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

D2.4 – Reference Scenarios and Use Cases – Version II

Revision Number	3.1
Task Reference	T2.2
Lead Beneficiary	FTS
Responsible	Jürgen Neises
Partners	GFT, ATOS, ARI, BS, UBS, SIA, BOI, LIB, NBG, AKTIF, BOS, PI, BPFI, DYN, GEN, JRC, PRIVE, WEA, FI, UNP, RRD, CRE, JSI, CTAG
Deliverable Type	Report (R)
Dissemination Level	Public (PU)
Due Date	2020-09-30
Delivered Date	2020-11-02
Internal Reviewers	ASSEN
Quality Assurance	ENG
Acceptance	WP Leader Accepted and/or Coordinator Accepted
EC Project Officer	Pierre-Paul Sondag
Programme	HORIZON 2020 - ICT-11-2018
	This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no 856632

Contributing Partners

Partner Acronym	**Role**[1]	**Author(s)**[2]
FTS	Lead Beneficiary	Jürgen Neises
ABILAB, AGRO, AKTIF ATOS, BANKIA BOC, BOS, BPFI CTAG, DYN, FI, FTS, FTSG, GEN, GFT, ISPRINT, JRC, JSI, NBG, PI, PRIVE, RRD, SILO, UNP, WEA	Contributors	
ASSEN, ENG	Internal Reviewers, Quality Assurance	

Revision History

Version	Date	Partner(s)	Description
0.1	2020-09-04	FTSG	ToC Version
0.2	2020-10-06	FTSG, ATOS, GFT	1st Draft
0.3	2020-10-21	FTS, AKTIF, ATOS, BANKIA, CP, GFT, JRC, PI, WEA	2nd Draft, Pilot contributions and clustering
0.4	2020-10-27	ABILAB, AGRO, AKTIF ATOS, BANKIA BOC, BOS, BPFI CTAG, DYN, FI, FTS, FTSG, GEN, GFT, ISPRINT, JRC, JSI, NBG, PI, PRIVE, RRD, SILO, UNP, WEA	3rd Draft Pilot contributions and clustering
0.5	2020-10-28	FTSG, WEA, BPFI, NUIG	Final Draft Pilot contributions and clustering Requirements consolidation Conclusion

[1] Lead Beneficiary, Contributor, Internal Reviewer, Quality Assurance

[2] Can be left void

1.0	2020-10-29	ASSEN	First Version for Internal Review
2.0	2020-10-30	ENG	Version for Quality Assurance
3.0	2020-10-31	FTSG, GFT	Version for Submission
3.1	2020-11-01	NUIG, FTSG	Updated Pilot #3 description

Executive Summary

The document provides use cases or **data-based reference scenarios** collected from the INFINITECH pilots. This deliverable is the final version of a total of two deliverables which are meant to provide the outcome of task T2.2.

The goal of task T2.2 "Smart, Autonomous and Personalized Services Specification" is the specification of innovative BigData/IoT based services for the sector, notably smart, autonomous, personalized and regulatory compliant services for most of the business processes of the sector (i.e. services featuring SHARP properties).

This 2nd version of the document describes, which **business services** are used in the pilot scenarios, and the **functional services** (often referred to solely as "services") are required for their fulfilment. The **SHARP properties** facilitated by the services are illustrated. **SHARP** is defined in the context of INFINITECH as **S**mart, **H**olistic, **A**utonomous, **R**egulatory compliant and **P**ersonalized. This way, this deliverable lists the functional requirements of the INFINITECH pilots.

In particular the deliverable contains

- A detailed specification of functional services required for the use cases and reference scenarios provided by the INFINITECH pilots.
- a clustering of the identified functional services alongside the INFINITECH RA.

Overall, the INFINITECH pilots' requirements reflect the State of the Art of the application of BigData, IoT and Artificial Intelligence (AI) in Financial and Insurance Services and contribute to the latest trends.

Table of Content

List of Figures

List of Tables

1 Abbreviations/Acronyms

Abbreviation	Definition
AML	Anti Money Laundering
API	Application Programming Interface
DL	Deep Learning
DoA	Description of Action
EO	Earth Observation
ERC20	Ethereum Request For Comments Standard 20 for fungible tokens
ES	Expected Shortfall
ID	Identity
IoT	Internet of Things
KYB	Know your Business Process
KYC	Know your Customer Process
MiFID	Markets in Financial Instruments Directive
MiFIR	Markets in Financial Instruments and Amending Regulation
ML	Machine Learning
NDA	Non-Disclosure Agreement
NLP	Natural Language Processing
NIS	Network and Information Systems
OES	Operators of Essential Services
PAN	Primary Account Number
PaaS	Platform as a Service
PCI DSS	Payment Card Industry Data Security Standard
PIA	Privacy Impact Assessment
PSD2	Payment Service Directive 2
PSP	Payment Service Provider

P2P	Peer-to-Peer
P2PP	Peer-to-Peer Payment
PSU	Payment Service User
RA	Reference Architecture
RTS	Regulatory Technical Standard
QTSP	Qualified Trust Service Provider
SA	Supervisory Authority
SCA	Strong Customer Authentication
SHARP	**S**mart, **H**olistic, **A**utonomous, **R**egulatory compliant and **P**ersonalized
SME	Small and Medium-Sized Enterprises
SECaaS	Security-as-a- Service
TI	Threat Intelligence
VaR	Value-at-Risk
3DS	Three-Domain Secure

1. Introduction

Task T2.2 provides the specification of innovative BigData/IoT based reference scenarios that underpin **business** services for the finance and insurance sectors. Notably it will describe smart, autonomous, personalized and regulatory compliant services for most of the business processes of the sector (i.e. services that feature **SHARP properties**). Based on existing services and processes (e.g., KYC/KYB, fraud detection, customer service, portfolio management, asset management and usage based insurance in the pilots' scenarios resulting from task T2.1, task T2.2 we identify at a high level how these services could become more autonomous, personalized and context-aware, while simultaneously taking into account the state of the art in BigData/IoT applications for the sectors. Finally, the main **functional services** required for the implementation of the business services in the pilots are defined and clustered along the INFINITECH RA (Reference Architecture) listing the requirements for further implementation.

1.1. Objective of the Deliverable

This, (version 2), deliverable is a revision and update of the specification of the innovative BigData/IoT based functional services involved within INFINITECH, described in the previous version of this deliverable D2.3 It is based upon the requirements resulting from the user stories, (see Task 2.1), and the analysis of the drivers in BigData/IoT applications for the finance and insurance sectors that took place in task T2.2. The deliverable reflects the progress in the evolving pilots and clearer business needs. Due to partner changes, some pilots (i.e. pilot #7 and pilot 15) may need further inspection.

The overall main objectives of this deliverable are to

- elicit the business services, which shall be covered in INFINITECH,
- provide a definition of functional service elements;
- describe the SHARP services which drive the pilots.

The deliverable especially shall provide:

- further insights on the state of the art in the pilots' clusters;
- an updated view on the reference scenarios and functional services;
- communalities of the functional services.

1.2. Insights from other Tasks and Deliverables

The deliverable extends the user stories developed in Task 2.1 with a services view towards a specification of reference scenarios and functional services in the pilots. Therefore, especially the pilot contributions may include some overlap with deliverable D2.3, e.g. in the use case descriptions.

Technical services have been defined in D2.13 / D2.17 on the INFINITECH Reference Architecture. These definitions will be updated in the future version of the INFINITECH Reference Architecture. Technical details of those services will be defined and described in the subsequent work packages WP3, WP4, WP5 and WP6. The application of those technical services in the pilots will be described in WP7.

1.3. Structure of the document

The deliverable is structured into six sections as follows:

- This current Section 1 describes the scope and content of the deliverable.
- In Section 2, we provide a brief overview on the drivers in the application of BigData, IoT and AI in the financial sector, based on analyst reports.
- The mapping of use cases to SHARP services is outlined in Section 3.

- Section 4 describes the pilots' abstraction from user stories to functional services.
- In Section 5, a requirements list along the INFINITECH-RA is described.
- Finally, Section 6 summarises the conclusions of the previous sections.

For easier understanding and better readability, e.g. due to completeness, it may partially contain content of the previous version. This content is marked grey.

2. Drivers for Application of BigData, IoT and AI beyond State of the Art in Financial and Insurance Services

Digitization has been a driver for change in Financial Services for many years from core banking to multichannel banking industry with different types of devices [1]. For several years now, the waves of digitization, financial technology (FinTech) and insurance technology (InsuranceTech) are rapidly transforming the financial services and insurance industries , [3]. For instance, this is illustrated by the rapid growth of FinTech start-ups. McKinsey tracked more than 2000 FinTech start-ups in 2016 with an assumption that there were actually a lot more going undetected [1]. Moreover, FinTech investments have grown from 1.8bn USD in 2011 to more than 30 bn USD in 2018 with a CAGR of ~ 50% p.a. [2].

FinTech is assigned to different views. They are disruptors to the status-quo and can also enhance the position of existing incumbents. . Furthermore, Fintechs are part of large ecosystems, e.g. within the Alibaba platform, and selling infrastructure, e.g. used for open banking [2]. In challenging times or the end of a market cycle that is moving into a downward trend , new measures are required to maintain steady growth. Analysts like McKinsey [4] suggest several levers for organic growth to be explored by banks.

(1) risk management based on powerful analytical tools to prepare for a downturn;
(2) productivity, using modular utilities to materially change cost structures; and
(3) revenue growth through an improved customer experience (CX), bringing a larger customer base and/or share of wallet.

Essential to exploiting these profitability levers are the critical enablers of advanced data analytics and talent. AI shows a great promise in this field especially with the progress in modelling techniques and methods. This will facilitate moving to new data sources e.g. IoT supplementing traditional big data analytics in FinTech.[2] A rapid scaling of advanced analytics and AI tools is a key to successful growth from McKinsey's point of view [4]. For instance machine learning models can improve predictive accuracy in identifying the riskiest potential customers by up-to 35%.

According to Juniper Research [5] "*Technologies such as machine learning and blockchain are having a transformative effect on fintech, fundamentally altering the way financial services are delivered and driving fintech platforms to become the 'new normal'. Such technologies will make new use cases mainstream, including smart contracts, loan underwriting using AI to analyse non-traditional data sources, and personalised insurance policies based on IoT-generated data.*"

Moreover, Regulation and Compliance, e.g. Financial Crime, Money Laundering, Fraud, includes strong opportunities to disrupt as Figure 1 shows.

As a side effect of the digital transformation in Financial Services, the trend towards persistent digital identities is accelerating. Indeed, *"this is due to multiple points of failure in conventional identification and verification processes, particularly for online payment details but also in a variety of other sectors. Passwords and centralised repositories have both been highlighted as the core issue within the growing problem of identity fraud, and a variety of approaches have arisen to combat this."* [6]

	Opportunities to Disrupt	Competition	Addressable Market	Expected Outcome
Banking	Challenger banks offer a better customer experience.	Incumbents improving their digital presence.	Limited appetite among consumers to switch providers.	Incumbents retain their role for providing current/checking accounts for consumers and businesses.
Payments	Appetite among consumers and businesses for quicker and more convenient payment types.	Fierce competition as technology firms enter the fray.	All providers looking to make a land grab; a case of building an installed base as soon as possible.	Technology firms add payments to their ecosystems by providing convenience and have a ready-made installed base of customers.
Lending & Financing	New and novel sources for assessing applicants.	Many new entrants targeting niches or improving the customer experience.	Suppliers keep expanding the market to include those previously excluded from financial services.	Fintech suppliers continue to keep ahead of incumbents by catering to niches.
Insurance	Good fit. Analytics technologies allied with consumers' desire for personalised service.	Incumbents investing heavily in new product areas, coupled with low barriers to entry for new start-ups.	Ever increasing numbers of niche areas to serve.	Insurtech business models become the new normal as incumbents able to replicate the business models of insurtechs.
Wealth Management	Appeal to millennials who are looking for new ways to look after their money.	Crowded market as consumer-orientated banks enter the market.	Unless suppliers can broaden the market by income, it will still be perceived as something for the rich. Given the nature of the application, trust will be critical and lends itself to those with a proven track record.	Standalone providers come under pressure from traditional banks as they invest in this new source of revenues.
Regulation & Compliance	Proven that financial firms cannot keep their house in order.	Technology arms race	Firms will need to grow engagements with accounts rather than skimming the market.	Fintech suppliers evolve into trusted business partners for banks and become the bank's data custodians.

Figure 1 Fintech impact on Financial Services (source: [5])

Within this deliverable, services related to the landscape of technology and data driven innovation of Financial Services are identified and clustered based on the broad use cases and reference scenarios provided by the INFINITECH pilots.

In the following sections the State of the Art and the business drivers for BigData, IoT and AI related to the 5 pilot clusters are described.

2.1. Pilot Cluster #1 - Smart and Reliable Scoring, Risk and Service Assessment

The three Pilots #1, #2, #15, within the Cluster #1 refer to three specific applications: Scoring, Risk and Service Assessment. They all implement ML algorithms to business cases, aiming to solve them with novel approaches. The first two pilots are data-driven and analyse different sources of data by pre-processing and converting such information into viable and effective data sources. The third one includes Natural Language Processing (NLP) technologies and AI algorithms applied to market-driven scenarios.

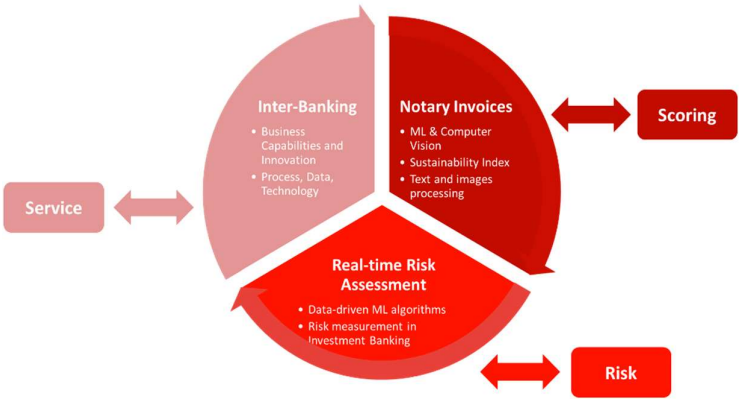

Figure 2 Cluster #1 Pilot applications

Pilot#1 (Invoices Processing Platform for a more Sustainable Banking Industry), better explained in Section 3.1, deals with the extraction of information from Invoices, running ML algorithms on such data to analyse them and compare all the different sources to come up with a Sustainability Index Scoring. The core part is the extraction, analysis and conversion of data intended as text (including tables) and images.

Pilot#2 (Real-time risk assessment in Investment Banking), better explained in Section 3.2, implements a real-time risk assessment and monitoring procedure of two risk metrics (VaR and ES) and market sentiment analysis to estimate market risks and allow updates with changing market prices and/or changes in portfolios in (near) real-time. Moreover, estimated changes in risk measures before a new trading position is entered will be implemented.

Pilot#15 (Open Inter-Banking Pilot), will be developed through a collaborative approach where requirements will be identified with the involvement of bank members of the AI Hub, headed by ABI Lab. More than 10 Italian Banks form such an environment and therefore the pilot is strongly market-driven.

2.2. Pilot Cluster #2 - Personalized Retail and Investment Banking Services

The four Pilots #3, #4, #5 and #6 within the Cluster #2, refer to four specific applications: KYC (Know Your Customer) Data exchange, Wealth management optimization, Business Financial Management and Personalized Investments proposal. The pilots use Open APIs, ML algorithms to business cases, aimed at developing novel ways for wealth, financial and investment management for retail customers and SMEs. All the pilots are using specific data sources for performing the necessary analysis, categorization and data exchange, as well as use additional sources for improving ML algorithms results.

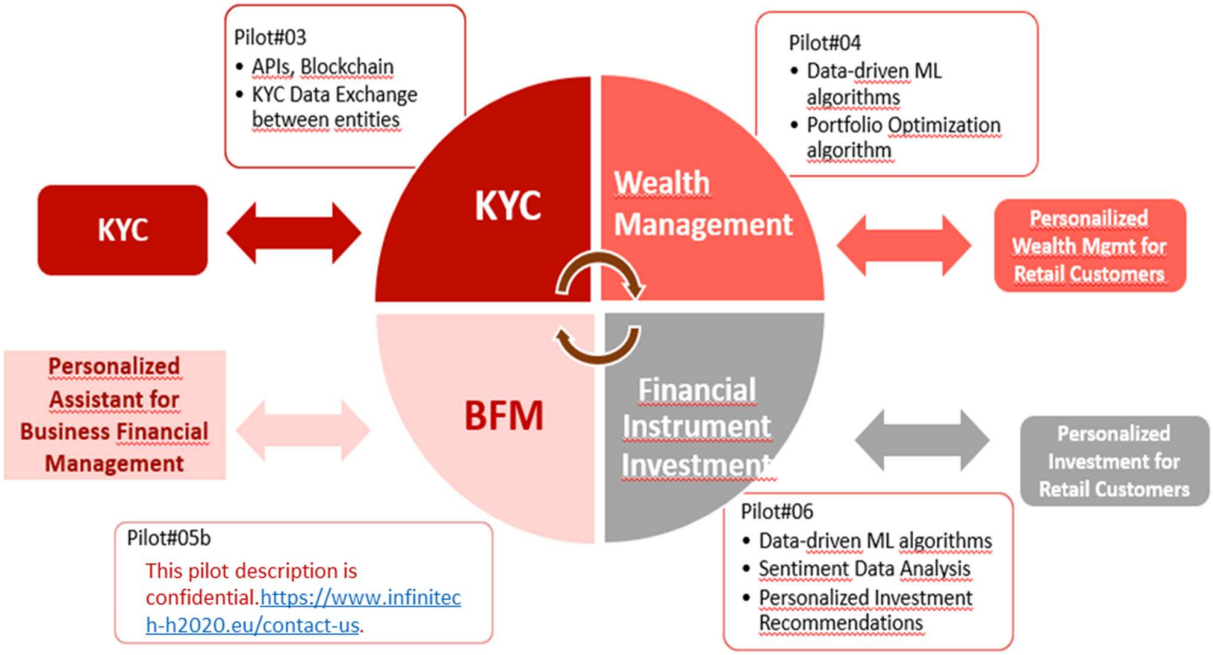

Figure 3 Cluster #2 – Pilots' business services

Pilot #3 (Collaborative Customer-centric Data Analytics for Financial Services), better explained in Section 3.3, aims to mimic a data sharing ecosystem by imitate participants in that ecosystem, providing rules of engagement and highlighting the value exchanges between participants.

Pilot#4 (Personalized Portfolio Management ("Why Private Banking cannot be for everyone?"), described in Section 3.4, will explore the possibilities of AI-Based Portfolio construction for Wealth Management processes, regardless of the amount to be invested (therefore the slogan "Private Banking could be for everyone"). The AI-Based Portfolio Construction will enable advisors and/or end-customers, to use the existing Wealth Management Platform "Prive Managers" and make use of its risk-profiling and investment proposal capabilities, starting from his/her personal risk-awareness.

Pilot#5 (Business Financial Management (BFM) tools delivering a Smart Business Advice): This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

Finally, Pilot#6 (Personalized Closed-Loop Investment Portfolio Management for Retail Customers) aims at using ML algorithms to provide personalized investment recommendations for the retail customers of a bank. This pilot will leverage large customer datasets and large volumes of customer-related alternative data sources (e.g., social media, news feeds, online information). Customer groups will be created, thus ML algorithm can be utilized to provide more targeted investment recommendations to retail customers being context-aware, and more effective. Depending on the state of the market these recommendations even may be tailored and recommendations can be generated automatically.

2.3. Pilot Cluster #3 - Financial Crime and Fraud Detection

The four Pilots #7, #8, #9, #10, within the Cluster #3 refer to four specific applications in the Banking context, relating to financial crimes (#7, #8) and frauds (#9, #10).

Pilot#7 (Avoiding Financial Crime): This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

Pilot#8 (Platform for AML supervision - PAMLS): This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

Pilot#9 (Analysing Blockchain Transaction Graphs for Fraudulent Activities), better explained in Section 3.9, aims to develop and deploy a scalable and high performance blockchain transaction graph analysis system for investigating whether customer blockchain account transactions can be traced to fraudulent activities or accounts. It will provide the following functions: i) Input blockchain transactions, ii) Form and dynamically update transaction graph involving transactions from addresses as well as ERC20/ERC721 token contracts, iii) Provide transaction parallel graph partitioning and graph traversals in order to achieve scalability, iv) Detect addresses and transactions linked to fraudulent addresses.

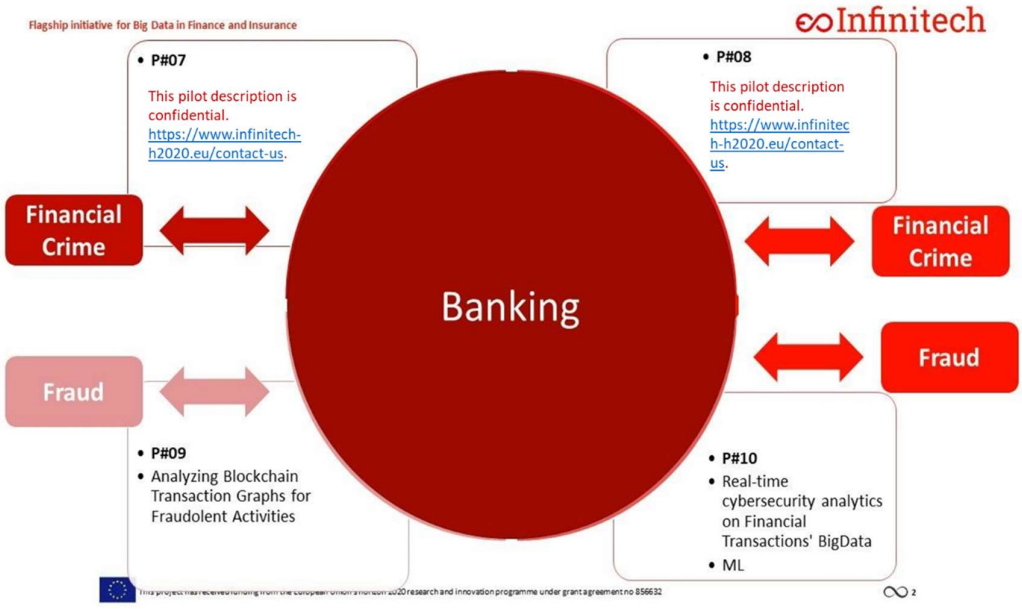

Figure 4 Cluster #3 Pilots' business services

Pilot#10 (Real-time cybersecurity analytics of financial transactions' data), better explained in Section 3.10, aims at enabling the identification of security-related anomalies on transactions while they are occurring, allowing proactive and prompt interventions on such potential threats. To achieve this goal, a sophisticated tool will analyse in real-time the financial transactions of a home and mobile banking system and will use machine learning models, in combination with high-efficiency, traditional analytics techniques, applied on high-volume - real data streams.

2.4. Pilot Cluster #4 - Personalized Usage-Based Insurance Pilots

The idea of customised insurance products adapted to the insured user needs and potential detected risks is the guideline of this cluster. In all cases, different ML/DL technologies will be analysed and put in place to define, develop and train specific AI models. These AI powered models will assist users' classification and risk detection processes, and, in turn, support the services offered to insured parties and insurance companies.

Cluster #4 comprises Pilot #11 and Pilot #12. More detailed descriptions of their objectives and involved user stories can be found in D2.1 and D2.3 whilst technical introductions are shown in D2.5 and D2.13. WP7 deliverables will provide updated overviews and evolution of the cluster.

Pilot#11 is oriented to the car insurance business, relying on rising connected car infrastructures and considering each connected vehicle as an IoT entity. Big Data, HPC and AI techniques and solutions will be applied here to identify and define diverse driving profiles and so, classify real drivers.

Pilot#12 focuses on the health insurance scheme and impact of *Real-World Data* (RWD), captured from users' smart devices (bracelets, smartphones, etc.). These datasets, managed within an eHealth specialised platform, will feed developed AI models to estimate the health-related risks of users and customise new insurance products.

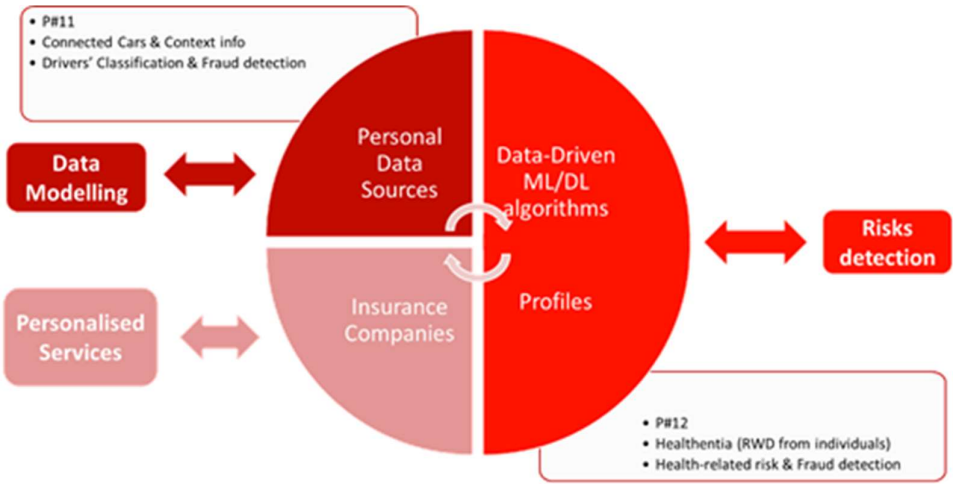

Figure 5 Cluster #4 Pilots' business services

As shown in the figure above, both pilots identify a set of commonalities on its development, particularised to each scenario:

- Data Modelling: vehicles and eHealth registries should be curated, homogenised and modelled, according to a set of common attributes. Infinitech Data Management and Data Processing tools will assist in this process. Both pilots are currently involved in this stage: collecting data from their

corresponding IoT infrastructures; defining the data models to be used; and implementing the data collection and storage processes.
- Data Anonymization: due to the type of data captured in both scenarios, this process should be considered. The same component, developed by Gradiant, will be adapted and integrated within each pilot architecture, implementing different anonymisation processes according to each pilot's requirements.
- Risk Detection: AI powered mechanisms, based on ML/DL technologies, are to be defined and deployed to identify and detect potential risks. IoT collected data will be used within this stage, as well as to train the selected models.
- Customised products: finally, customisation of the services offered to insured users will be their main outcome.

2.5. Pilot Cluster #5 - Configurable and Personalized Insurance Products

The idea of personalised insurance products adapted to the needs of companies, in particular SMEs, and to the possible risks detected is the aim of the first pilot of this group. In all cases, different ML technologies will be analysed and implemented to define, develop and train specific risk mapping models. The development of these models will help to offer personalized products so that underwriting is more efficient and automatic and the processing of claims and the assessment of compensation in agricultural insurance is much more accurate.

Cluster #5 comprises Pilot #13 and Pilot #14. More detailed descriptions of their objectives and involved user stories can be found in D2.1 and D2.3 whilst technical introductions are shown in D2.5 and D2.13. WP7 deliverables will provide updated overviews and evolution of the cluster.

Pilot#13 will implement an automation of the subscription process that helps the insurance company reduce costs. In addition, being able to verify that the data entered is correct with a double verification avoids possible errors in the cost of the insurance premium

The monitoring and identification of real-time risk changes allows the company to know if the insurance cost really corresponds to the real risk of the SME or if it should increase or decrease it to adapt it to its current situation.

Pilot#14 seeks to test specific services for the Agricultural Insurance sector in order to better protect agricultural assets by evaluating risks in a data-driven way:

1. Mapping of risks related to agriculture in predefined markets
2. The prediction and assessment of weather and climate risk probability
3. Damage assessment calculator for insurance companies.

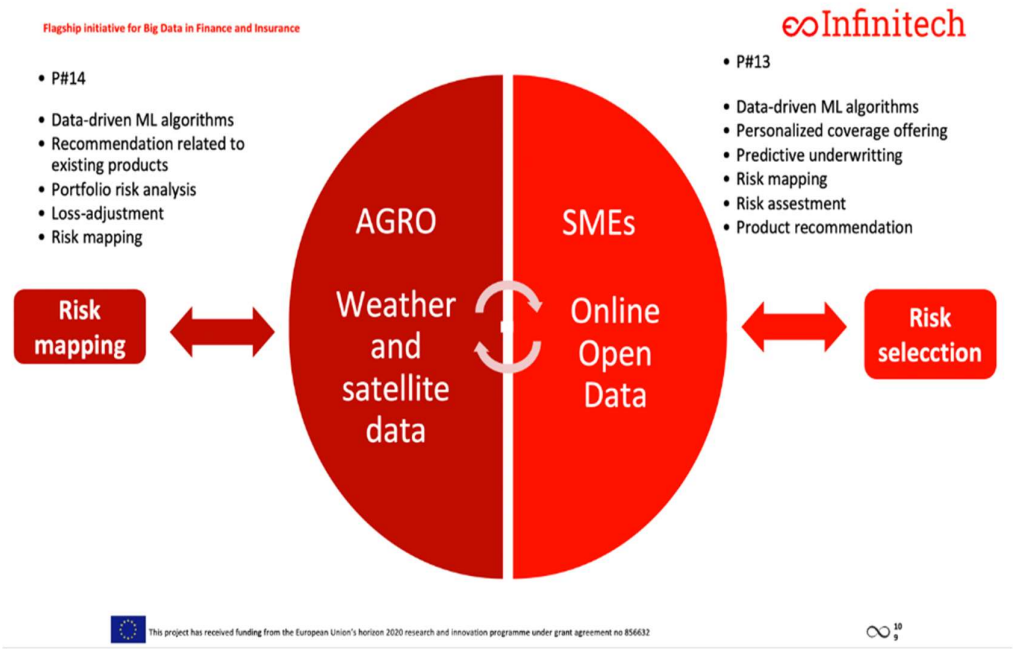

Figure 6 Cluster #5 Pilots' business Services

As shown in the figure above, the two pilots focus on three fundamental factors:

- The mapping of risks for a better knowledge of the client or the risks to be insured
- The incorporation of non-traditional data for the best selection of risks, such as satellite images or company information on their websites
- The increase of efficiency in the analysis of the claims that allows a fairer and faster payment of the compensations to the insured

The two pilots in cluster #5 have their proofs of concept ready and the biggest challenge is the handling of large amounts of data to perform the analysis on their respective platforms.

INFINITECH's contribution to these pilots is to provide them with technologies and platforms that allow them to manage these amounts of data in a more efficient, sustainable and fast way so that they can be served to the final customers.

3. Mapping Use Cases to SHARP Services

In this section, the changes compared to the first version of the deliverable are described. For sake of completeness and better readability sometimes, parts of the 1st version are included. These parts are marked grey.

3.1. Pilot 1

Pilot #1 is titled **"Invoices Processing Platform for a more Sustainable Banking Industry"** and belongs to the **"Smart and Reliable Scoring, Risk and Service Assessment"** category.

Use cases / data-based reference scenarios

Within this pilot digitization of the bank's invoicing process shall improve business sustainability.

Today notarial services for the contracting of several financial products with their customers (as mortgages, etc.) are required. Thus, many physical documents and a large number of redundant copies are generated. Moreover, each physical copy and its control causes significant costs over the period of the financial products lifetime.

However, a shift to digital invoices and subsequent processing using Artificial Intelligence technologies offers both cost-saving opportunities and also increased effectiveness. For example, AI can be leveraged to extract relevant indicators from digitized invoices, which in turn can be used to automatically and accurately rate notaries based on a sustainability index. Finally, contracting services and products will be simplified for clients as they can perform contracting online, while also needing to manage fewer copies.

This use case will be demonstrated by the Bankia Pilot, which will develop, integrate and deploy a data - intensive system to extract Information automatically from notary invoices in order to:

1. Provide to financial institutions the information (properly indexed) about the documents that are finally generated by notarial services required by the bank and the related cost. At present some of the physical copies produced are useless (for the bank and the end customer).
2. Promote notarial services from those with the higher sustainability score.
3. Establish the sustainability index of each notary based on the number of physical copies that are issued.

This use case will try also to discover behaviour patterns and implement statistical data analysis.

Business Service(s)

The pilot is closely related to customer service processes. It provides a business service serving digitized documents in the bank's workflow and a sustainability scoring of notary services. The pilot is aligned with financial processes in the bank that require notary services based on different products like mortgages and loans.

The business service relies on Artificial Intelligence to:

- Identify tables within invoices automatically.
- Extract tables from the invoices automatically.
- Extract relevant indicators from the invoices automatically that will be used to determine the sustainability score.

The pilot will enable digital invoice processing and management in order to achieve further Digitalization of the bank's processes and thereby ensuring a more sustainable business.

In the pilot, an MVP has been developed to test some different AI algorithms models trained with a dataset to process a sample of 150 notaries' invoices and it has successful results in terms of accuracy:

Table detection, OCR converter to extract the tables and the text elements, association of entities to get chargeable concepts, correlation of concepts and prices detecting rows and columns inside the table.

These excellent results lead us to address the following steps to validate and complete the system and extend these models towards a productive environment with a high volume of invoices.

Functional Service(s)
No changes compared to the previous version.

SHARP Properties
No changes compared to the previous version.

3.2. Pilot 2

Pilot #2 is titled **"Real-time risk assessment in Investment Banking"** and belongs to the **"Smart and Reliable Scoring, Risk and Service Assessment"** category.

Use cases / data-based reference scenarios

The scenario described in the first version of the deliverable remains valid, but will be augmented with a sentiment based decision support system for traders and risk managers. While VaR and ES are quantitative risk measures based on numerical price data, the market sentiment will be derived from financial and economic news data and social media channels.

Business Service(s)
In addition to the business services described in the previous deliverable a third service will be introduced that shall deliver an additional risk indicator based on news analysis. The indicator will show positive or negative market sentiment individually for the most important instruments, together with a score indicating the strength of the signal.

Functional Service(s)
Related to the description within the first version of the deliverable the utilization of DL/ML algorithms has changed. These will not be applied for the calculation of the quantitative risk measures (VaR/ES. Instead, these algorithms will be used to build models for sentiment analysis based on news and other textual data streams, followed by a subsequent rule-based creation of signals.
The required functional services are about:

- Seamless data management and querying of streaming data and data at rest.
- Data anonymization and governance.
- Modular, extensible real-time analytics technologies i.e. incremental and parallel analytics.
- web-scraping for data collection.
- text analysis / NLP.
- integration of a sentiment lexicon.
- DL/ML algorithms for text classification.
- data aggregation to time windows.
- Tracking and visualization of the following KPIs:
  - VaR figure with sensitivities;
  - ES figure with sensitivities;
- Pre-trade analysis.

SHARP Properties
No changes compared to the previous version.

3.3. Pilot 3

Pilot #3 is to evaluate how customer, account and transaction data is shared and analysed between banks and Fintechs using APIs to support customer-centric data services.

The Pilot #3 named as "Collaborative Human-centric Data Analytics using Financial Services" is evaluating how customer, account and transaction data is shared and analysed between banks and FinTechs using APIs to support customer-centric data services. We are also conducting additional analysis in relation to the applicability and impact of the financial-related activity to financial-related problems and illegal activities associated to them e.g. Human Traffic, Identity theft, others.

Financial institutions have nowadays access to very large amounts of customer-related data from many different data sources, including both banking systems and alternative data sources (such as open data and social media).

By aggregating, consolidating and sharing such data across institutions/organizations, financial organizations are having now unprecedented opportunities of increasing the quality of their services including automation, accuracy and credibility of customer-centric processes, enhancing KYC/KYB, services personalization, credit risk scoring, unusual financial activities identification etc.

Nevertheless, in order to support the data shared approach in this pilot it is required the design and implementation of solutions for linking and unifying the semantics of diverse data sources. We are using the principle of data sharing and data analytics tools for understanding on how transactions data and profile building capabilities can be used in the context of financial services activity to for example identify profiles that associate in one or another way financial activities to illegal actions. The stop the traffic initiative is an example where financial activities can determine the sources of payments or micro-payments associated to illegal activities, this initiative rely mostly in data sharing and analytics over global transactions data.

In previous version of this document i.e. D2.3 a general description and overview of the use scenarios was briefly described, in this version a more refined characteristics of the pilot and a list of potential use cases and scenarios in relation to the pilot focus area named as Human-Centric Analytics for Personalized Services are indicated, the scenarios have been now mapped into a more particular domain.

Use cases / data-based reference scenarios

The Pilot #3 activity will focus on developing a set of tools using intelligent (AI-based) support in order to support novel human-centric services that automate and solve concrete business problems and processes of the financial institutions. The tools will look at customer services using permission-based mechanisms for data sharing with associated analytics capability. The developed support tools will semantically link diverse data sources (Open Banking Data, Customer Data, Alternative Financial Data) and they will focus in three processes and associated solutions mainly as shortlisted as follow:

(i) Automation/improvement of KYC based on data sharing across institutions;
(ii) Credit Scoring based on data sharing; and
(iii) Anti-Money Laundering (AML) activities through intentionally or unintentionally identifying links to criminal elements associated to financial activities.

To support these processes, the pilot baseline main activity will be based on core activities that applicable independently of the specific use cases, such activities are listed as follow: a) Collaborative Data Sharing, including sharing of data regarding the individuals different roles and activities with and from third parties; for example a person can have different roles in different institutions e.g. owner of an SME, a director role in a Big Corporate. (b) Open Banking KYC APIs for access to datasets; for example at the services level there should be standard ways to access information related to customer, business and business Relations. (c) Evaluation of data sharing propositions with consumers and how these would work within the context of a Fintech ecosystem, for example KYC identifiers will be defined to easily process information related to particular financial process amongst the different services that can be used in the sharing data process; (d)

Methodologies supporting sharing of data directly, and the minimal mandatory and additional security extra controls be applied at behest of customer, for example security and Data Protection methods and protocols will be defined to avoid information thief and allow also data protection.

The following Figure 7 represent the human centric approach and the different roles a person can have in different organisations, this is the associated information that is planned to be use along with the financial transactions on each of those roles to identify particular financial activities.

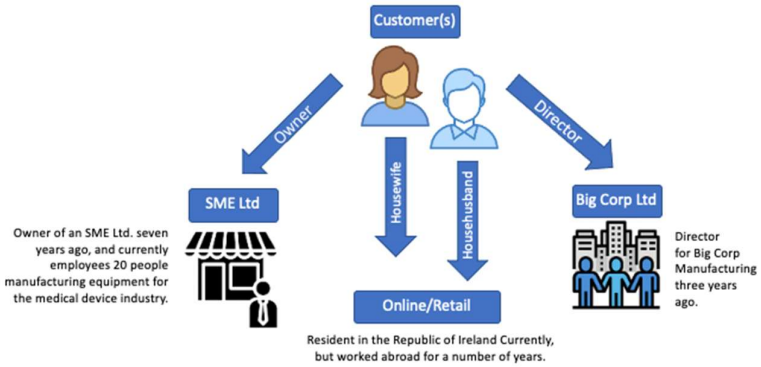

Figure 7 The human-centric approach and the different roles in different organizations

The goal is to provide privacy preserving, secure and trustworthy data sharing and analysis of financialinformation(customer,accountandtransactiondata)betweenbanks,theircustomers and other parties using APIs. This will support customer-centric data services for both data acquisition and collaborative data sharing, producing higher-quality data and also enabling analytics that deliver additional value to both banks and customers. The main factors to consider here are:

- The definition of a common model(s) for Data Sharing proposition, and define the characteristics and forms how these models would work within the context of Fintech eco-systems.
- The establishment and implementation of different data analytics methodologies e.g. ML/DL supporting the data sharing model directly, and the necessary extra security controls to be applied.
- The identification of a framework or a set of tools for banks that can be created to offer data custodian services where customers have control over the management, usage and sharing of their own banking data.

Business Services

The Pilot #3 is using the Capability Model for defining business Services. The capability model is a decomposition of what a business does and can do in a logical and granular grouping. The Capability Model is an integrated and comprehensive set of business activities defined as services. A business capability model is one of the critical business architecture deliverables and a bridge between business and IT, and a foundation for enterprise transformation. A business capability is an elemental building block of what a business does or can do. At its core, it is an abstraction of the underlying functionality and flows expressed as a noun form. A capability model provides a better way to:

- Organize how we think about a business
- Install & track business strategy & performance
- Communicate across disciplines (e.g., business & IT)
- Gather requirements & develop evolution roadmaps

The use of the Capability Model has result in the identification of two types of interconnected Business Services: Type 1) Bank (or Bank partner) data sharing services to support those customer facing services and Type 2) Customer facing data services and associated internal bank processes.

The quality of the customer facing services is impacted by the bank (or bank partner) data services. Project Pilot evaluates how customer, account and transaction data is shared and analysed between banks and FinTechs using APIs to support customer-centric data services.

A holistic view of how these services integrate and interact and best serve the customer and Banks is being investigated during the pilot. Primary service capability being investigated includes:

Type 1. Data sharing support services:

- Consent Management.
- Customer Digital ID.
- Analytics Services.
- P2P Data Sharing Applications – the use of a mobile application for provisioning and sharing service and collection of 'exhaust' graph data is being considered.
- Partner Authentication & Authorisation.
- Customer Authentication & Authorisation.
- API/Open Banking Services.

The following Figure 8 depicts the Entity Model used to enable data sharing capabilities in the pilot 3. Each of the "entityBank" can be associated to different organisations or belong to the same Organisation. Entity model is also named Party Model in the context of developing financial services following the so-called capability model used in Financial Sector.

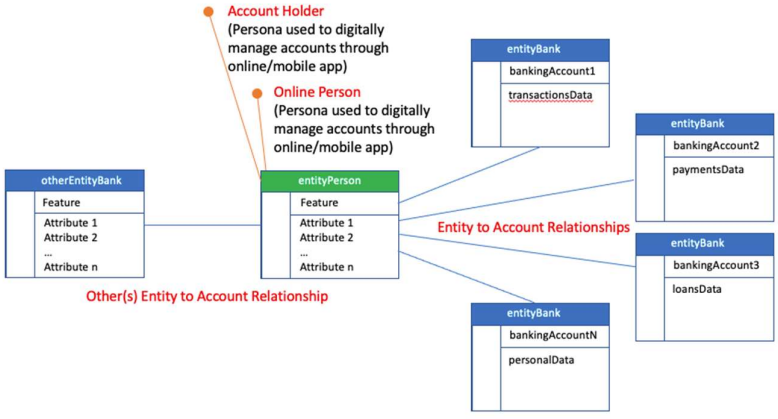

Figure 8 The Data Sharing Model following Data Representation by Entities

Type 2. Customer facing services.

- KYC/KYB.
- AML.
- Credit Scoring.
- Other customer areas where effectiveness and efficiency in data sharing processes may include Fraud detection

The scope of data envisaged is customer, account and transaction, plus potentially documentation, See Simplified Information Model below where different units of information (data) are connected by a blue arrow representing the use of information by different parties for different purposes and how established relationships can be defined in order to identify consequent activities in the banking process.

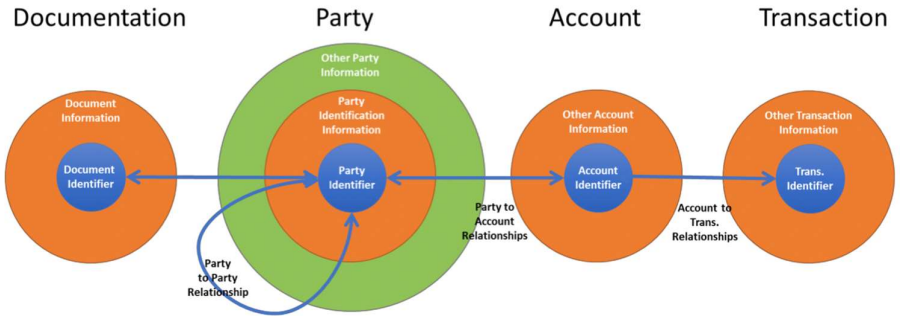

Figure 9 - Pilot #3 Information Model Based on Party (Entity) and Capability Models

Functional Service(s)

The following list is a summary of envisaged functional services in the Pilot #3. The Figure 9 depicts the service Landscape and how those functional services are mapped into the different stakeholders. At the same time a set of sub groups or layers are created to categorize the functional services according to common functions, each sub category include short sentences describing service functions needed for the business service deployment that is planned in the pilot.

Experience – Business and Financial Services Business

- Data Governance Building Blocks, such as eIDAS integration for client on-boarding;
- OpenBanking API enabling data exchange by several financial institutions
- Application development of data sharing and consent management interface s and associated UX (probably mobile)

Services – Government and FinTech

- Semantic interoperability and semantic linking technologies
- Data sharing, including technologies for trading and support for personal data markets.

Technology - Consortium

- Consent Management and Privacy Management technologies.
- Analytics based on large amounts of customer data from various sources or smaller richer datasets including graph data.
- A secured GDPR-compliant infrastructure for information sharing
- Data Processing Reference Architecture

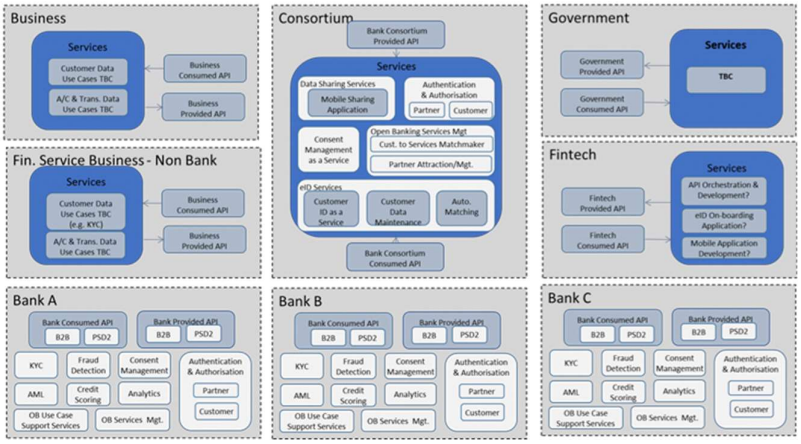

Figure 10 - Service Landscape

SHARP Properties

The identified Smart, Holistic, Autonomous, Regulatory, Personalized (SHARP) features, enabling the use case to have been identified and are included here as. Summary list

- Sharing financial data will enable KYC/ AML and Credit Scoring processes to become more intelligent and automated.
- Customer Profile Building processes as personalisation will aim improving customer experience.
- Holistic customer-centric analytics framework will support intelligent tools for new customer services using permission-based mechanisms for data sharing.

Note: The Smart, Holistic, Autonomy, Personalized and Regulatory Compliance is enabled from the underlying INFINITECH data sharing services. The Figure 11 depicts the three sub groups or layers that categorise the Pilot #3 functional Services, the figure also depicts the use of Platform as a Service (PaaS) a standard approach to identify different functional services and map the sharp properties described above mapping them to the three different PaaS layers.

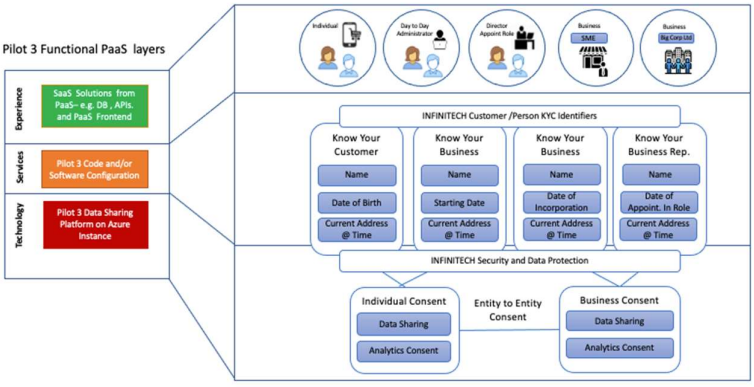

Figure 11 – Pilot #3 Functional Services – PaaS Representation

3.4. Pilot 4

Pilot #4 is titled **"Personalized Portfolio Management – Mechanism for AI based Portfolio Construction"** and belongs to the **"Personalized Retail and Investment Banking Services"** category.

Use cases / data-based reference scenarios

A new use case is assessed: Derive a risk, pricing, predictions on asset tickers based news feeds, in various languages

Business Service(s)
This additional use cases is considered to provide an AI based calculation or risk, pricing and prediction of financial assets, which can be used for portfolio optimization

Functional Service(s)
Newsfeed of selected tickers are gathered. News of portfolio related assets are analysed by AI regarding the impact on risk, pricing and potential prediction.

SHARP Properties
Enhancing the use case should underpin the SHARP features illustrated in the previous version of the deliverable

3.5. Pilot 5

Pilot 5a is no longer applicable. Since the piloting partner has left the project.

Pilot #5 (previously Pilot #5b) is titled "**Business Financial Management (BFM) tools delivering a Smart Business Advice**" and belongs to the "**Personalized Retail/SME and Investment Banking Services**" category.

This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

3.6. Pilot 6

Pilot #6 is titled **"Personalized Closed-Loop Investment Portfolio Management for Retail Customers"** and belongs to the **"Personalized Retail and Investment Banking Services"** category.

Business Service(s)

The pilot #6 business services remain as in the previous version of the deliverable.

Functional Service(s)

The functional services set that implement Pilot #6's scenario have not changed since the previous version of the deliverable.

SHARP Properties

The SHARP properties of Pilot #6 have not changed since D2.3.

3.7. Pilot 7

Pilot #7 is titled **"Avoiding Financial Crime"** and belongs to the **"Financial Crime and Fraud Detection"** category.

This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

3.8. Pilot 8

Pilot #8 is titled **"Platform for Anti Money Laundering Supervision (PAMLS)"** and belongs to the **"Predictive Financial Crime and Fraud Detection"** category.

This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

3.9. Pilot 9

Pilot #9, titled **"Analysing Blockchain Transaction Graphs for Fraudulent Activities"**, is under the **"Financial Crime and Fraud Detection category"**.

The description of Pilot 9 includes additional details on business and functional services and related adaptations in the overall scenarios compared to the 1st version

Use cases / data-based reference scenarios

The pilot is aligned with business processes involving fraud detection on publicly available Blockchain data. Customers as well as businesses may own assets issued on a public Blockchain such as crypto money and tokens representing stable coins (i.e. 1-to-1 fiat money such as USD, EURO or TRY). As a result, as part of the pilot's **reference scenario and use cases**, Blockchain transactions of customers and businesses owning such assets may need to be analysed to see if assets have been acquired using legal means. Assets that originate from sanctioned addresses or those involving fraud (i.e. blacklisted addresses) are not to be cashed out through the customer or the businesses that a bank serves. Since public Blockchains are global, it also provides a medium for money transfers among different jurisdictions. When an analysis is done about Blockchain activities of customers and businesses, a customer service will be available by generating easy to read reports, which represent obtained results in detail. The customers can read these reports and provide feedback if there are errors.

Business Service(s)

A **business service** will be provided that traces Blockchain transactions of the people to see if they originate from or relate to addresses involved in blacklists as a result of fraudulent activities. With this service, people with such activities can be blocked.

Functional Service(s)

The following three functional services will be provided by the pilot:

1. Blockchain Transaction Dataset Preparation
2. Scalable Transaction Graph Analysis
3. User Interface for Blockchain Transaction Reports and Visualization

The "Blockchain Transaction Dataset Preparation" functional service will operate on pre-processed high volume public Blockchain data that comes from multiple chains such as Ethereum (cf. https://ethereum.org/en/) and Bitcoin (cf. https://bitcoin.org/en/). As blocks are generated on the Blockchain, pre-processing tools will extract transactions involving crypto currency as well as ERC20 token transactions periodically. This service will prepare the anonymous massive dataset that will be analysed.

The "Scalable Transaction Graph Analysis" functional service will contain the full historical transaction data that will be loaded, operated on and analysed using parallel machines. Parallel processing of the data will ensure that the service will be scalable and hence sustain its performance as the Blockchain transaction data accumulates to huge volumes. Graph traversal and machine learning algorithms will be used to analyse the data. Parallel graph traversals will be used to extract features that are in the form of subgraphs. The analysis will be performed on massive transaction data that accumulates over time with higher and higher rates in

the future as Blockchain transaction throughputs increase. It will not be assumed that transaction data will fit on one node and hence it will be kept as partitioned data over several files. As a result, the codes developed will be distributed memory model based and will utilize MPI message passing libraries. A cluster with tens of nodes will be used. If accelerators such as GPUs are present, they can also be optionally made use of if the libraries used can make use of it but this will not be a requirement. Graph analysis service will be interacted through a message queue that will take commands in YAML format. The outputs of the service will be in the form of graph paths or subgraphs that show tracing of Blockchain addresses to blacklisted addresses.

The "User Interface for Blockchain Transaction Reports and Visualization" functional service will interact with the "Scalable Transaction Graph Analysis" and present results in a web browser. When subgraphs are returned that trace customer addresses to blacklisted addresses, these subgraphs will be output in vis.vj graph visualization software format for viewing in browsers.

The business service will be provided through a message queue such as RabbitMQ that will take commands in the YAML format. Visualization of transaction graph traces as well as a score that indicate the relationship to fraudulent will also be provided.

SHARP Properties

The system will provide automation in analysis of Blockchain transactions. The system will have holistic features since it will automate processing over the complete Blockchain and provide a whole pipeline that involves extraction of massive number of transactions from the Blockchain, construction of transaction graphs on a parallel HPC cluster and analysis via graph and machine learning algorithms and visualization. Being a parallel system, it will also scale by simply increasing the number of cluster nodes.

Not only parallel graph traversals and ML/DL algorithms but also scalable management of massive data will contribute towards provision of a system with smart features. Pre-processing of raw data coming from multiple Blockchain and token smart contracts will allow extraction, expression and management of transaction data in generic format. This will allow the developed system to operate on generic transaction data and provide support for a wide range of financial applications. For example since ERC20 token transactions will be supported, transactions from financial applications making use of tokenization can be handled by the system.

Even though all the Blockchain transactions will be available to process, tagging (labelling), masking and selection functionalities on transactions will be provided in order to do personalized graph analysis. Addresses stored on the public Blockchain are anonymous. If institutions want to use the system by linking identities, for example by tagging, then they can download the transaction data and open source system to their premises and use the system in a regulatory compliant manner.

3.10. Pilot 10

Pilot #10 is titled **"Real-time cybersecurity analytics on Financial Transactions' BigData"** and belongs to the **"Predictive Financial Crime and Fraud Detection"** category.

Use cases / data-based reference scenarios

The purpose of the pilot is to enable security related anomalies to be identified while they are occurring, if possible, by proactively doing and taking timely action on such potential security threats. This will build a sophisticated tool that will monitor in real time the financial transactions of a domestic and mobile banking system and will use machine-learning models, alongside and in combination with traditional high-efficiency analysis techniques, applied on high-volume - real data flows. In the pilot phase, the system will be tested against its ability to significantly improve the detection of fraud attempts while they are happening or about

to happen. Thus, the pilot will move from the current post-event detection approaches to a new real-time approach that will be based on BigData analytics technologies. It should also be noted that the repeated analysis of transaction flows will allow us to assess cyber security based on dynamic cyber risk assessment metrics.

Business Service(s)
For this pilot's reference scenario the business service to be delivered is related to a precise and fine grain financial fraud analysis and detection.

Such a business service will allow to meet two goals:

- The early detection of new and subtle types of frauds. Since fraudsters keep innovating novel ways to scam people and online systems, it becomes crucial to apply AI/ML methods to detect outliers in large transactional datasets and be robust to changing patterns.
- The reduction of the number of false positives which are usually analysed to understand if they are real fraud attempts or not. To this aim, it is very important to be able to train, validate and test ML models to make the most accurate ones operational.

For both goals ML algorithms and BDA (Big Data Analytics) stream and batch services will be used on data sets made at disposal.

Functional Service(s)
This pilot will require a tool for the design, execution and monitoring of the BDA service workflows to execute real time transaction data analysis. Several aspects will be considered that span from the design of an application, to the execution of this application and the delivery and visualization of the data resulting from the workflow execution by means of queries and graphs, in a meaningful way so as to support the decisions of stakeholders on issues related on the transaction fraud analysis. Thus, a set of parallel/incremental/declarative analytics services leveraging deep learning and big data processing engines (e.g. Spark - https://spark.apache.org/, Flink - https://flink.apache.org/, Storm - https://storm.apache.org/, Samza - http://samza.apache.org/, Tensorflow - https://www.tensorflow.org/), will enable the design and deployment of AI/ML workflows including data preparation/pre-processing, training and inference stages for Faster & Cost-Effective Analytics.
Such a tool will be based on the microservice approach to guarantee that each service, in charge of data processing and analysis, will run in its own process and communicate through lightweight mechanisms (e.g. HTTP resource API). Thus, services will be independently deployable by means of application container management technology adoption (Kubernetes, Docker, Kafka). As for scalable transaction data storage and management, the tool will support a set of cutting edge storage systems that cover the availability of both data streams and batch data sets, and address the variety of data formats and structures (HDFS, HIVE, Redis, Casandra, etc.).

SHARP Properties
The pilot will clearly result in a Smart financial fraud detection system able to proactively and autonomously detect and alleviate cyber-security and financial crime incidents.

3.11. Pilot 11

Pilot #11 is titled **"Personalized insurance products based on IoT connected vehicles"** and belongs to the **"Personalized Usage-Based Insurance Pilots"** category. It is focused on Data based risk assessment and pricing services, including also fraud detection mechanisms.

Use cases / data-based reference scenarios

The Pilot #11 use cases evolved from D2.3 keep the same approach and objectives, linked to the risk assessment in car insurance and supported by AI powered driving profiles. As an improvement in the

scenario description, since the first version of this document (D2.3) specific context data sources, not directly associated with the connected vehicle itself have been analysed and included in the data collection process: current weather information (precipitations, visibility, wind, atmospheric pressure, etc.) and traffic incidents (location, type, impact, etc.). These captured datasets will be used to enhance drivers' classification models.

Business Service(s)
The business services to be developed that support Pilot #11's evolved use case remains the same described in D2.3: Pay-as-you-drive service, assisted by driving profiles and drivers classification; and the Fraud-detection service that analyses traffic incidents. These two services will exploit AI models for driving profiling that will be enhanced by the inclusion of the weather and traffic incidents datasets.

Functional Service(s)
The functional service set that builds the Pilot #11 scenario has not been modified since the previous version. The driving profiling and the drivers' classification functional services that support the Pay-as-you-drive and the Fraud-detection business services rely on top of an AI framework that implements diverse ML & DL technologies to develop and train the models that define them. In turn, this AI Framework is fed by a data storage layer that keeps all the information collected from the connected (and simulated) vehicles, mapped as routes, plus the new included data sources (weather and traffic incidents), as described in D2.13 [7].

SHARP Properties
SHARP properties of Pilot #11 have not changed since the previous version

3.12. Pilot 12

Pilot#12 is titled **"Real World Data for Novel Health-Insurance Products"** and belongs to the **"Personalized Usage-Based Insurance Pilots"** category. It is focused on Data based risk assessment and pricing services, including also fraud detection mechanisms for health insurance.

Use cases / data-based reference scenarios

The pilot 12 use cases remain as in D2.3, improved only in terms of data collected. The exact measurements and reports to be collected are now consolidated, spanning physical activity, sleep, nutrition and liquids, alcohol, disease symptoms and instrumental activities of daily life. Apart from the data collected from real people, the pilot 12 participants, a simulator serving as a secondary source of data is nearly finalised.

Business Service(s)

The pilot 12 business services remain as in the previous version.

Functional Service(s)

The functional services set that implement Pilot 12's scenario have not changed since D2.3. Healthentia, the data collection platform has been modified to support the renewed data collection decisions described above, while decisions have been reached to utilise UBITECH's data collection middleware to fetch the data into our server and LeanXcale to store it.

SHARP Properties

The SHARP properties of Pilot 12 have not changed since the last version.

3.13. Pilot 13

Pilot #13 is titled **"Alternative/automated insurance risk selection- product recommendation for SME"** and belongs to the **"Configurable and Personalized Insurance Products"** category.

Use cases / data-based reference scenarios

The cases of use of the pilot 13 described in the first version of this document remain valid. As an improvement, data cleaning and updating has been included. We have detected that many times in order to build the risk prediction models or their evolution, the insurance companies do not have the updated data so we have added the cleaning and updating of the data as a new use case.

On the other hand, the scope of "coverage" has been changed to "insurance products", the use of coverage and not insurance products, so that it is no longer possible to compare between insurance companies

Business Service(s)

Customized risk profiles of each of the companies analysed will be generated, allowing not only to customize the product offer, permanent automated risk management, but also the dynamic individualized adaptation of coverage and pricing according to risk and market.

Functional Service(s)

The set of functional services for Pilot#13 has not been modified. The functioning is still the extraction of information in different open sources of SMEs, its dumping into a database for once there apply algorithms and techniques of ML

The result is a risk profile, its changes over time and the recommendation of insurance products suitable for your profile

In this regard, we should remember that, in accordance with D2.7 (compliance), no personal data are processed and therefore there has been no change

SHARP Properties

SHARP properties of Pilot#11 have not changed since version I.

3.14. Pilot 14

The pilot **"Configurable and Personalized Insurance Products for SMEs and Agro-Insurance''** is generally aligned with the business process of usage-based insurance (UBI).

Business Service(s)

The pilot 14 business services remain as in D2.3.

Functional Service(s)

The functional services set that implement Pilot 14's scenario have not changed since D2.3.

SHARP Properties

The SHARP properties of Pilot 14 have not changed since D2.3.

3.15. Pilot 15

Pilot #15 is titled **"Open Inter-Banking Pilot"** and belongs to the **"Smart and Reliable Scoring, Risk and Service Assessment"** category.

Use cases / data-based reference scenarios

Developed through a shared approach, the pilot is focused on service assessment application.

Due to its composition, the pilot is strongly market-driven and aims to implement the prototype of a solution based on Machine Learning and Natural Language Understanding paradigms.

This prototype will start from the analysis of a subset of process operating documents to attempt the classification of the information contained in them with respect to the ABI Lab taxonomy, used by Italian banks to build their business glossary and in general to support the Enterprise Architecture Modeling.

The technical and development aspects, in particular within the dedicated testbed, will be supported by GFT and HPE.

Business Service(s)

This pilot will allow the screening of extensive documentation in real time. This will be a starting point for the optimization of solutions that every single bank can possibly adopt and adapt in their own context.

The pilot will point out the tagging of documentation expressed in natural language and its classification in accordance with related taxonomies.

Functional Service(s)

The set of functional services for Pilot#15 is in progress. The general functioning is based on the extraction of data from internal processes in the area of data governance.

A high level view of the functional architecture is described below:

- A data storage layer, including tools and infrastructures aimed at data collection from different sources and in different formats, and their storage;
- A data ingestion/preparation layer, including technical components aimed at normalising and aggregating the data that we need for our specific analytical purposes, preparing the information to be processed by the Machine Learning tools;
- A machine learning engine layer, including Natural Language Understanding algorithms, opportunely configured for the use case purposes.

SHARP Properties

Increased Automation and Intelligence based on data processing leveraging data governance processes.

4. Service Communalities

In this section, the functional services are clustered along the layers of the BDVA [8] and the INFINITECH Reference Architecture [7]. As far as possible communalities are identified facilitating exchange of components and synergies in the project.

4.1. Pilot Cluster #1 – Smart and Reliable Scoring, Risk and Service Assessment

Table 1 Cluster #1 functional services

Functional Services	Cluster
All: Data Result Visualization, Document Digitization , Document Validator, Open Source Dataset	Pre-processing
Communalities: n.a.	
All: Invoices Database, Table Bank Dataset, News Article Data, Real-time market data, Electronic order platform, Historical data	Infrastructure
Communalities: n.a.	
All: Text extraction, Data windows, Portfolio input-stream, Data collector, Files (images, text, tables) extraction, pre-processing and converter	Data Management
Communalities: Data collector, Text extraction	
All: Anonymizer	Data Protection
Communalities: Anonymizer	
All: JSON data, data extraction, correlation matrix, scenario specifications, scenario generations	Data Processing and Architecture
Communalities: data extraction	
All: concepts price extractor, market sentiment extraction, analytics, configuration	Analytics
Communalities: n.a.	
All: JSON data, User interface	User Interaction
Communalities: n.a.	
All: License server, configuration, digital onboarding, authentication authorization accounting, process monitoring	Cross Cutting
Communalities: n.a.	

4.2. Pilot Cluster #2 - Personalized Retail and Investment Banking Services

Table 2 Cluster #2 functional services

Functional Services	Cluster
All: Bank Database extract, ERP Systems extract, KYC Data extract	Pre-processing
Communalities: Bank Database extract, Investment History extract	
All: Customer Data, Customer Transactions, Customer Historical Data, SMEs ERP data, Asset & Investment profile , Sentiment Analysis	Infrastructure
Communalities: Customer Data, Customer Transactions, Asset & Investment profile, Sentiment Analysis	
All: Data Cleansing, Data Enrichment, Harmonization, Join	Data Management
Communalities: Data Cleansing, Data Enrichment, Harmonization, Join	
All: Anomymizer	Data Protection
Communalities: Anonymizer	
All: JSON data, data extraction, Customer Profiling, Data Streaming, DBM, HTAP, Incremental Analytics,	Data Processing and Architecture
Communalities: data extraction, Customer Profiling, Data Streaming, Incremental Analytics	
All: Cash Flow Prediction, Budget Prediction, KPI Engine, Personalized Recommendation, Financial Instrument Sentiment	Analytics
Communalities: KPI Engine, Personalized Recommendation, Financial Instrument Sentiment	
All: JSON data, User interface	User Interaction
Communalities: n.a.	
All: License server, configuration, digital onboarding, authentication authorization accounting, process monitoring	Cross Cutting
Communalities: n.a.	

4.3. Pilot Cluster #3 - Predictive Financial Crime and Fraud Detection

Table 3 Cluster #3 functional services

Functional Services	Cluster
Blockchain Transaction Dataset preparation, Data preparation/pre-processing, Data Integration, Data preparation (enrichment)	Pre-processing
Communalities: n.a.	
SEPA, TARGET, FIU tr, High Risk countries, Bitcoin Transactions, Ethereum Transactions, Blacklist Addresses, Internal Transfer, Foreign Bank Transfer, Bank Transfer	Infrastructure
Communalities: n.a.	
Design, execution and monitoring of the BDA service workflows, Access transactions of FI/pattern according to criteria defined by human, Access transactions of BOS / FI / other Supervisory Authorities, Data ingestion (using pre-defined API), Filter, Join	Data Management
Communalities: n.a.	
Pseudo-anonymization	Data Protection
Communalities: anonymizer	
Behavioural analysis to update the customer risk profile, Near real-time data access of transactions and customer related data sources, Advanced customer risk model for risk scoring, Access transactions near real-time, Data extraction, Internal Storage, Joined Datasets	Data Processing and Architecture
Communalities: Data extraction	
Scalable Transaction Graph Analysis, Real-time transaction data analysis, Parallel/ incremental/ declarative analytics, ML/DL analytics to detect suspicious transactions, Risk Calculation, Feature Generation, Data representation, Data enrichment, Random Forest Prediction, Random Forest Model, Clustering, OneHotEncoder	Analytics
Communalities: n.a.	
User Interface for Blockchain Transaction Reports and Visualization, Visualization of the data resulting from the workflow execution by means of queries and graphs, Data visualisation / event generation, Visualization of search results, Risk assessment, Manual Risk Decision	User Interaction
Communalities: Data visualisation	

Identity Management System, Role Management System, Message Broker, Resource Management	Cross Cutting
Communalities: n.a.	

4.4. Pilot Cluster #4 - Personalized Usage-Based Insurance Pilots

Table 4 Cluster #4 functional services

Functional Services	Cluster
All: Technical connected vehicles (real assets) datasets; weather information; Traffic incidents; Urban traffic simulation; insurance profiles; activity data from users	Pre-processing
Communalities: Insurance datasets and (potential) insured records	
All: Connected Vehicles (CTAG); Traffic Simulator (SUMO); Weather Stations (AEMET); Insurance Databases (Car and Health); Helthentia Platform; Subject's records from Insurance DB	Infrastructure
Communalities: Insurance Databases (Car and Health); Synthetic Data	
All: Data Collectors/Ingestors	Data Management
Communalities: Data Collectors/Ingestors	
All: Anonymizer; Access Control; Regulatory Compliance Tools; Data Normalization	Data Protection
Communalities: Anonymizer	
All: Data Adaptors (modelling); Context Broker; Time Series (Historical datasets); Data Storage and Data Replicas	Data Processing and Architecture
Communalities: Data Storage and Data Replicas	
All: AI/ML Algorithms and Modelling; RT Analytics; Training tools; Subjects Profile Modelling; Subject Classifier	Analytics
Communalities: AI/ML Algorithms and Modelling: Subjects/Driving Profile Modelling; Subject/Driver Classifier	
All: Inferencers: Driving Profiling service, Driver/Subject Classifier service; Fraud Detection Service; Risk Assessment Service; Pay as You Drive Service	User Interaction
Communalities: Risk Assessment /Driver Classifier Service; Fraud Detection Service	
All: Regulatory Tools	Cross Cutting

Communalities: none	

4.5. Pilot Cluster #5 - Configurable and Personalized Insurance Products for SMEs and Agro-Insurance

Table 5 Cluster # Functional Services

Functional Services	Cluster
All: Data Result Visualization, Satellite, weather prediction , Open Source Dataset	Pre-processing
Communalities: n.a.	
All: Social media data, website data, Business directories data base, Climate data, Earth observation, weather predictions	Infrastructure
Communalities: n.a.	
All: Data aggregator, Data Manager	Data Management
Communalities: Data aggregator	
All: n.a.	Data Protection
Communalities: n.a.	
All: JSON data, data extraction, correlation matrix, scenario specifications, scenario generations, polyglot, geospatial DB, layer filters, SQL, Data manager	Data Processing and Architecture
Communalities: data extraction	
All: Algorithm, IA formulas, Semantic, Flash search, larabel API, WMS server	Analytics
Communalities: n.a.	
All: Analitycal tool, User interface	User Interaction
Communalities: User interface	
All: Role management, IAM Auth, acess control, APi rest, policy manager	Cross Cutting

5. Requirements Specification

The functional services defined in the previous section define functions required by the INFINITECH pilots. Thus, they specify the initial set of functional requirements. Within this section, the **functional** services defined per pilot are listed and clustered related to pilot groups and the INFINITEC-RA.

5.1. Requirements List

Requirements shall be classified in a regular scheme across the WP2. This scheme did not change and is listed for completeness.

- A requirements reference of type REQ-[P#]-[FN]-[GBTPR]-[MOPF]-[INCSEO]-Number
- Pilot number P#
- Functional / non-Functional [FN] –
  Functional: a functional requirement describes the behaviour of the system as it relates to the system's functionality.
  Non-functional: a non-functional requirement elaborates a performance characteristic of the system.
- Type: Type of Requirements chosen from [GBTPR]:
  General [G], Business [B], Technological [T], Physical [P], Regulatory [R]
- "mandatory" [MOPF] -
  Mandatory [M], Optional [O], Performance [P], Future [F]
- Category [INCSEO] -
  Interface [I], Networking [N], Capacity [C], Security [S], Ethical [E], Other [O]
- 3 digit requirement number #
- Title - Title of the requirement
- Description - Full detailed description of the requirement
- Affects (Reference, Architecture, other Tasks] - Block(s) of the RA or other Tasks/Elements affected by the requirement.
- Qualitative Criteria
- Quantitative Criteria

The updated requirements are listed in Appendix B

6. Conclusions

This deliverable is the final version of a total of deliverables D2.3 and D2.19 which are meant to provide the outcome of task T2.2.

In this task, latest business trends and drivers for BigData/IoT and AI application in the INFINITECH use cases have been elicited. This way **innovative services for the financial sector were identified and specified** in detail for each pilot cluster. Overall, the INFINITECH pilots' requirements reflect the business drivers beyond the State of the Art of the application of BigData, IoT and AI in Financial and Insurance Services and contribute to the latest trends and business requirements of the sector.

For each INFINITECH pilot in these clusters, use cases or **data-based reference scenarios were** collected. The **business services** used in the pilot scenarios are described and the **functional services** (often referred to solely as "services"), which facilitate the provisioning of the business services. The **SHARP properties** provided by the business services are illustrated.

Moreover, the functional services are assigned to the categories of the INFINITECH reference architecture and communalities within the pilot clusters are identified, which enable future synergies in the project.

Appendix A: Literature

[1] Miklos Dietz et. Al., “Cutting through the noise around financial technology”, McKinsey & Co, February 2016

[2] Jeff Galvin et. Al., “Synergy and disruption: Ten trends shaping Fintech”, McKinsey & CO, December 2018

[3] PwC, “Redrawing the lines: FinTech’s growing influence on Financial Services”, Global FinTech Report 2017, available at; https://www.pwc.com/jg/en/publications/pwc-global-fintech-report-17.3.17-final.pdf

[4] Chira Barua et. Al., “The last pit stop? Time for bold late-cycle moves - McKinsey Global Banking Annual Review”, McKinsey & Co, (2019)

[5] Michael Lamer: “The Future of FINTECH ~ The new Standard”, Juniper Research, May 2019

[6] James Moar: “Three Trends Accelerating the Growth of Digital Identity”, Juniper Research, July 2019

[7] INFINITECH: Deliverable D2.13 “INFINITECH Reference Architecture - I”, 07.2020

[8] BDVA: “BDV SRIA – European Big Data Value Strategic Research and Innovation Agenda”, Version 4.0, October 2017

Appendix B: Requirements List

The requirements list of the previous version is updated.

No longer valid requirements are marked ~~crossed out~~.

New requirements are marked red.

Requirements of Pilot 5 are still listed as requirements of Pilot 5b for simplicity of documentation.

Ref.	Title
REQ-[P#]- [FN]-[GBTPR]-[MOPF]-[INCSEO]-#	Title of the requirement
REQ-P1-F-B-M-O-001	Digitized documents will be generated by the bank's processes.
REQ-P1-F-B-M-O-002	A data lake will facilitate the secure storage of those digitized documents.
REQ-P1-F-B-M-O-003	Pre-processing will include privacy related measures with respect to the GDPR.
REQ-P1-F-B-M-O-004	Document analysis includes identification and extraction of tables from invoices, e.g. by a machine learning system.
REQ-P1-F-B-M-O-005	Extracted tables shall be analysed by regarding the included information.
REQ-P1-F-B-M-O-006	The extracted tables and information will be utilized to establish a sustainability scoring of the notary service.
REQ-P1-F-B-M-O-007	A visual console shows the results of the document processing and the sustainability score.
REQ-P2-F-B-M-O-008	Querying of streaming data and data at rest
REQ-P2-F-B-M-O-009	Data Anonymization and Governance
REQ-P2-F-B-M-O-010	Modular, extensible real-Time Analytics Technologies i.e. incremental and parallel analytics
REQ-P2-F-B-M-O-011	Tracking and visualization of: VaR figure with sensitivities; ES figure with sensitivities; Pre-trade analysis
REQ-P2-F-B-M-O-100	Web-scraping for data collection
REQ-P2-F-B-M-O-101	Text analysis / NLP
REQ-P2-F-B-M-O-102	Integration of a sentiment lexicon
REQ-P2-F-B-M-O-103	DL/ML algorithms for text classification
REQ-P2-F-B-M-O-104	Data aggregation to time windows
~~REQ-P3-F-B-M-O-012~~	~~Semantic interoperability and semantic linking technologies~~
~~REQ-P3-F-B-M-O-013~~	~~OpenBanking API~~
~~REQ-P3-F-B-M-O-014~~	~~Blockchain based data sharing, including technologies for trading and personal data markets~~
~~REQ-P3-F-B-M-O-015~~	~~AI analytics based on large amounts of customer data from various sources or smaller richer datasets including graph data.~~
~~REQ-P3-F-B-M-O-016~~	~~Data Governance Building Blocks, such as eIDAS integration for client on-boarding;~~
~~REQ-P3-F-B-M-O-017~~	~~Consent Management and Privacy Management technologies.~~
~~REQ-P3-F-B-M-O-018~~	~~Application development of data sharing and consent management interface and UX (probably mobile)~~
~~REQ-P3-F-B-M-O-019~~	~~A secured GDPR-compliant infrastructure for information sharing~~
~~REQ-P3-F-B-M-O-020~~	~~Data Reference Architecture~~
REQ-P4-F-B-M-O-021	Seamless Data Management and Querying
REQ-P4-F-B-M-O-022	Data Governance Building Blocks
REQ-P4-F-B-M-O-023	ML/DL Algorithms (in addition to AI capabilities of the PRIVE platform)

REQ-P4-F-B-M-O-024	Visualization of results for end-customers
REQ-P4-F-B-M-O-105	Gather selected newsfeeds on assets
REQ-P4-F-B-M-O-106	ML/DL Algorithm analysing newsfeed impact of risk and pricing
~~REQ-P5a-F-B-M-O-025~~	This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.
~~REQ-P5a-F-B-M-O-026~~	
~~REQ-P5a-F-B-M-O-027~~	
~~REQ-P5a-F-B-M-O-028~~	
~~REQ-P5a-F-B-M-O-029~~	
~~REQ-P5a-F-B-M-O-030~~	
~~REQ-P5a-F-B-M-O-031~~	
~~REQ-P5a-F-B-M-O-032~~	
~~REQ-P5a-F-B-M-O-033~~	
REQ-P5b-F-B-M-O-034	This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.
REQ-P5b-F-B-M-O-035	
REQ-P5b-F-B-M-O-036	
REQ-P5b-F-B-M-O-037	
REQ-P5b-F-B-M-O-038	
REQ-P5b-F-B-M-O-039	
REQ-P5b-F-B-M-O-040	
REQ-P5b-F-B-M-O-041	
REQ-P5b-F-B-M-O-042	
REQ-P6-F-B-M-O-043	Acquire data from various data-stores of the bank
REQ-P6-F-B-M-O-044	Data from alternative data sources
REQ-P6-F-B-M-O-045	Monitor the performance of the suggested and actually purchased portfolios
REQ-P6-F-B-M-O-046	Customer profiling under various criteria, such as his risk profile and appetite, demographics and behaviour.
REQ-P6-F-B-M-O-047	Present predefined investment portfolios tailored to the customer's profile, with a probability of acceptance from his side
REQ-P7-F-B-M-O-048	This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.
REQ-P7-F-B-M-O-049	
REQ-P7-F-B-M-O-050	
REQ-P7-F-B-M-O-051	
REQ-P7-F-B-M-O-052	
REQ-P7-F-B-M-O-053	
REQ-P7-F-B-M-O-054	
REQ-P7-F-B-M-O-055	
REQ-P8-F-B-M-O-056	This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.
REQ-P8-F-B-M-O-057	
REQ-P8-F-B-M-O-058	
REQ-P8-F-B-M-O-059	
REQ-P8-F-B-M-O-060	
REQ-P8-F-B-M-O-107	
REQ-P8-F-B-M-O-108	
REQ-P8-F-B-M-O-109	
REQ-P8-F-B-M-O-110	

REQ-P8-F-B-M-O-111	
REQ-P9-F-B-M-O-061	Access blockchain transactions of FI/pattern according to criteria defined by human
REQ-P9-F-B-M-O-062	Data Integration / Data protection (e.g. GDPR related)
REQ-P9-F-B-M-O-063	Blockchain Data preparation – gathering search results
~~REQ-P9-F-B-M-O-064~~	~~Visualization of search results~~
REQ-P9-F-B-M-O-112	Scalable Transaction Graph Analysis
REQ-P9-F-B-M-O-113	User Interface for Blockchain Transaction Reports and Visualization
REQ-P10-F-B-M-O-065	Access transactions
REQ-P10-F-B-M-O-066	Data Integration / Data protection (e.g. GDPR related)
REQ-P10-F-B-M-O-067	Data preparation (enrichment)
REQ-P10-F-B-M-O-068	ML/DL analytics (what do you analyse to detect risks?)
REQ-P10-F-B-M-O-069	Data visualisation / event generation (what happens, if a risk is detected by the tool?)
REQ-P10-F-B-M-O-070	Access transactions of BOS / FI / other Supervisory Authorities
REQ-P10-F-B-M-O-071	Data Integration / Data protection (e.g. GDPR related)
REQ-P10-F-B-M-O-072	Extract periodically transaction data from high volume Blockchains
REQ-P10-F-B-M-O-073	Analyse Blockchain transaction graphs with ML/DL
REQ-P10-F-B-M-O-074	Acquire real time transaction data
REQ-P10-F-B-M-O-075	Transaction fraud analysis using parallel/incremental/declarative analytics leveraging deep learning and big data processing engines
REQ-P10-F-B-M-O-076	Visualization of the data resulting from the workflow execution by means of queries and graphs, in a meaningful way
REQ-P10-F-B-M-O-114	Tool for the design, execution and monitoring of the BDA service workflows
REQ-P10-F-B-M-O-115	Set of parallel/incremental/declarative analytics services
REQ-P10-F-B-M-O-116	Scalable transaction data storage and management
REQ-P11-F-B-M-O-077	Collect real time driving data
REQ-P11-F-B-M-O-078	Collect external traffic information
REQ-P11-F-B-M-O-079	Collect other external sources
REQ-P11-F-B-M-O-080	Data anonymization
REQ-P11-F-B-M-O-081	Data Storage Framework
REQ-P11-F-B-M-O-082	Driver Profiling
REQ-P12-F-B-M-O-083	IoT devices (activity trackers) will be collected by the appropriately modified eClinical platform “Healthentia”
REQ-P12-F-B-M-O-084	Automated data privacy risk assessment and mitigation.
REQ-P12-F-B-M-O-085	AI based (ML/DL) algorithms will be applied on the collected data in order to extract the final risk model.
REQ-P13-F-B-M-O-086	Collection of large amounts of information from open sources and alternatives that are typically used by the insurance industry
REQ-P13-F-B-M-O-087	Parallelization of Stream Engines;
REQ-P13-F-B-M-O-088	High-Performance Analytics (Parallel & Incremental Algorithms); Declarative Analytics ML/DL Algorithms.
REQ-P13-F-B-M-O-089	Wenalyze big data analytics platform
REQ-P14-F-B-M-O-090	Estimation of a suite of crop-specific agroclimatic indicators in python, all of them computed on historical (reanalysis data ERA-5, ERA-5 Land, UERA) and seasonal climate forecasts from the multi-model ensemble system of Copernicus Climate Change Service.
REQ-P14-F-B-M-O-091	Tools (AI/ML libraries) to post-process these indicators

REQ-P14-F-B-M-O-092	Other data sources (e.g. past calamities data)
REQ-P14-F-B-M-O-093	Combine data sources
REQ-P14-F-B-M-O-094	More accurate and specific impact models, for each adversarial weather event.
REQ-P14-F-B-M-O-095	Temporal change detection methodologies in multispectral and SAR time-series satellite images
REQ-P14-F-B-M-O-096	Identify the area that impacted by hail, floods, drought and wildfires, as well as the severity of each peril.
REQ-P14-F-B-M-O-097	Tools (AI/ML libraries) to extend and enhance the capabilities of this service, by translating the observed damage to actual yield and money losses.
REQ-P14-F-B-M-O-098	Crop and farmer specific crop loss assessment tool.
REQ-P14-F-B-M-O-099	Computational and storage resources, to host the operation of the system to cover the needs for weather data
REQ-P15-F-B-M-O-117	Data storage layer, including tools and infrastructures aimed at data collection from different sources and in different formats, and their storage
REQ-P15-F-B-M-O-118	A data ingestion/preparation layer, including technical components aimed at normalising and aggregating the data that we need for our specific analytical purposes, preparing the information to be processed by the Machine Learning tools