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

Infinitech

D2.3 – Reference Scenarios and Use Cases – Version I

Lead Beneficiary	**FTS**
Due Date	2020-03-31
Delivered Date	2020-03-31
Revision Number	1.0
Dissemination Level	Public (PU)
Type	Report (R)
Document Status	Release
Review Status	Internally Reviewed and Quality Assurance Reviewed
Document Acceptance	WP Leader Accepted and Coordinator Accepted
EC Project Officer	Pierre-Paul Sondag

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

Acronym	Role[1]	Name Surname[2]
FTS	Lead Beneficiary	Jürgen Neises
AGRO, AKTIF, Atos, BANKIA, BOC, BOI, BOS, BPFI, BS, ENG, GFT, iSPRINT, JRC, JSI, LIB, NBG, PI, WEA	Contributors	
GLA, LXS	Internal Reviewers, Quality Assurance	

Revision History

Version	Date	Partner(s)	Description
0.1	2020-01-15	FTS	ToC Version
0.2	2020-02-26	FTS	Reviewed ToC, draft sections, request for contributions per partner/pilot
0.3	2020-03-09	FTS, AGRO, AKTIF, Atos, BANKIA, BOC, BOI, BOS, BPFI, BS, ENG, GFT, iSPRINT, JRC, JSI, LIB, NBG, PI, WEA	1st draft, Approach, business drivers and 1st pilot contributions
0.3	2020-03-17	FTS, AGRO, AKTIF, Atos, BANKIA, BOC, BOI, BOS, BPFI, BS, ENG, GFT, iSPRINT, JRC, JSI, LIB, NBG, PI, WEA	Introduction, 2nd pilot contributions
0.4	2020-03-23	FTS	Requirements list and classification based on T2.1, T2.3 and pilot contributions
0.5	2020-03-24	FTS	Document cleansing
0.6	2020-03-27	GLA, LXS	1st internal review
0.7	2020-03-29	FTS	Editing comments
0.8	2020-03-30	GLA, LXS	2nd internal review
0.9	20-03-31	FTS	Editing comments

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

[2] Can be left void

1.0	2020-03-31	FTS	Version for Submission

Executive Summary

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).

The version of the document provides use cases or **data-based reference scenarios** collected from all INFINITECH pilots. This deliverable is the first version of a total of two deliverables which are meant to provide the outcome of task T2.2. This version of the document describes, which **business services** are used in those scenarios, and the **functional services** (often referred to solely as "services") are required for their fulfilment. The SHARP properties facilitated by the services are outlined. This way, this deliverable provides an initial list of functional requirements of the INFINITECH pilots.

In particular the deliverable contains

- a first definition 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 BDV reference model.

Overall, the INFINITECH pilot requirements reflect the State of the Art of the application of Big Data, IoT and AI in the Financial and Insurance Services and contribute to the latest trends.

Based on the outcomes of this deliverable, in the next steps of task T2.2

- the state of the art in the pilots' reference scenarios will be elicited
- communalities of the services will be identified reducing the set of functional services
- technical details of the services will be specified

The work related to task T2.2 will continue until Month 12, when the 2nd version of this deliverable will be submitted (D2.4), with the updates on these next steps.

Table of Contents

List of Figures

List of Tables

Abbreviations

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
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
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
PSU	Payment Service User
P2P	Peer-to-Peer
P2PP	Peer-to-Peer Payment
RTS	Regulatory Technical Standard
QTSP	Qualified Trust Service Provider
SCA	Strong Customer Authentication
SME	Small and Medium-Sized Enterprises
SA	Supervisory Authority
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 services for the sector, notably smart, autonomous, personalized and regulatory compliant services for most of the business processes of the sector (i.e. services that feature SHARP properties). Indeed, 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, task T2.2 will 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 finance and insurance sectors.

1.1. Objective of the Deliverable

This (version 1) deliverable specifies the innovative BigData/IoT based services involved within INFINITECH on the Finance and Insurance sector. 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. Existing business processes (e.g., KYC/KYB, fraud detection and usage based insurance) are assessed based on the user stories of the INFINITECH pilots facilitating the development of use cases and data-based reference scenarios for SHARP (**S**mart, **H**olistic, **A**utonomy, **R**egulatory Compliance, **P**ersonalized) services. Moreover, these reference scenarios shall enable the definition of business services which shall be executed by BigData/IoT based services (i.e. services featuring these SHARP properties). In this way, the deliverable shall outline how BigData and IoT technologies can contribute to more autonomous, personalized and context-aware services.

The overall main objectives of this deliverable are:

- To elicit the business processes, which shall be covered in INFINITECH
- To identify related business services and provide an initial definition of functional service elements and probable building blocks

1.2. Insights from other Tasks and Deliverables

The deliverable extends the pilot descriptions and user stories developed in Task 2.1 with a services view towards a 1st description of functional services in the pilots. Therefore, especially the pilot contributions include some overlap with deliverable D2.1, e.g. in the use case descriptions.

Moreover, partners involved in Task 2.3 also contributed insights with respect to the INFINITECH background technologies.

1.3. Structure of the Deliverable

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 approach applied in this task is outlined in Section **Error! Reference source not found.**.
- Section 4 describes the pilots' abstraction from user stories to functional services.
- In Section 5, a requirements list per service and BDV reference model are described.
- Finally, Section 6 summarises the conclusions of the previous sections.

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

Digitization is changing the 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 and insurance services industry, [3]. For instance, this is illustrated by the rapid growth of FinTech start-ups. McKinsey tracked more than 2000 FinTech start-ups in 2016 expecting even many more 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 as well attackers as enhancements of 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 to a downward potential, 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 as 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.*"

The highest impact of technology application in Financial and Insurance Services shall be obtained in the focus areas of INFINITECH [4][5]:

- Lending & Financing, i.e. Credit Risks (Pilots 1 & 2)
- Wealth Management and Risk Assessment (Pilots 3 & 4)
- Customer Experience and Payment (Pilots 5 & 6)
- Usage based Insurance, which includes personalized insurance products (Pilots 11 to 14)

Moreover Regulation and Compliance, e.g. Financial Crime, Money Laundering, Fraud, includes strong opportunities to disrupt as Figure 1 shows. This area is covered by pilots 7 to 10.

	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 [5]

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]

This field is covered by INFINITECH technologies and considered e.g. in pilot 3.

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.

3 Approach

Initial information was retrieved from the pilot descriptions and the user stories documentation of task T2.1. This information was assessed and the result was given to the pilot partners for further detail. The pilots' contributions were assessed and an initial list of functional requirements was collected. These requirements were clustered according to the BDV reference model [7]. Figure 2 illustrates the approach taken to define the functional services of the INFINITECH pilots.

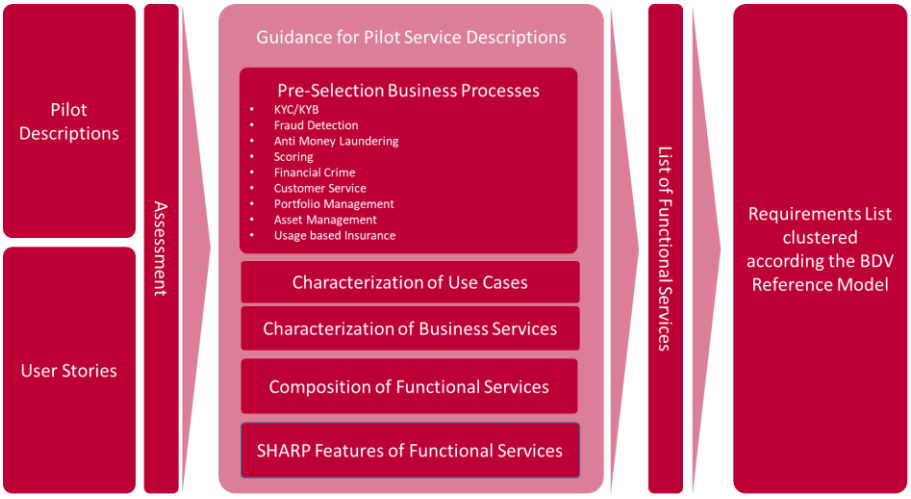

Figure 2 - Approach defining functional Services

Eliciting the functional services required by the pilots, we followed a four-step process to produce service requirements from the user stories developed in task T2.1 as follows:

1. First, the pilot related business processes, e.g. fraud detection or anti-money-laundering, were identified.
2. Second, the **user stories** of Task 2.1 were analysed to generate use cases/**data-based reference scenarios**. These scenarios were described by the pilot partners
3. Third, for each scenario, a list of **business and functional services** (often referred to solely as "services"), which are to be supported by INFINITECH were defined.
4. Finally, the **SHARP properties** facilitated by the services were detailed.

This way a common structure was specified facilitating a homogenized way of description of the pilots and the required functional services.

For step 1 the pilots should check the related business processes guided by a pre-selection from the categories

- KYC/KYB
- Fraud Detection
- Anti Money Laundering
- Scoring
- Financial Crime
- Customer Service
- Portfolio Management
- Asset Management
- Usage based Insurance

For step 2 the user stories were analysed and use cases were characterized by short keywords giving guidance to the description by the pilots. Analogously keywords for business services representing those use cases and an initial composition of functional services providing were proposed.

The pilot leaders and their technical advisors contributed this information and described the services that support the pilot applications.

4 Mapping Use Cases to SHARP Services

4.1 Pilot 1

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

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

Functional Service(s)

From a technical perspective, the system will be comprised of 4 main blocks. The first one is a data lake to facilitate the secure storage of digitized documents, the second one is a system to parallelize different jobs that compose the AI pipeline (pre-process, process and post-process the invoices), the third one is a computer vision system to identify and extract tables from invoices, and finally a machine learning system to extract the relevant information from the identified and extracted tables. The overall process is as follows:

1. Digitized documents will be generated by the bank's processes.
2. A data lake will facilitate the secure storage of those digitized documents.
3. These documents shall be processed in parallel by a pipeline (involving pre-processing, analysis and post-processing).
    - Pre-processing will include privacy related measures with respect to the GDPR.
    - Document analysis includes identification and extraction of tables from invoices, e.g. by a machine learning system.
4. The "Data Types" define the type of data which is used in the BigData application as additional data related pillars.
5. Finally, a visual console will show the results of the document processing and the sustainability score.

Sharp Properties

Digitization of contracting and invoicing processes increases the smartness and automation of the bank's services while adhering to regulatory compliance. Moreover, an automated analysis of digitized documents enable a smart and autonomous scoring for notary services. To rate notaries

based on a sustainability index will provide a new criterion to be applied when contracting these services impacting positively in the short and long-term in the amount of paper used and the economic fees applied. This system will be aimed to be adopted by the whole banking industry as a means to ensure more sustainable businesses.

4.2 Pilot 2

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

Use cases / data-based reference scenarios

In today's high risk business environment, one of the main challenges in asset management is to provide detailed risk information in a timely manner, i.e. in real time. Financial organizations and asset managers are therefore seeking novel ways for overcoming the current practice of risk applications relying on batch processing to produce aggregated reports. Rather traders, risk managers and sales negotiator should be supported on the fly, which requires real-time analytics technologies to process live operational data.

Risk assessment and monitoring procedures use two standard risk metrics – VaR (Value-at-Risk) and ES (Expected Shortfall) - for estimating market risk and to allow updates with changing market prices and/or changes in the bank's portfolio in (near) real time. Both can be applied for measuring various types of risk, above all, market risk of portfolios of assets. VaR is the standard risk measure within the financial sector today (and meanwhile also used by industrial and commercial enterprises for risk measuring), and it is requested by the regulatory authorities.

Business Service(s)

The pilot provides two business services for Asset Management. One service comprises risk metrics for estimating market risk and allow updates with changing market prices and/or changes in the bank's portfolio in (near) real time. Moreover, the evaluation of what-if-scenarios allowing pre-trade analysis, i.e. estimating changes in risk measures before a new trading position is entered, shall facilitate traders' decision making.

Functional Service(s)

The foreseen risk assessment is based on:

- 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 -
- DL/ML algorithms –
- Tracking and visualization of the following KPIs:
    - VaR figure with sensitivities;
    - ES figure with sensitivities;
    - Pre-trade analysis

Sharp Properties

Increased Automation and Intelligence based on near real-time data processing leveraging on-line data will facilitate smart and more accurate risk assessment. Moreover, it will support regulatory compliance to new regulations like BASEL IV or FRTB.

4.3 Pilot 3

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

Use cases / data-based reference scenarios

The goal is to provide privacy preserving, secure and trustworthy data sharing and analysis of financial information (customer, account and transaction data) between banks, their customers 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:

- Data sharing proposition with consumers, and how these would work within the context of Fintech eco-systems.
- Methodologies supporting sharing of data directly, and the necessary extra security controls to be applied.
- How a framework for banks can be created to offer data custodian services where customers have control over the management, usage and sharing of their own banking data.

This will result in two types of **interconnected Business Services** –

- Customer facing data services and associated internal bank processes
- Bank (or Bank partner) data sharing services to support those customer facing services

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:

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.

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

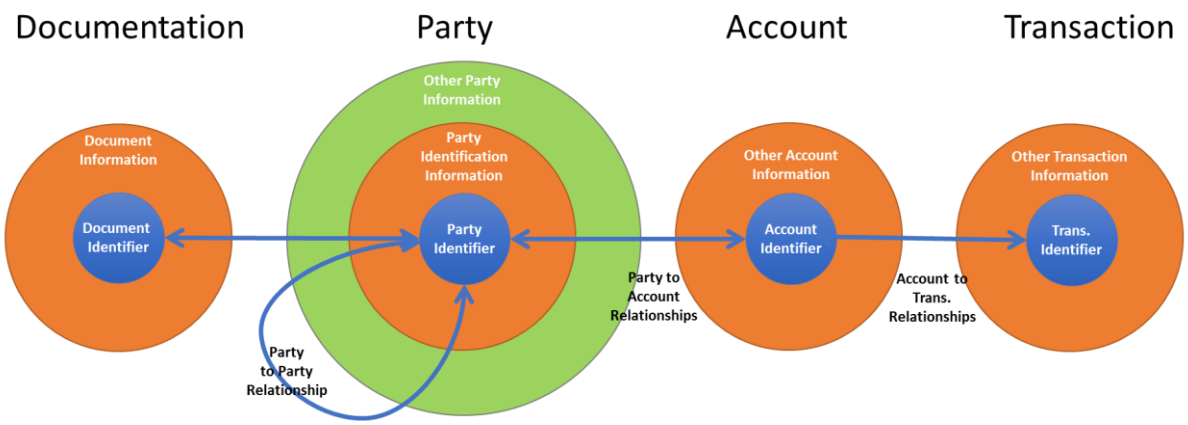

Figure 3 - Information Model

Functional Service(s)

- Semantic interoperability and semantic linking technologies
- OpenBanking API enabling data exchange by several financial institutions
- Blockchain based data sharing, including technologies for trading and support for personal data markets
- AI analytics based on large amounts of customer data from various sources or smaller richer datasets including graph data.
- Data Governance Building Blocks, such as eIDAS integration for client on-boarding;
- Consent Management and Privacy Management technologies.
- Application development of data sharing and consent management interfaces and associated UX (probably mobile)
- A secured GDPR-compliant infrastructure for information sharing
- Data Reference Architecture

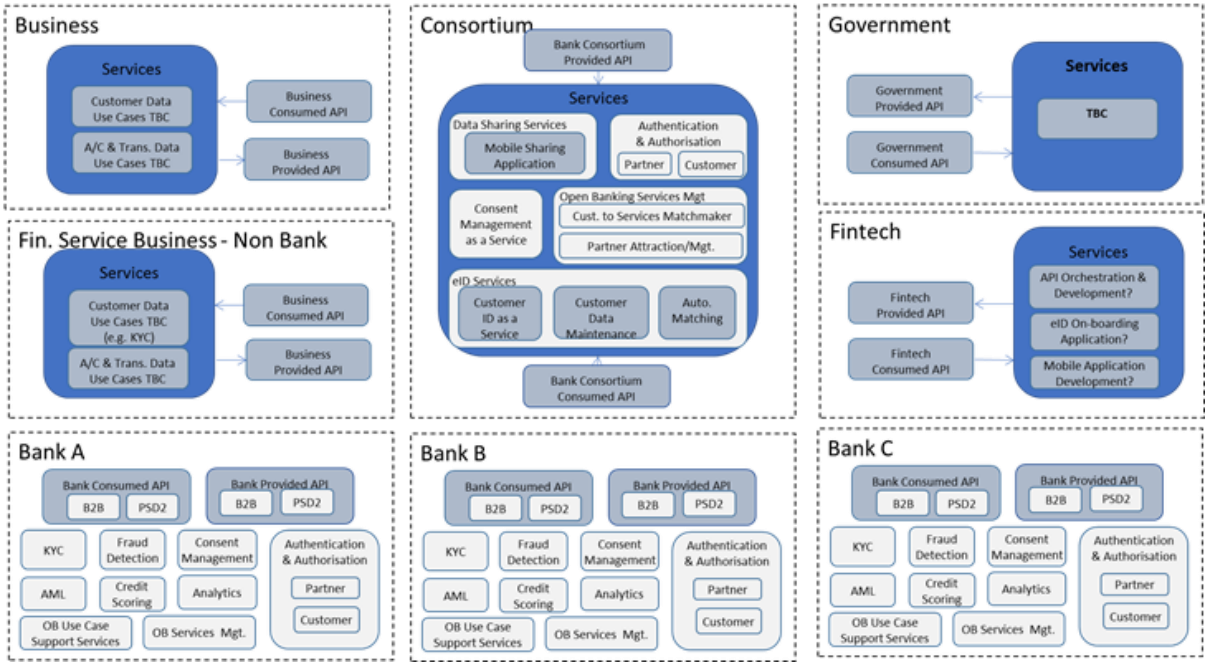

Figure 4 - Service Landscape

SHARP Properties

- Sharing financial data will enable KYC/ AML and Credit Scoring processes to become more intelligent and automated.
- KYC processes will be personalized aiming at improved customer satisfaction.
- A holistic customer-centric analytics framework will be produced that is able to support intelligent tools for new customer services using permission-based mechanisms for data sharing with associated analytics capability.

Note: The Smart, Holistic, Autonomy, Personalized and Regulatory Compliance is enabled from the underlying data sharing services.

4.4 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

The goal of this pilot is to explore the possibilities of AI Based Portfolio construction for Wealth Management in general regardless which amount is to be invested. The result could be that "Private Banking could be for everyone" which would lead to positive socio-economic benefits in terms of getting better and well-managed portfolios as well as benefits to the financial industry by having more and satisfied customers/investors.

The AI Based Portfolio Construction shall enable interested advisors or end-customers, after an initial "customer on boarding" (KYC, Risk Profiling) to upload relevant personal portfolios and start a portfolio optimisation process, where the AI Based portfolio construction is started together with a "genetic portfolio optimisation methodology". In several steps of portfolio calculations the "fittest" portfolio construction – based on risk appetite and defined risk limitations - shall be identified.

The optimization algorithm is based on increased using & sharing of data for portfolio construction. Among those data are

- assets under administration on Privé platform,
- customer based quantities,
- news feeds & blogs

Business Service

The underlying business service of Personalized Portfolio Management, which includes portfolio design and optimization, shall be automated based on customer behaviour and risk appetite. Based on analytics technology targeted recommendations of individual portfolio management shall be provided.

Functional Services

The business service Personalized Portfolio Management will be supported by

- Seamless Data Management and Querying;
- Data Governance Building Blocks (in addition to AI capabilities of the PRIVE platform)
- ML/DL Algorithms;
- Visualization of results for end-customers.

Sharp Properties

Broader availability and applicability of smart and personalized portfolio management shall have a positive socio-economic in terms of getting better and well-managed portfolios for a wide range of the society. This way the use case will enable increased/high personalization for a wide range of portfolios, including smaller ones.

This will be achieved due to an increase of automation of wealth management processes offered to end-customers and thus the reduction of cost & increase portfolio performance over the total population.

4.5 Pilot 5

4.5.1 Pilot 5a

Pilot #5a is titled **"Smart and Personalized Pocket Assistant for Personal Financial Management"** and belongs to category **"Personalized Retail and Investment Banking Services"**.

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

4.5.2 Pilot 5b

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.

4.6 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.

The pilot is aligned with the following **business processes**:

Know your Customer:

Various customer profiles will be created, such as risk, demographic and behavioural profiles, by leveraging on data from the various data-stores of the bank (i.e. customer data warehouse, customer interactions data in operational databases) and of data from alternative data sources (i.e. customer related social media activity, news and other on-line feeds). The KYC processes will be enhanced by on-line customer behaviour analytics that will allow the deeper understanding of customer's needs.

Customer Service:

The customer will experience a better and more trustful relationship with the Bank by receiving more tailored and acceptable recommendations.

Portfolio Management:

A number of alternative portfolios tailored to the customer's risk profile and behaviour will be constructed automatically, based on a recommendation engine. The NBG investment consultants will interact with the customer in order to offer them the recommended portfolios.

The system will continually monitor the performance of the suggested and actually purchased portfolios, as a means of tracking the effectiveness of the recommendation processes, but also as a means of continually improving the recommendations process.

A **use case scenario** that involves all types of users (investment advisors, retail customers and the Bank) would be the following:

The case:

A high-value customer visits an NBG branch to be informed about the terms in the time deposit schemas that the Bank is currently offering.

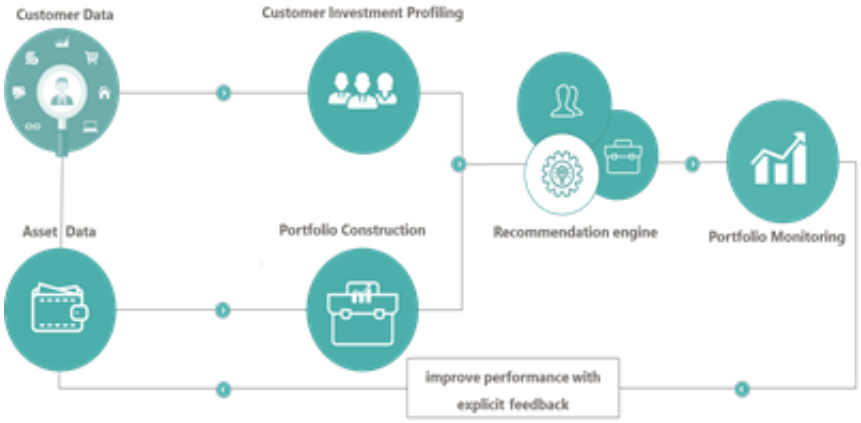

Figure 5 - Pilot 6 High Level Scenario

The background:

In the context of the pilot, the Bank will have categorized this customer under various criteria, such as his risk profile and appetite, demographics and behaviour. In order to incorporate the market stance on the investment products that comprise the portfolio, the Bank will also consider social media activity and news feed related to these products.

The investment products data will be combined and confronted against customer profiling information towards identifying investment prospects.

The result of the aforementioned steps is a set of predefined investment portfolios tailored to the customer's profile, accompanied with a probability of acceptance from his side.

The customer experience:

The branch will dedicate an investment professional to the customer of our case, to discuss upon the predefined proposed portfolios with him.

Not only will the customer be introduced to more sophisticated investment capabilities that couldn't access through the Bank until now, but also he will realize that the bank has proactively addressed his needs and optimal investment strategy.

The value for the Bank:

The Bank will develop a better and more trustful relationship with the customer, who hopefully will gradually turn exclusively to NBG for the entire spectrum of financial advice, products and services. The Bank will also increase its trading volumes. The investment consultants will see their productivity improving.

The **initial technology building** blocks are:

- Infrastructure (Databases with historical data and Real time data)
- Data Management
- Data Protection
- Data Processing
- Data Analytics
- User Interaction
- Visualization

Sharp Properties

The investment portfolio management will be personalized based on smart data acquisition and by customer behaviour profiling.

4.7 Pilot 7

Pilot #7 is titled **"Operation Whitetail – 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.

4.8 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 restricted to the consortium. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.

4.9 Pilot 9

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

The pilot is aligned with business processes involving KYC/KYB, fraud detection, AML and customer service. Customers as well as businesses may own assets issued on 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 addresses 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 are not to be cashed out through the customer or the businesses that a bank serves. Since public Blockchain are global, it provides also a medium for money laundering among different jurisdictions. When an analysis is done about Blockchain activities of customers and businesses, customer service in the form of generating easy to read reports that explain how the results obtained must be provided. The customers can read these reports and provide feedback if there are errors.

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. The business service will be provided through OpenAPIs. Visualization of transaction graph traces will also be provided.

Functional Service

The Functional service will operate on pre-processed high volume public Blockchain data that comes from multiple chains such as the Ethereum and Bitcoin. As blocks are generated on the Blockchain, pre-processing tools will extract transactions involving crypto currency as well as ERC20 token transactions periodically. The pilot sandbox 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.

Blockchain transaction graph will be analysed using parallel graph traversals and ML/DL algorithms. 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 but this will not be a requirement.

An interface in the form of command line and an OpenAPI interface will be developed. The outputs of the service may be in the form of graph paths or subgraphs that show tracing of Blockchain addresses to blacklisted addresses. These subgraphs will be output in graph visualization software format such as vis.vj for viewing in browsers.

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. This will allow the system to have holistic characteristic.

SHARP Properties

The system facilitates automation in processing of Blockchain transactions by a holistic & dynamic processing across Blockchain. It couples and automates several stages involving accumulating big raw data collection and pre-processing, scalable and fast analysis algorithms and easy to understand outputs with visualization. This will enable the system to have autonomous characteristic.

Even though all the Blockchain transactions will be available to process, tagging (labelling), masking and selection functionalities on transactions will be provided to users in order to do personalized graph analysis. Addresses stored on the public Blockchain are pseudo-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 regulatory compliant manner.

4.10 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 to assess cyber security based on dynamic cyber risk assessment metrics.

Business Service

For this pilot's reference scenario the business service to be delivered is related on 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 Services

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 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, Flink, Storm, Samza, Tensorflow), 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 tool will be based on the micro-service approach so 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.

4.11 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 Case / Data based Reference Scenario: Data based risk assessment and pricing

Current insurance policies are based on statistical analysis that exploit drivers' historical data and professional experience to categorize their clients. The datasets and the mechanisms used to classify the drivers are too static and neither include the particularities of each client nor provides enough flexibility to consider the different situations that happens in a driver's daily journeys. On the contrary, driver's score historically considers age, gender, car's colour etc.

Here, new services will be developed to adapt the driver's policy to his/her daily driving, in a modern way of service's consuming fashion: you pay depending on how you drive. In this sense, we find two main beneficiaries, the insurance company and the (good) drivers, penalizing bad habits and behaviours: as an example, vehicles and/ or drivers with no speeding fines that move in areas with low accident rates and accelerate and brake in reasonable way, will be rewarded with a reduction in tariffs. On the contrary, there will be an extra charge for more aggressive drivers.

In addition, information collected from connected cars will be also used to analyse traffic incidences, enabling possible fraud detection, which also benefits the insurance company and the rest of the (good) drivers. For instance, real data collected from the vehicle and other external sources could show that vehicle's regular circulation area is not in the declared region, or the age of the vehicle and/ or the driver are not as stated by the insured. If this is the case, there will be penalty in tariffs. Moreover, in case of accident reported by the insured, the insurance company may deny the claim if connected vehicle's data do not verify the accident and its location.

In this sense, the initial focus of this pilot is set on Smart Connected Vehicle entities and the datasets collected from these devices. The pilot plans to periodically gather information about the status of a

wide number of connected cars and so register historical data of their behaviour. The collected information will be analysed to define and extrapolate different driver's profiles that later assists AI and cloud computing technologies to classify a given driver according the way he/she drives. This dynamic classification will derive in new adaptive insurance usage-based products. In addition, data captured from smart vehicles may also feed and enhance roadside assistance and fraud-detection tools, providing relevant context information related to a traffic accident or incident.

The pilot's reference scenario is a risk assessment service based on real data collection (including IoT) and enhanced by AI and Cloud Computing. Its implementation will develop initially around two **business services**:

- Pay-as-you-drive: a used-based payment service that defines and infers (using different AI/ML, BigData and IoT platforms) different drivers' profiles based on the way the driver drives and so usage-based prices.
- Fraud detection: as a mechanism that analyses real time car data, traffic information and other context datasets to properly depict and relate traffic incidents.

Functional Service - From an overall perspective, this pilot will require, as a first component, vehicle smart objects. These will be provided by the partners' infrastructures in two different datasets: as real IoT devices (cars); and simulated smart objects that enhance the training processes. Other external traffic data sources will complement the analysed information. The data anonymization building block will guarantee that only traffic and car related information is used to define and train the general drivers' profiles, without putting their privacy at risk. Finally, the AI framework will use different AI technologies to, assisted by Cloud Computing and HPC components, infer drivers' profiles by processing real time connected cars' collected data. Figure 6 shows an approach of the building blocks that compose this pilot #11.

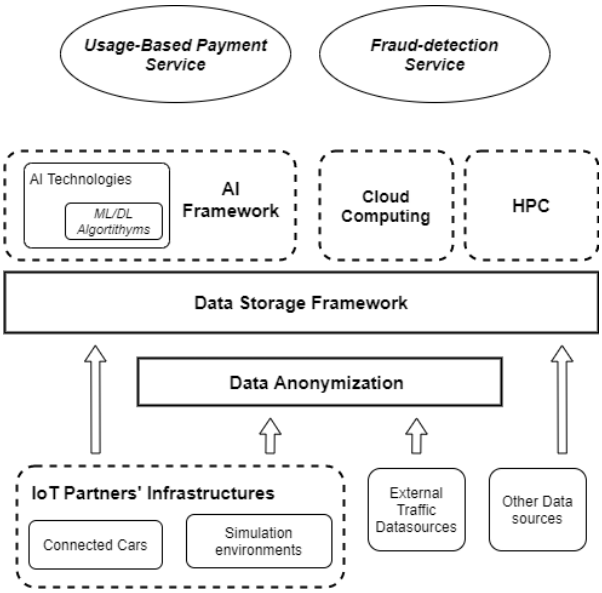

Figure 6 - Components of Pilot #11. Overall approach.

SHARP Properties

The pilot will clearly result in a more personalized and smarter insurance system, which will also create a basis for automating processes like fraud detection and claim management.

4.12 Pilot 12

Pilot#12 is titled **"Real World Data for Novel Health-Insurance Products"** and belongs to the **"Personalized Usage-Based Insurance Pilots"** category.

Use cases/ data-based reference scenarios

When an individual wants to buy a health insurance program Insurance Company's underwriting team receive his/her application and classify him/her in some predefined category in order to give him/her an offer. The classification is based on some predefined categories that the have been created based on historical data and statistical models. This approach is not flexible and does not take advantage of a wide variety of data that are available today. Classical statistical approaches for health insurance underwriting results in some generic categories for new client risk category estimation. For example, age, gender and the absence or presence of some particular disease are three main characteristics that are considered when a new client will be categorized in a particular risk category. This rule-based underwriting is static and may lead to misclassification of a new client. Furthermore it does not facilitate personalized insurance offers, risking to either displease the client or incur loss to the company.

In order to have a personalized risk assessment model this pilot focuses on health and behaviour-related data collection and analysis. Volunteer users will be using a customer kit based on a variant of the Healthentia eClinical platform to collect their physiological data. These are objective measurements of some vitals and mostly physical activity together with subjective data reposted by the users on their quality of life and nutrition. The collected physiological data will coupled with the more traditional demographic ones and assessed from a privacy perspective, to be later analysed using machine learning algorithms to produce clusters of people. A risk model will be build based on these clusters, aiming to then be able to classify new clients in the appropriate risk category and adjust their premiums according to their risk estimation. Thus, users that have healthier habits will have reduced premiums and Insurance Companies will be able to give more attractive offers to new clients according to their needs.

That demand transforms into the following **business services**:

- Risk management: Dynamic individualized adaptation of coverage and pricing according to clients' behaviour and automated risk calculation.
- Automated data privacy risk assessment and mitigation.
- Fraud Detection as a mechanism to alert insurance companies of erratic/ fraudulent behaviour.

Functional Services - This pilot will require IoT devices (activity trackers) that will be used by volunteer users. Data produced by these devices will be collected by the appropriately modified eClinical platform "Healthentia" and assessed from the privacy's perspective. Finally AI based (ML/DL) algorithms will be applied on the collected data in order to extract the final risk model.

SHARP properties: Based on the business services described pilot#12 will deliver in an Autonomous way Smart Personalized health insurance programs based on a Holistic account of clients behavior, covering both the traditional demographics data but also on measured and reported Real World Data, while observing Regulations on data privacy.

4.13 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 pilot will develop an insurance product configuration platform for SMEs, which will leverage large amounts of digital data in order to compute the offering. The results product configuration systems will support:

(i) **Configuration based on characteristics**, i.e. learning to select the appropriate coverages and services based on the mass of customers or SMEs;

(ii) **Configuration based on grouping** i.e. dividing SMEs into groups based on their characteristics through unsupervised learning algorithms, aligning them with existing products and determining the degree of adjustment of existing products in the market to these groups;

(iii) **Configuration based on collaborative filtering** i.e. configuring products that were interesting to SME's with profiles of similar customers, based on the preferences expressed by the SME's and their risk profiles.

Main use cases as output of the pilot would be:

1. Automate underwriting
2. Detect and forecasting changes in the risk needs of SME
3. Customize insurance products offering

Business Service

By obtaining the data in open sources and the application of machine learning the pilot will be able to monitor the changes in the risks, so we will be able to radically improve the risk management that companies face in the development of their daily activity. The indicators will be based on information from each of the companies about their activity, business volume, participation in social networks, number of employees, opening of new offices or offices, etc.

In this way, 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 Services

The process is based on the collection of large amounts of information from open sources and alternatives that are typically used by the insurance industry, such as companies' websites, official records, social networks, opinion forums, etc. After collecting the information by the robots, the algorithms that give us the outputs that are expected for the results of the project are applied as can be seen in the following graph.

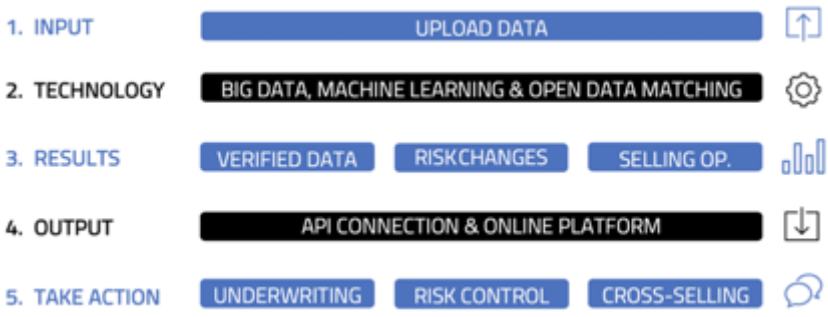

Figure 7 - Pilot 13 Functional Services

The initial technology that would be use are High-Performance Analytics (Parallel & Incremental Algorithms); Declarative Analytics and Parallelization of Stream Engines; Wenalyze big data analytics platform; ML/DL Algorithms. See chart

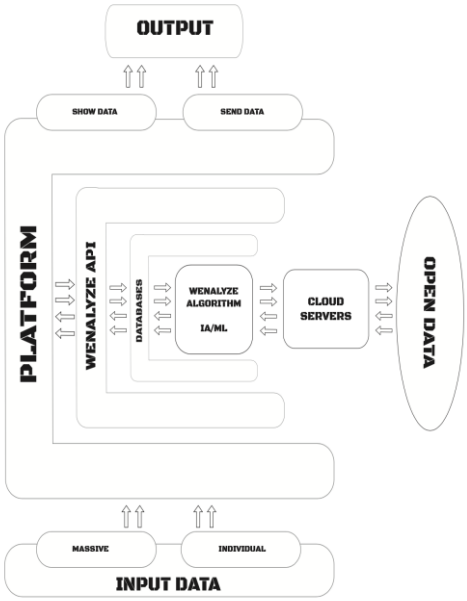

Figure 8 - Pilot 13 - High Level Architecture

Sharp Properties

The pilot, the technology and the results to be obtained are innovative since there is no automated system that recognizes, recommends and serves as a control for the risks that SMEs face in a personalized way. Recommends the appropriate products and coverages and their prices. In addition to enabling regulatory compliance control such as the GDPR.

4.14 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). Although the term UBI originally refers to car insurance programs (pay-as-you-go, pay-as-you-drive) based on data collection powered by telematics, the same approach in terms of using technology to gather data on-field/farm level can be applied to agricultural insurance in order to a) empower the Agri actors to operate under less uncertainty and lower risk; but also b) enable them to offer more personalized insurance products to farmers and other Agri-end-users (e.g. farmers cooperatives). Furthermore, the whole supply chain of Agri-Insurance provision can benefit from the solution developed within the INFINITECH project and defined in the User Stories (T2.1), not exclusively covering Reinsurers, Underwriters, Sales Agents and Brokers etc.; and can significantly improve the business process of customer service and/or management of their insurance portfolio.

Because the User Stories for Pilot 14 cover a wide-angle of the agricultural insurance process (every step in the AgI insurance supply chain), a minimum of four **use cases / data-based reference scenarios** are to be considered:

Case 1: Actuaries and underwriters will benefit from data-based risk assessment, pricing and product development.

Case 2: On the ground, loss adjusters will be supported by automated and remote earth observation (EO) data analysis. This process will also lead to more efficient claims management as losses can be reported in less time.

Case 3: Based on collected risk-relevant information about the focus market(s), sales agents can easily identify target regions and contribute to risk diversification within the insurance portfolio starting from the point of sale.

Case 4: Brokers and consultants can use the data and information for their advisory services and a more comprehensive approach to achieve individual risk transfer solutions for insurance companies on the international reinsurance market.

In order to accomplish the four described reference scenarios above, the solution must include the following services:

Service 1: Climate Risk assessment & seasonal Climate prediction for extreme weather events;

Service 2: Remote Crop-Damage assessment (drought, hailstorms, floods, wildfires);

Service 3: Verification and forecasting of extreme weather events;

That demand is transformed into the following **business services**:

Biz. Service 1 Risk management: Dynamic individualized adaptation of coverage and pricing according to risk and market, including dashboard (information on risks according to data driven intelligence), statistical data for further individual analysis (risk, market);

Biz. Service 2 Claim handling: Remote Crop-Damage assessment through change detection, image analysis.

Biz. Service 3. Support risk Analysis - Market, Portfolio: Risk diversification based on portfolio data against the background of market risk information

The envisioned services will be classified in INFINITECH under "High variety of data sources, multivariate analytics and AI, multi-level risk assessment scoring" in order to satisfy the manifold requirements of configurable and personalized insurance products for SMEs and Agro-Insurance companies. The usage of EO data, weather intelligence data and monitoring of agroclimatic indicators for a multi-level risk monitoring and damage assessment requires the identification of a variety of data sources, data pre-processing and feature extraction processing, post-processing of EO data and spatial analysis.

Functional Services - As it was referred previously, the solution that will tested in pilot 14 will offer 3 services:

Functional Service 1 is based on the 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. INFINITECH will offer the tools (AI/ML libraries) to post-process these indicators and combine them with other data sources (e.g. past calamities data) and develop more accurate and specific impact models, for each adversarial weather event.

Functional Service 2 is based on temporal change detection methodologies in multispectral and SAR time-series satellite images to identify the area that impacted by hail, floods, drought and wildfires, as well as the severity of each peril. INFINITECH will offer the tools (AI/ML libraries) to extend and enhance the capabilities of this service, by transforming it from a simple change detection/disaster monitoring service to a crop and farmer specific crop loss assessment tool, translating the observed damage to actual yield and money losses.

Functional Service 3 is based on AgroApps Weather Intelligence Engine that is offering very high resolution advanced atmospheric data assimilation products for extreme weather events identification and numerical weather predictions in convective scale. INFINITECH will offer the computational and storage resources, to host the operation of the system to cover the needs for weather data for the pilot 14.

SHARP Properties - Based on the business services described, the solution will enable a more detailed and smart calculation of insurance premiums, based on a holistic approach that addresses all relevant risk parameters impacting agricultural production. Additionally, increased efficiency and automation along the insurance process chain (claims management, loss adjusting, sales activities,

and advisory services) will extend the digital opportunities of agricultural insurance towards a more personalized product development and market penetration within the regulatory framework.

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 services defined per pilot are listed and clustered related to pilot groups and the BDV reference model.

5.1 Requirements List

Requirements shall be classified in a regular scheme across the WP2. This scheme includes

- 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 Functional Services are listed as functional requirements in an initial and preliminary list of requirements (See Appendix B "Requirements List"). However, this list is intended to collect these functions as a 1st overview, which will be reduced in a process of classification and detailing of the INFINITECH services in the subsequent steps of the task T2.2.

5.2 Requirements Clustering

Pilot are clustered in pilot groups. These pilot groups are

(i) Smart, Reliable and Accurate Risk and Scoring Assessment Pilots including pilots 1 and 2
(ii) Personalized financial products and services for both investment and retail banking, including various personalized services based on customer centric analytics and personalized digital assistants including pilots 3-6
(iii) Predictive financial crime and fraud detection applications including pilots 7-10
(iv) Personalized Usage based insurance products based on IoT data collection including pilots 11 and 12
(v) Configurable insurance products and services based on alternative data including pilots 13 and 14

The following clustering of requirements shall consider the pilot groups as this should lead to similar requirements per group facilitating group wise abstraction.

The Functional services and thus requirements listed in the previous section shall be structured in a scheme inspired by the BDV reference model Figure 9 facilitating identification and clustering of the service's components.

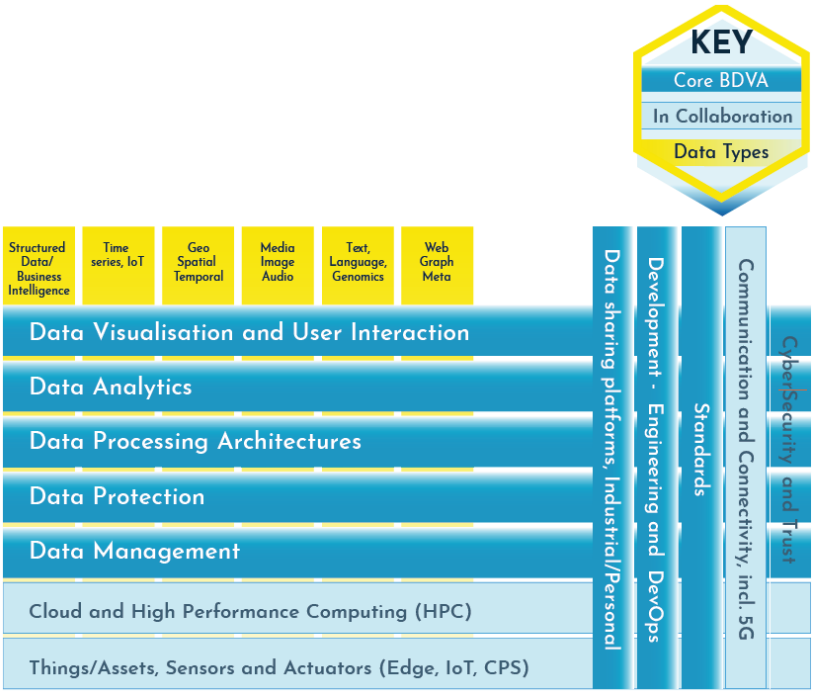

Figure 9 - BDV reference Model [7]

The BDV Reference Model [7] comprises the core elements of the BDVA, features in collaboration of the BDVA with other European activities and types of data, which are utilized in the model.

Stacked horizontal layers illustrate the data processing chain from data collection to data visualization, user interaction or data related events, which shall be processed by other pilot systems. However, these layers shall not imply a layered architecture according to [7] as e.g. visualization might be applied directly to collected data without further data processing or analytics.

The vertical pillars illustrate cross-cutting areas, which may affect several or even all horizontal layers. However, the Cybersecurity and Trust pillar might be emphasized more than in the figure with regard to compliance and regulatory issues, which shall be addressed in the INFINITECH project.

The "Data Types" define the type of data which is used in the BigData application as additional data related pillars.

Finally, it should be noted that the BDV Reference Model has no ambition to serve as a technical reference structure. It shall serve as a guidance for initial characterising and clustering the functional requirements.

In the following table an initial clustering related to the BDV reference model is listed. This will be refined in the subsequent steps of the task T2.2 including contribution by other tasks, e.g. related to the data types.

Note: In the following, the layers of the BDVA reference model are shown starting with the lowest layer (data processing) in the direction of the read sequence for easier understanding.

Table 1 - Initial Clustering of Functional Services along BDV Reference Model

Pilot	Functional Service	Data Types	Horizontal Layer	Vertical Layer
Pilot Group 1: Smart, Reliable and Accurate Risk and Scoring Assessment Pilots				
1	Digitized documents will be generated by the bank's processes.		Data Processing	
	A data lake will facilitate the secure storage of those digitized documents.		Data Management	
	Pre-processing will include privacy related measures with respect to the GDPR.		Data Protection	
	Document analysis includes identification and extraction of tables from invoices, e.g. by a machine learning system.		Data Analytics	
	Extracted tables shall be analysed by regarding the included information.		Data Analytics	
	The extracted tables and information will be utilized to establish a sustainability scoring of the notary service.		Data Analytics	
	A visual console shows the results of the document processing and the sustainability score.		Data Visualization	
2	Querying of streaming data and data at rest		Data Management	
	Data Anonymization and Governance		Data Protection	
	Modular, extensible real-Time Analytics Technologies i.e. incremental and parallel analytics		Data Analytics	
	Tracking and visualization of: VaR figure with sensitivities; ES figure with sensitivities; Pre-trade analysis		Data Visualization	
Pilot Group 2: Personalized Retail and Investment Banking Services				
3	Semantic interoperability and semantic linking technologies	All data groups	Data Management	
	OpenBanking API		Data Management	

	Blockchain based data sharing, including technologies for trading and personal data markets			Data Sharing Platform
	AI analytics based on large amounts of customer data from various sources or smaller richer datasets including graph data.		Data Analytics	
	Data Governance Building Blocks, such as eIDAS integration for client on-boarding;			Cybersecurity & Trust
	Consent Management and Privacy Management technologies.		Data Protection	
	Application development of data sharing and consent management interface and UX (probably mobile)		Development Engineering & DevOps	
	A secured GDPR-compliant infrastructure for information sharing		Data Protection	
	Data Reference Architecture		Data Management	
4	Seamless Data Management and Querying	Structured Data (assets under administration on Privé platform, customer based quantities) Web Data (news feeds & blogs)	Data Management	
	Data Governance Building Blocks		Data Protection	
	ML/DL Algorithms (in addition to AI capabilities of the PRIVE platform)		Data Analytics	
	Visualization of results for end-customers		Data Vizualization	
5a	This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.			
5b	This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.			
6	Acquire data from various data-stores of the bank	customer data warehouse, customer interactions data in operational	Data Management	

		databases	
	Data from alternative data sources	Web data (i.e. customer related social media activity, news and other on-line feeds).	Data Management
	Monitor the performance of the suggested and actually purchased portfolios		Data Management Data Processing
	Customer profiling under various criteria, such as his risk profile and appetite, demographics and behaviour.		Data Analytics
	Present predefined investment portfolios tailored to the customer's profile, with a probability of acceptance from his side		Data visualization
Pilot Group 3: Financial Crime and Fraud Detection			
7	This pilot description is confidential. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.		
8	This pilot description is restricted to the consortium. In case of specific interest, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us.		
9	Extract periodically transaction data from high volume Blockchains	Structured Data	Data Management
	Analyse Blockchain transaction graphs with ML/DL		Data Analytics
10	Acquire real time transaction data	Structured Data	Data Management
	Transaction fraud analysis using parallel/incremental/declarative analytics leveraging deep learning and big data processing engines		Data Analytics
	Visualization of the data resulting from the workflow execution by means of queries and graphs, in a meaningful way		Data Visualization
Pilot Group 4: Personalized Usage Based Insurance Products			
11	Collect real time driving data	IoT Data	Data Management
	Collect external traffic information		Data Management
	Collect other external sources		Data Management

	Data anonymization		Data Protection
	Data Storage Framework		Data Processing
	Driver Profiling		Data Analytics
12	IoT devices (activity trackers) will be collected by the appropriately modified eClinical platform "Healthentia"	IoT Data	Data Management
	Automated data privacy risk assessment and mitigation.		Data Protection
	AI based (ML/DL) algorithms will be applied on the collected data in order to extract the final risk model.		Data Analytics
Pilot Group 5: Configurable and Personalized Insurance Products			
13	Collection of large amounts of information from open sources and alternatives that are typically used by the insurance industry	All data groups	Data Management
	Parallelization of Stream Engines;		Data Processing
	High-Performance Analytics (Parallel & Incremental Algorithms); Declarative Analytics ML/DL Algorithms.		Data Analytics
	Wenalyze big data analytics platform		Data Visualization
14	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.	Structured data	Data Management
	Tools (AI/ML libraries) to post-process these indicators		Data Processing
	Other data sources (e.g. past calamities data)		Data Management
	Combine data sources		Data Processing
	More accurate and specific impact models, for each adversarial weather event.		Data Analytics
	Temporal change detection methodologies in multispectral and SAR time-series satellite images	Images	Data Management

Identify the area that impacted by hail, floods, drought and wildfires, as well as the severity of each peril.	Images	Data Processing
Tools (AI/ML libraries) to extend and enhance the capabilities of this service, by translating the observed damage to actual yield and money losses.		Data Analytics
Crop and farmer specific crop loss assessment tool.		Data Visualization
Computational and storage resources, to host the operation of the system to cover the needs for weather data.		Infrastructure

This way, an overview of service components related to the BDV data categories and horizontal and vertical layers is obtained. This view facilitates the specification of service modules in a later step of the task.

The table illustrates, that a refinement of this clustering could lead to a classification of service modules enabling its reuse in several pilots. Moreover, it reveals that analogous to the BDV reference model description, a business service not necessarily will comprise elements of all layers.

However, with this initial clustering of functional services and requirements the vertical issues like data sharing and security & trust need to be elicited more intensely in collaboration with the pilots. As usual the awareness of security and compliance issues needs to be fostered at the beginning of an innovation project. It is assumed that with the progress in Task T2.4 Security and Compliance the awareness of those topics will increase a lot.

6 Conclusions and next steps

Scope of task T2.2 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 with the SHARP properties).

In challenging times at or the end of a market cycle moving to a downward potential, new measures are required to change rapidly. Scaling of advanced analytics and AI tools is key to success Analysts point of view.

Overall, the INFINITECH pilot requirements reflect the State of the Art of the application of Big Data, IoT and AI in the Financial Services and contribute to the latest trends.

Within this deliverable, services related to the landscape of technology and data driven innovation of Financial Services are identified based on the broad use cases and reference scenarios provided by the INFINITECH pilots. Based on existing services and processes (e.g., KYC/KYB, fraud detection, customer service, portfolio management, asset management, usage based insurance) it has been outlined in the pilots' scenarios how these could become more autonomous, personalized and context-aware. In addition the pilot user stories have been abstracted to business services and the required functional services for their fulfilment.

These functional services defines an initial list of functional requirements for the INFINITECH pilots.

Finally, a 1st clustering of functional services alongside the BDV reference model facilitates the next steps within the task T2.2.

Based on the outcomes of this deliverable, in the next steps of task T2.2

- the state of the art in the pilots' reference scenarios will be elicited
- communalities of the services will be identified reducing the set of functional services
- technical details of the services will be specified

7 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] BDVA: "BDV SRIA – European Big Data Value Strategic Research and Innovation Agenda", Version 4.0, October 2017

8 Requirements Table

In the following section the initial functional services are listed as INFINITECH functional requirements. This list will be revised and detailed in the subsequent steps of task T2.2. An updated version of this list will be included in the next version deliverable D2.4.

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-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

This project has received funding from the European Union's horizon 2020 research and innovation programme under grant agreement no 856632

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-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
REQ-P7-F-B-M-O-049	

REQ-P7-F-B-M-O-050	https://www.infinitech-h2020.eu/contact-us.
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 restricted to the consortium. 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-P9-F-B-M-O-061	Access 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	Data preparation – gathering search results
REQ-P9-F-B-M-O-064	Visualization of search results
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-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