific typologies will be updated in a centrally accessible repository for authorized Financial Institutions to augment their internal KYC processes as they relate to Human Trafficking transaction monitoring.

Next steps and timeline

The next steps in the development of the pilot include:

- **Multi-Stakeholder workshop (Q4 2021):** Roundtable discussion with the Pilot team as well as several other banks and Consultancies to verify the selected use cases as being representative of the problems to be solved.
- **HT Typology Identification Model (Q1 2022):** In the scope of this phase, the pilot will focus on developing & training a custom NLU-based model to identify red-flag transactions & generate typologies from the TAH negative news data sources which include sources such as newspapers, blogs, websites and forums. The model will also provide an additional typology strength indicator, which will be influenced by a combination of the types of red-flag indicators included in the typology and the frequency with which the typology is identified. We will then review the approach and the initial results generated (typologies) with industry stakeholders through a series of workshops & demonstrations and gather stakeholder feedback as we progress.
- **Assist Bank of Ireland to integrate the HT Typologies in their FIU processes (Q2 2022):** We will work with Bank of Ireland, as our primary industry partner in Pilot 3, to consume the typologies generated within the existing Financial Investigations Unit (FIU) processes. Through this activity we will establish the typical real-world limitations/challenges which industry partners may encounter in leveraging the HT Typologies and use these findings to refine the artefacts generated.

3.4 Pilot #4 Personalised Portfolio Management

As initially planned and described in several deliverables, Pilot #4 targets a “private banking like” offering for a broader audience including the development and adoption of Portfolio Optimization algorithm “AIGO” within SaaS based Privé Managers Wealth Management Platform. The improvement and expansion of the algorithm’s capabilities will be also further pursued, having always real customer demand in mind. The use case with this artificial intelligence engine will be aimed at supporting better investment propositions for retail clients, which can be based on personal investment preferences like Performance or Risk-Control, but also ESG can be set as a preference there.

Business Innovation

Privé has conducted business and market analysis that resulted in fruitful results which are of a high importance to the business activities. The derived findings highlighted the following aspects: a) first, the current market landscape for using AI portfolio construction tools in a broader scaling is in its nascent stages and many competitors from Asia and North America are vying to exploit this opportunity; b) second, current economic conditions and, as a result, market trends highlight an interesting tendency. Mainly, it is implied that clients’ values have been drastically impacted by uncertainties caused by the pandemic with the growing attention granted to digital tools and solutions. This is why the need for digitisation, personalisation and more risk appetite become clear and is and has been the core-focus of Pilot #4 within INFINITECH. All these market trends are of high relevance to the project-scope.

Apart from conducting further market analysis and continuous market trends’ monitoring initiatives, Privé’s upcoming focus will be focusing on the newly-developed personalised Risk-Profiling and the enhancement of the Portfolio Optimisation (AIGO) functionalities to new and existing customers (mainly Banks, Insurers and Asset Managers) to create and identify potential selling and cross/up-selling opportunities. This includes also additional fitness factor implementation for new data sources, like Social Media activities (based on Infinitech partner solution Reportbrain).

Next steps and timeline

In terms of further progress within the Infinitech project, the most recent update on the Pilot 4 is the decision to expand Pilot's 4 coverage and integrate a first "Blockchain Use Case", which would be an extension of the original Pilot 4 connected with additional development effort. As a part of this Blockchain extension, Privé will create a managed single-node blockchain use-case which persists securities/asset-transactions and financial asset-positions in a standard format defined by Privé. This will be a first-step implementation that will be targeted within the Infinitech Project. As this is a first step within Pilot #4 no further consumption of the created blockchain storage is currently planned. This could be potentially added and done in a next step, which is not yet covered by the Infinitech 2020 project scope. Regarding the Blockchain extension, Pilot #4 roadmap update has been conducted with the probability of further updated specifications.

In terms of other activities, Privé has contributed to the most recent Infinitech Newsletter – September 2021 Edition – by submitting an article "Using AI-driven Portfolio Optimisation to Consider ESG Preferences in Portfolio Creation". Here Privé examined the importance of socially-responsible investing practices that have become a priority and regulatory requirement for many clients across the globe. In this regard, the use of AI-driven portfolio optimisation can help advisors to factor in a myriad of sustainable preferences effortlessly. Privé also focused on the ability of 'the 'AIGO'' algorithm to cover sustainability as a preference (fitness factor). In this contribution we explained how to use different fitness factors for an overall sustainability rating. In addition, Privé's further research in this field was mentioned. These future findings could be used to build fitness factors and to provide portfolio optimization solutions based on such sustainability and ESG factors.

3.5 Pilot #5b Business Financial Management (BFM) tools delivering Smart Business Advice

The recent progress and effort of Pilot #5b focuses on three main pillars, with distinctive goals and tasks related to each of them. All three are equally important for the pilot's successful development and are presented below.

Pilot's deployment and testbed

Setting the pilot's incumbent testbed on Bank of Cyprus: as the pilot has been successfully deployed in the shared testbed following the "INFINITECH way" of development and working as a blueprint (alongside Pilot #2) for the rest of the pilots, the main focus has shifted to completing all the legal, technical and business requirements connected with the acquiring and operation of the AWS testbed by the Bank of Cyprus. With additional resources and support from involved partners, the pilot's assets (testbed specs and all relevant source code) are expected to be migrated soon to the bank's own cloud environment. The pilot shall be the first to be replicated with the support of WP6 partners, implementing the script developed and acting as a reference for the rest of the pilots with incumbent testbeds' needs.

Continuous development of BFM toolkit analytical services: the introduction of the hybrid transaction categorization and cash flow prediction analytical engines has enabled the development of further analytical services included in the Business Financial Management toolkit, exploiting the results of the already developed components and offering additional value to the SME end-users. The Benchmark Engine back-end will soon be finalized and included in the pilot's cloud testbed. However, it is noted that for the integration of data streams and the development of some BFM components where such streams are required, the migration from the project's shared testbed to the bank's cloud environment is required.

Trusted AI: Explainable, ethical, and unbiased predictions

Apart from the summed progress highlighted above, the need for improved comprehension of the developed AI models and especially the hybrid transaction categorization engine has risen. XAI has been introduced in model development, leveraging LIME and SHAP frameworks in the Transaction Categorization Engine.

Stakeholders' meetings and events

Pilot 5 team has already been conducting one-on-one workshops with SMEs and internal bank stakeholders, both providing valuable input and shaping the pilot's end services. As the in-depth workshops with SMEs cannot be conducted with a statistically significant number of SMEs based on the total SME end-users, alternative approaches are also being considered to involve SMEs and creating synergies with related BOC business operations and actions. Moreover, with the imminent testbed migration and the platforms front-end development, additional pilot members from involved partners have been introduced.

3.6 Pilot #6 Personalized Closed-Loop Investment Portfolio Management for Retail Customers

The progress and effort of Pilot #6 for the past months focuses on four main pillars: a) Asset Categorization; b) Customer Investment Profiling completion; c) Sentiment Analysis; and d) GUI implementation. Below you may find a short description for each of them:

Business and competitiveness analysis

Asset Categorization: the main goal in this part of the implementation is to create clusters of investment products. This classification will be accomplished according to their levels of risk, returns, liquidity and maturity.

Customer Investment Profiling: we are aiming in creating a mechanism that predicts customer's behaviour based on investor's deposits, time horizon, liquidity, risk aversion and risk appetite. All customers will be categorized, with minimum risk of misclassification, into the appropriate risk profiles. This way we will be able not only to depict their willingness and ability to invest, but also to understand their potential balance to be invested.

Sentiment Analysis: with a contribution of our partner, results from Sentiment Analysis of comments related to investment products will be incorporated. On top of Recommendation Engine information, this will enhance the final ranking of the proposed assets to the customers.

GUI Implementation: In parallel with the implementation described above, we are currently working alongside CrowdPolicy to develop a user-friendly GUI: a proper tool for Bank representatives which will demonstrate the results of our analysis fast and transparently.

Stakeholders' meetings and events

All the above achievements will be demonstrated to internal stakeholders to get feedback. The demonstration will take place in the week of 18-22/10.

Next steps and timeline

The final goal in the period Q4 2020 to Q1 2022 is to create a list of investments proposal per customer, ranked by the degree of matching to the Customer Investment Profiling.

3.7 Pilot #7 Operation Whitetail - Avoiding Financial Crime

Pilot #7 has defined one use case: **Improve the detection of Financial Crime using refreshed customer data and AI-based analysis of customers' transaction behaviour.**

We have selected a kind of transaction that fraudsters are using prior to committing fraud, which exploits security weak points in systems used by "pay-loan companies" to make instant loans to people who run out of money before their next pay check arrives; for example, fraudsters can use false credentials to set up thousands of loan accounts and immediately each loan is authorised they can empty that loan account and commit large-scale

fraud. This is not new but now is something quite common and we want to apply INFINITECH's innovations to accomplish:

- Customers' behavioural data collection.
- Customer risk profiling.
- Pilot use case: Immediate Loans' fraud risk reduction.

Business Innovation

The adoption of AI is a priority trend included in the CaixaBank's Digital Security department's strategic plan for the next 3 years. We have foreseen a main objective and a secondary one, based on increases in the number and the of sophistications of attacks we are suffering year after year.

The main use case's objective is to introduce AI-powered techniques throughout the whole Security configuration described above, in two ways:

1. How CaixaBank can take advantage of AI techniques to enrich and improve our current tools and controls, basically in two ways, first one by improving the quality of the results of our Data Analytic tools, that is to say detecting the majority of the attacks whether they are known or unknown. Putting special emphasis on the unknown attacks because the current techniques based on rules and analytics are working fine when we have known patterns but are not so able to detect and respond to new patterns of unknown attacks.
2. How CaixaBank can be protected against new kinds of attacks that can use new and sophisticated AI techniques in order to avoid our current controls.

Minimize the number of alerts that have to be analysed by the human team, basically reducing the false positives, but also being able to take decisions autonomously in case there are no people in the SOC's team able to analyse alerts (due to an increase of attacks or other reason), apart from this it would be good if AI techniques could be applied to provide insightful information on the cases and alerts that could help and facilitate analysts to decide how to deal with an incident in less time.

Trusted AI: Explainable, ethical, and unbiased predictions

To carry out the pilot we will use an algorithm trained with CaixaBank's dataset which could have intrinsic bias (i.e. based only on CXB clients' profile), However, CaixaBank's dataset does not include explicit personal features that could be used to discriminate. And in any case human analysts are expected to be in charge of all steps in the process (selection of the dataset, selection of the fields, verification of the results).

In relation to making AI explainable, we expect to include a good degree of XIA, in order to know what are the features which had most influence on each detected fraud, and which ones can be discarded (i.e. if discrimination to a specific collective arises).

Explainable AI is basic to be aligned with current CaixaBank's fraud detection and prevention strategy because our main focus is not to detect a particular fraud but to detect a pattern of a fraud attack, for this reason, our analysts are analysing cases to determine which fields are important to take into account and which of them are conforming to a pattern, to do this is mandatory if we are to be able to explain why a particular case is fraud, and facilitate analysts to find a pattern and other instances of fraud related to the former one.

Next steps and timeline

The next steps in the development of the pilot include:

- **Q4 2021 – Supervised analysis and extraction of “immediate loans” fraud detection model.** To be able to see if there is a pattern in the immediate loans which could let us know if this transaction is genuine or it will be followed by a fraud.
- **Q1 2022 – Evaluate the model with new datasets in CXB’s Information Security Garage Lab.** We will install INFINITECH’s tools in a secure environment inside CaixaBank to evaluate it with more Datasets.
- **Q1 2022 – Evaluate requirements for the integration in CXB’s operations and production systems.**

3.8 Pilot #8 Platform for AML supervision

Tools within PAMLS as a SupTech solution would be of great interest especially to the relevant supervisory bodies, financial intelligence units and other supervisory authorities supervising financial institutions (hereinafter: FI) in their responsibility to comply with Anti-money Laundering and Countering Terrorist Financing (hereinafter: AML/CFT) provisions.

Business and competitiveness analysis

There are several AML/CFT related software solutions on the market, that support FI to comply with AML/CFT regulations in their daily processes: data management, transaction monitoring, PEP screening, KYC, ID verification (so called RegTech). In the meantime, European supervisory authorities just started to promote the need for digitalization and innovative technology in supervision. Consequently, there is still a lot of room for development of innovative technology to support supervisors in their supervisory activities (so called SupTech), especially in the field of AML/CFT supervision. Based on market analysis in the field of AML/CFT some SupTech does exist, however it is in its developmental or experimental stage. In addition, identified AML/CFT SupTech solutions enable only Data analytics or only Data collection. Therefore, PAMLS provides innovative combinations of RegTech and SupTech not only as one of the few AML/CFT SupTech solutions, but also as it combines Data collection and Data analytics features for more transparent and effective supervisory planning and execution.

Trusted AI: Explainable, ethical, and unbiased predictions

In Pilot #8 we are using AI in a limited way – we are applying AI methods in a discovery phase of the AML supervision process. In Pilot #8 we are developing the PAMLS platform, which is empowering supervision analysts to streamline the supervision process of financial institutions. During the PAMLS development AI methods (unsupervised Machine Learning) are applied in the discovery phase of the Screening tool – detecting pre-defined scenarios to identify unusual patterns and relationships among data, that could indicate typologies and risks of money laundering/ financing terrorism (hereinafter: ML/FT). We are also using AI approaches and methodology to detect previously un-noticed unusual patterns and relationships. The context of the problem at hand calls for unsupervised learning techniques, therefore we are not foreseeing additional explainability measures.

We have however encountered several potential biases in the data. Available data is historical and pseudoanonymized, which is grouped from three different sources and different time windows. Even if all the three sources will be available, they will not cover the whole picture. Therefore, the data is incomplete. Next type of possible bias is labelled data. BOS randomly selected several suspicious behaviour examples and labelled them.

Possible measures for mitigating biases are:

- Awareness (creating awareness of PAMLS users to be aware of possible biases and pre-requisites of PAMLS usage)
- Education of PAMLS users – in terms of assumptions, which were used in PAMLS development and especially in how to interpret results

As part of PAMLS development, Pilot #8 will provide brief documentation with suggested guidelines on how to mitigate biases and how to interpret the PAMLS results.

Stakeholders' meetings and events

In cooperation with JSI, BOS organized two workshops in October, where Risk Assessment tool functionalities and Screening tool developments were presented. One workshop was with Slovenian Financial Intelligence Unit and the second was with Securities Market Agency. Participants in both workshops were AML experts, IT and legal experts. There has been interest in the presented tools and BOS and JSI will be holding meetings with both supervisory bodies in future. In July, BOS also presented PAMLS to the Chair of the ECB Supervisory Board during his virtual meeting with representatives and reported in EBA survey on financial innovation.

Next steps and timeline

Next Figure 2 represents the current status and the main next steps for the near future of Pilot #8.

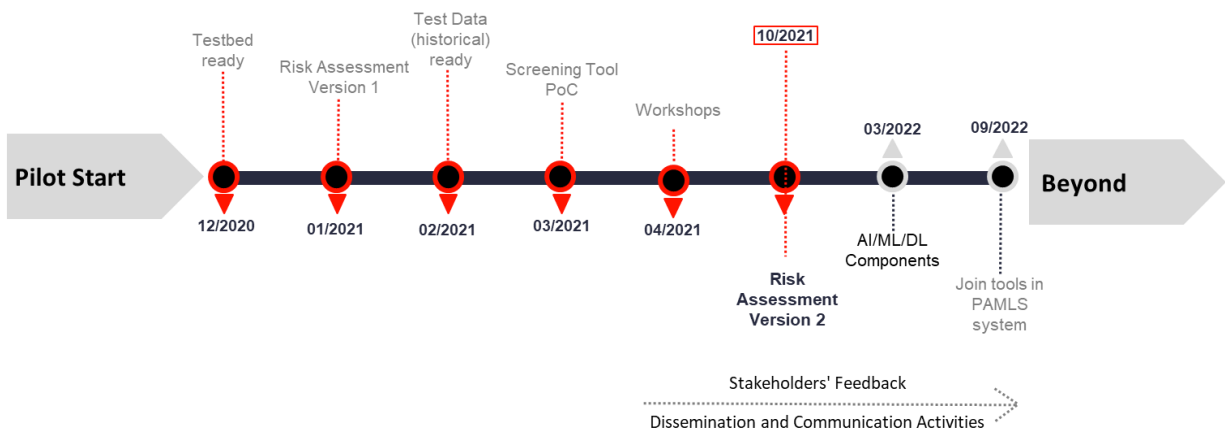


Figure 2: Pilot #8 summarised roadmap

- Complete the second version of the Risk assessment tool [October 2021],
- continue to develop the Screening tool components [start in November 2021],
- dissemination and communication activities,
- Stakeholders' feedback (working workshops).

3.9 Pilot #9 Analysing Blockchain Transaction Graphs for Fraudulent Activities

During the last semester of the project, the following activities were achieved:

- Analysis of Bitcoin transactions were carried out in addition to Ethereum data.
- A database of all Bitcoin and Ethereum addresses was constructed and feature tagging capability was developed.
- Evolving developments in the area of regulations were followed closely.
- Development of machine learning based analysis of blockchain transactions was initiated.
- Presentation about INFINITECH and Pilot #9 at “EBINTEC Banking Innovation Conference EMEA”.

Business and competitiveness analysis

Companies that have recently started offering blockchain fraudulent transaction tracing services have been researched. Some of these companies like Chainalysis and Elliptic entered this market early and hence, they have accumulated more historical data about fraudulent activities. They also have established relationships with exchanges and government agencies which may provide them with customer data. Since availability of blacklisted addresses are limited to what we can find publicly on the Internet, our pilot is differentiating and positioning itself as a service that is far less human-intensive but rather as algorithmic intensive and that builds on automated HPC tools and advanced parallel algorithms which enables it to offer scalable compute-based fraudulent-related automatic feature extraction from large volumes of Ethereum, Bitcoin and token transactions. We believe that there will be big market demand for fully automated algorithmic solutions, since new proof-of-stake based blockchains will have thousands of transaction throughputs which will require parallel automated procedures for their analysis. Our pilot is also being interfaced to Hyperledger Chain and can analyse transactions coming from token contracts. Ability to perform transaction graph analysis on Hyperledger transactions is not currently offered by other companies.

Trusted AI: Explainable, ethical, and unbiased predictions

Pilot blockchain fraudulent activity analysis techniques utilise mainly graph algorithms. Machine learning algorithms have also been initiated during the second period of the project. Predictions obtained can be checked and reasoned about by traversal-based graph algorithms. Public blockchain data and addresses are used without using any identity information enabling us to avoid privacy issues.

Technical progress

The following technical works were carried out:

- Scikit-learn and Imbalanced-learn libraries were used to apply machine learning algorithms to small Ethereum datasets.
- Development of Integer programming based convex subgraph finding algorithm that generalize detection of potential fraudulent path-based peel chain activity to subgraphs.
- Bitcoin and Ethereum address database management with feature labelling capability.

Next steps and timeline

Currently, we are working on organizing a workshop about Pilot#9 in November 2021. Internal Aktif Bank personnel will attend the workshop. Digital Currency team of Central Bank of Turkey has also been invited to the workshop who replied positively about attendance. Topics about crypto regulations in Turkey will be discussed and demonstrations as well as presentations about Pilot #9 will be made in the workshop.

3.10 Pilot #10 Real-time cybersecurity analytics on financial transactions' data

Frauds on financial services are an ever-increasing phenomenon and cybercrime generates multi-million operational costs, therefore even a small improvement in fraud detection rates would generate significant savings.

Pilot #10 goal is to improve significantly the detection rate of malicious events (i.e. fraud attempts) and enable the identification of security-related anomalies while they are occurring, by the analysis in real-time of the financial transactions of a home and mobile banking system. It aims at guaranteeing a real time detection of

suspicious fraudulent transactions by adopting a Machine Learning model periodically retrained based on always-fresh coming data.

Business Innovation

The limitations of the rule-based systems in blocking potentially fraudulent financial transactions are well known: this approach to fraud detection provides low coverage: only a few rules show accuracy and effectiveness and very low flexibility and adaptability to an ever-changing fraud scenario is shown. The need to overcome these limitations is partially addressed by novel AI based systems which anyway suffer from high false positive and false negative rates too. However, a different approach in managing ML and data models can address many of these issues, and more effectively identify risky transactions. Pilot #10 developed a novel AI-based fraud detection system – built over Data Science and Machine Learning – which runs the pre-processing of transaction data and model training in a batch layer (to periodically retrain the predictive model) 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 and for automating the fraud detection processes.

Major expected benefits from this approach:

- **Lower costs** through efficiencies generated by higher automation, reduced errors rates, and better resource utilization
- Ability to address major growth areas as **new types of frauds** are expected
- **Minimizing Disruption for Legitimate Customers**
- **Increase Client trust and security**
- **Competitors'** offerings are strongly affected by false-positives and false-negatives, and their solutions are **inelastic**: they do not fit easily to new threats and fraud scenarios
- **Brainstorming** with fraud experts in Poste Italiane defined **user-requirements** and drove the development. Forthcoming stakeholder's **workshop** will validate the solution.

Stakeholders' meetings and events

Stakeholder engagement is the key for a fruitful product development: during the initial development phase brainstorming sessions with fraud experts have been arranged to collect users' needs and requirements from the field and which drove the development phase. As the solution becomes more and more stable, stakeholders' workshops will be arranged in order to validate the solution and its potential for the market. Relevant stakeholders from the banking sector will be involved.

Current plan for workshops:

- 1st Stakeholders' workshop -> **Q4 2021**
- 2nd Stakeholders' workshop -> **Q1 2022**

Trusted AI: Explainable, ethical, and unbiased predictions

The Pilot leverages the ALIDA platform capabilities for ML pipeline automation. It will be able to ensure the execution of model training automatically and the continuous training of the model whenever new data is available. Doing that, model degradation will be avoided, and biases will be monitored. Moreover, Explainability AI techniques could be used in order to better understand what features to use for the fraud detection model's training. In particular we could take advantage of SHAP (SHapley Additive exPlanations) techniques to find the most important features and how to pre-process raw data to extract the highest amount of information.

Next steps and timeline

The next steps in the development of the pilot include:

- System integration and first end-to-end test (First iteration) (Q4 2021)

3.11 Pilot #11. Personalized insurance products based on IoT connected vehicles.

Pilot #11 belongs to the insurance cluster where IoT infrastructures are used to collect real time information from insured clients and so, by applying AI methodologies, identify and assign behaviour models. These models will later be used to infer accurate and updated risk estimations that support insurance companies to develop customised services and products. The baseline for Pilot #11 has been described in D7.1 and D7.12 and relies on a connected cars infrastructure, as its IoT source, provided by CTAG. This infrastructure captures technical information from cars and uploads this into a specific framework, the Smart Fleet supported by ATOS. This framework is based on the NGSI standard and FIWARE data models, and here, car datasets are homogenised and merged with additional related information, such as traffic alerts, roads, and weather information. Then, this is linked and served to the pilot's AI layer, where it is curated, aggregated, and formatted to identify and define data clusters as precursors of driving profiles. These profiles, in turn, are the core of the use cases proposed by this pilot: a) the Pay as You Drive (or Pay How you Drive) service, which uses these driving profiles to analyse the driving behaviour of the insured client and so, propose updated and adapted premiums that reward good drivers and penalize bad ones; and b) the Fraud detection enhancement, that by applying the driving profile model, insurers can infer who were the actual drivers involved in a traffic incident and what were its context conditions. These improvements related to risk and fraud estimations will finally impact on car insurance companies, represented here by Dynamis, and their economic outcomes, avoiding unfair and periodically rising prices.

Business Innovation and market analysis

Pilot #11 directly aims at the risk assessment on car insurance. This is traditional done by analysing static data mainly from a) the driver (age, experience, gender, marital status, etc.); b) the vehicle itself (age, value, reported usage, brand, power, colour, etc.) and the area where it is been driven (city limits, countryside, etc.); and by comparing this with statistical analysis. In this sense, the market introduces the concept of the Usage Based Insurance (UBI) that pursues a risk assessment based on the actual usage of the insured vehicle. Current solutions rely on an external device that captures information about the driving style, this is, e.g. using the sensors of the mobile phone of the driver (reporting estimated vehicle's acceleration, location, and speed) or an installed tracker that reports current location of the vehicle. This information is collected and used to define two main models of services: a) a Pay As You Drive (PAYD) service that consider distance covered to recalculate primes; and b) a Pay How You Drive (PHYD) model that considers also speed (and over-speeding), braking and acceleration to provide feedback about the driving style, prevent traffic incidents and recalculate prices. The business innovations introduced by Pilot #11 are aligned with this UBI concept expanding it in two ways: a) by collecting data from the Controller Area Network (CAN) Bus of the vehicle. This is, directly gathering the speed, acceleration, steering wheel position and status actually reported by the vehicle, which anticipates the Vehicle to Infrastructure (V2I) paradigm; and b) by aggregating driving context information (traffic, roads, weather) and develop AI powered models that identify and define enhanced drivers' profiles.

Stakeholders' meetings and events

The pilot refers mainly to a specific stakeholder: the insurance companies, as these are the main business actors that put together the whole solution. As a stakeholder themselves, they provide the insured clients, the policies and the insured products that can be developed on top of the driving profiles, whilst as end users, they can benefit from the new risk assessment and fraud detection services. Derived from this, the insured drivers can be added as end users of the novel products these insurance companies would provide.

On the other hand, and looking at the data gathering and homogenisation process, an extra stakeholder has been identified in terms of vehicles' manufacturers and the autonomous driving research. The standard APIs and data models developed within this pilot plus the framework for this aggregation and homogenisation are interesting for these actors to develop new V2I services, as well as to expand the Pilot 11 solution by adding more and more connected vehicles.

In this sense, Pilot #11 is working on new workshops to enrol these stakeholders and get their feedback about the innovations presented. A joint Infinitech clusters' workshop was recently (Oct. 12th, 2021) hosted online by the University of Glasgow where Pilot #11 was introduced to Fintech and university student attendees, with an AI and technical scope. A new workshop is planned before end of 2021.

Trusted AI: Explainable, ethical, and unbiased predictions

In terms of Explainable AI (XAI), Pilot #11 will use decision tree algorithm, as an explainable machine learning algorithm all by itself. This is a relatively simple model, used widely for feature importance of linear and non-linear models, which is easily explained by visualizing its corresponding branches. For Python powered models, SHAP (SHapley Additive exPlanations) libraries will be used to support an explainable AI framework derived from the shapley values of the game theory. This will provide end users with graphical visualizations of the relevance of each variable (speed, acceleration, precipitations, visibility, etc.) on each identified cluster.

To guarantee ethics in AI, the pilot will put in practice the principles of ATOS' (partner developing AI models in this pilot) Code of Ethics and Ethical Digital Vision, a commitment with a human-centric approach for the adoption of AI and digital technologies which is part of the company's "Sense of Purpose"³

Next steps and timeline

This pilot's detailed roadmap so far and for the next year will be included in the next D7.13. Some remarks with most relevant points are mentioned here:

- **Data gathering and homogenisation process (Q2 2021):** this stage has provided the Smart Fleet platform, currently deployed and collecting data from involved sources since April 2021. An instance of this is also deployed in the INFINITECH testbed.
- **Driving Profile AI model (Q4 2021):** a first functional version of the AI powered model to identify the profile of a selected driver will be released. This first version will use data collected from the real connected vehicles provided by CTAG.
- **Pay as You Drive & Fraud Detection services (Q2 2022/Q4 2022):** Functional/Final versions of the end user services built on top of the AI models and developed/provided by the insurance company

3.12 Pilot #12 Real world data for novel health insurance products

Pilot #12 belongs to the insurance cluster where wearable devices and self-assessments are used to collect real time information from insured clients and so, by applying AI methodologies, identify and assign behavior models. These models will be later used to infer accurate and updated risk estimations that support insurance companies to develop customized services and products. The baseline for Pilot #12 has been described in D7.1 and D7.12 and relies on collecting information from people and use this for building models at the pilot's AI layer, where it is curated, aggregated and formatted to identify and define data clusters as precursors of behavioral profiles. These profiles, in turn, are the core of the use cases proposed by this Pilot #12: a) the Risk Assessment service, which uses these profiles to analyze the behaviour of the insured client and so, propose updated and adapted premiums that reward good lifestyles and penalize bad ones; and b) the Fraud detection service, which processes

³ <https://atos.net/en-nl/lp/digital-vision-ethics>

the collected signals for outliers, but also utilizes explainable AI techniques to determine important factors in the behaviour of people, and offer them back this in the forms of insights. The anti-fraud strategy thus involves a combination of deterrence (detecting abnormalities) and persuasion (on the usefulness of acting truthfully).

Business Innovation and market analysis

Health insurance products are today static in terms of customers' health evaluation. Personalized health insurance products need to be dynamic, employing a continuous risk assessment of the customer. Medical history of insurance customers can be scarce, and anyway only partly determining health. While the continuous estimation of risk factors is well-known in medicine, it is not widely used to personalize insurance products. Such personalized products start appearing as digital risk assessment platforms based on data and the continuous monitoring of customers' lifestyle and behaviour start transforming insurance.

In this pilot we define with insurance experts the data to be collected, and we employ the Healthentia system for data collection. Risk assessment services can be provided to health insurance professionals by training machine learning (ML) models for predicting important health parameters. The usage of ML in insurance is not new. ML has been used to analyse insurance claim data. Instead of an analysis of data at the end of the insurance pathway, after the event, this pilot focuses on the continuous analysis of data to modify the insurance pathway by personalizing the insurance product.

Insurance companies benefit from personalized dynamic product offerings, as they can be competitive with lower prices for low-risk customers. However, to obtain their customers' consent to monitor them, insurance companies also need to persuade their customers about the benefits for them. Customers will potentially consent to two types of rewards: On the one hand, monetary rewards stem from receiving personalized offers with reduced premiums due to the lower risk of their healthy behaviour. On the other hand, coaching for well-being is an indirect reward that can be offered by employing explainable AI techniques in the classifiers utilized by the risk assessment service.

Stakeholders' meetings and events

Pilot #12 refers to health insurance companies, as these are the main business actors that put together the whole solution. As a stakeholder themselves, they provide the insured clients, the policies and the insured products that can be developed on top of the behavioural profiles, whilst as end users, they can benefit from the new risk assessment and fraud detection services. Derived from this, the insured public can be added as end users of the novel products these insurance companies would provide.

Pilot #12 is collaborating within cluster 4 on new workshops to enrol these stakeholders and get their feedback about the innovations presented. A joint Infinitech clusters' workshop was recently (Oct. 12th, 2021) hosted online by the University of Glasgow where the more technical aspects of Pilot 12 were introduced to Fintech and university student attendees. A new workshop is planned before end of 2021.

Trusted AI: Explainable, ethical, and unbiased predictions

Explainable AI (XAI) is used in Pilot 12 to identify the attributes playing important role in the model decisions. To this extent SHAP (SHapley Additive exPlanations) provides insured clients with the most important elements in their behavior leading to positive or negative outcomes and are coached to strengthen the former while avoiding the latter.

Next steps and timeline

The pilot is currently at the stage of initiating the study with actual users for collecting data and training the models. The detailed timeline is as follows (Figure 3):

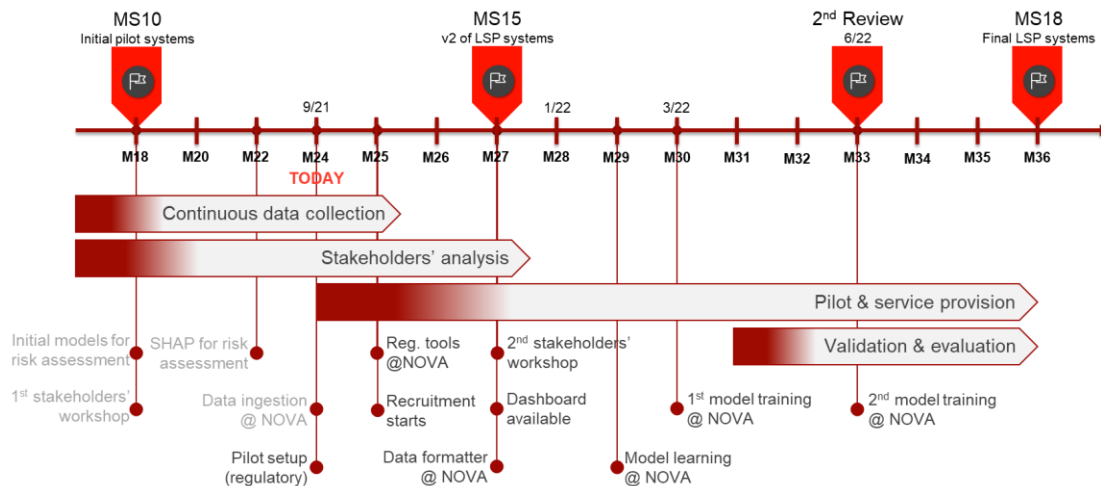


Figure 3: Pilot #12 summarised roadmap

3.13 Pilot #13 Alternative and automated insurance risk selection and insurance product recommendation for SME's

Capture from open sources non-traditional data from SMEs to automate risk rating and insurance product recommendation and exploit AI technologies to create SMEs risk profiles and automate underwriting processes

The activities carried out in recent months in pilot #13 have been directed towards three verticals: the technological and product development of the pilot, improving communication of Infinitech and the developments of pilot #13, raising awareness of the need and advantages of using the platform, and meetings with insurance and reinsurance companies to increase the marketing funnel of the pilot #13 product. About product actions, progress has been made in the following areas:

- Implementation in Nova's infrastructure
- Improved data loading capabilities from input cards.
- Introduction of the family business determination model algorithm.
- Introduction of variables related to environmental sustainability

Business and competitiveness analysis

During the last few months an analysis of the competitors of pilot #13 has been carried out, comparing a total of 8 competing companies with similar services aimed at the insurance sector. The main competitors come from the USA and the UK.

There are four competitive advantages over our competitors. Flexible connectivity: Wenalyze can exchange data with any system the insurer has, regardless of its technological level. Automated implementation: the solution is plug-and-play and in only 15 days it is ready to use in any setting, financial institution, or country of the world. The platform can work with any size of clients regardless of setup costs, since there are not, therefore it is absolutely scalable. Finally, Wenalyze is the first and only analytics platform in the world to provide insurers with sustainability indicators for their corporate clients.

Stakeholders' meetings and events

In terms of communication, the following actions have been carried out (only some of them are shown in brief in Table 1).

Table 1: Pilot#13 meetings and events with stakeholders

DATE	SOCIAL NETWORK	LINK	TYPE OF CONTENT
06/04/21	Linkedin	Post	Educational
14/04/21	Linkedin	Content access	Promo
29/04/21	Youtube	Content access	Educational (webinar)
03/05/21	Blog	Post	Educational
06/05/21	Linkedin	Content access	Educational (infography)
02/06/21	Youtube	Content access	Interview
10/06/21	Blog	Post	Interview
26/07/21	Blog	Post	Educational
30/09/21	Youtube	Content access	Educational (webinar)

It is important to highlight the workshop held in March, which together with the pilots 11, 12 and 14 brought together more than 60 attendees from potential clients. Finally, meetings and demos have been held with the following insurance companies: Hiscox Spain and Germany, CGPA Europe, Banco Caminos, Banca March, Zurich Italia, Zurich Italia, Zurich Italia, Banco Caminos, Banca March.

Trusted AI: Explainable, ethical, and unbiased predictions

Given that the models are being developed in pilot #13, we cannot speak of conclusions regarding the bias of the algorithms being developed; what we can say is that they are not using black box modelling and always based on the XAI.

Next steps and timeline

The next steps of the road map are clear, given the integration into Nova's servers, the completion of the testing and commissioning of the algorithms and the implementation of the connectivity of the data output through APIs.

3.14 Pilot #14 Big Data and IoT for the Agricultural Insurance Industry

Business Innovation

Pilot #14 has successfully secured one more pilot partner, so that the INFINITECH AgI toolbox is currently being tested – by two insurance companies in two countries, namely **Wiener Städtische osiguranje (WSO Serbia) – Serbia** and **Croatia osiguranje d.d. (CO) – Croatia**. In both cases the insurance companies use their current and past multiple peril crop insurance (MPCI) contracts with the portfolio of insured agricultural land and experienced losses to benchmark “*business as usual to the Infinittech way*”. Furthermore, the uptake of INFINITECH AgI toolbox is enabling them to deploy new parametric insurance products in their market. For this reason, **Genillard & CO**, will design two new insurance products for those pilot partners, an **Index-based insurance product for the Serbian insurer** and a **Frost Parametric Insurance product for the Croatian insurer**, based on the data and services the INFINITECH AgI toolbox is providing. After implementing our pilot partners’ feedback and findings, we will continue our co-design process by approaching further insurance companies for validation during the last phase of the project.

Stakeholders’ meetings and events

Following the INFINITECH Stakeholders Workshops Series on “Artificial Intelligence and Big Data analytics applied to Personalised, Usage Based and Configurable Insurance Products”, a number of contacts have been initiated

with a broad number of Insurance actors globally. The aim is to inform them of INFINITECH and Pilot #14 activities, to acquire feedback and understand their operational pain points and present the first version of INFINITECH AgI toolbox capabilities. Through those iterations - meetings we confirmed our initial understanding with regards to the sectors' pain-points and their expectations from a commercial tool that would address their needs.

Trusted AI: Explainable, ethical, and unbiased predictions

Pilot #14, piloted and evaluated a number of ML/AI models in an attempt to select the ones most applicable in terms of efficiency, performance and accuracy of the results. Following this evaluation, a supervised ML solution was selected as the most prominent one, considering in situ data about crop yield /crop damage, and training ML algorithms with EO layers as predictors. The ML framework consists of:

- Outlier Detection: a detailed analysis of outlier events on the in situ dataset.
- Model Training: a Nested Cross Validation approach was adopted in order to obtain unbiased Hyperparameter Tuning/Feature Selection and get an honest model performance estimate. The selected ML model used an SVM algorithm for regression/classification tasks.
- Model Testing: An ultimate evaluation of the selected model that was performed, on "unseen" data.

Technical progresses

Pilot #14 infrastructure is set and is currently operational on AgroApps infrastructure, while at the same time the deployment of Weather Intelligence engine has been initiated in the UNINOVA's infrastructure.

AgroApps' Weather Intelligence engine as described in previous deliverables (see D6.10) is the only set of Pilot #14 software hosted on the NOVA testbed due to an extrinsic constraint. The instances of AgroApps Weather Engine are launched as Statefulsets while all the other components will only have Kubernetes Services of kind ExternalName. These services, initialized into Pilot #14 namespace, will work as network references to AgroApp's managed premises, and allow every sandbox module to deliver requests and receive responses from the remote infrastructure.

The Weather Intelligence service provided by AGROAPPS is produced by operating the Weather Research and Forecasting Model (WRF), a meso-scale numerical weather prediction model. Under Pilot #14 namespace in Kubernetes orchestrator, 12 nodes with 244 cores will be utilized to run Weather forecasting model. Currently the images needed for the smooth operation of the weather forecasting model have been uploaded and Message Passing Interface (MPI) compatibility with Kubernetes (MPI-cubeflow) and the forecasting model is being tested.

Next steps and timeline

Technical: Finalization of the current tests running on the UNINOVA's infrastructure and fully operational deployment of Weather Intelligence service on the new infrastructure.

Operational:

- Refinement of the INFINITECH AgI toolbox on the basis of Pilot partners' feedback.
- Evaluation of pilot KPI's.
- Population of provided services with more indices.
- Extending coverage in terms of calamities.
- Enabling the deployment of two new insurance products based on the AgI toolbox.
- Preparing the second pilot run-iteration, according to the pilot countries cultivating periods and crops.
- There is the possibility (undergoing discussion) to further extend the Pilot partners, including an additional market-country.

Business related:

- Iterative process of feedback gathering from pilot insurance partners creating value for them and their agro-insurance processes.
- Developing the business scheme for a new parametric insurance product based on the INFINITECH AgI toolbox.
- Benchmarking the INFINITECH AgI toolbox with other products in the European market.

3.15 Pilot #15 Inter-Banking Open Pilot

Extraction of complex metadata from raw bank documents is paramount to support intelligent data indexing, to face the challenge of sharing info in an effective way within large organisations.

Banking language is very specific and rather different from common language. General purpose semantic engines may be ineffective in understanding banking-related concepts. This evidence raises the need to develop innovative solutions for metadata extraction.

Our pilot is based on a weakly-supervised neural methodology for creating semantic metadata from bank documents. It exploits a neural pre-training method optimized against legacy semantic resources able to minimize the training effort.

The pilot aims at innovating the Italian banking sector in a specific and quite complex domain such as the analysis of unstructured data.

The pilot exploits thus a neural pre-training method optimized against legacy semantic resources able to minimize the training effort. We studied an application to design and process management tasks in banks. Furthermore, we tested the method on documents from the Italian banking community.

Business innovation and market analysis

ABI Lab as the research and innovation centre works in a pre-competitive field, with the aim to favour innovation in the Italian banking sector. The business relevance of this initiative therefore is based on the mission to accelerate banking experimentation of the most innovative technologies.

The business case of pilot #15 was developed with the banks part of the AI Hub, chaired by ABI Lab. The market research was carried out together with the representative of the Italian banking sector, and it led to the conclusion that there were no similar solutions ready made in the market, focused on the corpus of texts produced by banks. Therefore, we acted as first-mover and we have developed a new experimental prototype starting from available models.

Although general-purpose tools are easily available and neural techniques demonstrates very accurate language modelling and inference capabilities (e.g., BERT or GPT-3), straightforward applications of such neural methods in business process mining scenarios are still limited.

The pilot comprises 5 different institutions, acting as early adopters of the Italian Banking sector, which are part of the ABI Lab AI Hub.

We leverage the concept of cooperation – collaboration within the competitors through a shared governance.

Trusted AI: Explainable, ethical, and unbiased predictions

- **ASSESS QUALITY OF THE INPUT DATA:** The raw documents, used for the scope of the experimentation, passed through an evaluation step (executed by each of the 5 banks that participate to the initiative) assessing the quality and reducing risk.

- **BALANCED SAMPLING:** to avoid Bias, it was ensured that the input data was sufficiently representative of the banking realities and sufficiently focused on the most relevant documental areas.
- **TEXT DRIVEN PROCESS MINING:** The approach used here to classify documents is based on a specialization of BERT model, using a Transformer-based architecture, representing the state-of-the-art in NLP field.

Next steps and timeline

In the last three (3) months, several activities were performed to monitor the progress of the solution, and to gather feedback from the key stakeholders:

- In June and July three meetings with early adopters were organized to monitor the progress of the solution, assess the requirements and to share the results.
- In last three months three (3) meeting with the AI Hub took place, to gather input and feedback from the stakeholders and to encourage the dissemination of the experimental findings.
- A restricted group of the early adopters were involved to support the refinement phase of the experimental phase, which started last July, and it is still ongoing.

The next planned activities are summarized as follows:

- Enhancement of the Solution, going further with the refinement of the AI Model.
- Analyse results and identify key issues.
- In December it is planned a workshop with the key stakeholders, to disseminate the first results and collect feedback.
- In 2022 the Pilot will proceed according with the identified priorities.

3.16 Pilot#16: Data Analytics Platform to detect payments anomalies linked to money laundering events

The pilot will build a data analytics platform to help Nexi's AML team to discover, monitor and analyse suspicious scenarios related to money laundering through digital card payments. Nexi, as the Italian paytech leader, owns and manages a large, big data ecosystem, which includes information regarding cardholders, merchants, organizations, and digital payment authorizations and transactions.

Two main challenges are met by the AML team:

- The difficulty to detect complex money laundering scenarios, like self-financing with several payments' actors, money transfers patterns to evade AML tracking, and many other cases.
- Lack of appropriate tools that allow AML operators to analyse large volumes of complex and interrelated data, such as those related to payments.

Business Innovation and market analysis

The innovation potential of the current pilot lies in introducing novel technologies like, machine learning, artificial intelligence, graph database to detect anomalous scenarios, which allows us to automatically detect complex anomalous money-laundering scenarios. Thanks to those technologies, we will see an impact on quality and productivity of AML operations teamwork.

We expect that the adoption of pilot platform will improve quality and efficiency of AML users work and, at the same time, will concur in reducing risk of unmatched scenarios related to money laundering events. Nexi AML regulatory compliance obligations include the notification to the competent authority (FIU) enhanced due diligence and suspicious transactions report linked to money laundering. Failing to communicate these reports

on time can lead to monetary sanctions and Nexi reputation damage. Both an easy to use and advanced data platform adoption are critical elements for speeding up the search and notification of suspicious anomalies for AML users and create appropriate reports. Moreover, Nexi pilot platform will remove time of AML analysts dedicated to manually wrangle and clean data with inappropriate tools like spreadsheets, allowing them to focus more on value added activities like: analysing suspicious scenarios create reports and regulatory authorities' notification. The Nexi AML team purpose is to preside anomalous scenarios linked to money laundering, adhering to European AML regulatory compliance policies, by notifying detected cases to the Italian Financial Intelligence Unit (FIU).

Next steps and timeline

The next steps in the development of the pilot include:

- **Anomaly detection algorithms (Q1 2022):** In the scope of this phase is to develop and test a first version of the anomaly detection algorithms to monitor suspicious scenarios that are automatically matched in a straight way. This will be supervised and unsupervised by machine learning algorithms that are able to detect advanced money laundering scenarios by matching complicated patterns, that instead are hard to be identified through standard AML screening processes.
- **Diligence analysis (Q2 2022):** We plan to use the test on the anomaly detection algorithm to explore and drill down each suspicious operations, to get the more details needed for their due diligence analysis. To do so, a graph database will be used to fit to model and represent interrelated and correlated data as the ones of payments.
- **Data visualization (Q2 2022):** Advanced and dynamic data visualization that will be developed in the data platform frontend allowing business users, like the ones in AML team, to explore and find information autonomously without the need to know how to query a NO-SQL database.

4 Conclusions

The first report (D7.1 – M14) was centred on providing the layout of each of the pilots, summarising the use cases, end users and functional requirements, as well as the technologies, architecture and datasets used to support their objectives. This baseline in most pilots haven't changed substantially, so the focus of this second report is business innovations brought by pilots' development of solutions and the activities carried out during the last 10 months, plus those already planned for the near future to enrol stakeholders and collect their valuable feedback. Fostered by the EC reviewers and their comments during the M20 project's review, each pilot provided their reinforced business description, complemented by a market analysis (or competitiveness) that point at the differences of their developing approaches and the existing solutions when facing a specific challenge. In this sense, the relying innovation idea behind most of the INFINITECH pilots is to explore diverse AI approaches, involving different ML/DL technologies (selecting those most suitable according to their corresponding context) to revamp the way the risks on behaviours, transactions and money movements are calculated. The pilots combine AI, Big Data and HPC to create novel models that manage huge datasets (e.g. bank transactions) and exploit new relevant ones (such as real world data, social media or IoT sources). These models will be able to support enhanced risk estimations, even updated in real time, using live data against current common solutions based on statistical studies, whilst reinforcing attack-detection and fraudulent-activities identification. The pilots will use these models to create different profiles for customers, investors, or insured clients to analyse the risks derived from their activities and customise offers, protection levels and products, as well as for transaction streaming to avoid fraud, prevent attacks and avoid money laundering. The pilots introduced here some of the different ML/DL techniques they are using, together with the datasets involved on each solution, as the key points for further description on their corresponding cluster's report (M27). And this is just the competitiveness pointed on most of the market analysis carried out within each cluster: the novel AI-powered algorithms that enhance the current data analytics mechanisms. These allow the expansion of the considered data sources and the development of faster and accurate solutions to evolve the fintech and insurance market. On the other hand, Pilot #15 explores a different but complementary usage of AI: it works on the analysis of the unstructured data to support new intelligent data indexing mechanisms, currently unavailable in the fintech market, that enables the data sharing between big organisations.

As mentioned, the usage of AI technologies is one of the key points for all the pilots (and so, for the relationship with WP5 outcomes), and this was one of the action points reinforced by most of them during this second stage. The outcomes on this sense are still preliminary, as they are currently evaluating different algorithms to develop accurate models and define their corresponding AI framework. The techniques reported include DeepVaR⁴, Nested Cross-Validation⁵ or Neural Networks approaches to implement specific designed models but, according to reviewers' recommendations, pilots put an extra effort on Explainable AI and bias' avoiding. Aligned with the first, several pilots from different clusters are betting on SHAP (SHapley Additive exPlanations)⁶ frameworks; SHAP provides tools (libraries) to explain the predictions obtained by the developed AI models.

Other relevant activity carried out during this second stage has to do with the dissemination and evaluation of pilots' achievements, by organising dedicated workshops to enrol specific stakeholders from the markets linked to each corresponding pilots or cluster. The feedback collected from them will be relevant to understanding how the solutions must evolve to better adapt to the market expectations. In this sense, the progress within the pilots

⁴ https://marketplace.infinitech-h2020.eu/infi_assets/deepvar-value-at-risk-prediction-leveraging-deep-learning

⁵ <https://www.analyticsvidhya.com/blog/2021/03/a-step-by-step-guide-to-nested-cross-validation/>

⁶ <https://shap.readthedocs.io/en/latest/>

varies, due to, mainly, the difficulties in identifying and reaching the proper stakeholders, depending on a given pilot's scope. While some pilots have already managed to set up different extra activities with their stakeholders (e.g. Pilot #13), some others (e.g. Cluster 4) are still organising the layout for new ones. As regards the general approach, during the first half of 2021 (M16 – M20) there were several workshops, fostered by the INFINITECH coordination team that, combined, involved all active pilots. In October 2021 (M25) the University of Glasgow set up a joint INFINITECH workshop⁷ where all clusters, and so, all pilots were introduced. Most plans for all pilots take into account the organisation of at least one activity involving stakeholders before the end of this year (M27) or a relevant workshop during the first quarter of 2022.

In terms of technical progress, most pilots have already adopted the INFINITECH Way and are deploying their corresponding components (those components agreed) in their assigned testbeds' environments.

⁷ Infinittech Banking services Workshop (sibpages.com)