PART II:

Methods and Design Principles for Financial Innovation, Explaining the Supply Side for Interoperability in Finance- and Insurance-Tech

The INFINITECH Applications and Solutions

Martín Serrano, Bardia Khorsand, John Soldatos, Ernesto Troiano, Juergen Neises, Pavlos Kranas, Kostis Perakis, Alessandro Mamelli, Ignacio Elicegui, Dimosthenis Kyriazis, George Makridis, Gisela Sanchez and Marina Cugurra

now
the essence of knowledge

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

ISBN: 978-1-63828-230-3
E-ISBN: 978-1-63828-231-0
DOI: 10.1561/9781638282310

Suggested citation: Martín Serrano, Bardia Khorsand, John Soldatos, Ernesto Troiano, Juergen Neises, Pavlos Kranas, Kostis Perakis, Alessandro Mamelli, Ignacio Elicegui, Dimosthenis Kyriazis, George Makridis, Gisela Sanchez and Marina Cugurra. (2023). *INFINITECH Book Series – Part 2. Methods and Design Principles for Financial Innovation, Explaining the Supply Side for Interoperability in Finance- and Insurance-Tech: The INFINITECH Applications and Solutions*. Boston–Delft: Now Publishers

Table of Contents

Preface

The data is changing the way society and technology evolves, with the advent of IoT, Big Data, ML and AI, a rapid development in technology towards more human-centric applications has been envisaged. The finance and insurance sectors are not an exception and developments in FinTech and insurance-tech are in a phase of developing unique offerings.

It is very important to have a common understanding of the actual conditions in the financial and insurance sectors and how the technology can help to advance and evolve those conditions in a positive manner. By discussing the principles of the modern economy that make the modern financial sector and FinTech the most disruptive areas in today's global economy, a better understanding and knowledge will be acquired.

The use of data-driven approaches envisions many opportunities emerging for activating new channels of innovation on the local and global scale while at the same time catapulting opportunities for more disruptive human-centric services. Data-driven human-centric applications are at the same time the result of a shared vision from a natural evolution of technology and society. Experts in the financial and insurance sectors are looking at a dramatic change in how people think about global economy and at the same time the technology is facilitating the instruments for new ways of understanding, providing a common vision and identifying impacts in finance and insurance.

The INFINITECH book series is focused on addressing the need for clear information for better understanding of the foundations, principles and technologies for experts and non-technical experts that participate in the financial and insurance process and the constant need for innovation and new services across banks and insurance organizations.

The Editors and Contributors of this INFINITECH book series would like to thank the European Commission and the Science Foundation Ireland for their support in the planning and execution of the INFINTECH project that resulted in the preparation of this book. The recommendations and opinions, the provided and developed technologies alike experiences described in this book are those of the editors and contributors, and do not necessarily represent those of the European Commission or the Science Foundation Ireland.

Dr. J. Martin Serrano O.
INFINITECH Scientific Manager and Data Scientist
Adjunct Lecturer and Senior SFI Research Fellow at University of Galway
Data Science Institute – Insight SFI Research Centre for Data Analytics
Unit Head of Internet of Things, Stream Processing and Intelligent Systems Research Group
University of Galway, www.universityofgalway.ie | Ollscoil na Gaillimh
<jamiemartin.serranoorozco@universityofgalway.ie>
<martin.serrano@insight-centre.org>
<martin.serrano@nuigalway.ie>

Who Should Read This Book?

Financial & Insurance Regulators

The unique offering for non-technical experts but that participate in the financial regulatory process and of the core service to enable the sharing of innovation and new services across banks and insurance without exchanging any customer data.

General Public & Students

The power of understanding the future of FinTechs, their services and their ability to identify different methodologies indicators from a human perspective.

Entrepreneurs and SMEs

The most powerful tools to innovate, increase opportunities and increase the power of innovation into small and entrepreneurs to meet its full potential if there is good participation across the banking and insurance sector.

Technical Experts & Software Developers

The guide for technologies and legacy open and non-open sources as a guidebook for including the most recent experiences in Europe towards innovating technology for the financial and banking sectors.

What is Addressed in the Book Series?

"Concepts and Design Thinking Innovation addressing the Global Financial Needs"

In the first part of the INFINITECH book series we begin by discussing the principles of the modern economy that make the modern financial sector and FinTech the most disruptive areas in today's global economy. INFINITECH envision many opportunities emerging for activating new channels of innovation on the local and global scale while at the same time catapulting opportunities for more disruptive user- centric services. INFINITECH is at the same time the result of a shared vision from a representative global group of experts, providing a common vision and identifying impacts in the financial and insurance sectors.

"Methods and Design Principles for Financial Innovation, Explaining the Supply Side for Interoperability in Finance- and Insurance-Tech"

In the second part of the series we review the basic concepts for Fintech referring to the diversity in the use of technology to underpin the delivery of financial services. The demand and the supply side in the financial sector are demonstrated, and further discussed is why FinTech is the focus of industry nowadays and the meaning for waves of digitization. Financial technology (FinTech) and insurance technology (InsuranceTech) are rapidly transforming the financial and insurance services industry. We provide an overview of Reference Architecture (RA) for BigData, IoT and AI applications in the financial and insurance sectors (INFINITECH-RA). Moreover, this book reviews the concept of innovation and its application in INFINITECH, and innovative technologies provided by the project for financial sector practical examples.

"Technical Financial Innovation, Solving the Interoperability Problems of Europe"

The third book begins by providing a definition for FinTech as: The use of technology to underpin the delivery of financial services. This book further discusses why FinTech is the focus of industry nowadays as the waves of digitization and the way financial technology (FinTech) and insurance technology (InsuranceTech) are rapidly transforming the financial and insurance services industry. In this book technology assets that followed the Reference Architecture (RA) for BigData, IoT and AI applications are introduced. Moreover, the series of assets includes the domain area where applications from the INFINITECH innovation project and the concept of innovation for the financial sector are described. Further, we describe INFINITECH Marketplace and its components including details of available assets. Next, we provide descriptions of solutions developed in INFINITECH.

What is Covered in this INFINITECH Part II Book?

"Methods and Design Principles for Financial Innovation, Explaining the Supply Side for Interoperability in Finance- and Insurance-Tech"

In this second part of the series we review the basic concepts for FinTech referring to the diversity in *The use of technology to underpin the delivery of financial services*.

The demand and the supply side in the financial sector are demonstrated, and further discussed is why FinTech is the focus of industry nowadays and the meaning for waves of digitization.

Financial technology (FinTech) and insurance technology (InsuranceTech) are rapidly transforming the financial and insurance services industry.

We provide an overview of Reference Architecture (RA) for BigData, IoT and AI applications in the financial and insurance sectors (INFINITECH-RA). Moreover, this book reviews the concept of innovation and its application in INFINITECH, and innovative technologies provided by the project for financial sector practical examples.

Acknowledgements

To our families for their incomparable affection, jollity and always understanding that scientific career is not a work but a lifestyle and encouraging us to be creative, and for their enormous patience during the time away from them, invested in our scientific endeavours and responsibilities and for their understanding about loving too much our professional life and its consequences, we love you!.

To all our friends and relatives for their comprehension when we have no time to spend with them and we are not able to join in time because we are in a conference or attending yet another meeting, for their attention and the interest they have been shown all this time to keep alive our friendship; be sure our sacrifices are well rewarded.

To all our colleagues, staff members and students at our respectively institutions, organisations and companies for patiently listening with apparent attention the descriptions and progress of our work and for the great experiences and the great time spent while working together with us and the contributions provided to culminate this book series project. In particular thanks by the support and confidence from all people that believed this series of books would be finished in time and also to those that didn't trusted on it, because thanks to them we were more motivated to culminate the project.

To the scientific community, this is our family when we are away and working far from our loved ones, for their incomparable affection, loyalty and always encouraging to be creative, and for their enormous patience during the time invested in understanding, presenting and providing feedback to new concepts and ideas, sincerely to you all, thanks a million!

Martín Serrano on Behalf of All Authors

Contributing Authors

Achille Zappa
NUIG-Insight, Ireland

Adrien Besse
FTS, France

Aikaterini Karamargiou
NBG, Greece

Akshay Shetty
PRIVE, Germany

Alain Vailati
HPE, Italy

Alberto Crespo
ATOS, Spain

Alberto Danese
NEXI, Italy

Aleksandra Cargo
BOS, Slovenia

Alessandra Forlano
ENG, Italy

Alessandro Amicone
GFT, Italy

Alessandro Mamelli
HPE, Italy

Alessio Del Soldato
NEXI, Italy

Alex Acquier
NUIG-Insight, Ireland

Alexander Kostopoulos
RB, UK

Alper Sen
BOUN, Turkey

Andrea Becerra
CTAG, Spain

Andrea Grillo
PI, Greece

Andrea Toro
HPE, Italy

Andreas Politis
DYN, Greece

Angeliki Kitsiou
CP, Greece

Anja Rijavec Ursej
BOS, Slovenia

Ann Smith
BPFI, Ireland

Anna Semeniuk
PRIVE, Austria

Annalisa Ceccarelli
PI, Greece

Anne Elisabeth Lenel
ORT, France

Antonis Litke
INNOV, Cyprus

Antonis Skarpelis
CP, Greece

Ariana Polyviou
INNOV, Cyprus

Ariana Polyviou
INNOV, Cyprus

Aristodemos Pnevmatikakis
ISPRINT, Belgium

Baran Kilic
BOUN, Turkey

Barbara Cacciamani
ABILAB, Italy

Bardia Khorsand
NUIG-Insight, Ireland

Beatrice Paolone
PI, Greece

Bjoern Torkar
PRIVE, Germany

Borja Pintos Castro
GRAD, Spain

Brigitte Benerink
RRD, Netherland

Bruno Almeida
UNP, Portugal

Bruno Lepri
FBK, Italy

Can Ozturan
BOUN, Turkey

Candeago Candeago
FBK, Italy

Carlos Albo
WEA, Spain

Carmen Furquet
INSO, Spain

Carmen Perea
ATOS, Spain

Chi Hung Le
NUIG-Insight, Ireland

Christian Hanley
PRIVE, Austria

Christiane Grunloh
RRD, Netherland

Christina Katsikari
RB, UK

Christoforos Symvoulidis
SILO, Greece

Christopher Genillard
GEN, Germany

Chrysostomos Symvoulidis
UPRC, Greece

Claudia Amador
UNP, Portugal

Claudia Mertinger
FTSG, Germany

Craig Macdonald
GLA, UK

Cyril Armange
FI, France

Danae Lekka
ISPRINT, Belgium

Dario Francés
WEA, Spain

David Delgado
WEA, Spain

Davide Dalle Carbonare
ENG, Italy

Davide Profeta
ENG, Italy

Dejan Adamic
BOS, Slovenia

Diego Burgos
LXS, Spain

Dimitrios Kotios
UPRC, Greece

Dimitrios Miltiadou
UBI, Greece

Dimitris Drakoulis
INNOV, Cyprus

Dimitris Dres
INNOV, Cyprus

Dimosthenis Kyriazis
UPRC, Greece

Diogo Inácio
UNP, Portugal

Domenico Costantino
HPE, Italy

Domenico Messina
ENG, Italy

Dominik Hedderich
GEN, Germany

Dominique Faessel
ORT, France

Dustin Ciccardini
JRC, Germany

Ehsan Arefifar
CTAG, Spain

Elena Battistini
GFT, Italy

Elena Femenia
INSO, Spain

Eleni Mavrogalou
CP, Greece

Eleonora Ascolani
PI, Greece

Eoin Jordan
NUIG-Insight, Ireland

Erdem Oguz
AKTIF, Turkey

Ernesto Troiano
GFT, Italy

Eva Sotos Martinez
GRAD, Spain

Evelina Peristeri
SILO, Greece

Eymard Hooper
BOI, Ireland

Fabiana Fournier
IBM, Israel

Fabio Dezi
NEXI, Italy

Fabio Magrassi
GFT, Italy

Farid Meinkohn
ORT, France

Fethi Ata
AKTIF, Turkey

Filip Koprivec
JSI, Slovenia

Filipa Sousa
UNP, Portugal

Gabriele Gamberi
ABILAB, Italy

Gabriele Santin
FBK, Italy

Gary Thompson
BOI, Ireland

Gavin Purtill
BPFI, Ireland

George Fatouros
INNOV, Greece

George Giaglis
UNIC, Cyprus

George Karamanolis
CP, Greece

Georgios Makridis
UPRC, Greece

German Herrero
ATOS, Spain

Giacomo Toselli
SIA, Italy

Giancarlo Sfolcini
SIA, Italy

Giorgia Gazzarata
GFT, Italy

Giorgio Dabormida
GFT, Italy

Giorgio Roffo
GLA, UK

Giovanni Di Orio
NOVA, Portugal

Gisela Sanchez
FI, France

Giuseppe Avigliano
PI, Greece

Gokcehan Kara
BOUN, Turkey

Gregor Krzmanc
JSI, Slovenia

Gregor Zunic
JSI, Slovenia

Grigoris Mygdakos
AGRO, Greece

Guilherme De Brito
NOVA, Portugal

Harm op den Akker
ISPRINT, Belgium

Harm Opdenakker
RRD, Netherland

Hermie Hermens
RRD, Netherland

Hoan Nguyen
NUIG-Insight, Ireland

Iacopo De-angelis
PI, Greece

Iadh Ounis
GLA, UK

Iago Abad Fernandez
GRAD, Spain

Ian Godfrey
FTS, France

Ian Shiundu
RB, UK

Ignacio Elicegui
ATOS, Spain

Ilesh Dattani
ASSEN, Ireland

Ines Ortega-Fernandez
GRAD, Spain

Inna Skarbovsky
IBM, Israel

Irene Zattarin
GFT, Italy

Javier Rodriguez Viñas
GRAD, Spain

Javier Sanz-Cruzado Puig
GLA, UK

Javier Yepez Martínez
GRAD, Spain

Jelena Milosevic
BOS, Slovenia

John Soldatos
INNOV, Cyprus

Jonathan Gay
ASSEN, Ireland

Juan Mahilo
LXS, Spain

Juergen Neises
FTSG, Germany

Julian Schillinger
PRIVE, Germany

Julien Mousset
PRIVE, Germany

Klaudija Jurkosek-Seitl
BOS, Slovenia

Klaus Brisch
DWF, Germany

Klemen Kenda
JSI, Slovenia

Konstantina Kostopoulou
ISPRINT, Belgium

Konstantina Tripodi
JRC, Germany

Konstantina Zafeiri
NBG, Greece

Kostas Perakis
UBI, Greece

Lambis Dionysopoulos
UNIC, Cyprus

Lena Neidhardt
GEN, Germany

Lex Vanvelsen
RRD, Netherland

Lilian Adkinson Orellana
GRAD, Spain

Luca Latella
NEXI, Italy

Lucile Aniksztejn
FI, France

Lukas Linden
GEN, Germany

Maanasa Srikrishna
GLA, UK

Machi Simeonidou
AGRO, Greece

Mads Tingsgard
CPH, Denmark

Magdalena Schmid
GEN, Germany

Maja Skrjanc
JSI, Slovenia

Manolis Syllignakis
NBG, Greece

Manuela Masci
PI, Greece

Marc Meerkamp
DWF, Germany

Marcio Mateus
UNP, Portugal

Marco Avallone
PI, Greece

Marco Crabu
ABILAB, Italy

Marco Muller Terjung
DWF, Germany

Marco Pistore
FBK, Italy

Marco Rotoloni
ABILAB, Italy

Marco Spallaccini
PI, Greece

Marcos Alvarez Diaz
GRAD, Spain

Marcos Cabeza
CTAG, Spain

Margarita Khokhlova
FTS, France

Maria José Poveda
WEA, Spain

Maria de Vries
GEN, Germany

Maria Smyth
NUIG-Insight, Ireland

Marian Hurmuz
RRD, Netherland

Marianna Charalambous
UNIC, Cyprus

Mariarosaria Russo
ENG, Italy

Marina Cugurra
GFT, Italy

Marina Rodriguez Hidalgo
LIB, Spain

Marinos Xynarianos
CP, Greece

Mario Maawad Marcos
CXB, Spain

Mario Trinchera
ABILAB, Italy

Marko Grobelnik
JSI, Slovenia

Marta Sestelo
GRAD, Spain

Martin J. Serrano Orozco
NUIG-Insight, Ireland

Massimiliano Aschi
PI, Greece

Massimiliano Aschi
PI, Greece

Massimiliano Hocevar
PI, Greece

Matej Koletnik
BOS, Slovenia

Matteo Falsetta
GFT, Italy

Matteo Gerosa
FBK, Italy

Maurizio Ferraris
GFT, Italy

Maurizio Megliola
GFT, Italy

Maximilien Nayaradou
FI, France

Michael Concannon
BPFI, Ireland

Michael Michalakoukos
DYN, Greece

Michael Psalidas
CP, Greece

Misu Helal Ali
FTS, France

Mitja Jermol
JSI, Slovenia

Mojca Trstenjak
BOS, Slovenia

Nadia Roberti
PI, Greece

Napoleon Liontos
CP, Greece

Neil Giles
TAH, UK

Nial O'Brolchain
NUIG-Insight, Ireland

Niarchos Vasilios
NBG, Greece

Nicola Masi
ENG, Italy

Nikolaos Kapsoulis
INNOV, Cyprus

Nikos Drosos
SILO, Greece

Nikos Droukas
NBG, Greece

Nuria Ituarte Aranda
ATOS, Spain

Oliver Sjastedt
CPH, Denmark

Omerbora Zeybek
AKTIF, Turkey

Orkan Metin
AKTIF, Turkey

Pablo Carballo
PRIVE, Germany

Padraig Flannery
BOI, Ireland

Palmira Aldeguar
LIB, Spain

Paolo Testa
NEXI, Italy

Patrick Karlsson
RB, UK

Patrizio Sangermano
PI, Greece

Paul Lefrere
CCA, France

Pavlos Kranas
LXS, Spain

Pedro Malo
NOVA, Portugal

Perdikouri Eleni
NBG, Greece

Petra Ristau
JRC, Germany

Phil Atherton
TAH, UK

Prokopaki Georgia
NBG, Greece

Qaiser Mehmood
NUIG-Insight, Ireland

Raman Kazhamiakin
FBK, Italy

Ramon Martin de Pozuelo
CXB, Spain

Rebeca Jiménez
WEA, Spain

Rene Danzinger
PRIVE, Austria

Ricard Bruguera
WEA, Spain

Ricardo Jimenez-Peris
LXS, Spain

Richard McCreadie
GLA, UK

Richard Walsh
BPFI, Ireland

Rishabh Chandaliya
NUIG-Insight, Ireland

Roger Ferrandis
WEA, Spain

Roland Meier
PRIVE, Austria

Roman Benito
LIB, Spain

Sabina Podkriznik
BOS, Slovenia

Sara El Kortbi Martinez
GRAD, Spain

Saso Crnugelj
BOS, Slovenia

Silvio Walser
BOC, Cyprus

Simon Schou
CPH, Denmark

Simone Centellegher
FBK, Italy

Sofoklis Kyriazakos
ISPRINT, Belgium

Spyros Spanos
ISPRINT, Belgium

Stamatis Pitsios
UBI, Greece

Stathis Kanavos
ISPRINT, Belgium

Stefano Gatti
NEXI, Italy

Stelios Kotsopoulos
AGRO, Greece

Stelios Mantas
NBG, Greece

Stelios Pantelopoulos
SILO, Greece

Stephanie Jansen-kosterink
RRD, Netherland

Susanna Bonura
ENG, Italy

Susanna Bonura
ENG, Italy

Tanja Zdolsek draksler
JSI, Slovenia

Teoman Onat
PRIVE, Germany

Teresa Spada
ABILAB, Italy

Theodoros Kotzastavros
CP, Greece

Theodoros Arnaoutoglou
CP, Greece

Thomas Diesinger
GEN, Germany

Thomas Krogh
CPH, Denmark

Thorsten Jansen
DWF, Germany

Tiago Teixeira
UNP, Portugal

Vaia Gousdova
SILO, Greece

Vasilis Koukos
SILO, Greece

Vasilis Koukos
UPRC, Greece

Vicent Sebastiá
WEA, Spain

Vicky Foteinou
CP, Greece

Victoria Michailidou
RB, UK

Vito Morreale
ENG, Italy

Vittorio Monferrino
GFT, Italy

Yasar Khan
NUIG-Insight, Ireland

Ziga Bucaj
BOS, Slovenia

Abstract

The large number of emerging FinTech companies and the transformation that financial corporates i.e. banks, credit and insurance companies are suffering as consequence of the new wave of disruptive human-centric services is changing the landscape of the global financial economy, making the financial sector evolve very fast.

The use of emerging technologies like BigData, Machine Learning Frameworks and Algorithms and Artificial Intelligence within the financial sector are only examples of how the technology can catapult a large number of new user-centric services.

In this second part of the series we review the basic concepts for FinTech referring to the diversity in *the use of technology to underpin the delivery of financial services*.

The demand and the supply side in the financial sector are demonstrated, and further discussed is why FinTech is the focus of industry nowadays and the meaning for waves of digitization.

Financial technology (FinTech) and insurance technology (InsuranceTech) are rapidly transforming the financial and insurance services industry. We provide an overview of Reference Architecture (RA) for BigData, IoT and AI applications in the financial and insurance sectors (INFINITECH-RA).

Moreover, this book reviews the concept of innovation and its application in INFINITECH, and innovative technologies provided by the project for financial sector practical examples.

DOI: 10.1561/9781638282310.ch1

Chapter 1

FINTECH Services

1.1 FINTECH Services

This book series aims to specify different aspects of each large-scale pilot: readiness; development; and validation of different services and components. Validation is a core pillar, as one of the main objectives of INFINITECH is to test innovative (IoT, BigData, AI, ML, Blockchain and more) technologies towards improving business services in the Financial and Insurance sector. Specifically, the present deliverable reports on the readiness of the various pilot sites to test the INFINITECH innovative AI, IoT and BigData technologies into the testbeds/sandoxes that are developed during the project, while validating their ability to improve the business processes of end-user organizations (i.e. financial organizations, banks, and FinTech firms). [D7.1]

In summary, this deliverable reports for each one of the pilots sites the following information:

- A General overview of the status of the pilot, including its main business and technical objectives.
- The development status of the different components and services that comprise each pilot system.

- The status of the integration of a subset of their components as part of a Proof-of-Concept (PoC) pilot system.
- Information on the availability and deployment status of the testbed/sandbox, where the pilot' final infrastructure will be deployed and validated.

Pilots have already contributed to other previous deliverables and tasks (requirements, user stories, security, policies, technologies, services, RA, etc). Therefore, this deliverable builds on top of these contributions. However, it also integrates and extends them, through illustrating how individual technical activities are enabling the integration and deployment of a complete pilot system with relevance for the end-users (i.e. financial organizations, banks, FinTechs). Overall, the present deliverable focuses not only on individual contributions, but rather on the overall readiness of the pilot sites and the pilots' frameworks as a whole. [D7.1]

The book presents the status of an initial PoC implementation for each pilot. This PoC enables a first demonstration of the viability and applicability of the various INFINITECH technologies that support the pilots. The various pilots PoCs demonstrate the different developments accomplished up to date and serve as a basis for ensuring that the pilots' developments are on the right track, while identifying points that need attention where required.

The current status of testbeds/sandboxes is also an important part of the deliverable, because it directly affects the readiness and demonstrability of each pilot. Therefore, a quick overview of each testbed/sandbox is covered by each pilot.

The book aims to aid an understanding of the overall progresses of pilots, serving as an index to go deeper into different pilots' achievements. Further, the book have started work on synergies and KPIs. Therefore, a preamble of this work is included. Finally, the book will make a first introduction to actions related to the main suggestions coming from the Review Report. [D7.2]

This book includes the outcomes of the project in relation to pilot systems and pilot activities in the INFINITECH project. The information included reflects the activities and the operations conducted for the pilots of the project related to smart and reliable scoring, risk and service assessment. [D7.3]

The current book clusters the activities, organization, and deployment of three different pilots, namely Pilot#1, Pilot#2, and Pilot#15. These pilots refer to specific application fields: Scoring, Risk and Service Assessment. They all implement ML algorithms to business cases, aiming to enter the market with necessarily-novel approaches. Those pilots are data-driven and analyse different sources of data by pre-processing and converting such information into viable and effective data sources. [D7.3]

This book contains an overview of each of the pilots listed above addressing several technical and operational aspects. Each pilot is briefly introduced, highlighting

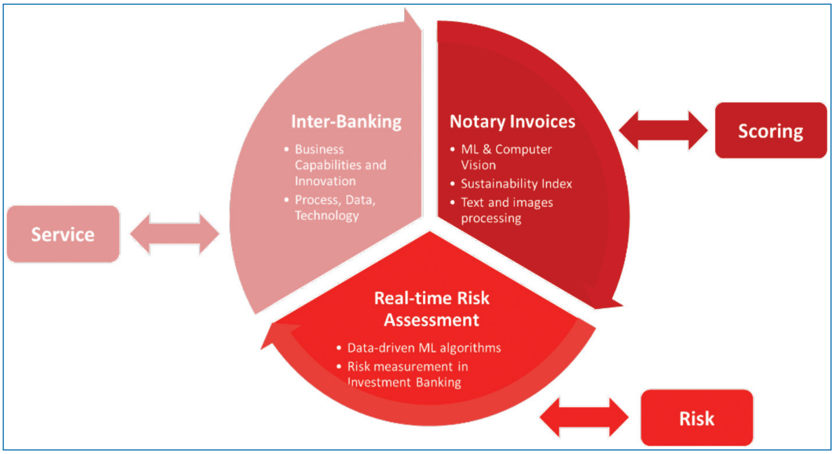


Figure 1.1. Cluster #1 Pilot applications.

three key questions: What is the problem it is intended to solve?; How is it solved?; and What are the main benefits? Subsequently, the pilots are described in terms of two main streams: the pilot systems and the pilot activities. [D7.3]

The pilot system is intended to report the business services and each pilot provides explanations of their innovation, and the list of technologies and components used while mapping them with the services. The pilot activities describe the operations and roadmap towards the first cycle of development and beyond, along with descriptions and visualizations of the actual status of implementation, concluding with information about the performed or planned validation workshops. [D7.3]

We report of the activities of Cluster #1, which is devoted to developing, conducting and operating the pilots of the project related to smart and reliable scoring, risk and service assessment (Figure 1.2). The pilots feature similarities in terms of their characteristics, yet they will be deployed using different technologies and based on different sandboxes. The pilots will be deployed and validated in three iterations, that will gradually advance the maturity of each of the pilot deployments. This approach is taken to ensure the proper technical and business validation of the pilot systems. [D7.3]

The three Pilots #1, #2, #15, within the Cluster #1 refer to three specific applications: Scoring, Risk and Service Assessment. All the pilots exploit ML algorithms, Big Data, and other technologies to address business cases, aiming to penetrate the market with needfully-novel approaches. Figure 1.1 below maps each pilot with the specific application field and shows graphically what are the major components of each of them. [D7.3]

Pilot #1 (Invoices Processing Platform for a more Sustainable Banking Industry), deals with the extraction of information from Invoices, running ML algorithms on such data to analyze them and compare all the different sources to come up

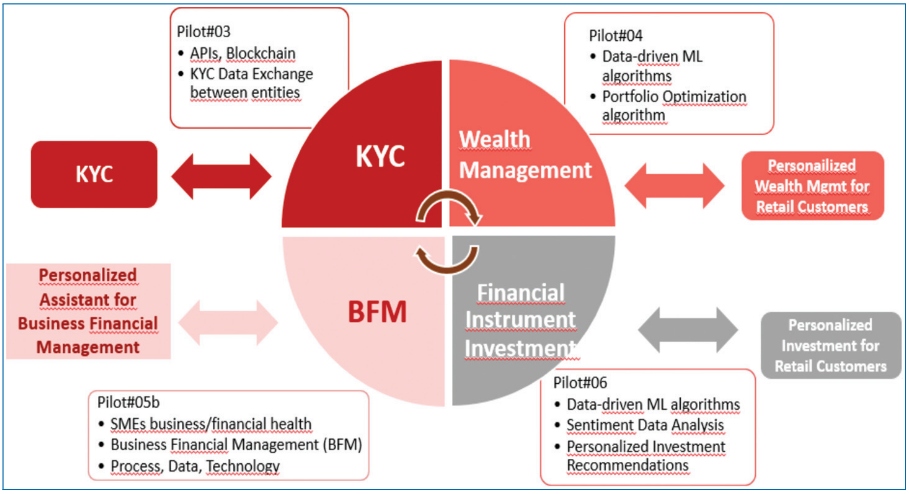


Figure 1.2. Cluster #2 Pilot applications.

with a Sustainability Index Scoring. The core part is the extraction, analysis and conversion of data intended as text (including tables) and images. AI technologies are applied to both the scanning of physical documents, and the development of automatized sustainability index scoring; as a consequence, this approach leads to cost saving and increased efficiency. [D7.3]

Pilot #2 (Real-time risk assessment in Investment Banking), 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) realtime. Moreover, estimated changes in risk measures before a new trading position is entered will be implemented. Several stakeholders would benefit from such a pilot, mainly because it's risk-driven and processes data and provides results in either real-time or near real-time. [D7.3]

Pilot #15 (Open Inter-Banking Pilot), classifies the information contained in a subset of process-operating documents used by Italian banks, to build a business glossary respecting the ABI Lab taxonomy, thus supporting the Enterprise Architecture Modelling. This pilot will allow the screening of extensive documentation in real-time, addressing a business pain shared by several banks, therefore being pre-competitive and strongly market-driven. [D7.3]

This book also aims to specify different aspects of large-scale pilots that involve personalized recommendations to customers and customer centric analytics about: development, deployment, extension and validation of different services and components, that will be developed or used as part of cluster #2 for the INFINITECH Project. [D7.6]

The pilots feature similarities in terms of their characteristics, yet they will be deployed in different sandboxes and based on different technologies. All pilots will be deployed in three iterations, that will gradually advance the maturity of each of the pilot deployments, while at the same ensuring the proper technical and business validation of the pilot systems. [D7.6]

The relative pilots listed below, intended to provide personalized financial products and services for both investment and retail banking, include various personalized services based on customer centric analytics and personalized digital assistants, based on:

Pilot #3: Customer Centric Analytics & KYC: Use the customer-analytics solution in additional banking/finance processes commonly named Know Your Customer/Know Your Business (KYC/KYB); Move the solution to pre-production deployment for selected services (Partners: BPFI, NUIG, BOI, IBM).

Pilot #4 Personalized Asset Management: Integrate more data in the PRIVE's solution and expand the recommendations to additional products; Include in PRIVE's consulting services (Partners: PRIVE, RB).

Pilot #5b Smart and Personalized Pocket Assistant for PFM & BFM tools delivering a Smart Business Advise: Run a larger-scale pilot with the engagement of more customers of the bank; Offering respective functionalities also to corporate customers and moreover through integration with other third parties; Receive feedback, improve usability and deploy in productions; Deploy similar solution to other banks and financial institutions (Partners: BOC, GFT, UPRC, CP).

Pilot #6 Personalized Investment Recommendations for Retail Clients: Following business validation, prepare the solution for use in the NBG's portfolio of retail solutions. Exploit the tailored sandbox of the solution for driving additional innovation inside the bank (Partners: NBG, CP, RB, UBI, LXS, GLA).

In particular, we intend to provide an overview for each of the pilot above, answering to the following questions:

- What is the problem and how the pilot development is addressing it?
- What is the innovation that pilot brings either for business or technology? What are the technologies developed within or outside INFINITECH that are used for this pilot?
- What is each pilot's development roadmap?
- Which are the workshops planned with internal or external stakeholders for validation of the expected results and outcome for each pilot?

We also specify different aspects of INFINITECH cluster #3 (Figure 1.3), a cluster of four pilot systems that involve Predictive Financial Crime and Fraud Detection. It focuses on the development, deployment and validation of different services and components of the pilot systems. [D7.9]

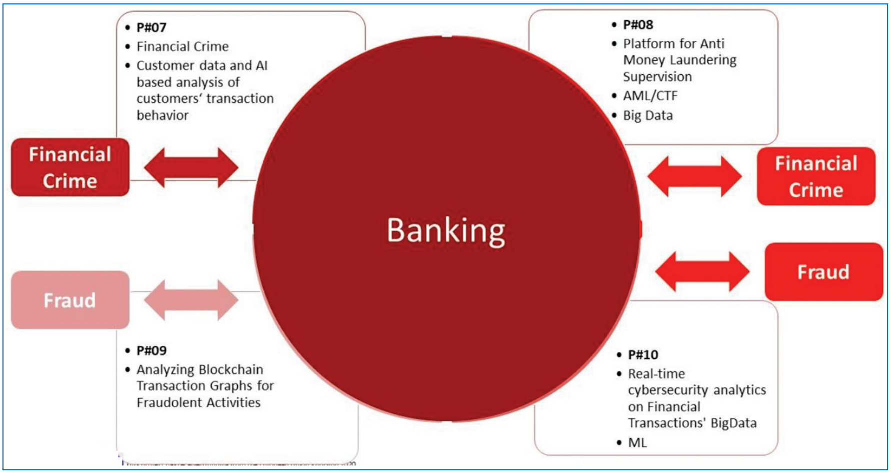


Figure 1.3. Cluster #3 Pilot applications.

The four pilots of the cluster feature similarities as far as their characteristics are concerned, yet they are deployed in different sandboxes and based on different technologies. All pilots are developed and deployed in three iterations, that will gradually advance the maturity of each of the pilot deployments, while at the same time ensuring the proper technical and business validation of the pilot systems. [D7.9]

The related pilots intend to provide advanced financial products and services for banks, supervisory authorities, financial institutions, and governmental agencies, aiming to prevent and protect against financial crimes and fraudulent activities, as follows: [D7.9]

Pilot #7 Avoiding Financial Crime: explore more accurate, comprehensive and near real-time representations of suspicious behavior in Financial Crime, Fraud, and cyber-physical attacks with the final objective of stealing bank customers' identity and money. (Partners: CXB, FTS, FBK)

Pilot #8 Platform for Anti Money Laundering Supervision (PAMLS): development of a Platform for Anti Money Laundering Supervision (PAMLS), which will improve the effectiveness of the existing supervisory activities in the area of anti money laundering and combating financing of terrorism by processing large quantity of data owned by the Bank of Slovenia and other competent authorities. (Partners: BOS, JSI)

Pilot #9 Analyzing Blockchain Transaction Graphs for Fraudulent Activities: the aim of the pilot is to detect fraudulent activities monitoring blockchain transactions. (Partners: AKTIF, BOUN)

Pilot #10 Real-time cybersecurity analytics on Financial Transactions' BigData: improved detection of cases of suspected fraudulent transactions, to enable the identification of security-related anomalies while they are occurring, by the analysis in real-time of the financial transactions of a home and mobile banking system. (Partners: PI, ENG)

Pilot#16 Data Analytics Platform to detect payments anomalies linked to money laundering events: development of a data analytics platform, based on Machine Learning and graph database composition, to help the NEXI AML team to discover, monitor and analyze suspicious scenarios related to money laundering through digital card payments. (Partners: NEXI, GFT)

We describe the pilot prototype, the technologies that underpin each pilot system, as well as the innovative characteristics of each pilot system. It also elaborates on each pilot's development roadmap. Specifically: [D7.9]

- Pilot#7 has already identified a valid dataset where all the transactions of the type "Immediate loans" are included from October 2020 to March 2021, in this dataset where fraudulent transactions have been tagged. Also, the dataset has been treated so as to anonymize the fields that could otherwise reveal personally identifiable or confidential information. The dataset has been shared with the technical providers of the core technologies to support the pilot, FTS and FBK, with successful outcome resulting in an improvement of the efficiency for the fraud discovery in this type of transactions by applying AI to the data and process.
- Pilot#8's Risk Assessment tool is now in its final stage of development and is already in test and verification phase on Pilot#8 testbed. The Screening tool is in its main developing phase. Another important component of PAMLS platform is the pseudo-anonymization component, which enables regulation compliant data pseudo-anonymization in a way that the analytical results still represent valueadded information. Pilot#8 also contributes three components: Stream Story, Pattern discovery and matching, and Anomaly detection and prediction. While Anomaly detection and prediction component is already developed, Pattern matching and discovery component is currently in the main phase of the development, while Stream Story component is in early stage of the development.
- Pilot#9 is currently using its HPC based scalable system to analyze massive real Bitcoin and Ethereum cryptocurrency and ERC20 token transactions, in order to trace fraudulent activities. It is also able to take token transactions as input from permissioned ledgers such as the Hyperledger Fabric. A market analysis has been performed for the Pilot#9, attempting to assess size and nature of an industry as well as competition and regulations. It was also found that the Pilot#9 graph analysis system can have potential uses in the area of Central Bank Digital Currencies (CBDCs). In Period 2, implementation

of machine learning based analysis of blockchain transactions started. The objective is to predict fraudulent blockchain addresses and hence the problem was posed as a supervised ML problem.

- Pilot#10 arises from the need to overcome the limitations of rule-based systems to block potentially fraudulent transactions and the need to exploit the ML capabilities to more effectively identify new kinds of risky transactions. In the period M18-M27 of the project, the fraud detection system architecture was redesigned to more effectively support the continuous batch ML model retraining, according to the ML Ops best practices, which aim to deploy and reliably and efficiently maintain machine learning models in production. Two machine learning models are now adopted, one trained in an unsupervised fashion and one in a supervised one.
- Pilot#16 has entered the INFINITECH project in September 2021, so the level of detail about the pilot status in this document is not at the same degree of the other pilots. The book features information about the use case/ data-based reference scenarios and business services, as well as the technology component foreseen for the pilot development and a roadmap.

The book is devoted to developing, deploying, extending, and validating the pilots of the project that involve Predictive Financial Crime and Fraud Detection. The pilots feature similarities in terms of their characteristics, yet they will be deployed in different sandboxes and based on different technologies. All pilots are deployed in three iterations. The clustering of the pilots is aimed at facilitating synergies between them, including knowledge exchange and best practice sharing. Pilot#16 has started lately, joining the project in September 2021. [D7.10]

We also describe category/cluster 4, that involves personalized usage based insurance products and lists all the activities (development, deployment, extension and validation) related to the components, architecture, services and dissemination activities carried out by the cluster's 4 pilots. [D7.10]

Cluster 4 (Figure 1.4) comprises two pilots oriented to the insurance sector that exploit IoT based infrastructures to gather real world and real time data and develop AI powered services to enhance risk profiling. These pilots are building their own infrastructure (according to INFINITECH-Reference Architecture) by combining INFINITECH and pilot's specific technologies, configure their corresponding sandboxes and run their testbeds, all within the INFINITECH framework. They are deployed in three iterations, which will gradually advance the maturity of each of the pilot deployments and the validation of their brought business innovations. [D7.10]

Category 4 pilots are intended to provide personalized insurance products and services based on IoT connected devices, including various personalized services

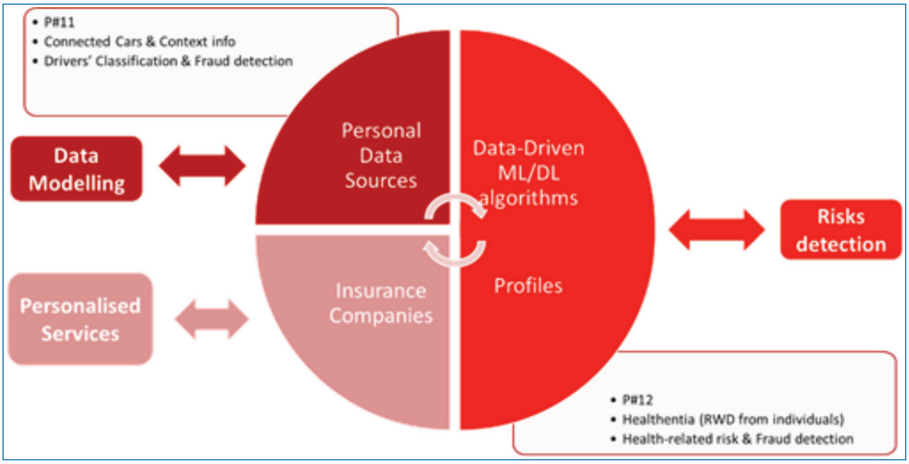


Figure 1.4. Cluster #4 Pilot applications.

based on customer centric analytics and personalized digital assistants. These are: [D7.12]

Pilot #11: Personalized insurance products based on IoT connected vehicles: Improve the risk insurance profiles using the information collected by connected vehicles and applying IoT, HPC, Cloud Computing and Artificial Intelligence technologies: drivers' classification and fraud detection. (Partners: ATOS, CTAG, GRAD, DYN) [D7.12]

Pilot #12 Real World Data for Novel Health Insurance: Improve the risk insurance profiles using the information collected by activity trackers & questionnaires and applying IoT & ML technologies (Partners: SiLO, iSPRINT, RRDD, GRAD, ATOS, DYN). [D7.12]

The book contains an overview and status report for each pilot listed above, introducing the identified problem/s (as the seed/s of the business models), the way it/they is/are addressed, the involved innovations (technical and business ones) and the corresponding roadmap. In the scope of the present deliverable are also included the main achievements of each one of the usage-based insurance pilots, covering: [D7.12]

- The design of the first prototype of each pilot system, in-line with the INFINITECH Reference Architecture.
- The development of initial prototypes for both pilots, centered in data gathering and integration, and leveraging INFINITECH technologies (implemented in WP3/WP5), as well as pilot specific technologies.
- The organization of a workshop for the validation and reception of stakeholder feedback about the usage-based insurance pilot systems.

The book provides the development, integration, and validation of the INFINITECH IoT-based pilots, oriented to the insurance sector and the personalized usage-based services and products design. The work addressed by this task involves the end users' engagement, deployment and integration of the required IoT systems, collection, storage and classification of data, identification and evaluation of proper ML/DL algorithms, applications/services development and deployment, stakeholders' identification, technical and business innovations dissemination and a final business validation of the pilots involved in the INFINITECH Cluster/Category 4: Personalized Usage-Based Insurance Pilots.

This book series focuses on the progress and achievements related to the pilot's userstories, technologies and architecture which have already been referenced in that initial version. With this in mind, this document covers three main parts for each pilot: [D7.13]

- Technological updates and new features implementation.
    - Both pilots (Pilot #11 in Section 2.2.1 and Pilot #12) have developed and evolved their AI models for Driving profiling and Risk Assessment and introduced XAI methodologies. Also, their testbeds (following the Infinitech way) have been updated. TRLs status are also shown.
- End-users' engagement and workshops
    - Stakeholder activities, workshops, and early adopters' progresses summarized within a table. Both pilots have worked together to present new achievements to the stakeholders.

This book series describes the use cases, pilots, and technical achievements of the personalised insurance scenarios (Cluster 4), covering pilots 11 (Motor Insurance) and 12 (Health insurance). It contributes with the final versions of the systems and applications developed within Cluster/Category 4 pilots, showing the final PoCs achieved by each pilot and sharing their components through the Infinitech Marketplace. It also presents the final progresses in terms of technologies, services and outcomes offered to end users and to the FinTech marketplace. [D7.14]

This document describes and presents in detail the following end-user services:

- Pay How You Drive is the service developed by the insurance company utilizing the drivers' profiling and classification in order to adjust motor insurance premiums according to driving behaviour.
- Fraud detection is the service for the insurance companies providing them with the real circumstances of an accident, so as to support detection of fraudulent acts against them.
- Health risk assessment is the service provided to insurance professionals that utilizes the learnt health outlook models on the data of insured individuals,

to facilitate the professionals' decision on the modification of their clients' health insurance premiums.

- Health fraud prevention is the service provided to insured individuals, that analyses the decisions of the above model to offer actionable advice to them, hopefully persuading them to use the provided measurement system truthfully.

In order to validate the cluster results, the insurance services developed within the cluster were presented to relevant sector's stakeholders, including internal actuaries and external insurance companies to evaluate the listed services and orient their further evolution and exploitation.

Pilot #11 Personalized insurance products based on IoT connected vehicles: Improve the risk's profiles in motor insurance using the information collected by connected vehicles and applying Artificial Intelligence technologies: development of drivers' classification and fraud detection services. (Partners: ATOS, CTAG, GRAD, DYN)

Pilot #12 Real World Data for Novel Health Insurance: Improve the risk insurance profiles using the information collected by activity trackers & questionnaires and applying IoT & ML technologies (Partners: SiLO, iSPRINT, RRDD, GRAD, ATOS, DYN).

Specifically, this third report is focused on Usage-Based Insurance (UBI) services, as Cluster #4, covering the pilot's #11 and #12 Proof of Concepts. Pilot #11, centred on motor insurance applications, and pilot #12, dealing with health sector exploit the AI technologies applied to the real-world data captured from the insured clients (using IoT frameworks) to evaluate the real risk associated to the individuals and so, create specific product lines that can be customised according to defined profiles and/or specific individual's behaviour. The target is to evolve the way the insurance's companies calculate their client's premiums, changing classical statistical techniques by real time data from users. [D7.14]

This book series also focuses on category/cluster 5, that involves configurable and personalized insurance products based on alternative and automated insurance risk selection and insurance product recommendation for SME's and "Big Data and IoT for the Agricultural Insurance Industry and lists all the activities (development, deployment, extension and validation) related to the components, architecture, services and dissemination activities carried out by the cluster 5's pilots. [D7.15]

Cluster 5 is composed of two pilots that base their analysis on big data from different sources, both open and online and from satellite imagery to gather real world and real time data and develop AI powered services to enhance risk profiling (Figure 1.5). These pilots will develop their own architecture by combining

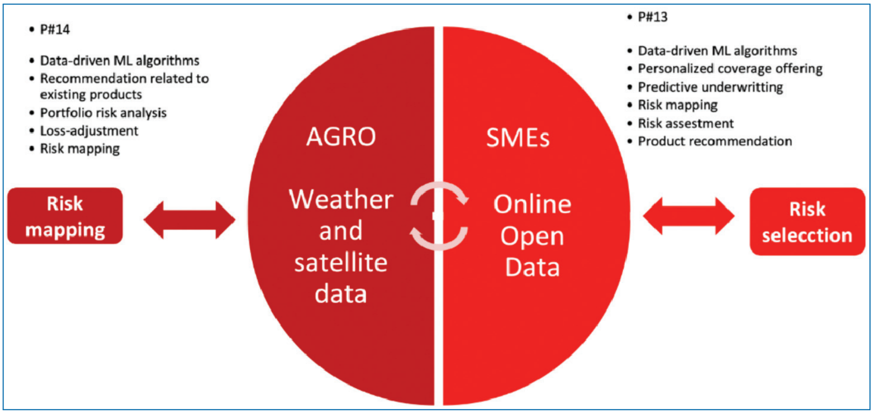


Figure 1.5. Cluster #5 Pilot applications.

INFINITECH and pilot's specific technologies, configure their corresponding sandboxes and run their testbeds, all within the INFINITECH framework. These will be deployed in three iterations, that will gradually advance the maturity of each of the pilot deployments and the validation of their brought business innovations. [D7.15]

Category 5 pilots are intended to provide configurable and personalized insurance products based on alternative data sources and big data, including various personalized services based on customer centric analytics and personalized digital assistants. These are: [D7.15]

Pilot #13: Alternative and automated insurance risk selection and insurance product recommendation for SME's: Focuses 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.

Pilot #14 Big Data and IoT for the Agricultural Insurance Industry: Provide Insurance companies with a robust and cost-effective toolbox of functions and services- allowing them to alleviate the effect of weather uncertainty when estimating risk of AgI products, reduce the number of on-site visits for claim verification, reduce operational and administrative costs for monitoring of insured indexes and contract handling, and design more accurate and personalized contracts.

The objective of this book series is to identify possible Pilot synergies in order to detect the same pattern of problems and/or situations between the Pilots. By identifying these synergies, it enables the Pilots to collaborate with each other, transmitting their knowledge and visions to overcome problems and/or situations that arise. [D7.18]

To identify synergies, it was necessary to analyse the User Stories that were established. The identification of these synergies is also very relevant to the work related to the entire WP7 and its respective deliverables, such as the collaboration between Pilots in solving problems and/or use of technologies. [D7.18] The synergies listed in the book series have been divided into categories. These categories were defined at the beginning of the project, whose order was maintained to divide the Pilots by these categories and by synergies. [D7.18]

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

It is therefore illustrated the status of the Pilots' KPIs, as to embody an initial snapshot from which to base the evaluation, as well as the methodology that will be used to carry out the future monitoring process (which belongs to the second phase of the framework). [D.20]

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

DOI: 10.1561/9781638282310.ch2

Chapter 2

Applications & Services for the Financial Sector

2.1 Applications and Services for the Financial Sector

The Interoperable Data Exchange and Semantic Interoperability focuses on establishing the foundation for common, shared meaning across the several data sources and message and event feeds within the INFINITECH platform while facilitating the technical implementation of the INFINITECH principles. In this landscape, there are defined a set of objectives: [D4.1]

1. Defined shared semantics (ontologies) for semantic interoperability of BigData and IoT streams in the finance/insurance sectors;
2. Provide the means for scalable the massive analytics over linked semantic streams;
3. Provide a permissioned blockchain solution for exchange data across different organizations in the finance and insurance supply chains;
4. Enhance the permissioned blockchain of the project with tokenization functionalities, as means of enabling digital assets trading; and
5. Implement techniques for secure querying of encrypted personal data over a blockchain.

Taking into account the overall objectives, the following set of activities are identified as main outcomes of the INFINTECH project objectives:

- Shared Semantic for BigData and IoT Streams: This activity specifies models and ontologies for semantic interoperability of diverse applications in the finance and insurance sectors. It extends and integrate ontologies such as Financial Industry Business Ontology (FIBO)/Financial Instrument Global Identifier (FIGI) with additional concepts associated with INFINITECH applications and testbeds. The task produced the project's ontology for semantic interoperability, which will provide the concepts needed for annotating and linking diverse data streams.
- Massive Distributed Processing of Semantically Linked Streams: This task provides a prototype implementation of the Super Stream Collider (SSC) engine, called SeSA-ME (Semantic Stream Analytics Engine/Middleware) that enables analytics for semantically linked streams (linked data). The engine is scalable and suitable for massive parallelization in cloud environments. It is implemented on top of SSC component, which is customized to support linked data in-line with the shared semantics specified in INFINTECH priorities.
- Distributed Ledger Technologies for Decentralized Data Sharing: This task implements permissioned blockchain infrastructures based on Corda R3 and/or the open source Hyperledger Fabric project. These blockchains are customized to support the requirements of the financial sector, including data models, authentication and authorization mechanisms, as well as APIs for implementing Ledger Clients for financial/insurance sector applications. The infrastructure is integrated to existing BigData/IoT platforms in the testbeds, based on appropriate ledger clients.
- Tokenization and Smart Contracts Finance and Insurance Services: This activity enhance the permissioned blockchain with cryptographic tokenization features, as a means of enabling assets trading. Likewise, the activity specifies and implement Smart Contracts for adding and retrieving information on the tokenized blockchain for all the essential data exchange use cases of the project's pilots. The applications provide the means for trading access to data and information through the permissioned blockchain. The activity specifies and implement ledger protocols for the financial/insurance applications, including trading protocols.
- Secure and Encrypted Queries over Blockchain Data: This activity implements and provide a framework for querying encrypted data over the project's permissioned blockchain infrastructure. It exploits and customize algorithms from the OPAL project, based on Multi-Party Computation (MPC) and

Linear Secret Sharing (LSS) schemes (i.e. homographic encryption). The mechanisms implemented resemble Enigma's (enigma.io) Personal Data Management infrastructure, through the integration of consent mechanisms that enable consumers/customers to provide consent for access to their data through the blockchain. In conjunction with the trading and tokenization functionalities of the blockchain, this activity created the foundation for a personal data market where customers are able to trade their data in exchange for tokens on other assets.

- Situation Awareness Front-End over Aggregated Information: This activity provides a web-based framework for the visualization of the aggregated results of analytic algorithms developed in the scope of the INFINITECH project, and more generally of all information of relevance. The framework is based on the community edition of Knowage, an OS solution for BI, which is part of the OW2 community. The Knowage suite ia extended and customized in order to support specific data models and persistence technologies. The visualization functionality allow users to assemble personalized dashboards for situation awareness, wiring together related information from different sources. Special emphasis was paid in visualizing information from distributed ledgers.

Background and Related Works

This section is intended to frame the research realized under the scope of the Shared Semantic for BigData and IoT Streams [D4.1 to D4.3]. It establishes a common ground and a necessary foundation to support the design and definition of the proposed methodology for developing INFINITECH models and ontologies for semantic interoperability while avoiding any misunderstanding regarding INFINITECH main concepts.

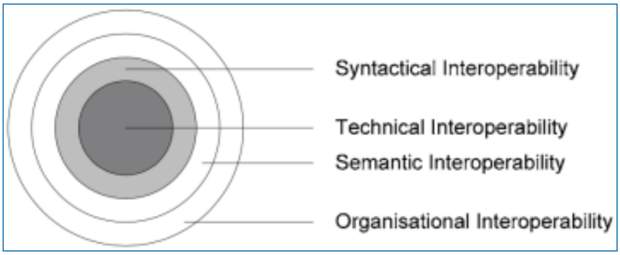


Figure 2.1. Different Interoperability levels according to [6].

Concepts and Definitions

Interoperability

There is no unique definition of interoperability in the literature since the concept has different meanings depending on the context. As a matter of fact, according to

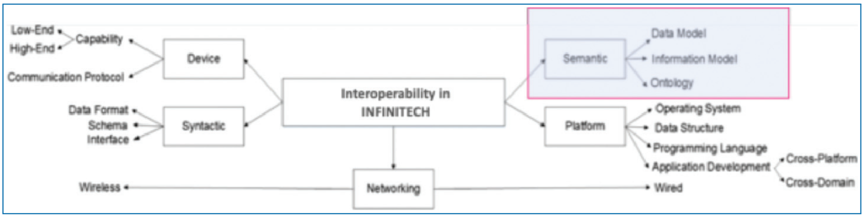


Figure 2.2. Interoperability in INFINITECH perspectives and Task 4.1 main focus [8].

ISO/IEC 2382-01 [1] interoperability is: "The capability to communicate, execute program, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units". According to Next Generation Networks (NGN) from ETSI's technical committee TISPAN [2], interoperability is: "the ability of equipment from different manufacturers (or different systems) to communicate together on the same infrastructure (same system), or on another". EICTA defines interoperability as [3]: "the ability of two or more networks, systems, devices, applications or components to exchange information between them and to use the information so exchanged". Although the particular definition of interoperability is always about making sure that systems are capable of sharing data between each other and to understand the exchanged data [4]. In this scenario the word "understand" includes the content, the format, as well as, the semantic of the exchanged data [5]. Interoperability ranges over four different levels [6] namely:

I. physical/technical interoperability: concerns with the physical connection of hardware and software platforms;
II. Syntactical interoperability: concerns with data format, i.e. it relates on how the data are structured;
III. Semantic interoperability: concerns with the meaningful interaction between systems, devices, components and/or applications; and
IV. Organizational interoperability: concerns with the way organizations share data and information.

Interoperability in INFINITECH

The first three interoperability levels are part of the INFINITECH platform and handled in Task 4.1. INFINITECH Semantic models and Ontologies are the final result of an exercise that takes as inputs physical and syntactical interoperability aspects already analysed in WP2 Task 2.1 – User's Stories and Analysis of Stakeholders' Requirements, Task 2.5 – Open Banking APIs, Testbeds and Data Asset Specifications, Task 2.6 – Specification and Design of Integrated Data Models and Task 2.7 – Reference Architecture for BigData, AI and IoT in Financial Services Industry.

As stated in [7], nowadays ICT solutions – in the most desperate context of application from e.g. manufacturing, healthcare, automotive, white goods, logistics, finance, etc. – comprise several distinct elements – e.g. devices, communication infrastructures, services, applications etc. – typically distributed and heterogeneous that need to cooperate and communicate with each other. However, communication between two systems is more than the particular network protocol to be used. Several aspects need to be considered whenever a communication channel between two systems needs to be established. As a matter of fact, the information flow within an ICT system and/or platform ranges from information detection from the data extraction, data transformation, data provisioning, data processing and data usage. In such a context, interoperability represents the enabler and the facilitator for this flow. As shown in Figure 1.3, interoperability can be seen from different perspectives, however Task 4.1 is restricted to discussing the semantic interoperability and thus data models, information models and ontologies.

Semantic Interoperability

Semantics plays a main role in interoperability for ensuring that exchanged information between counterparts are provided with sense. For Computer Systems, this notion of Semantic Interoperability translates in the ability of two or more systems to exchange data between them, by means of adopting it with precise unambiguous and shared meaning, therefore allowing its readily access and reuse.

Since around the nineties of the past century, the emerging concept of Semantic Web [9], coined by World Wide Web (WWW) founder Tim Berners-Lee, has been conducted by an exhaustive research and industry applicability, turning itself has base fundamentals to Semantic Web Services and the latest Semantic Internet of Things (IoT) concepts [10–12]. All of them aim to carry out collaboration across semantically heterogeneous environments, contributing to a connected world of consuming and provisioning devices that can potentially exchange and combine data to potentially offer new or augmented services. However, accomplishing this vision has raised several challenges due to the varied standards, legacy systems constraints, tools, etc. currently in use worldwide.

The Semantic interoperability process can, therefore, focus on different viewpoints of semantic aspects, such as the exchanged data description or the systems interaction terms. As example, the interoperability specification beside defining the meaning of a given sensor, it can also provide information on the units of such value or what protocols to use in order to connect and extract the value from the provider device.

Semantic Models

The provision of semantic information modelling can be granted with several types, including key-value, mark-up scheme, graphics, object-role, logic-based and ontology-based models [13]. From this set, the keyvalue type offers the simplest data structure but lacks expressivity and inference. On the other hand, the ontology-based model provides the best way to express complex concepts and inter-relations, being therefore the main trend model used for elaborating semantic models.

Ontologies

Since semantic web has started to gain shape, its inherent semantic interoperability has been mostly grounded on the use of ontologies for knowledge-representation basis. In this sense, usually there exists a top-level ontology (or domain ontology), and multiple sub-domain ontologies, each one representative of a more specific domain. With the use of ontologies, the entity is provided with comprehension [14].

Semantic Annotations

Semantic annotation is the process of attaching additional information to any element of data encompassed in some sort of document. Ontologies on their own are not sufficient to fulfil the semantic interoperability requirements to enable data readability by machines, as there may be differences and inconsistencies. Semantic annotation has been widely used to fill this gap by creating links between the disparate ontologies to the original sources [15].

DOI: 10.1561/9781638282310.ch3

Chapter 3

INFINITECH Implemented Solutions

3.1 Configurable and Personalized Insurance Products

Pilot #13: Configurable and Personalized Insurance Products for SMEs

The pilot's objective is to obtain forms of profiling of insurance needs for small and medium enterprises, in order to know their risks better and in this way to be able to offer a selection of products and coverage in a personalized way. The pilot 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 (Figure 3.1). 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. This is based on the collection of information from these companies in open and alternative data sources to those traditionally used by insurers. To achieve it, the platform's robots filter and track the company's fingerprint on the internet in various open sources, ranging from social networks, public official records, opinions web, the company's own web pages, etc.

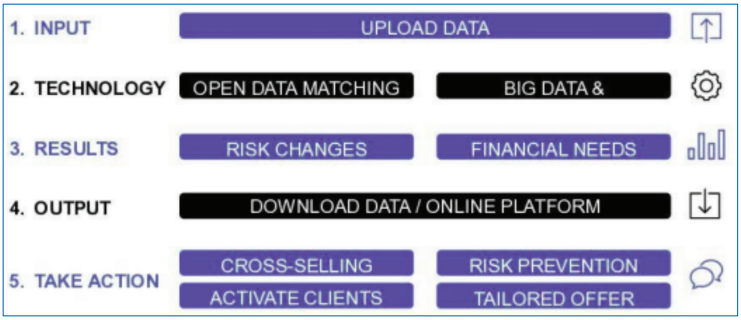


Figure 3.1. Pilot #13 summary.

Pilot #13 will monitor risk's changes, so it will be able to radically improve the risk management that companies (SMEs) face in the development of their daily activity. The indicators will be based on information from each of the companies coming from online sources that will give information about the digital presence and activity os those companies like activity, business volume, participation in social networks, number of employees, use of ecommerce, payment platform etc, etc (Figure 3.2). The company to be analysed does not need to provide much information, developed tools are in charge of searching and gathering information related to his company from many sources. In this way, risk profiles of each of the companies analysed will be generated, allowing to customize the product offering and to make a permanent automated risk management. But this is not the only usage of data, insurance companies will use this information, resulting on better customized products.

Technological components and Services

This book series links the shown software components with the corresponding RA layers, providing some details about their implementation (Figure 3.3).

- Data Sources (infrastructure). To obtain the data from the information sources we will use the automatons developed by Wenalyze, based on extensions and instances.
- Data Management (Data Collection and aggregation), For data management we will rely on LeanXcale with its non-relational databases, its data manager and its polyglot.
- Analytics, For the application of the models developed in ML we will use the Wenalyze platform that will connect with the testbed in NOVA through Javasceript.
- Connectivity, Finally the connectivity is foreseen through an API rest, but to facilitate the realization of the PoC also has developed connectivity through the use of browser, access by users and password or the upload of files in CSV format.

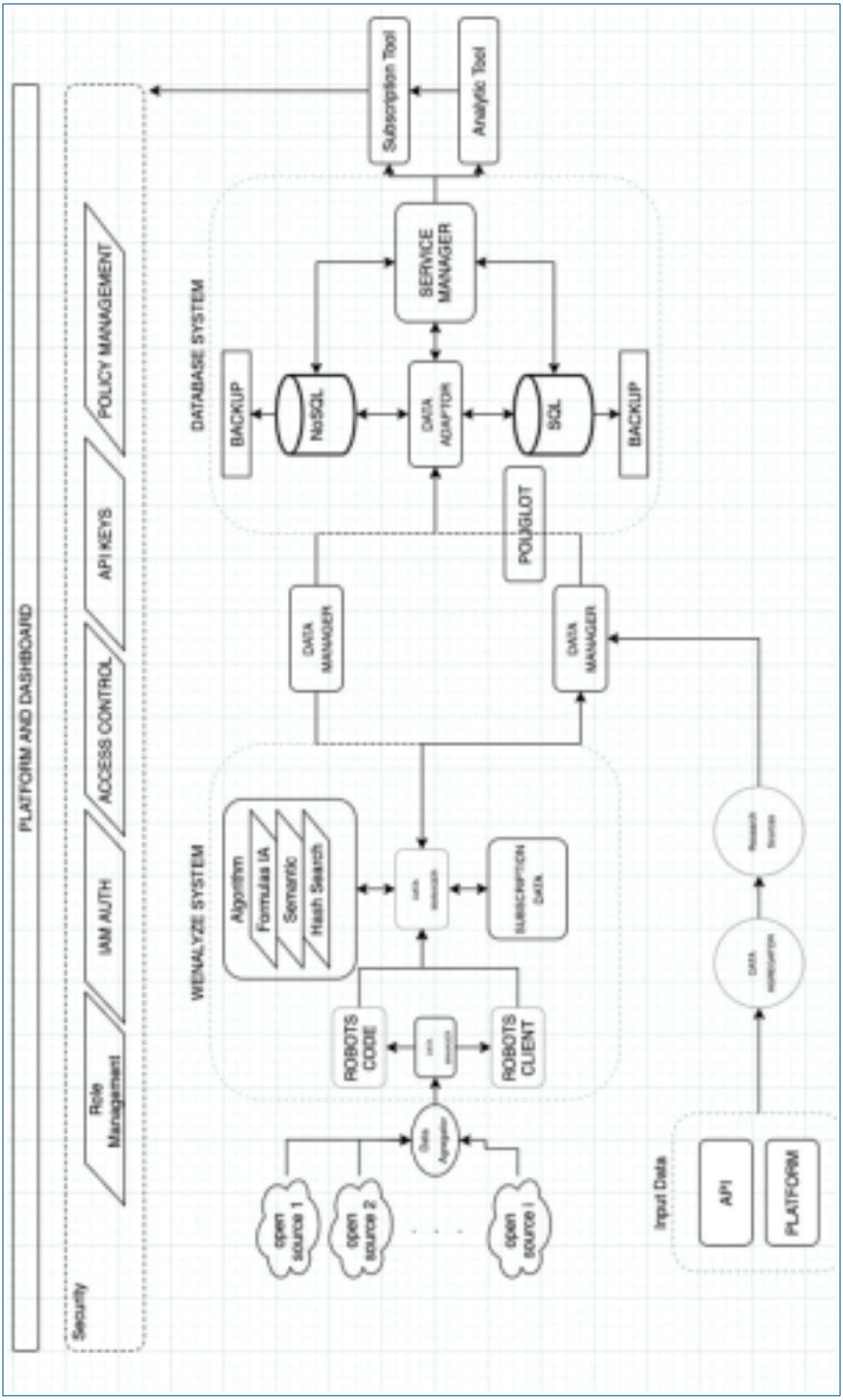


Figure 3.2 Pilot #13 Platform schema.

Hw Requirements	CPU´s 8 cores	Architecture, X86_64	Memory, 32 Gb of RAM	Storage, 500 Gb	Storage type, SSD	Network performance 25 Gigabit
Sw Requirements	Linux operating system (Debian like Ubuntu, no GNU)	SSH access	Node.js and NPM	GITHub	NginX server	

Figure 3.3. Pilot #13 hardware/software requirements.

Testbed

The 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. The infrastructure that will be used will be place in UNINOVA.

- In Nova hosting just will be implement the data base system by LeanXcale.
- The solution consists in make transactions and be storage and manage in non-relational databases.

Other non-technical requirements

A wealth and variety of important data is essential for the proper functioning of pilot #13. The data are obtained from open sources through which we can obtain different information related to the real time activity of the companies. We will use two sources of data, one that is available internationally and a second that must be incorporated in each of the countries. These secondary data sources per country are not always available with the same information and this, although avoidable, can complicate the development of the pilot.

Implementation of a first Proof of Concept

This service makes it possible to monitor the risks of SMEs now and in the future and therefore improves the control of the accident rate, the renewal of insurance policies and offers personalised insurance cover (Figure 3.4).

The RA that will be used are Data processing and data analytics. Related to the information that will be use as a input in the proof of concept will be, SMEs website data, opinion platforms, business directories, social media and ecommerce platforms (Figure 3.5). The PoC will be run based on data recovered from Spanish companies.

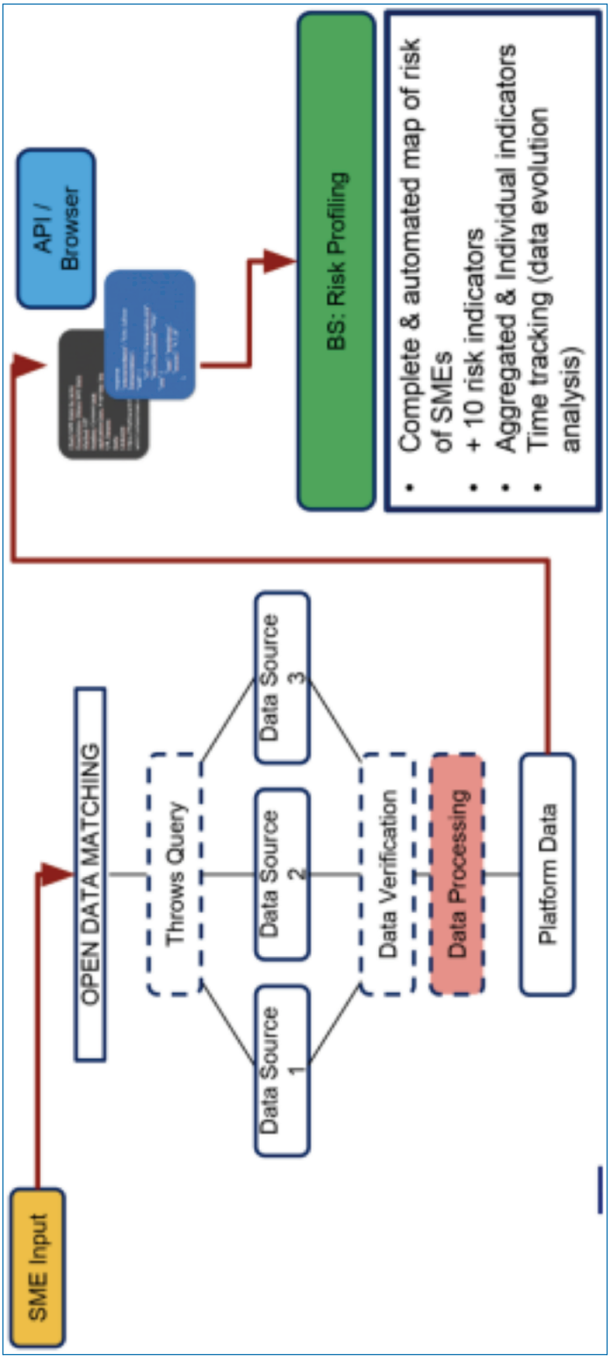


Figure 3.4 Pilot #13 implemented proof of concept.

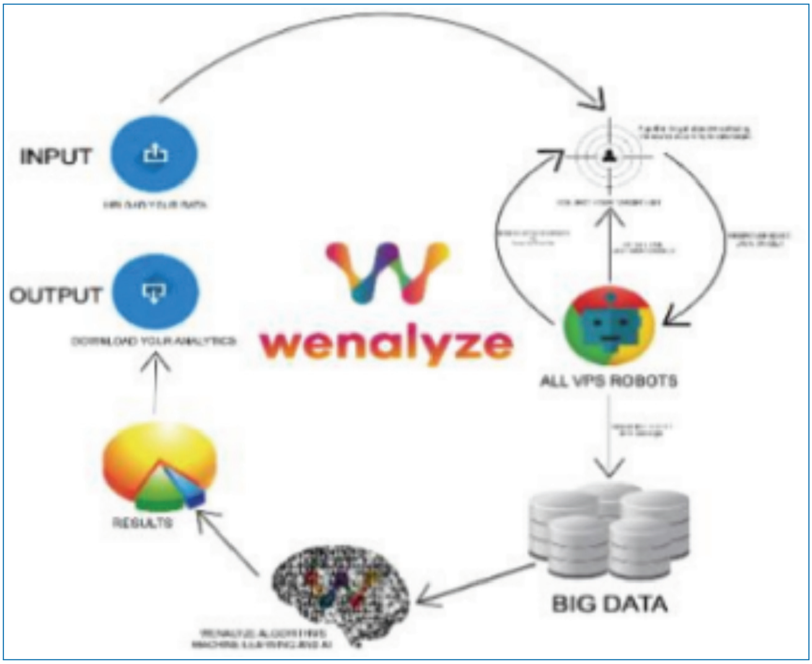


Figure 3.5. Data analyzing routine for Pilot#13.

Expected Outcomes

- Better knowledge of the behaviour of SMEs in relation to the risks they face in their activity.
- Reduction of the necessary information of introduction manual for the quote of policies.
- Increase the automation of the level of risk determination and of the coverage and services that are adapted to the needs of each SME based on their activity and risk.
- Design of insurance products adapted to the needs of SMEs.

Datasets

Data will be extracted from open sources such as company websites, official registers, social networks, opinion forums, etc. Data will include 150.000 SME targets with 50.000.000 data fields.

- SMEWIF: SMEs website information and functionalities. Description of the text containing in the website of the companies, services and structure of the company.
- ROPS: Review and opinions platforms. Reputation information and opinions of clientes about productos ans services.

- EUBD: European SMEs Business Directories. Oficial and legal information about the comanies, social object, activitie, other companies where they have equity.
- GIO: SMEs geolocation information and characteristics, images and geographical information.
- S MSIP: Social media SMEs information and presence.
- I&R: Key performance indicators and insurance needs.

The Pilot will also use synthetic data. P13-Alternative/automated insurance risk selection – product recommendation for SME SMEs synthetic raw data.

Data Produced

The output will be the ERP (Enterprise Risk Profile) and EIAU (Enterprise Insurance Automated Underwriting).

With these two outputs, not only the risk profile and its levels of any SME company are obtained, but also the information for the automation of the subscription and the application of rules to obtain the price automatically.

Explainable Workflow

Pilot #13 is a "Big Data" data analysis platform applying ML (Machine Learning) and AI (Artificial Intelligence) technologies to better predict the insurance needs of SMEs.

Well, this system must be prepared to offer a commercial use to the companies, so it must have a user interface so that they can manage the information (Frontend) and a management layer at a logical level (Backend).

The companies (enterprises) will access our platform through a registration process and subsequent validation by assigning a package of number of customers, the basic and commercial information will be recorded in Amazon Cognito and the logical information of the company will be recorded in a table of DynamoDB called Enterprises.

With regard to the use of the information by the companies, the user must load the information they have stored in their systems in our platform, this will receive the name of raw data (crude-data). The raw data will be uploaded to the platform as structured information in CSV format. The companies that use our service will have a limited amount of clients loaded in crude-data, for this, the fields of the Enterprises table, limit, clients_uploaded, total_clients_uploaded will be used in a monitored way.

Each row of this document will identify a client, which can be target in different sources of information on the Internet and other open sources in real time, depending on the information available (the quality of information depends on the company), which will be recorded in the DynamoDB Targets table (Infrastructure).

Once the robots can obtain the least updated target for a particular source, they will proceed to obtain the information and subsequent storage in big data, this information container will be in Amazon S3 in a loop called big-data as well as stored in the folder with the name of the company's identifier. The information obtained from the source will be stored in the folder mentioned above in a JSON document whose name will be the user's identifier. (Data Management) From this layer the analysis algorithms will be applied and the results of the analyzed companies will be shown with the indicators of risk levels and the configuration and automation of the underwriting obtaining the ERP (Enterprise Risk Profile) and EIAU (Enterprise Insurance Automated Underwriting) (Data Processing and Visualization).

It is important to note that the quotes to services provided by AWS are for illustrative purposes and provided they have the same technical and technological characteristics they can be replaced by another supplier, as in the case of the project under consideration may be NOVA and LeanXcale.

Logical Schema

The following Figure 3.6 provides an INFINITECH-RA compliant logical view of the logical architecture of the pilot.

Pilot's Reference Architecture and main data flows have been presented. In summary, the main components to be developed in this pilot are:

- Data Sources layer, that through information collection, select and obtain the information from dozens of sources in an efficient way, minimizing the necessary computation.
- A Data Management layer, that selects, captures, and curates the data sources required to implement the pilot's functionalities. This information management allows data collection sources to efficiently dump the data into non-relational streaming databases.
- An Analytics block, fed by the data layers, where different ML/DL technologies and visualization tools will enable data monitoring, analysis, and exploitation. Two main AI models will be developed to cover pilot's uses cases.

In pilot 13 the main participants are Wenalyze and LeanXcale. LeanXcale will provide the Data Management and Data Processing components, while Wenalyze will provide the infrastructure for obtaining data from open sources, the Analytics part, User interaction and the Visualization part. The end users of the information obtained and processed by the platform are insurance and reinsurance companies and banks. Banks regard the sale, underwriting and control of risks from their business clients and SMEs. First contacts with different insurance companies have already been made. The comments have been very positive, and the pilots would

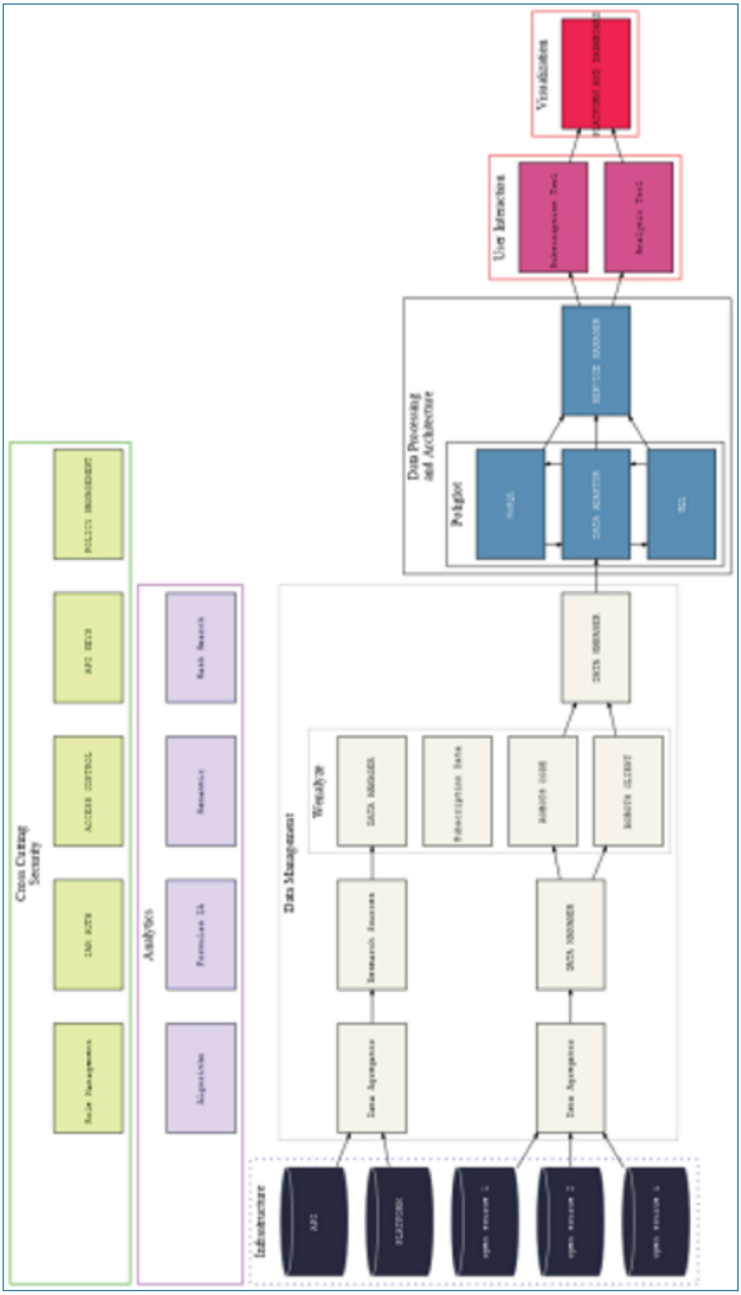


Figure 3.6 Alternative/automated insurance risk selection- product recommendation for SME pilot in-line with IRA.

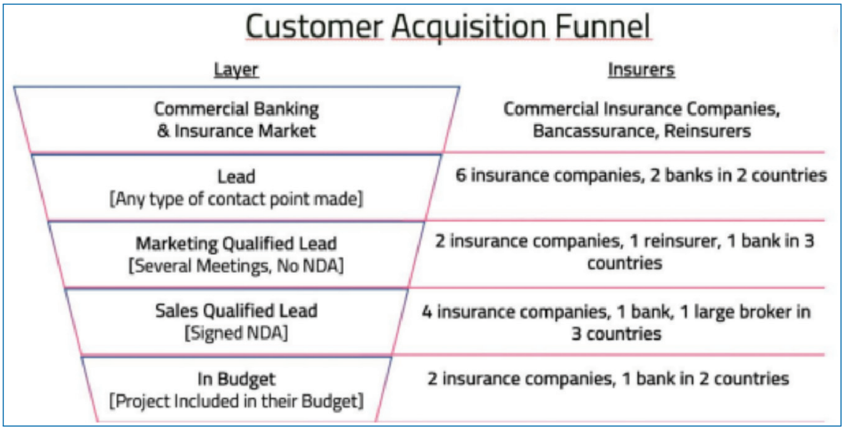


Figure 3.7. Pilot #13 customer acquisition funnel.

start once the algorithms and the platform are implemented. Also, a communication and conversion funnel has been created and is being distributed in both European Union countries and the United States. At present, 16 insurance and banking entities from eight countries are in this funnel and are in the process of marketing qualified lead to sales qualified lead. Different proofs of concept have been already agreed. The actual conversion funnel regarding end user is (Figure 3.7):

The development of the pilot is very positive, and it is expected to be completed in the time foreseen by the consortium.

Components

The following INFINITECH component will be used as part of this pilot:

- Data Acquisition Layer (Data Ingestion in RA); The data acquisition layer is composed of microrobots that roam the data sources. The deployment of the micro-robots is discretionary depending on speed and analysis needs, being a flexible and scalable deployment;
- INFINISTORE (HTAP data store and the polyglot engine) (Data Management in RA);
- Analytics Layer (Analytics and Machine Learning in RA);
- Connectivity layer through API-Rest.

Conclusions – Issues and Barriers

At this moment, the Pilot #13 is progressing according to the plan. The intake of information is neither presenting any problem, nor the construction of the models.

The only delay with respect to the plan is the transfer of the AWS development to the Nova testbed. For the time being, this is being managed by LeanXcale, this delay should not impact on the deadlines set for the pilot.

Finally, to point out the efforts on the commercial promotion of the tool, starting with communication in forums of the sector in the European Union and obtaining the first leads for the conversion funnel.

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

The objective of Pilot #14 "Big Data and IoT for the Agricultural Insurance Industry" is to deliver a commercial service module that will enable insurance companies to exploit the untapped market potential of Agricultural Insurance (AgI), taking advantage of innovations in Earth Observation (EO), weather intelligence & ICT technology. EO will be used to develop the data products that will act as a complementary source to the information used by insurance companies to design their products and assess the risk of natural disasters. Weather intelligence based on data assimilation, numerical weather prediction and ensemble seasonal forecasting will be used to verify the occurrence of catastrophic weather events and to predict future perils that could threaten the portfolio of an agricultural insurance company. The INFINITECH AgI-module derived indices will allow and enable the agricultural insurance industry to enlarge its market, while delivering a larger portfolio of products at lower costs and serve areas, where classical insurance products could not be delivered (Figure 3.8).

Also, the aim is to define, structure and pilot test specific services for the Agricultural Insurance sector in order to better protect agricultural assets by evaluating risks in a data-driven way and to improve the business process of agricultural insurers and clients (farmers). The services tested will be (1) a mapping of risks related to agriculture in predefined markets, (2) the prediction and assessment of weather and climate risk probability and (3) a damage assessment calculator for insurance companies.

Technological components and Services

Based on the reference architecture the following components and services will be deployed and used as part of the pilot:

ICT Modules

- Octopush EO Service: Octopush EO Service is an integrated satellite derived software service, which collects earth observation, geospatial, in-situ and other geo-referenced data, it applies appropriate processing algorithms and returns the results in a ready-to-use format.
- AgroApps Weather Intelligence Engine (AgroApps WIE): The WIE is an integrated weather derived software service which collects weather information from several resources and along with the georeferenced data, it applies

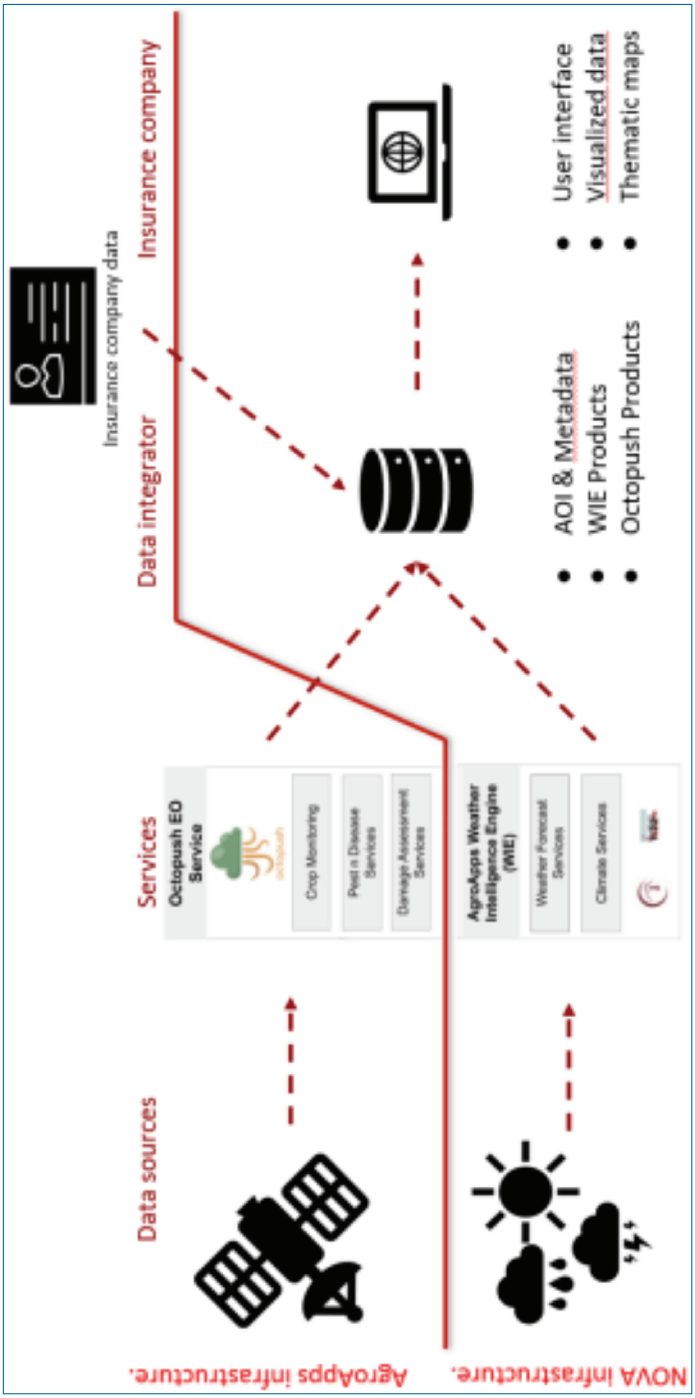


Figure 3.8 Pilot #14 testbed.

appropriate processing algorithms and returns the results in a ready-to-use format.

- Data integrator: The Data Integrator acts as a bridge between the WebGIS subsystem, Octopush EO service and WIE. It is responsible for performing the essential scheduled calls to the data providers in order to fetch and process the desired EO and weather information. It is able to run calls on demand or daily data integration tasks by retrieving EO data and weather products from Octopush EO service and WIE and transforms, binds, injects those into the WebGIS database.
- Business and Geospatial DB: Business DB offers a storage layer essential to carry the business logic and relevant information/data stored and managed by API. It also stores, retrieves and provides information related to user accounts, settings, actions and preferences. The geospatial data storage and data persistence mechanisms allows the storage of the geometries and zonal statistics and provides the essential functionality for querying and retrieving data via an API or WMP server components.
- Web Map Server (WMS Server): WMS is responsible for rendering and serving of the GIS layers to the User Interface.
- RESTful API: The API will act as a communication and data exchange bridge, that allows the platform to share processed and structured content internally, between the different components.
- User interface: The front-end user interface is the gateway responsible to present all the system data through user-friendly controls and web mapping interfaces.

Testbed

All modules of Pilot 14 services will be hosted in AgroApps premises, except the weather intelligence engine that will be deployed in UNINOVA's infrastructure. In this sense, server specifications will be defined at a later stage.

Others non-technical requirements

Besides the technical requirements for the pilot, there are also other non-technical requirements in order to test the application successfully. These requirements mainly relate to the data provided by the agricultural insurance company:

- In the past, we observed that the quality of the data provided by agricultural insurers was often poor. This is mainly due to the IT structure of the insurers, which often does not allow targeted queries at short notice.
- However, in order to apply the structured and unstructured data provided in the shared testbed by AgroApps and UNINOVA (Earth Observation (EO),

Numerical Weather Prediction Data, Reanalysis Data and Seasonal Climate Forecasts) to the insured/to be insured regions, the data provided by the insurance company needs to be accurate, timely and on a correlating spatial resolution.

- This applies not only for the clear identification of a region/field by coordinates or IDs from national databases, but also to the existing/desired form of insurance cover, the crop to be insured, average yield values and a (potential) loss history.
- If the data quality is insufficient, the national statistics office could also be consulted, e.g. for average yield data.

Furthermore, Pilot #14 make use of the respective national network of Weather stations for collecting data used in the Weather Intelligence Engine to predict weather and climate patterns.

Implementation of a first Proof of Concept

The first Pilot #14 Proof of Concept (PoC) will focus on a data processing architecture and a data analytics infrastructure to create an Area Risk Profile for the defined Area of Interest (AOI) in order to assess the risk of natural disasters and to develop a pricing framework for a drought index product.

Therefore, EO data derived from satellites and weather intelligence based on data assimilation, numerical weather prediction and ensemble seasonal forecasting will be used to verify the occurrence of catastrophic weather events and to predict future perils that could threaten the portfolio of an agricultural insurance company (Figure 3.9).

In this first phase of the pilot site preparation, the focus will be on providing solutions for users situated in agricultural insurance companies (Actuaries, Underwriters, Sales Agents, Loss Adjuster) as described in the User Stories #14.01-14.08).

By combining the components developed in the AgroApps and UNINOVA Infrastructure and the data set from the insurance company, the respective user application can be set up and tested. The results of this first PoC will help to improve the data flow and data analytics processes for the Pilot's final services (Figure 3.10).

Expected Outcomes

- Identification of areas within the large-scale pilots where crop productivity and catastrophe probability are high based on intelligent risk mapping.
- Creation of additional datasets with high predictive value for improving underwriting of agricultural risks with regard to weather and climate risk probability.
- Improved damage assessment and claims handling procedures for the insurance industry to increase the efficiency of calculating indemnity pay-outs.

Figure 3.9 Pilot #14 proof of concept.

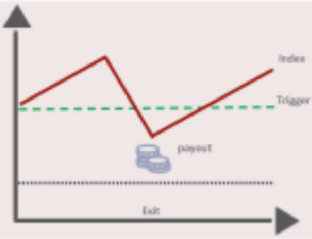
Term	Index
Insured value	**Index value,** derived from historical average (high correlation with field variable(s) (e.g. **weather or yield data**)
Client risk retention	% of historical average (area above trigger)
Payout limit	Maximum payout: % of historical average (area between trigger and exit)
Framework for index insurance pricing	

Figure 3.10. Pilot #14 Data requirements for developing a pricing framework.

Figure 3.11 shows the workflow for configurable and personalized insurance products for SMEs and agro-insurance.

Datasets

The main data source for the pilot is produced by satellite and a weather intelligence engine. The Earth Observation (EO) data will be derived from the satellites Sentinel-1,2,3, LandSat-8, MODIS and PROBA-V). Also, numerical weather predictions for the pilot areas (gridded data) are generated each day and will replace the previous prediction. Lastly, gridded climate indices based on ERA-5 Land and ERA-5 Reanalysis Data will be used for the pilot.

Following, the list of datasets is presented (GEN):

- Gridded Climate Indices (1/1/1979 to 31/12/2019): Climate Indices based on the ERA-5 Land and ERA-5 Reanalysis Data.
- EO Data: Earth Observation Data (Sentinel-1,2,3/LandSat-8, MODIS, PROBA-V) for remote damage and crop loss assessment.
- Numerical Weather Predictions: Very High-Resolution Weather Predictions for the Pilot Areas.
- Temperature, precipitation, evapotranspiration, soil moisture, crop growth data, crop water requirements, wind speed, relative humidity,solar radiation, snow cover, snow depth, etc.
- Hail data.
- Loss history.

Data Produced

The data produced will result in a solution for Agricultural Insurance companies allowing them to efficiently couple EO satellite data and weather/climate data

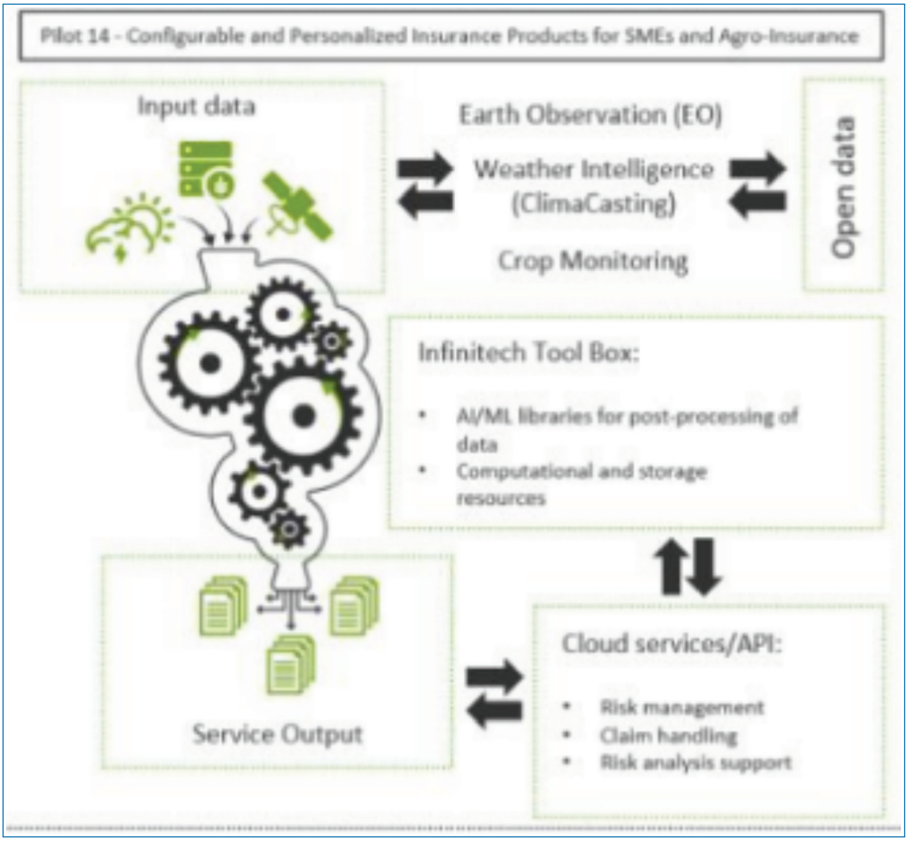


Figure 3.11. Configurable and personalized Insurance product for SMEs and agro-insurance.

with any type of complementary data (from separated drone shots to ultra-high-resolution SAR imagery). The INFINITECH AgI module will enable Insurance companies to alleviate the effect of weather uncertainty when estimating risk for AgI products, reduce the number of on-site visits for claim verification, reduce operational & administrative costs for monitoring of insured indexes and contract handling, & design more accurate & personalized contracts. By deriving impartial indices on top of a multitude of data, the module will allow insurers to reduce significantly the time needed for handling and verification of claims and the costs imposed by fraud, moral hazard and adverse selection.

Explainable Workflow

The data produced by the Octopush EO Service (Crop Monitoring, Pest & Disease Services, Damage Assessment Services) and the Agro Apps Weather Intelligence Service (Weather Forecast Services, Climate Services) is pipelined into a Data integrator. The integrator feeds back Metadata and information on the Area of Interest

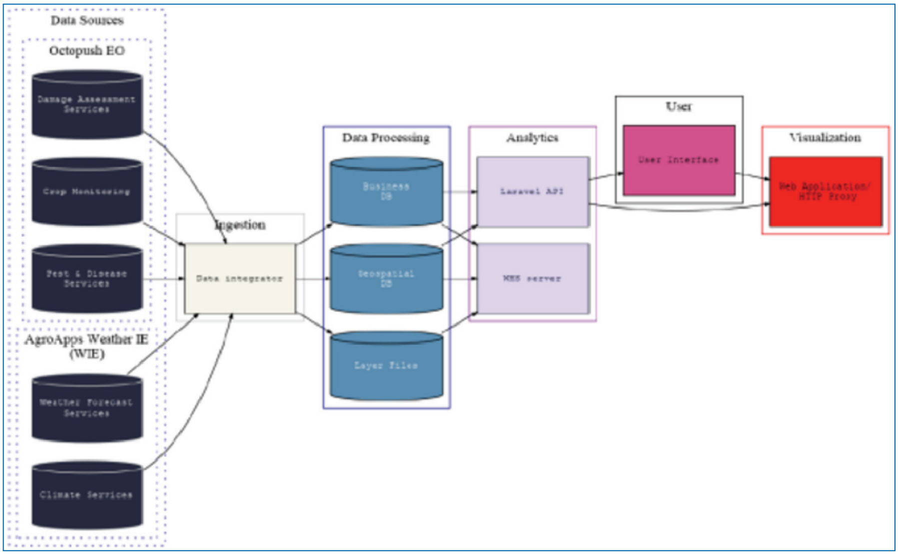


Figure 3.12. Bigdata and IoT for agricultural insurance industry pilot pipeline in-line with the IRA.

(AOI) to the data producing services. The integrator itself is retrieving the AOI & Metadata from a Business DB storage layer. The geospatial data storage and data persistence mechanisms allows the storage of the geometries and zonal statistics and provides the essential functionality for querying and retrieving data via an API (alerts) or WMS server components (vectors source). The WMS server is then responsible for rendering and serving of the GIS layers to the User Interface. The restful API will act as a communication and data exchange bridge, that allows the platform to share processed and structured content internally, between the different components. The front-end user interface is the gateway responsible to present all the system data through user-friendly controls and web mapping interfaces.

Figure 3.12 depicts the logical schema for Bigdata and IoT for agricultural insurance industry pilot pipeline in-line with the IRA.

Logical Schema

AgroApps is developing the entire infrastructure for the pilot #14 data products, based on the reference architecture starting from data collection from different sources over processing and analytics to user interface & data visualization. The ongoing development of the service module is based on scientific research in the field of agricultural insurance, climate & weather risk modelling and the most recent evolutions in the area of remote sensing technologies. The reason for this is that these three areas will play a crucial role for the future of agricultural insurance providers in order to tap new markets, provide better risk transfer solutions and make insight-based strategic decisions. To meet the demands of this rapidly evolving field, it is necessary to follow these current developments.

As described in the User Stories, the service module is mainly designed for staff working in the underwriting and sales department of agricultural insurance companies (majority of User stories serves this group of end-users). However, within this departments, there are several roles who can benefit from the services provided by Pilot #14. First of all, Actuaries (business professional/mathematician who analyses the financial consequences of risk by using statistics) are able to improve their data set for risk pricing and product development based on the data retrieved from the service module. Based on this information, Underwriters can better evaluate the risk and exposure of potential clients (crop monitoring) and hence make the overall insurance portfolio more resilient by at the same time increasing the outreach to clients (farmers). Additionally, Sales Agents can identify areas where to prioritize sales activities without increasing the cumulative risk since they are aware of e.g. regional risk profiles.

Lastly, with the support of data derived from the Octopush EO (damage assessment services), loss adjusters have additional information to make the on-farm process of loss adjusting more efficient and for certain perils conduct this process remotely via the service module (without visiting the farm/respective field).

In addition to the implementation within insurance companies, at a later stage of the project other users in the insurance value chain can also be considered as end users.

A first contact inside an insurance company in the Area of Interest (Serbia) has been made and immediately generated interest because of the benefits the Pilot #14 service module has to offer. The feedback on the presented services was very positive, just a final decision by the management is outstanding.

As in this very first stage of the preparation of the pilot site the receiving of an appropriate and high-quality dataset from the pilot insurance company and the application of the services described in Pilot #14 have highest priority, there are no training plans developed for deployment to the final user yet.

However, for the internal deployment at the final pilot site, Pilot #14 can provide an independent web-based user interface for the end users to access the service module via their browser.

Components

The pilot comprises ICT modules and services for the insurance sector.

- ICT Modules:
  - Octopush EO Service (Data Source in RA): Octopush EO Service is an integrated satellite derived software service, which collects earth

observation, geospatial, in-situ and other geo-referenced data. It applies appropriate processing algorithms and returns the results in a ready-to-use format.

- AgroApps Weather Intelligence Engine (AgroApps WIE) (Data Source in RA): The WIE is an integrated weather derived software service which collects weather information from several resources and along with the geo-referenced data, it applies appropriate processing algorithms and returns the results in a ready-to-use format.
- Data integrator (Data Ingestion in RA): The Data Integrator acts as a bridge between the WebGIS subsystem, Octopush EO service and WIE. It is responsible for performing the essential scheduled calls to the data providers in order to fetch and process the desired EO and weather information. It is able to run calls on demand or daily data integration tasks by retrieving EO data and weather products from Octopush EO service and WIE and transforms, binds, injects those into the WebGIS database.
- Business and Geospatial DB (Data Management in RA): Business DB offers a storage layer essential to carry the business logic and relevant information/data stored and managed by API. It also stores, retrieves and provides information related to user accounts, settings, actions and preferences. The geospatial data storage and data persistence mechanisms allows the storage of the geometries and zonal statistics and provides the essential functionality for querying and retrieving data via an API or WMP server components.
- Web Map Server (WMS Server) (Analytics and Machine Learning in RA for Geoserver and Interface for Apache Tomcat and RESTful API): WMS is responsible for rendering and serving of the GIS layers to the User Interface.
- RESTful API (Interface in RA): The API will act as a communication and data exchange bridge, that allows the platform to share processed and structured content internally, between the different components.
- User interface (Interface in RA): The front-end user interface is the gateway responsible to present all the system data through user-friendly controls and web mapping interfaces.

- Services for the Insurance Sector:
  - Remote Damage Assessment for drought and hail
  - Flood and wildfires mapping
  - Short and medium range weather forecasts
  - Seasonal Climate Forecasts of Agroclimatic Indicators
  - Climate Risk Assessment

Conclusions – Issues and Barriers

Based on the work done so far, we are aware of the following challenges:

- Provision of UNINOVA IT infrastructure to run the testbed as foreseen in WP6 to enable enough computing capabilities to run the Weather Intelligence Engine.
- Receiving appropriate and high-quality dataset from insurance company for the PoC and ongoing activities.
- Identifying the right correlations between the data provided via the testbed and the dataset for the respective AOI in order to draw the right conclusions for the are risk profile and hence the insurance pricing for a drought index insurance product.

To conclude the status of the pilot site preparation it can be stated that the partners involved in Pilot #14 (AGRO,GEN) are in close contact with Nova as the Testbed provider and are awaiting their notification of a successful set-up of the shared testbed infrastructure in the coming weeks.

Furthermore, a good relationship was established with two agricultural insurance companies which would be able to provide the for this pilot required insurance company data for the defined AOI.

Both insurance companies approached are composite insurance companies, hence not only focusing on agricultural insurance. The Agricultural Line of Business (LOB) of insurance companies in most markets is not the most profitable one. On the one hand, the service module developed in Pilot #14 will contribute to exploiting untapped market potential and new/innovative business and product opportunities, on the other hand though, it is difficult to convince the Management and the Underwriting Departments of all benefits.

Therefore, GEN is using its business relationships to directly talk to potential decision makers. To convince those decision makers, GEN has pitched the overall goal of the INFINITECH project together with the objectives of Pilot #14, the structure of the pilot in general terms, data requirements and lastly the benefits in the short, medium, and long term for the pilot user (as defined in the user stories for agricultural insurance companies).

In a next step, decision-makers will be given time for feedback and questions. Afterwards, a meeting together with the Tech-Proxy of Pilot #14 (AGRO) will be organized to dive deeper into the set-up of the service module, the capabilities of the module to provide additional data and to discuss the data requirements to be derived from the insurance company.

This process is essential for reaching out to potential pilot users in order to test and evaluate the added value of the service module (based on defined user requirements) developed for the specific business processes in agricultural insurance.

3.2 Personalized Retail and Investment Banking Services

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

This pilot would examine how banks and FinTech(s) in collaboration with research organisations and NGOs can develop an AI driven capability using transactional data generated by the financial activities that identifies money-related profiles based on the transactional data generated. Data profiles e.g. from social media then can be associated to human profiles base on their financial activity. These profiles will be built into the available AI engine and will be combined with existing technology and data sourced from the TAH human trafficking platform. The results will produce a complete picture of people profile, people trafficking routes and the corresponding money flows back to the criminal organizations.

This pilot will utilize a combination of open-banking, social and internal-bank generated data sources to establish a high-volume and high-quality view of the customer to be used for a range of data analytics performed on big data platforms (Figure 3.13). The use of analytical methods could include link analysis to support permission-based customer relationship analytics on behalf of customer and bank, or transaction monitoring to support credit risk management for bank, but also that provide value for customers.

The Pilot#3 will need to simulating a data sharing ecosystem by mimicing participants in that ecosystem and provide rules of engagement and highlighting the

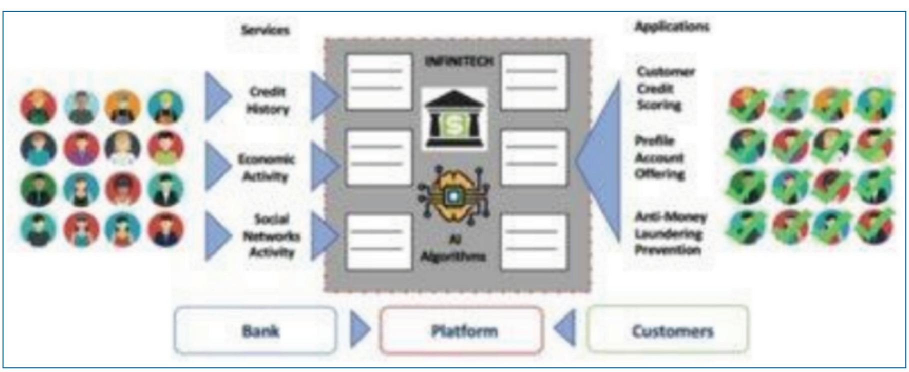


Figure 3.13. Pilot#3 workflow.

value exchanges between participants. A digital ecosystem framework is described here to articulate testbed components required.

Expected Outcomes

- The pilot will produce three data intensive systems, including a KYC system based on data sharing, a credit scoring system and an AML system operating based on semantic technologies and blockchain based data sharing. The pilot evaluation will consider KPIs associated with the speed of the processes (i.e. KYC), customer satisfaction and customer engaging in sharing data. The workflow is described in Figure 3.9.

Datasets

- Customer and Account Data (Bank Data).
- Customer to Customer Relationships (Bank Data).
- Customer Account Data (Open Banking Data).
- Other Open Customer Data (Social etc.).

Pilot #3 will consider two sources of data:

- Operational Data Sources – We will not using existing BOI Ops. Data sources, because of confidentiality issues even if anonymized and also data consistency issues. Instead Proof of concept data sources are 'synthetic' customer, account and transactions data designed to mimic real world data scenarios from financial services.
- Captured data from data entry in application including consent or metadata exhaust from sharing process.

Data Produced

- For Pilot 3 a better representation of the data lifecycle might be as follows:
- Data utilized/transferred (E.g. data sharing payload – Customer/KYC, Account or Transaction Data),
- Data transformation (e.g. any data changes),
- Data produced (e.g. new data) &
- Data deletion (e.g. revoked consent) etc.

Explainable Workflow

The whole premise of the pilot purposed is to enable unlimited use cases between any participants via a single application creating a single ecosystem. Specific back end data services might be built to support a particular use case e.g. KYC. Below, Figures 3.14 and 3.15 illustrate a data flow of KYC use case in terms of business process and data flow and technical data flow, respectively.

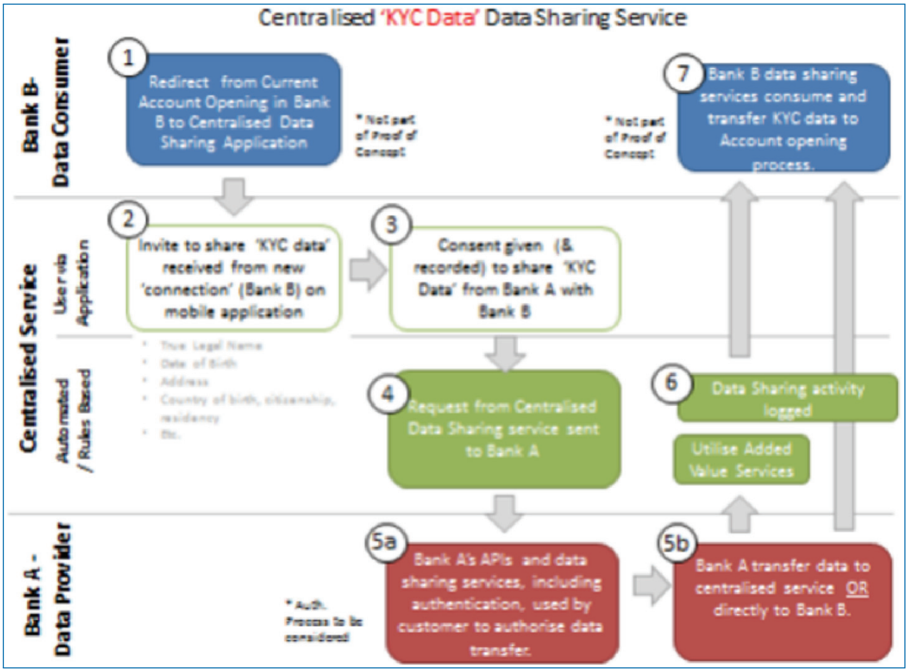


Figure 3.14. Customer-Centric data analytics pilot workflow – KYC data sharing process – business workflow.

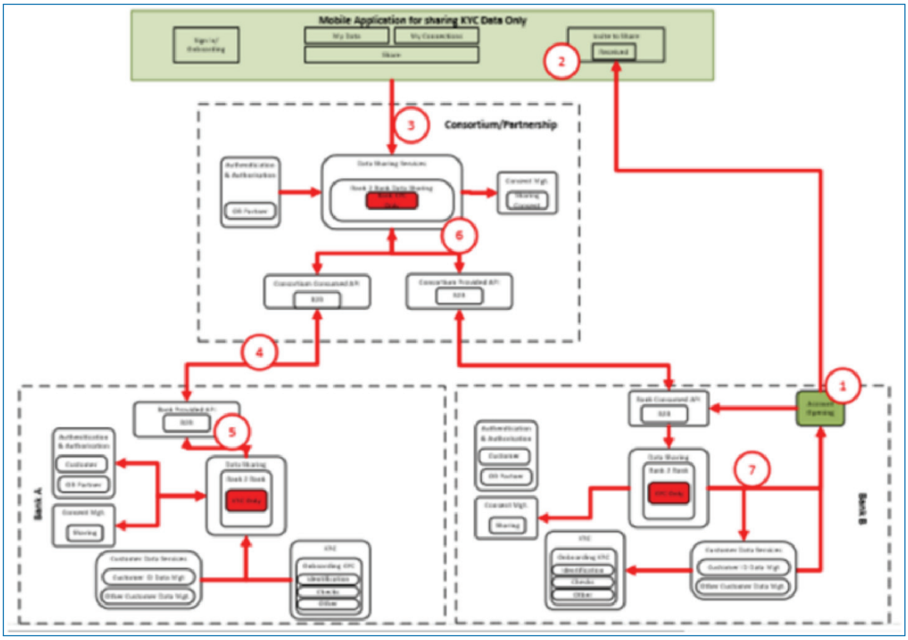


Figure 3.15. Customer-Centric data analytics pilot workflow – KYC data sharing process – technical workflow.

Logical Schema

The following figure refactors the components of the above-listed workflows towards illustrating the pilot logical architecture in-line with the INFINITECH-RA.

Pilot #4: Personalised Portfolio Management – Mechanism for AI Based Portfolio Construction

The main goal is to develop and adapt within SaaS based Privé Managers Wealth Management Platform a Portfolio Optimization algorithm (further on called Privé Optimizer or "AIGO"), as well as improving and expanding its capabilities as an artificial intelligence engine to support better investment propositions for retail clients.

This pilot 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 (Figure 3.16). AIGO allows for a variety of use cases which cater to the needs of financial advisors, end-clients and financial services companies. The innovative AIGO genetic algorithm can be used for proposing investments and evaluating them given an easy-to-use, personalizable set of criteria, in the form of so-called fitness factors (Figure 3.17). These fitness factors will be used to generate "health" scores for portfolios, which are used to define the "fittest" investments.

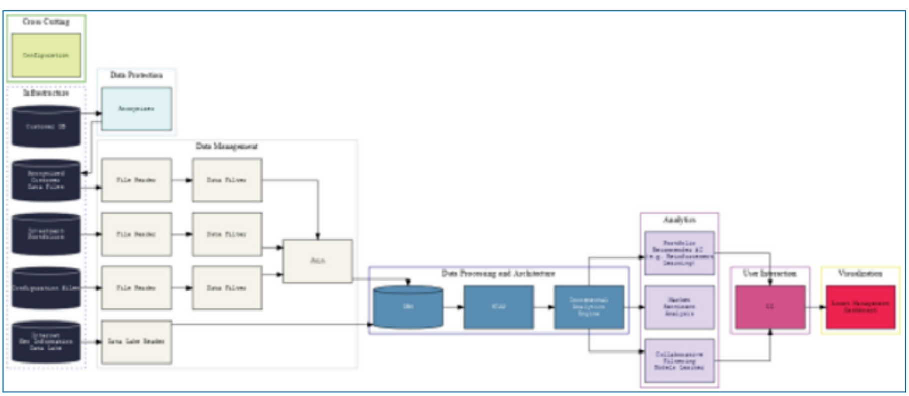


Figure 3.16. Customer-centric data analytics pilot pipeline in-line with IRA.

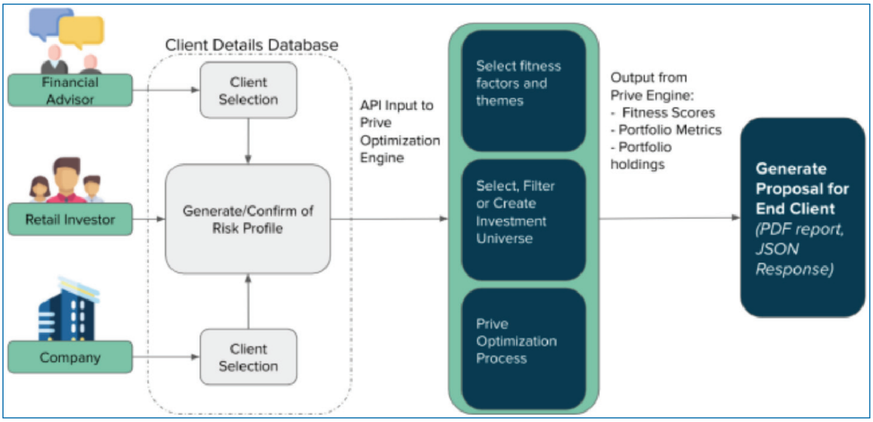


Figure 3.17. Pilot #4 roles and services.

Starting from a client's cash pool or current investments/portfolios, the user will select the fitness factors and constraints or preferences to perform the portfolio construction, based on the client's risk profile and preferences. The optimisation tool that will be developed from the Pilot, will run on a pre-set universe of assets taking into account all the input data and constraints. The AI genetic algorithm will generate a new proposal, where the selected preferences and risk parameters have been recognised. The optimisation tool can be run multiple times, after the necessary changes in initial parameters are made. In this context the main innovation of the pilot lies on the applicability of AI technologies to build customized portfolios (Private Banking for everyone).

Technological components and Services

The High-Level Architecture presented in Figure 3.18 presents the software components that build the Pilot's use cases. This figure has been used in D6.1 Testbeds Status and Upgrades to identify hardware requirements, and in D2.5 Specifications of INFINITECH Technologies to describe the technologies behind its principal components.

This book series links the shown software components with the corresponding Reference Architecture layers, providing some details about their implementation. In this sense:

- Data Collection (Data Management layer of the RA) based on customers cash pool or current investments/portfolios data.
- Customers' & investments/portfolio Data quality check (Data Processing layer in RA): according to specific data models, in order to perform the data preparation for portfolio construction, based on the client's risk profile and preferences.

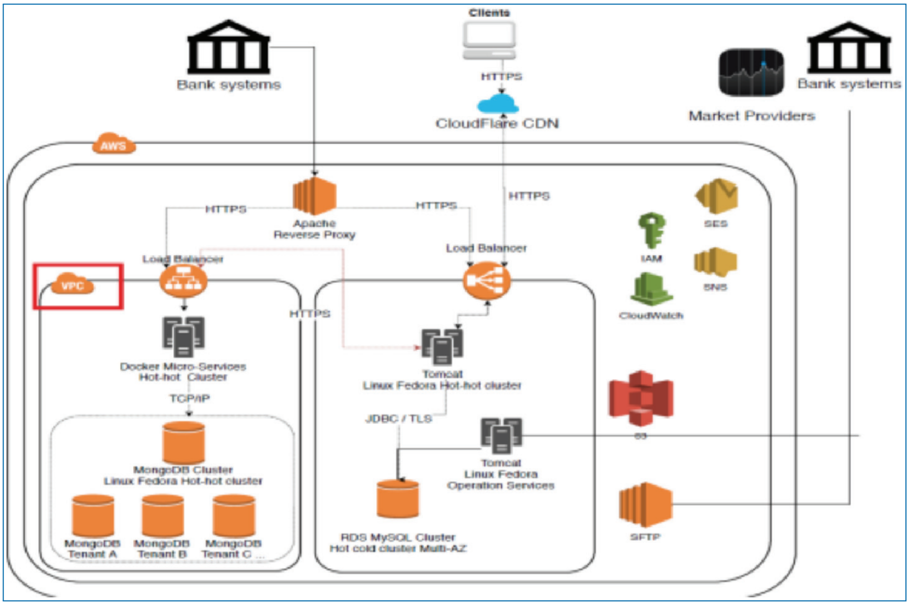


Figure 3.18. Pilot #4 high-level architecture.

- AI Based Portfolio Optimization Process (AIGO) that will be developed from the Pilot based on AI Algorithm, will run on a pre-set universe of assets taking into account all the input data and constraints, generating a new proposal, where the selected preferences and risk parameters for a specific customer.
- New proposal for the personalized portfolio will be visualized through a PDF report generation or a JSON extract that will be able to be imported in any relevant portfolio management tool.

Testbed

As indicated below, Privé will be storing its Testbed on its own Amazon Cloud in AWS with an architectural setting as indicated below.

Technical Specifications from Privé's internal Testbed

Hardware Specifications (in case of Cloud Installation, include the relative cloud configuration) 3 VM instances, each with the following: CPU: Intel Xeon 3 GHz or faster Core: minimum 2 Core 4 threads Memory: 32 GB DDR4 1600 or 1866 Hard Disk: 16 GB SSD.

Technical Architecture Diagram

Software Specifications including for each module the relative software stack that is used (e.g. Operating System, Layered Software, Application Software, Development Platform, Database, etc).

Operating System/Application Software/Layered Software

The SaaS platform runs in multiple data centers with active-active setup to achieve high availability. Privé has the following environments: DEV, SIT, UAT and PROD. Data can be transferred via SFTP, FIX or API. Most Privé APIs are REST, but SOAP and GraphQL are also supported. The architecture is based on microservices.

Operating System: Ubuntu 18.04 LTS Framework: SpringBoot: 2.2 Application Server: Tomcat: 7.0.103 Database: MySQL: 5.6.47 Database: MongoDB: 3.6 Language Runtime: Java: OpenJDK 8u242.

Development Platform (Figure 3.19)

We use html5/ReactJS for frontend. Our platform is written in Java, with Spring MVC, Spring boot, and hosted with Apache Tomcat.

Note that the requirements for Hardware specifications (e.g. RAM, No of CPUs, etc) will be required to be defined based on the requirements of the relative technology solutions (e.g. Data Management & Processing, Analytics & AI, etc) that will be used for each Testbed and relative sandboxes, in cooperation with the relevant Technical Partners.

Privé testbed is ready to be used and the tests conducted on our own AWS cloud were successful. The proof of concept will be delivered and presented showing the current API functionalities and results stored in this testbed.

Implementation of a first Proof of Concept

As a minimum-viable-product or better first Proof of Concept, Privé will be presenting the back-end /calculations capability from the AI GO (Artificial Intelligence

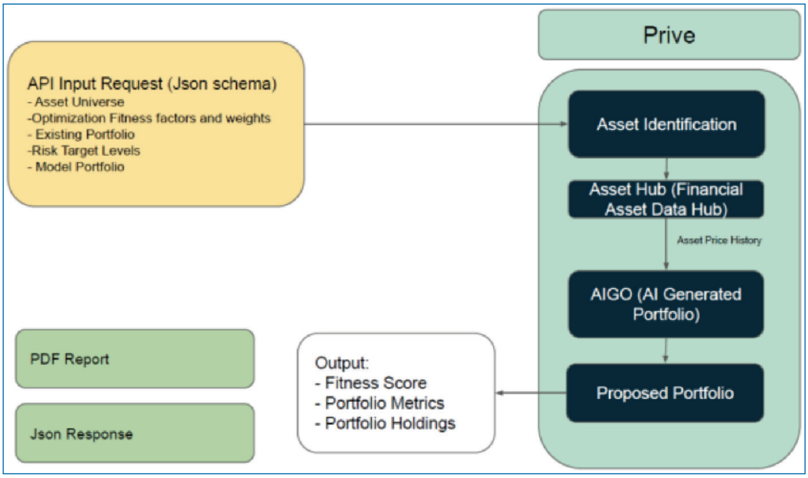


Figure 3.19. Pilot #4 components.

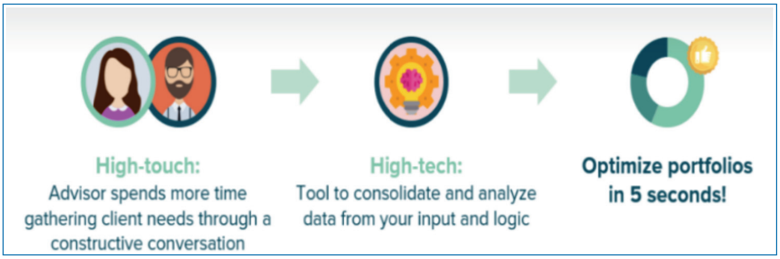


Figure 3.20. Portfolio optimization procedure.

Portfolio Construction Optimizer) via API Calls. This will consist of the optimization process presented for an example-portfolio via a couple of so-called fitness-factors which will allow to optimize a pre-given portfolio with a pre-determined investment universe. For the first proof of concept not all services or data sets described in the user stories will be implemented. The figure below highlights the PoC main components (Figure 3.20).

In this case the input will consist of an investment universe of 50 European stocks and a pre-defined portfolioexample and the out-put will result into an optimized portfolio based on the selected user preferences (fitness-factors functions for the optimizer for that matter). Both data sets and testbed infrastructure have been described in more detail above. All the inputs and outputs will be callable via API.

Expected Outcomes

- The AI Based Portfolio Construction shall enable interested advisors or end-customers, after an initial "customer onboarding" (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 following figure illustrates the portfolio optimization procedure in pilot #4. The first step is gathering client needs though constructive conversation with advisors. Then, analytical tools are applied to analyze data from input and logic. Last, the portfolio is optimized in 5 seconds.

Datasets

The data to be used by this pilot will be:

- Customer Transactions Data: customer securities and cash transactions through their deposit accounts. They are fetched directly from the Bank or an Asset Manager;

- Financial Market Price Data: price data for Stocks, Bonds, Mutual Funds and or other assets like certificates/warrants. They are fetched from several Market Data Providers;
- Financial Market Asset Master Data: asset related characteristics (e.g. expiration date, minimum investment amount, asset class breakdowns). They are fetched from several Market Data Providers;
- Customer Risk Profile Data: customer Risk Profile Data through their account data and profiling, based on B2B customers parameters. They are fetched directly from the Bank or an Asset Manager;
- Mutual Fund, ETF and Structured Products Breakdown: asset breakdowns based on bank data or market data providers breakdown. They are fetched from several Market Data Providers:
- Customer Economic Outlook: they are fetched directly from the Bank or an Asset Manager based on questionnaires and Customer Profiles;
- Single Account & Investors Data: 19484 accounts for about 15400 investors (live data) 94.407 different securities available; Investors serviced by 309 different advisor companies; Accounts in 28 different custodian banks (Data from 2019). All datasets will be stored within Privé SaaS solution in a cloud setup. Asset data and Client data are fetched from 3rd party databases and partially from selected market-data providers. Risk metrics are calculated in the historical backtesting component for each single portfolio. A Genetic Algorithm component evaluates different Fitness Factors and generates a customised portfolio proposal.

Data Produced

JSON files will be produced from Privé API (if other 3rd party solutions address to this Portfolio Optimisation functionality, and PDF files can be generated for UI display and customer documentation.

The output data consists of the single portfolio holdings, their weights and amounts to decide about the Proposed Portfolio. Fitness Factors Scores and Total Fitness Score will be output for both the current and proposed (optimised) portfolio. For both Portfolios also Risk and Return metrics will be shown: 5 year annualized return, volatility and sharpe ratio.

Explainable Workflow

Starting from a client's cash pool or current investments/portfolios, a risk profile is created or an existing one is updated (Steps 1 to 3 on Figure 3.21). Then the user will select the fitness factors and constraints or preferences to perform the portfolio construction, based on the client's risk profile and preferences (Step 4). The optimisation tool will run on a pre-set universe of assets taking into account all the input

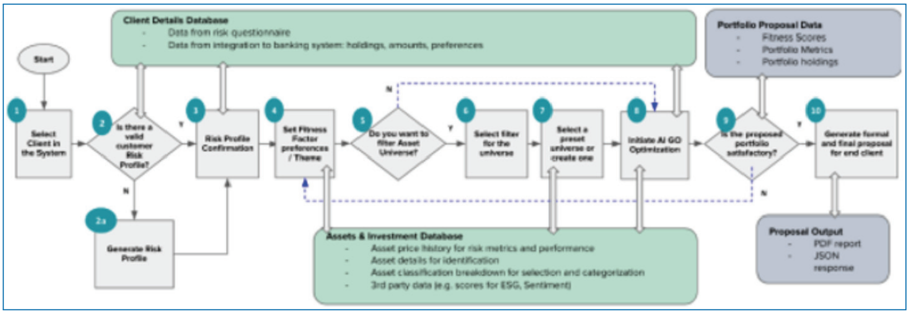


Figure 3.21. Personalized portfolio management pilot workflow.

data and constraints (Steps 5 to 7). The AI genetic algorithm will generate a new proposal, where the selected preferences and risk parameters have been recognised (Step 8 and 10). The optimisation tool can be run multiple times, after the necessary changes in initial parameters are made, based on that the proposed portfolio is satisfactory or not (Step 9). This process can result in a UI proposal or a PDF generated investment proposal.

Both inputs and outputs will be stored in Privé own cloud. Al fitness-functions within the AIGO will be callable via API based on the initial user preferences inputs. All the datasets will also be stored on Privé side for both inputs and outputs for the algorithm.

Logical Schema

An initial mapping of the explainable workflow of the pilot to the INFINITECH-RA layers and constructs is depicted in the following Figure 3.22.

Pilot's Reference Architecture can be simplified considering:

- A Data Management layer, that performs data ingestion based on cash pool or current investments/portfolios data, quality checking and harmonization of the data provided in order to be imported into the datastore for use of the pilot's functionalities.
- A Data Processing Layer in charge of homogenise and store all data collected, according specific data models to perform the data preparation for portfolio construction, based on the client's risk profile and preferences.
- An Analytics layer, that will be based on the AIGO optimisation tool that will be developed from the Pilot, which will run on a pre-set universe of assets considering all the input data and constraints. The AI genetic algorithm will generate a new proposal, where the selected preferences and risk parameters, based on the data provided from the customer and the relative investments/portfolios, available.

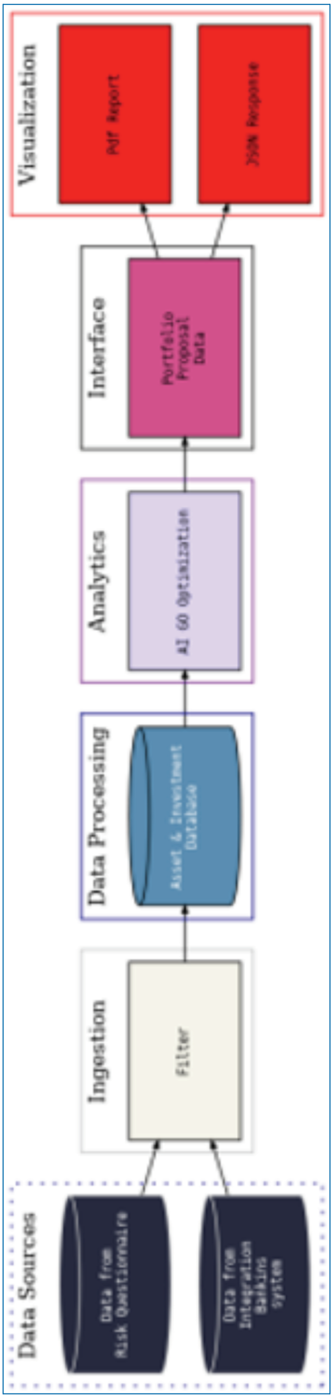


Figure 3.22 Personalized portfolio management pilot pipeline in-line with the IRA.

- Finally, a visualization layer will provide the proposed portfolio suitable for the specific customer through a report in PDF format or JSON response.

Privé external stakeholder regarding AIGO is currently Report Brain. Privé will provide the technology for the optimization process. On top of that Reportbrain will support Privé with their own specific dataset. In that way, the development will be carried out by Privé with the support of Reportbrain. The end user will be advisors, asset managers, insurance companies, banks, family offices or their end-users/clients.

Data Components

This section links the software components with the corresponding Reference Architecture layers, providing some details about their implementation. In this sense:

- Data Collection (Data Management layer of the RA) based on customers cash pool or current investments/portfolios data.
- Customers' & investments/portfolio Data quality check (Data Processing layer in RA): according to specific data models, in order to perform the data preparation for portfolio construction, based on the client's risk profile and preferences.
- AI Based Portfolio Optimization Process (AIGO, Analytics layer in the RA) that will be developed from the Pilot based on AI Algorithm, will run on a preset universe of assets taking into account all the input data and constraints, generating a new proposal, where the selected preferences and risk parameters for a specific customer.
- New proposal for the personalized portfolio will be visualized through a PDF report generation or a JSON extract that will be able to be imported in any relevant portfolio management tool.

Conclusions – Issues and Barriers

After development started, Privé successfully finished implementing a First Proof of Concept in a UAT Environment stored on our own AWS Cloud. The pilot testbed is already set up and available via SaaS access.

The main challenges consisted of the Market Data Availability Setup on our UAT Environment, as will be required the relative datasets to be enhanced either with more customer portfolio data, or with more variety of financial instruments data available from various sources that will affect the fitness factors and constraints to perform better proposed portfolio construction.

Similar challenges could arise in the future as other investment universes or market providers are made available for the optimization process.

Also, the integration of a so-called new fitness-factor based on Reportsbrain Market Sentiment Factor as an external provider via API could bring up some challenges too, as will require further exploration of AIGO optimisation tool capabilities in order to provide better results for personalized proposed portfolio taking in account also sentiment analysis factor.

In general the outcome of this pilot will be develop and adapt within SaaS based Privé Managers Wealth Management Platform a Portfolio Optimization algorithm AIGO (or Privé Optimizer), as well as improving and expanding its capabilities as an artificial intelligence engine to support better investment propositions for retail clients, that can be used as SaaS service through an API for other interested parties (investment firms, private banks, wealth management firms, etc).

Pilot #5A: Smart and Personalized Pocket Assistant for Personal Financial Management

This pilot will build a personal pocket assistant for clients of the bank, based on the development of an AI-enabled personal financial management (PFM) software. The assistant will process large about of data concerning the full range of an individual's or an enterprise's interaction with the bank based on a variety of different analytics techniques, including external data from other entities, predictive analytics and machine learning technologies. Its main characteristic will be its ability to make comparisons between clients with similar profiles, launch custom offers for every client and predict and alert end-users on future activities.

Expected Outcomes

- Patterns Detection Engine (including fraud).
- Recommender Engine.
- A Mobile App as UI for customer interaction.

Figure 3.23 summarizes the pilot workflow where inputs include customer profile, customer transactions. LIB transactions and open data sets. After applying the functionalities, the outputs include recommendations/ customer offers, alerts, chatbot/intelligent interaction.

Datasets

- Customers & Retail Customer
- LIB clients transactions
- Selected assets
- Thousands of Profile Data/Photos

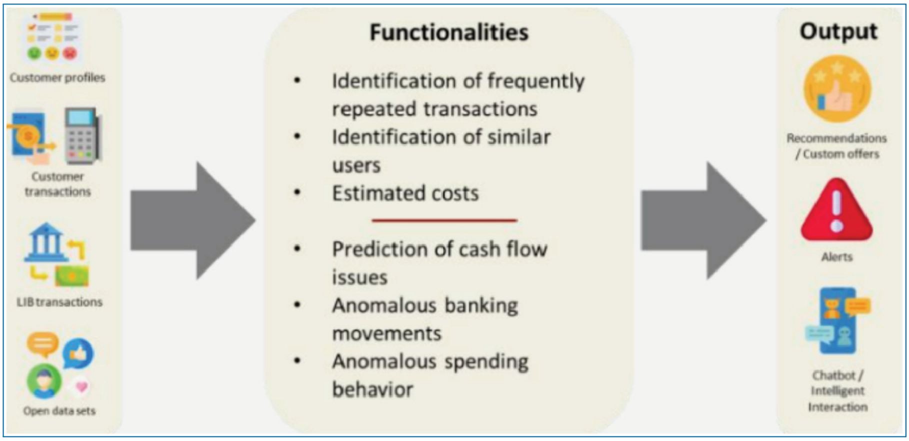


Figure 3.23. Pilot #5A: smart and personalized pocket assistant for personal financial management.

Pilot #5B: Business Financial Management (BFM) Tools Delivering a Smart Business Advise

Most of today's Financial Management tools for Small Medium Enterprises (SMEs) are geared towards analysing only past transactions, making such tools inadequate in today's world. Today, SMEs and their customers alike demand just-in-time processing, transparency and personalized services to assist SME owners not only in understanding better their SME business/financial health but also to be able to decide on the next best action to take. Thus, Pilot#5b aims to assist SME clients of Bank of Cyprus (BOC) in managing their financial health in the areas of cash flow management, continuous spending/cost analysis, budgeting, revenue review and VAT provisioning, all by providing a set of AI powered Business Financial Management tools and harnessing available data to generate personalized business insights and recommendations. Machine learning algorithms, predictive analytics and AI-based interfaces will be utilized to develop a kind of smart virtual advisor with the aim to minimize SME business admin effort, to focus on growth opportunities and to optimize cash flows performance.

Main stakeholders of the pilot development include Bank of Cyprus (BOC) and University of Piraeus Research Centre (UPRC). BOC is providing a variety of data mainly regarding its SME clients and their respective transactions, while also being the key driver in designing the Business Financial Management toolkit, which will generate valuable insights and add value to the existing online services for SME beneficiaries. UPRC is working closely with BOC in designing all provided services. It is responsible for the development of all required ML/DL algorithms of the pilot and the technical support of the pilot's implementation throughout the project.

The pilot aggregates a variety of data related to SMEs accounts from Bank of Cyprus' operation data warehouse, which include: (i) account, (ii) customer and (iii) transaction data. Moreover, (iv) open banking data will be utilized to provide a holistic approach, as well as (v) invoice data from the SMEs in order to provide accurate reconciliation services.

Technological components and Services

All services developed focus on providing valuable business insights and recommendations to the SMEs, empowering them to effectively monitor cash flow, budgeting, revenue and perform reconciliation activities, all leading to improved business management and data-driven decision making. The services provided are depicted in the Figure 3.24:

The figures show a set of different services/components/engines. Each one, in a different development stage.

An early version of the Transaction Categorization Engine, which is considered a key component, has been developed. This component is in charge of labelling the transactions of selected SME customers of Bank of Cyprus into 20 main categories (with around 80 respective subcategories to be implemented soon). This first version has been implemented combining rule-based classification and ML algorithms.

The development of the Cash Flow Prediction component has also been initiated, exploring a variety of ML models to predict the expenses of certain categories of a given account in a short period of time.

These two have already started the development and will be included into the fist PoC. The development process will include/add new components:

- Budget Prediction engine that allows setting easily budget targets through the provision of suggested target values as well as simple budget monitoring.
- KPI engine leading to valuable insights on the SME financial health and performance.
- Transaction monitoring engine that watches out for potential anomalies and savings.
- Invoice Processing engine that generates meaningful invoice background info to other components (e.g. Cash Flow Prediction) and SMEs. This applies if respective data can be obtained from SME relative ERP system.
- Benchmark engine supporting comparisons to other SMEs with similar profiles and
- Recommender engine generating actionable insights for a SME that will allow to perform better.

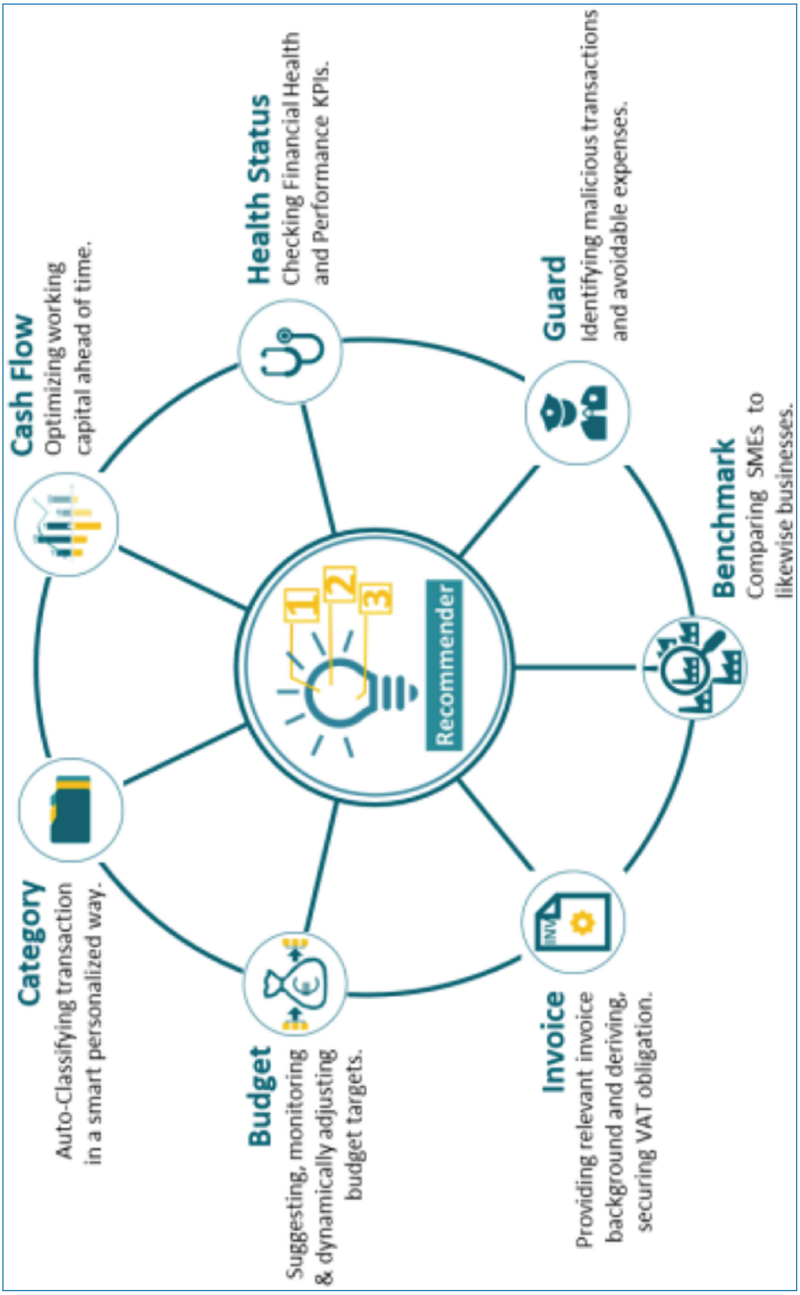


Figure 3.24 Pilot #5 services.

Testbed

Bank of Cyprus (BOC) is developing an AWS testbed, based on the technical requirements and guidelines of the relevant partners, and tailored for the unique pilot's components and the required data ingestion. As the testbed's specifications have not yet been finalised and certain bank processes require time, until the bank's AWS ecosystem is available the pilot's first components will be hosted in GFT's AWS environment.

Other non-technical requirements

The pilot's component providing competitive advantage among other available BFM tools is considered to be the Smart Virtual Advisor that leverages extensively supervised and unsupervised machine learning, takes into consideration the output from all BFM tools to come up with a holistic view of the SME business corresponding accurate business advise and reconciliation all fostering an optimal day to day business operation. The main non-technical requirement to achieve this will be solving all consent and data protection issues arising from including such enterprise data.

Implementation of a first Proof of Concept

The Proof of Concept is aiming to establish the foundation for the various smart Business Financial Management (BFM) engines. To achieve this, the design, development and implementation of a Transaction Categorization engine is prioritized as it maintains a vital role for the development and interconnection of all other components (Figure 3.25). To demonstrate the integration between the various engines, a basic Cash Flow engine will also be implemented.

The pilot's testbed will be accommodated by Bank of Cyprus, which is going to provide an AWS environment for the various pilot's components and operation. As the testbed development has not yet been completed, the PoC version represents a static development approach, where data have been collected and preprocessed by BOC and then sent to UPRC, where the Transaction Categorization and the Cash Flow prediction components are developed locally at the university's premises in an offline environment. Once development of the testbed is completed, the two components will be migrated to the INFINITECH ecosystem and will be fine-tuned accordingly.

Pseudonymized data have already been transferred to UPRC in .csv format to initiate the development of the two main components of the PoC version. Those data include:

- Customer Data from BOC: Data regarding selected SME BOC clients that will be the first.

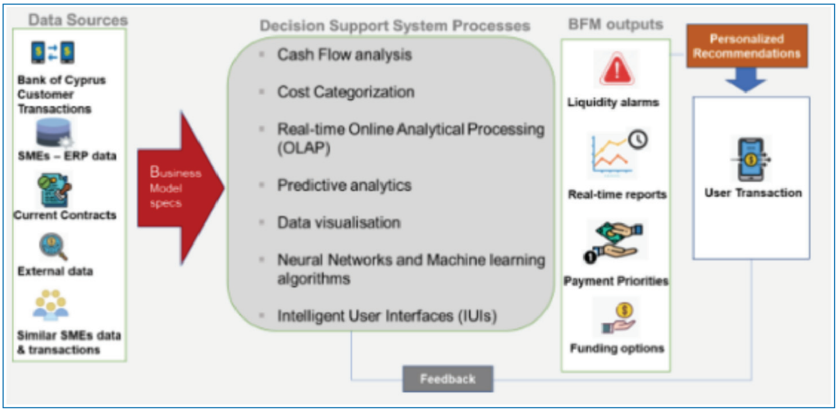


Figure 3.25. Pilot #5B: business financial management (BFM) tools delivering a smart business advise.

- Account Data from BOC: Information regarding more than a thousand accounts linked to the abovementioned selected SME clients.
- Transaction Data from BOC: Dataset with approximately the transactions of the selected SME BOC clients over the last three years. The dataset is considered the main source for developing the first two pilot's components.

Rest of the pilot's data will be utilised to enrich and refine the Categorization and Cash Flow Prediction components included in the PoC and will also be crucial for the development of the rest of the components.

Expected Outcomes

The BFM tools will drive the SME digital adoption rate as well as pave the ground for reduced credit risk, lowering amount of Non-Performing Loans (NPLs) and moreover for vital needed improved/streamlined SME lending. The expected deliverables are:

- AI driven Transaction Categorization Engine.
- AI driven Financial Business Advice (Insights) Engine.
- Chatbot for BFM (crowd policy).

The following figure illustrates the general workflow of Pilot #5B. It includes data sources, decision support system processes, BFM outputs and personalized recommendations.

Datasets

The following data sources will be integrated and used in the pilot:

- Transaction Data from BOC: a .csv file with around 500MB and 3.5 millions of transactions between 2018 and 2019;

- Transaction Data from Open Banking (i.e. PSD2 data);
- Transaction Data from SMEs (optional);
- Other Data (Market);
- Other Data from SMEs (optional);
- Accounts Data from BOC: maps accounts with the transactions;
- Accounts Data from Open Banking;
- Customer Data from BOC: links customer to accounts and the available NACE code is used in the transactions' categorization model;
- Direct Input from SMEs (e.g. feedback loop for transaction categorization).

Data Produced

The pilot will combine the abovementioned diverse datasets in order to produce personalized business insights and recommendations for SME customers of BOC. Output data, as shown below, will be generated by the various engines in relation to cash flow predictions, budgeting, KPIs, benchmark(s) and transaction monitoring and categorization. The data will be stored in the common datastore and be available to the end user (SME) via the Infinitech Reference Architecture (IRA) gateway (and the banks middleware). To this end a pilots-specific REST API will be developed leveraging different endpoints for each specific service. The output data/endpoints include:

- A JSON containing the obtained insights and recommendation to be provided to the respective SME.
- A JSON containing the obtained cash flow related data to be provided to SME directly or indirectly.
- A JSON containing the derived budget target for each category used by the respective SME.
- A JSON containing results on Financial Health and Performance matrix.
- A JSON containing results on abnormal transactions and suspicious expenses.
- A JSON containing Matrix with invoice information and payment prioritization.
- A JSON containing benchmarks that allow the SME to compare to likewise businesses.

Explainable Workflow

Some of the available datasets require real time data collection, while in others historical data collection is sufficient to provide actionable business insights. In detail, transaction and account data related to the respective SME will be drawn from BOC's repository by a real time/historical data collector as well as transaction and account data from Open Banking (PSD2), as well as BOC customer data, will utilize a historical data collector. Furthermore, a way of handling batch of data is needed to provide as there should be an option of pushing data to the Infinistore once a day by the bank (e.g. in cases where the real-time connection is lost or for the purpose of uploading history data). To this end, the bank IT team will be capable of uploading a batch of data in CSV format directly to the pilot specific cloud sandbox. IIn addition, an external data collector will also be used in order to integrate other related Open Banking/macroeconomic data. The SMEs data source (e.g. ERP/Accounting system) utilization remains optional as consent is required for the collection and processing of such data and its cloud availability being required. All data except external macroeconomic data will be pseudoanonymized (by tokenization) before being uploaded to the IRA. The cloud Data Repository (within IRA) will then store all collected data, along with the generated insights, past SME financial actions (to measure at what degree the SME actions reflect the recommended insights), as well as minimum user input which is required. A continuous data streaming will connect the Data Repository with the various deployed BFM tools (machine learning algorithms), which would allow the retraining of the respective AI models and the generation of useful insights and recommended actions. A reverse data pseudoanonymization will then be applied before the processed data move to the bank middleware component that contains composite APIs and produces push notifications, all which will be offered to the SMEs via Android, iOS and web apps. Upon SME user login the IRA is also accessed, insights/recommendations picked up from the cloud data repository and provided to the SME user. To this end, a prototype component will be developed in order to digest and properly present the results of the corresponding analytics components. The pilot's workflow is depicted in Figure 3.26.

Logical Schema

The following figure illustrates a logical view of the pilot system architecture in-line with the INFINITECH-RA.

The datasets used, as well as the pilot's RA is illustrated in Figure 3.27. All personal and sensitive data related to SME customers of BOC will be pseudonymized at the bank's premises using a tokenization approach before streamed to the INFINITECH ecosystem to ensure the protection of vital SME data. A reverse pseudonymization will be applied before presenting the data to the SME end user. The RA of the pilot, as included in D2.13 and depicted below:

The various components will be containerized using Docker, and a LeanXscale database will be used to store and query the results of the analytics processing, as well as insights generated by the recommender engine. Most of the data analytics components are developed using Python data analytics and ML/DL libraries, i.e. Numpy, Pandas, ScikitLearn and Tensorflow, where data streams required in some components for real time analytics will be handled with Apache Kafka. For the time,

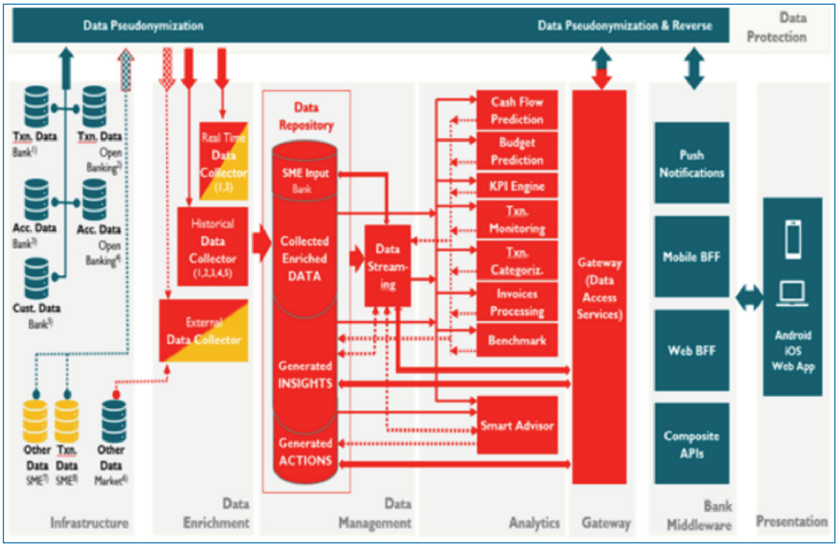


Figure 3.26. Business financial management (BFM) pilot workflow.

a static approach has been followed and all development progress has been done in offline mode in University of Piraeus premises, with all progress being migrated to the INFINITECH ecosystem once the pilot's AWS testbed is set.

Components

The following components will be deployed and used in the pilot pipelines:

- Transaction Categorization Engine (Analytics layer in the RA): key component in charge of labelling the transactions of selected SME customers of Bank of Cyprus into 20 main categories (with around 80 respective subcategories to be implemented soon);
- Cash Flow Prediction component (Analytics layer in the RA): based on a propabilistic Deep Neural Network (implementation of DeepAR model) to predict the expenses of certain categories of a given account in a time horizon of 12 weeks;
- Budget Prediction engine (Analytics layer in the RA): allows setting easily budget targets through the provision of suggested target values as well as simple budget monitoring;
- KPI engine (Analytics layer in the RA): leading to valuable insights on the SME financial health and performance;
- Transaction monitoring engine (Analytics layer in the RA): watches out for potential anomalies and savings; To this end Graph analysis apporaches is being explored and implemented;

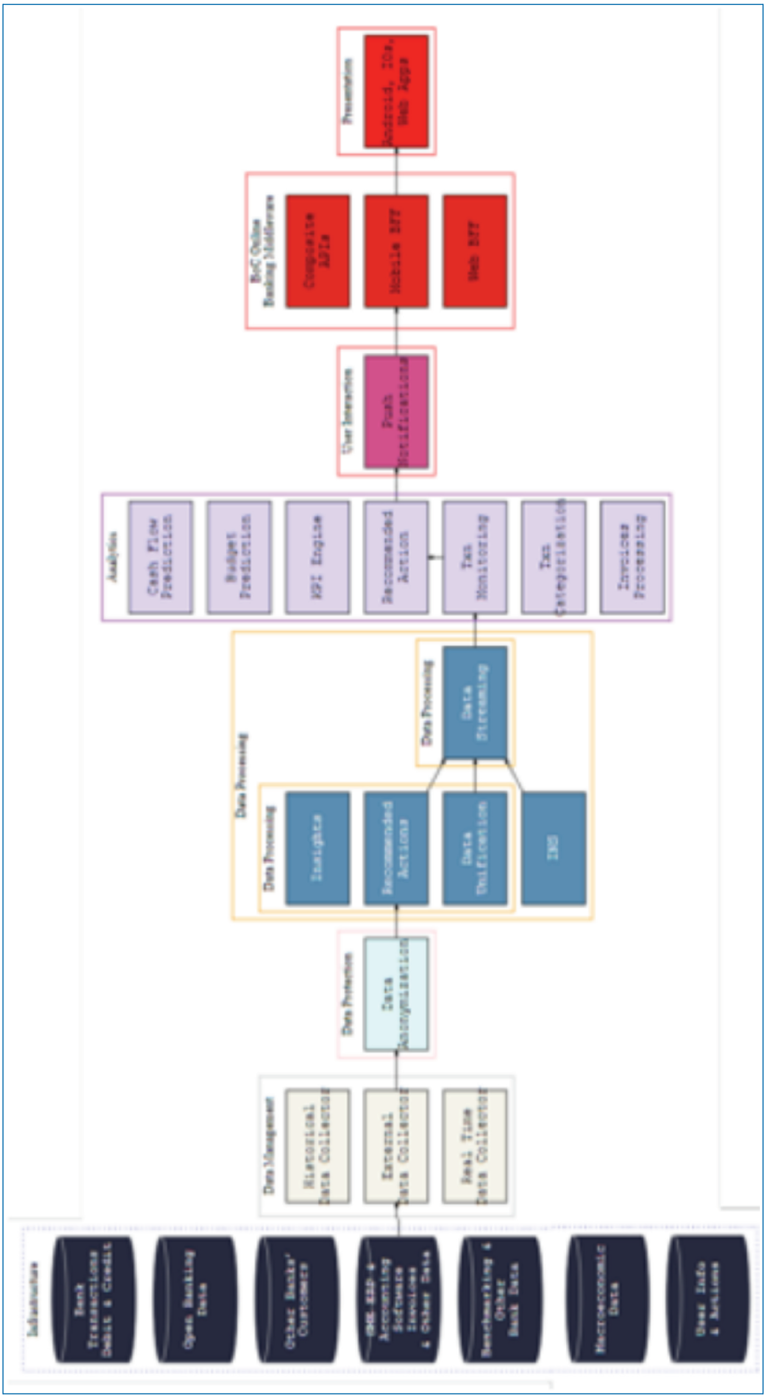


Figure 3.27 Business financial management pilot pipeline in-line with the IRA.

- Invoice Processing engine (Analytics layer in the RA): generates meaningful invoice background info to other components (e.g. Cash Flow Prediction) and SMEs. This applies if respective data can be obtained from SME relative ERP system;
- Benchmark engine (Analytics layer in the RA): supporting comparisons to other SMEs with similar profiles;
- Smart Advisor (Analytics layer in the RA): generating actionable insights for a SME that will allow to perform better.

Conclusions – Issues and Barriers

Concluding the Pilot's development is progressing based on the project's timeline already establishing the Transaction Categorization and Cash Flow Prediction components that are considered the foundation for designing and developing the rest of the AI powered components included in the BFM toolkit that will be the outcome for SMEs. Next pilot's milestone is moving all development progress to the cloud environment and setting the required data streaming/data collection mechanisms. Main challenge is the AI powered Business Financial Management tools development and their efficiency, as will be based on the availability of all the required data for SMEs from BOC or from the SMEs in order the final goal to be achieved.

Main goal of the Pilot aims to assist SME clients of Bank of Cyprus (BOC) in managing their financial health in the areas of cash flow management, continuous spending/cost analysis, budgeting, revenue review and VAT provisioning, all by providing a set of AI powered Business Financial Management tools and harnessing available data to generate personalized business insights and recommendations. Machine learning algorithms, predictive analytics and AI-based interfaces will be utilized to develop a kind of smart virtual advisor with the aim to minimize SME business admin effort, to focus on growth opportunities and to optimize cash flows performance.

Pilot #6: Personalized and Intelligent Investment Portfolio Management for Retail Customer

The goal of this pilot is to create a system for personalized investment recommendations for the retail customers of the bank. NBG will leverage large customer datasets and large volumes of customer-related alternative data sources (e.g., social media, news feeds, on-line information) in order to make the process of providing investment recommendations to retail customer more targeted, automated, effective, as well as context-aware (i.e. tailored to state of the market).

Pilot #6 focuses on providing personalized investment recommendations for the retail customers of the bank. National Bank of Greece (NBG) will leverage large

Figure 3.28. Pilot #6 personalized closed-loop investment portfolio management for retail *customers*.

customer datasets and large volumes of customer-related alternative data sources (e.g., social media, news feeds, on-line information) in order to make the process of providing investment recommendations to retail customer more targeted, automated, effective, as well as context-aware (i.e. tailored to state of the market). The latter is the main innovation of the pilot. An overview of Pilot #6 is given in the Figure 3.28:

Technological components and Services

Going a step beyond the Pilot's RA towards the functional overview shown in Figure 3.29, the High-Level Architecture presented presents the software components that build the Pilot's use cases. This figure has been used to identify hardware requirements and describes the technologies behind its principal components. This document links the shown software components with the corresponding RA layers, providing some details about their implementation. In this sense:

- NBG supply raw datasets required for the implementation of the final services. The pilot has already identified the relative customers portion of data that will be utilized, based on the existing DWH (Data Ware House).
- Data Collection and Data Normalization components (Data Management and Protection layers of the RA): based on Icarus from UBI, define the rules to (Data Processing layer in RA): process and harmonize, cleanse and anonymize data from NBG and insert them in a datastore available from LXS.
- Customer Risk Profile Cluster implemented by ML/DL Algorithm developed by NBG that cclassify customers into 4 risk profiles: Conservative, Income Seeking, Balanced, Growth Seeking. The algorithm is applied to both investors (having answered the MiFID questionnaire) and non-investors.

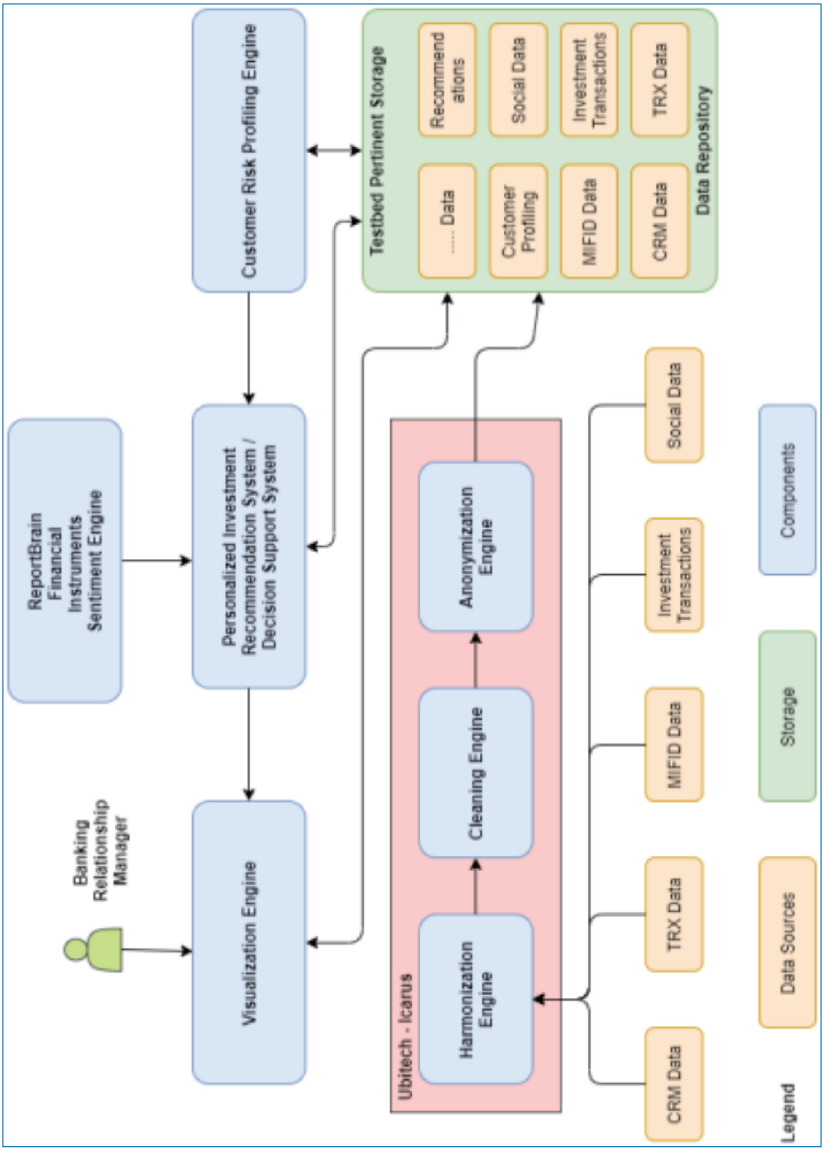


Figure 3.29 Pilot #6 high level architecture.

Component	vCPUs	RAM	Storage	Software
UBITECH Harmonization Tool	vCPUs: >= 4	>= 32GB	>= 250 GB	Java, Docker, Linux OS
UBITECH Cleansing Engine	vCPUs: >= 4	>= 32GB	>= 250 GB	Java, Docker, Linux OS
UBITECH Anonymization Engine	vCPUs: >= 4	>= 32GB	>= 250 GB	Java, Docker, Linux OS
LeanXcale Datastore (2 Data Nodes, 1 Query Data Node, 1 Metadata Node)	vCPUs: >= 4	>=8GB min >=32GB max	>=40GB	Docker, Linux OS
Visualization API & Application (2 Servers)	vCPUs: >= 4	>=16GB	>=50GB	Java Spring Boot, Angular, Express, Node JS, Docker, Linux OS
Retail Customers Risk Profiling	vCPUs: >= 4	>= 32GB.	>= 250 GB	Python Libraries
Personalized Investment Recommendations/ Decision Support System	vCPUs: >= 4	>= 32GB.	>= 250 GB	Python Libraries

Figure 3.30. Pilot #6 hardware/software requirements for the testbed.

- Personalized Investment Recommendation AI engine, that will also utilize sentiment analysis data from RB, will produce the recommended instruments for investment.
- Customer initiation and personalized recommendation is obtained through a visualization application developed by CP. This application will also orchestrate the processes of analysis, initiation, execution and processing, when a new customer or new data are available.

Testbed

Pilot's #6 final deployment relies on MS-Azure cloud infrastructure that NBG will provided. Further details of the software/hardware first analysis and their results can be found in that document, but are summarised in the Figure 3.30:

NBG MS-Azure infrastructure it's currently deployed from NBG IT team in order to accommodate the first Proof of Concept being under development and based on the Pilot's development progress will be adjusted in terms of resources to accommodate any additional requirements related to resources.

Other non-technical requirements

Besides the technical requirements that compose the core pilot's platform, the AI technologies deployment and the data collection, we have not identified any other non-technical requirements that may affect the best outcomes for the Pilot.

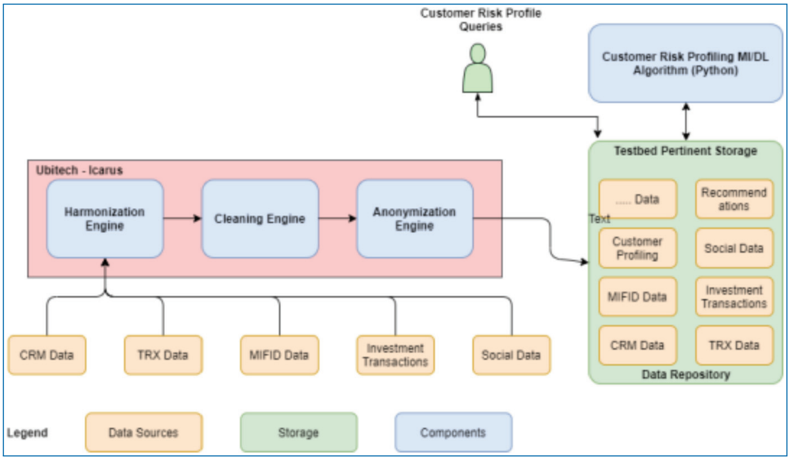


Figure 3.31. Pilot #6 Data collection PoC architecture.

Implementation of a first Proof of Concept

First Pilot #6 demonstrator (PoC) is focused on processing a subset of Cards and Deposit Accounts Transaction Data extracted from bank's Operational DWH. Figure 3.29 presents the functional diagram of the developed PoC.

Based on the raw datasets available from NBG for retail customers, a first version of the relative ML/DL algorithm that will be implemented from NBG, will provide the Customer Risk Profile clustering. The Risk Profile will be one of the different inputs to feed the final core component: Personalized Investment Recommendation AI engine. The AI Engine is not available in the first PoC.

Based on the algorithm results for Customer Risk Profile clustering a web page will be provided as dashboard for visualization of the results as a way of making first demos of the PoC (Figure 3.31).

Proof of Concept execution, will provide valuable feedback for the AI approach that will work better for the Pilot execution, as well as create the common ground for the future of the development that will be required in order the full scope of the Pilot to be realised.

Expected Outcomes

- BigData/AI system for personalized investment recommendations for the retail customers of the bank.
- Development of a closed-loop system that continuously learns, improves itself and provides better recommendations.
- The system shall improve productivity of investment consultants of the bank, through enabling them to access faster recommendations tailored to their retail customer needs.

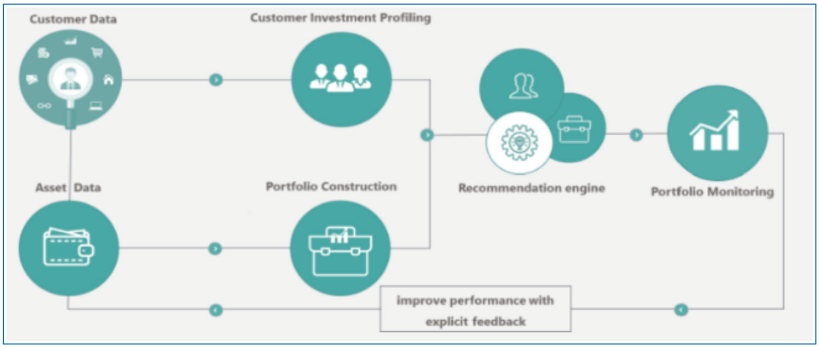


Figure 3.32. Pilot #6: personalized and intelligent investment portfolio management for retail customer workflow.

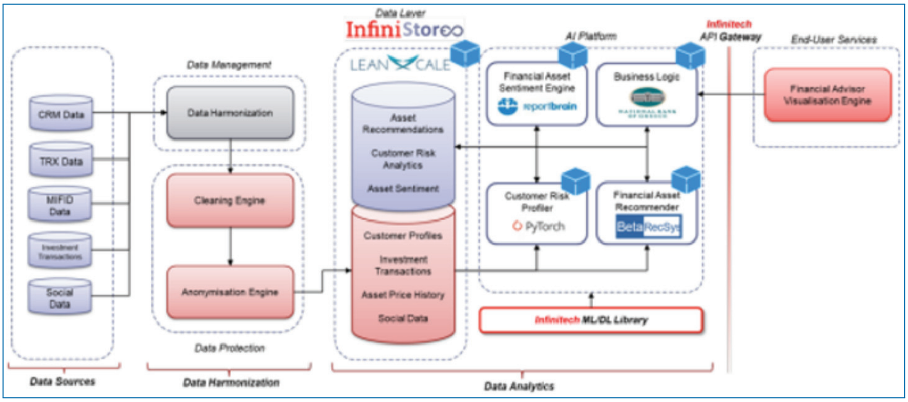


Figure 3.33. Personalized closed-loop investment portfolio management pilot pipeline in-line with IRA.

The Figure 3.33 depicts Pilot #6 workflow for personalized and intelligent investment portfolio management for retail customer.

Datasets

Data that will be used for this pilot will be extracted and anonymized in CSV files from NBG Datawarehouse and several data sources:

- Deposit Account Transactions: Data of Deposits accounts transactions for retail customers are extracted for the last two (2) years (8,91G).
- Cards Transactions: Data of Transactions related to Cards for retail customers for the last two (2) years (7,3GB).
- Instruments Historical Prices: Data for Instruments Historical Prices for the last two (2) years (0,23GB).
- Investment Related Transactions: Data of Investment Related Transactions for last two(2) years (0,3GB).

- Instruments Characteristics: Data for Instrument characteristics for matching with customers profiles, including asset class, currency, ISIN, maturity etc. (0,01GB).
- CRM Data: 150.000 Customers related data like demographics, product ownership and responses to MIFID questionnaires (0,05GB).
- Sentiment Analysis for each instrument proposed from Data Analysis as recommendation using RB information from the news or/and social media to provide to NBG customers with clearer and real-time risk results.

Data Produced

Personalized investment recommendations for the retail customers of the bank, based on their Risk and transactions profiles. Banks relationship managers based on each customer risk and transactions profile, will be able to propose the possible alternatives of financial instruments that a customer will be interested to invest, with the relative prioritization. The proposed recommendations will be based on the instruments available from the bank, with the necessary input data for sentiment analysis for each financial instrument, based on the news & social feed, for the specific instrument (e.g. stock, bond, etc), or the relative instrument category.

Existing Landscape in Financial Institutions and particularly Banks has set as priority the identification of targeted Customer propositions and especially in investments sector. Driven both by Competition as well as Customer needs, depiction of each Customers potential and risk appetite in combination with interesting for the Customer recommendations, may lead in the increase of each Customer's share of wallet and at the end increase of Bank's Market share in the specific Sector.

Explainable Workflow

Data from NBG Datawarehouse related to Investment Products Retail Clients (CRM Data, Deposit Account Transactions, Cards Transactions, Investment Related Transactions), will be extracted in CSV files and utilizing the relative tools for data processing, anonymization and quality checking and cleansing will be imported to Leanxcale Datastore. Based on the data extracted for NBG Clients, through the Customer Risk Profile engine using data analysis tools, will be able to divide all customers in specific profile clusters based on the investment & banking behaviour (MIFID questionnaires), deposit and card transactions, as well as investments transactions. Also, NBG will provide for each investment customer cluster profile the relative instruments that will be suitable for investment.

Logical Schema

A first approach to mapping the pilot architecture to the INFINITECH-RA layers and pipelines approaches is illustrated in the following figure.

Components

The following components will be deployed and used in the pilot:

- DataStore (Leanxcale) (Data Sources in RA).
- NBG Datasets (Data Sources in RA).
- Data Collection (UBI Icarus) (Data Management in RA).
- Data Normalization (UBI Icarus) (Security in RA).
- Customer Risk Profile Cluster (Analytics in RA): classify customers into 4 risk profiles: Conservative, Income Seeking, Balanced, Growth Seeking.
- Personalized Investment Recommendation AI engine (Analytics in RA).
- Customer initiation and personalized recommendation UI Application (Presentation in RA).

Conclusions – Issues and Barriers

Based on the work done so far, the Pilot it seems that is on track, following the implementation plan already agreed with all the contributing partners and utilize the available technology components already available or will be, as part of the INFINITECH project.

The main foreseen challenges would include:

- Implementation of the best performed ML/DL Algorithms, for this purpose we have started to evaluate some of the algorithms already available from INFINITECH partner University of Glasgow (GLA).
- Setup of the relative testbed based on the blueprint reference architecture (as will be hosted on MSAzure).
- Calibrate the ML/DL Algorithms to provide best results for investment recommendations.

As the outcome the Bank will develop a better and more trustful relationship with its customer base, who hopefully will gradually turn exclusively select the specific bank 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.

Pilot #15 Open Inter-banking

Pilot 15 main objective is to deliver a prototype to address and tackle business pains shared within banking institutions leveraging Machine Learning and Natural Language Understanding paradigms. The model aims at reading and analyzing extensive internal documents of banks in real time to highlight the main concepts and compare them with a reference taxonomy to build a common business glossary in order to:

- Provide banks with a tool able to standardise the documentation analysed;
- Increase Automation and Intelligence based on data processing leveraging data governance processes;
- Improve the analysis and comprehension capabilities of internal documents and contents.

The Inter-Banking Open pilot, as explicated by its name, is the result of an Open Call to shared business pains among several Banks, and its objective is to develop a solution that could address and tackle such pains in a pre-competitive environment. 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 Modelling.

ABI Lab is the Banking Research and Innovation Centre founded and promoted by the Italian Banking Association (ABI). Through research and advocacy, ABI Lab promotes innovation as a mean of growth and reinforcement of the banking system. To support digital transformation, ABI Lab has created the AI Hub, a centre of excellence to discuss over the AI application in the banking and financial sector.

Within the AI Hub, the objective of the pilot is to promote the development of a common use case, which will involve different banks through a shared research approach. The use case will be developed following two steps, as described in Figure 3.34.

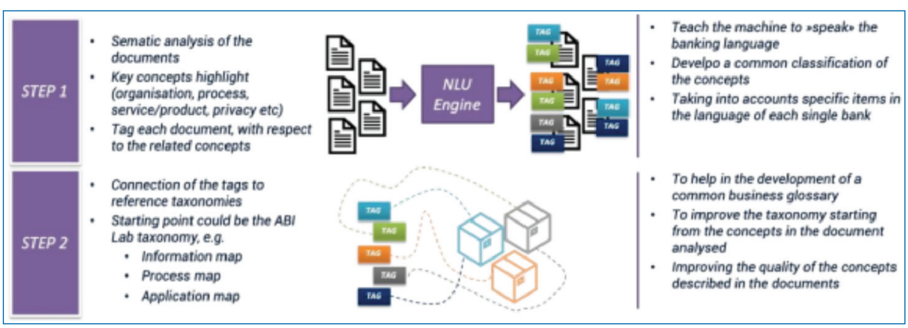


Figure 3.34. Pilot #15 steps and main objectives.

Technological components and Services

The main objective is to build an AI tool able to read the internal documents of a bank to highlight the main concepts and compare them with reference taxonomies to build a common business glossary.

Technological components and services will be defined according to the pilot objectives.

Testbed

The technical and development aspects, in particular within the dedicated testbed, will be supported by GFT and HPE. The pilot will be hosted and deployed on the Testbed blueprint that will be developed accordingly to the pilot requirements.

Data Sources

Data will be extracted from a large set of bank's internal documents in pdf, word and/or txt format, provided by the banks involved in the pilot. The documents will focus on the following areas: KYC, entering into a relationship with the customer and the Markets in Financial Instruments Directive (MIFID). In addition to documents related to the three specific areas, other data sources includes:

- Additional documents relating to different areas and identified randomly within the document base:
- Internal dictionaries, internal glossaries, internal taxonomies useful for the development of metadating techniques.
- ABI Lab architectural framework (reference taxonomy)

Data Produced

The advanced document processing will allow real-time useful information via searching semantically relevant text according to the semantic metadata, increasing automatisation, easiness of use and usefulness of outcomes.

Explainable Workflow

STAGE 1 – study and research @ ABI Lab controlled environment

- Data ingestion/preparation, 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;
- Data storage, including tools and infrastructures aimed at data collection from different sources and in different formats, and their storage;

- Machine learning engine optimisation, enabling continuous Natural Language Understanding algorithms optimisation, following the use case experimental purposes
- Semantic model design A data visualisation layer, including tools and methods to display results to different users and stakeholders.

STAGE 2 – test and validation @ Infinitech testbed (Model based on BDVA RA)

Logical Schema

The following figure illustrates the logical architecture of the pilot in-line with INFINITECH-RA constructs and approach. Pilot's Reference Architecture (Figure 3.35) and main data flows is presented in detail. This RA can be outlined in three main layers to be implemented through different software components. These main three layers are:

- A Data Management layer, that performs data quality checking and harmonization of the data provided from NBG and imported them into the datastore for use of the pilot's functionalities. On this first stage, the data that will be used transactions data of deposit accounts, cards, investments and CRM data for a small subset of NBG Customer will be used.
- A Data Protection and Data Processing Layers in charge of cleanse, homogenise and store all data collected, according to specific data models provided from NBG operational DWH, so these are available for the analytics processes. Here are also included all the operations needed to anonymise (if required) the captured data and protect this information from unauthorised access.
- An Analytics block, fed by the data layers, where different ML/DL technologies and visualization tools will enable data monitoring, analysis, and exploitation. Two main AI models will be developed here, the Customer Risk Profile Clustering and Personalized Investment recommendation decision support, that will utilize also the Sentiment Analysis data provided from RB relative engine, in order to provide for a customer, the recommended products to invest through a visualization application (in the RA's Visualization layer).

The main stakeholders for this pilot are the account officers of a bank, who will be able to provide personalized investment recommendations for customers. Recommendations based on customer (risk) profile, as well as with the relative sentiment analysis data from the news, social media, and other resources on the internet. In Pilot #6 these stakeholders will be represented by the bank, NBG (National Bank of Greece), that provides the user stories.

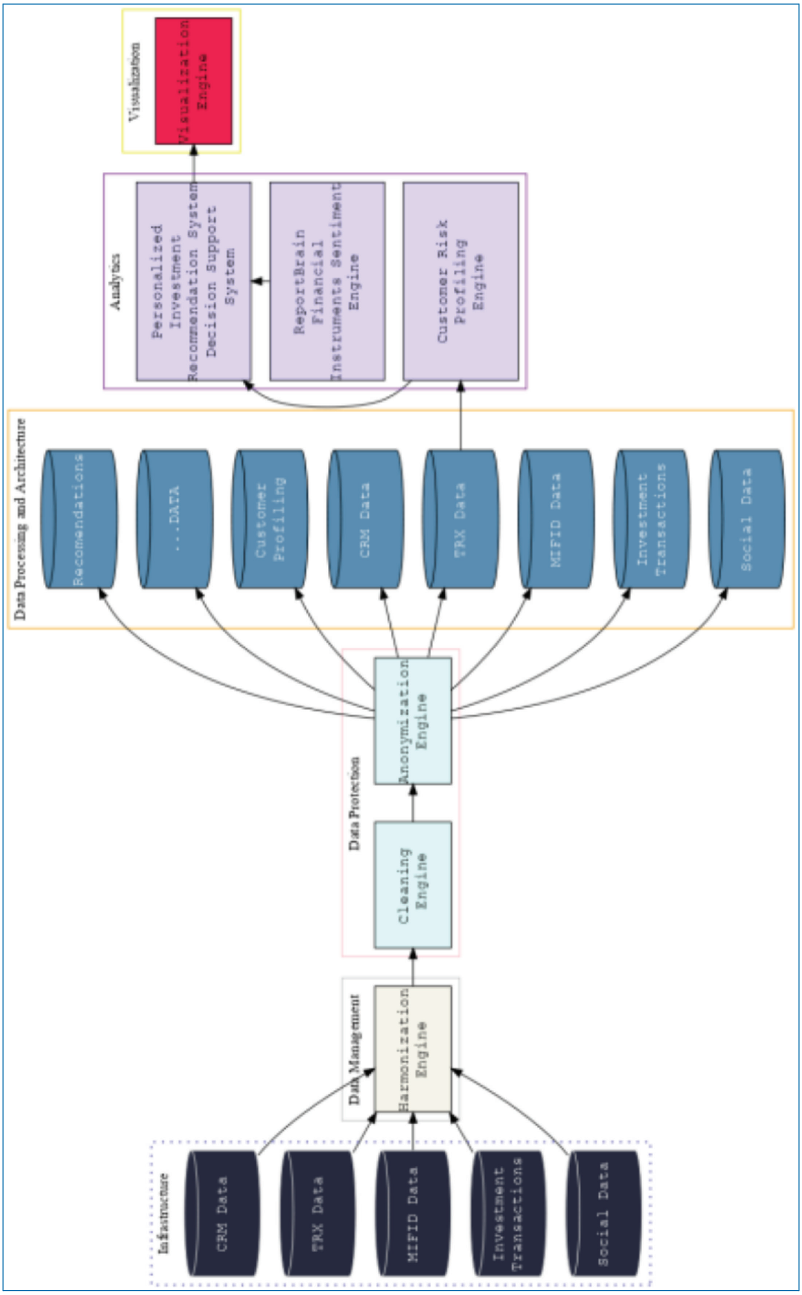


Figure 3.35 Pilot #6 reference architecture.

The configuration and roles by each partner in this pilot consists of: NBG (as Bank and Business Owner) provides customer's data, UBI(Ubitech) process the data through the Data Management and Data Processing layer, UBI inserts this data into the datastore software provided by LXS (LeanXcale). AI algorithms (NBG), utilizing sentiment analysis data by ReportBrain(RB). University of Glasgow is now also participating enhancing AI algorithms. Finally, a final user application developed by CP (Crowdpolicy) will show the desired information and recommendations.

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 needed for this specific analytical purpose, 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.

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 involve a community of banks, which will:

- Provide data-set related to internal documentation.
- Provide information and addressing issues around the usage of common taxonomy or glossaries to build a classification ad analysis model.
- Participate to the requirement identification and service evaluation.

The development (and also the training plans for the AI models) will be driven by ABI Lab, supported by the members of the AI Hub community.

The banks will be the final users, keeping into consideration that the objective of the pilot is to develop an experimental prototype that will be the object of further analysis by the participant stakeholders.

Components

The main technological components that will be implemented and integrated as part of this pilot are:

- 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.

Conclusions – Issues and Barriers

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. Indeed, the scope of the pilot could arise some foreseen challenges, mentioned below:

- Put together multiple banking, technical and academic stakeholder to achieve shared objectives.
- Harmonization of semantic representational models in the context of financial services.
- Exploitation of data assets.

3.3 Personalized Usage Based Insurance Products

Pilot #11: Personalized Insurance Products Based on IoT Connected Vehicles

In a few words, this pilot aims to develop new services for driving insurance companies, based on the information gathered from a connected vehicle, as an IoT ecosystem. Current driving insurance services try to reward good drivers against the “bad one”, but based on very static or historical information: your age, colour of your car, incidents by year, etc. A new approach, more dynamic, adapted and custom services are needed. You pay as you drive, in a similar approach to a cloud word, where you pay as you consume. Complementary to this, a second service will help to detect possible fraud’s situation. Fraud causes not fair costs to the company that would affect indirectly to the good/ honest drivers.

In both use cases the underline technology is based on connected vehicles, IoT and BigData, because of the expected amount of data to be managed. The business analysis part, which determines how good driver you are, and the detection of possible frauds, will be based on AI and ML techniques. Due to the personal data managed in the pilot, security and privacy will be also a technology challenge to achieve.

This pilot focuses on car insurance and risk analysis by developing two AI powered services: Pay as you Drive, that allows the insurance company to adapt prices by classifying the driver according to the way he/she drives; and the Fraud Detection

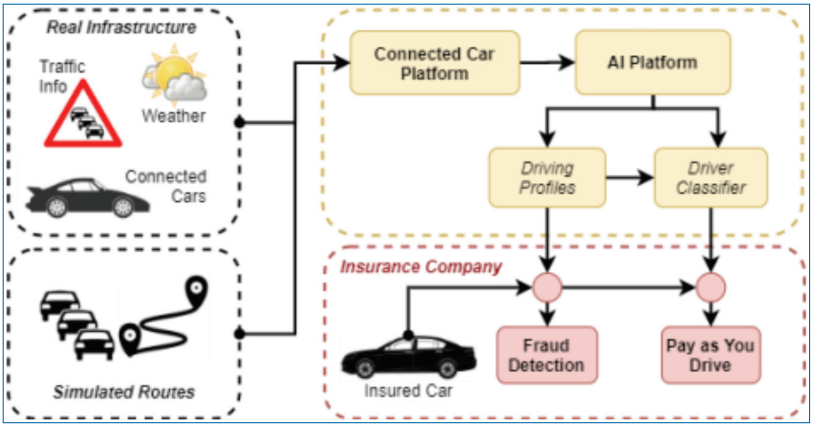


Figure 3.36. Pilot #11 Personalized insurance products based on IoT connected vehicles overview.

which helps to identify the actual driver of a vehicle involved in an incident. These two services rely on a driving profiling tool that requires datasets from connected vehicles to define, identify and train the different profiles as ML models. Other external data sources, such as traffic incidents or weather, will be used to classify the driver, contextualizing its assigned driving profile. An overview of Pilot #11 is given Figure 3.36.

Technological components and Services

Going a step beyond the Pilot's RA towards the functional overview shown in Figure 3.37, the High Level Architecture represents the software components that build the Pilot's use cases. This figure has been used to identify hardware requirements and to describe the technologies behind its principal components. This document links the shown software components with the corresponding RA layers, providing some details about their implementation. In this sense:

IoT infrastructures supply raw datasets required for the implementation of the final services. The pilot has already identified and linked connected vehicles (real and simulated); weather stations (from AEMET); roads (from OpenStreetMap) and traffic alerts.

Data Collection & Aggregation and Data Normalization components (Data Management and Protection layers of the RA): based on NGSI-LD and FIWARE Data models, define the rules to ingest data from IoT infrastructures. First functional versions for the identified IoT sources are deployed and ingesting data. Remark here the work done to integrate the Simulation of Urban Mobility (SUMO) tool with the Pilot's framework, following the NGSI guidelines. Also in this layer, Gradiant's Anonymizer tool analyses and anonymises (when required) the collected data before being uploaded.

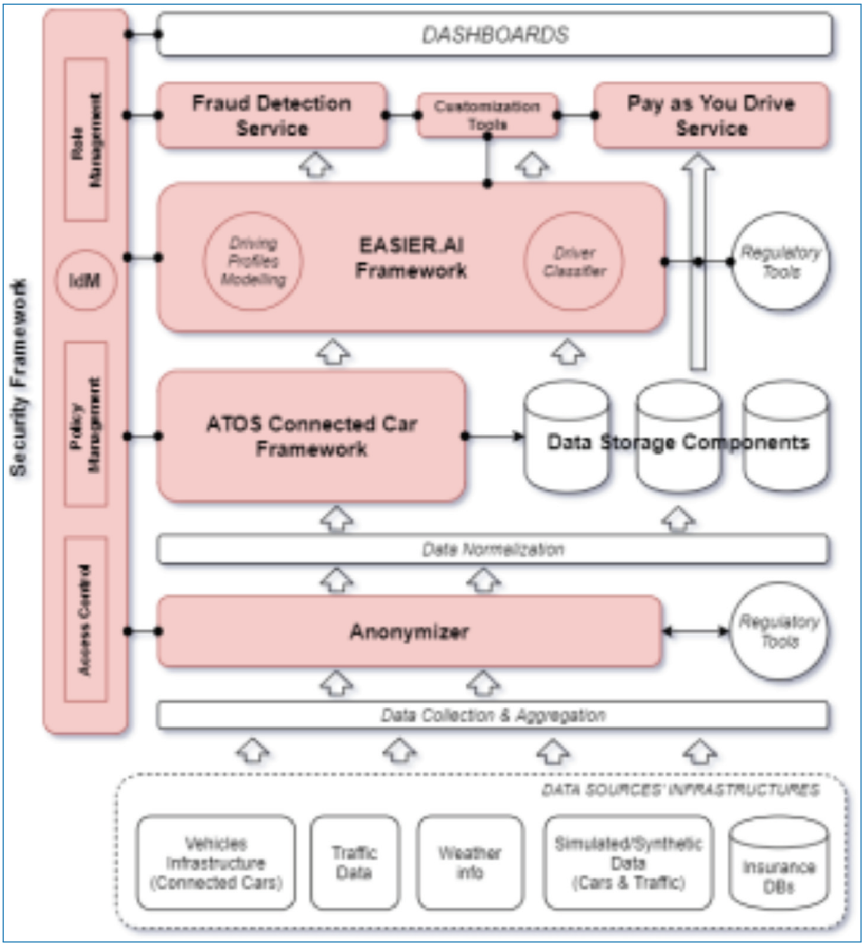


Figure 3.37. Pilot #11 high level architecture.

Connected Car framework (Data Processing layer in RA): composed by the FIWARE Orion Context Broker, that supports all context management functionalities (context information broker), and an instance of the FIWARE QuantumLeap General Enabler (context information persistence) that supports historical information management. A first instance, covering these two components, has been implemented and deployed in Atos' infrastructure.

EASIER-AI component (RA Analytics layer) is a Hybrid (Cloud/Edge) under development framework that facilitates to develop, measure, monitor and deploy customised AI models. It is built on top of the Elastic Search, Kibana and TensorFlow slate of three and enables different ML/DL technologies deployment. On top of this, Pilot #11 is developing (and will train) the Driving Profiling and Driver Classification inferencers (User Interaction RA layer) that will support the "Pay as You Drive" and the "Fraud Detection" services (Visualization RA Layer).

The access to these frameworks (Connected Car and EASIER-AI) is protected by an OAuth identification and authentication component that relies on the FIWARE KeyRock IdM. SSL/TLS is used to protect communications. This is deployed and integrated with the Connected Car framework.

Testbed

Pilot's #11 final deployment relies on UNINOVA infrastructure. Further details of the software/hardware first analysis and their results can be found in Figure 3.34.

UNINOVA infrastructure it's currently being dimensioned to provide support to several clusters, so it is not still available for deployments. Pilot #11 demonstrator is being deployed within ATOS premises. All first versions of the P#11 components follow an approach combining docker and kubernetes for their deployment to make easier the migration to the final testbed location.

Others non-technical requirements

Besides the technical requirements that compose the core pilot's platform, the AI technologies deployment and the data collection, an additional and relevant requirement has been identified to obtain the best outcomes. This is related to the availability of enough data sources (Figure 3.38).

Anonymous Connected Cars vehicles, that will provide the routes (and vehicle data) needed to define, train and evolve the different AI models (and ML/DL technologies). The more vehicles enrolled, the better models obtained, but, on the same side, the more vehicles reporting around the same area, the better traffic models can be created and so, better driver classifications can be performed. In this line, the pilot will get 20 connected vehicles, mounting an smart on board unit that captures data from the CAN bus of the vehicle (technical vehicle data, such as speed, acceleration, systems status, etc.) plus an NMEA unit to capture GPS vehicle's location. These vehicles will start driving, supported by CTAG infrastructure,

Technical Components					
Component	Nodes	Cores by Node *	Memory (RAM) **	Storage ***	Special Hardware
EASIER.AI	1 Small + 2 Medium + 1 Large + 1 XXL	1x1 + 2x4 + 1x6 + 1x6 = 21 Cores	1x4 + 2x8 + 2x16 + 1x16 = 68 GB	1x100 + 2x500 + 1x2TB + 1x10TB = 13,1 TB	Some nodes would have access to GPU, FPGA, TPU, or other special AI hardware. AVX2 Instructions set support
Connected Car	1 Small + 2 Medium	1x1 + 2x4 = 9 Cores	1x4 + 2x8 = 20 GB	1x100 + 2x500 = 1,1 TB	
Anonymiser	1 Medium	1x4 = 4 Cores	1x16=16GB	1x1TB = 1TB	
Trainners/Inferencers AI	TBD later, in a first stage we will use the EASIER.AI infrastructure to also run trainers and inferencers				
TOTAL	9 Nodes	34 Cores	104 GB	15,2 TB	GPU, FPGA, TPU and AVX2 instructions support

Figure 3.38. Pilot's #11 hardware testbed.

next Feb. 2021 and it is planned to report connected vehicles' datasets for 4 hours a day and for at least 1 year long.

Implementation of a first Proof of Concept

First P#11 demonstrator is focused on data collection and homogenisation process, in order to identify any potential issue (or required data set) that may impact on the subsequent Pilot's steps. This will also provide with fundamental elements the AI modelling stage. Figure 3.39 presents the functional diagram of the developed PoC.

As centred in data gathering, the ATOS Connected Car framework will be the core component to test and evolve. As mentioned above, this components' set is mostly deployed in Atos Infrastructure, with support from CTAG to build and deploy their own data adaptors for their vehicles. In this sense:

- Data adaptors' first versions (based on NGSI and FIWARE Data models) are deployed, ingesting data from the painted data sources.
- Connected Car core framework (Context Broker and Historical Repository based on FIWARE) is also ready, managing ingested context information. An NGSI-LD REST API is ready to access collected data.
- Identification and Authentication layer, based on FIWARE KeyRock IdM, is, in turn, managing Oauth tokens to grant access to the framework.

With all these components up & running, some dashboards are being developed in order to present the collected data and to start the data analytics processes (Figure 3.40). These will lead to identify the best AI approach to work on the Pilot's final services.

Expected Outcomes

- Provide personalized insurance plans: Pay as you Drive/Usage-based insurance.
- Collect additional data about the status of the connected vehicle and reaction of the drive (in case e.g. an accident), improving the capabilities of fraud detection.
- Provide a more effective and dynamic billing system.

The following figure illustrates the schematic overview of the Pilot #11 workflow.

Datasets

The main data source in the pilot is produced by the connected vehicle, with about 20 vehicles. It is under study, the inclusion of some historical data provided by vehicles from other previous project; if legally possible. The data produced by the connected vehicle includes: CAN data, traffic events, gps, speed, etc. Complemented with data provided by the city of Vigo.

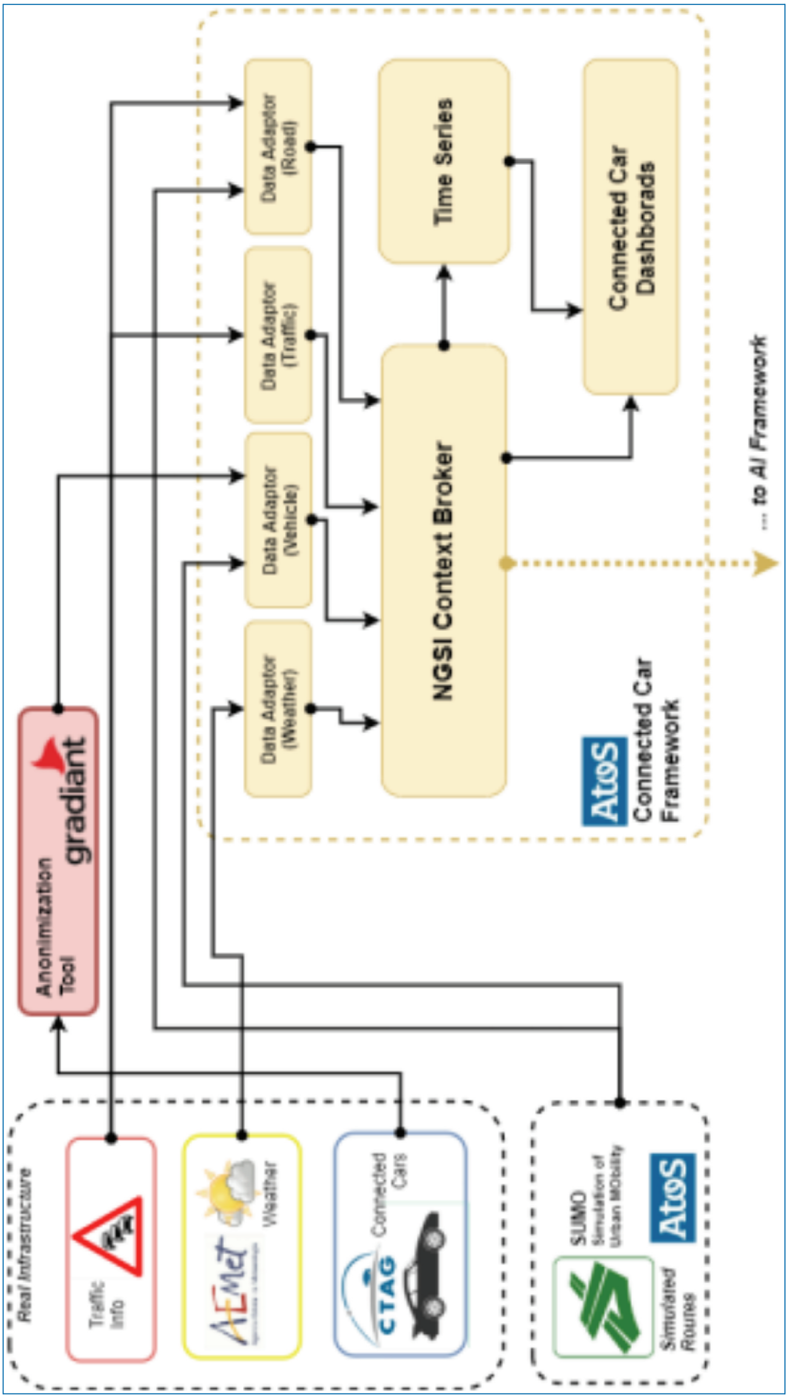


Figure 3.39 Pilot #11 Data collection PoC architecture.

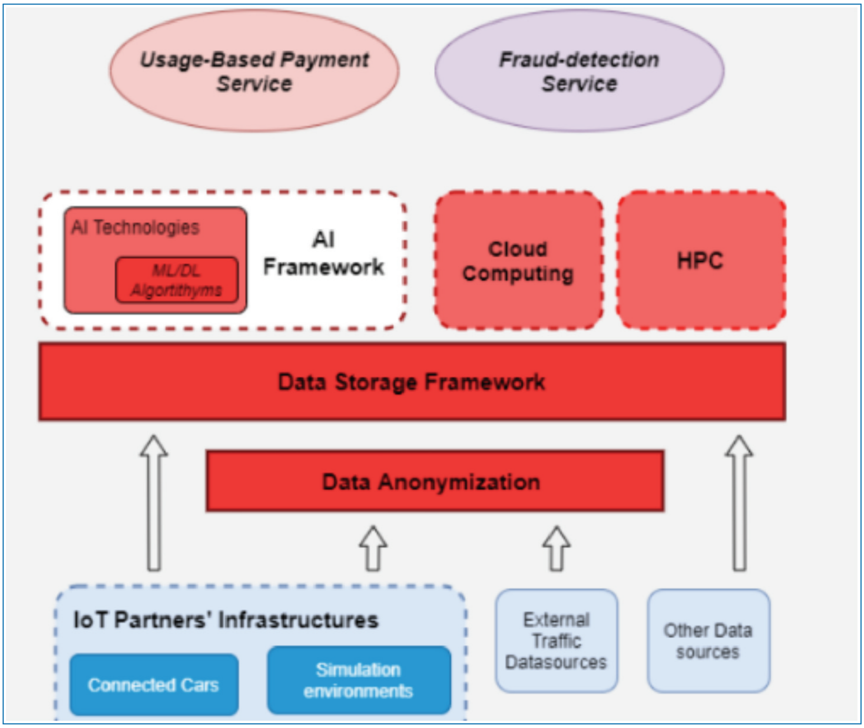


Figure 3.40. Pilot #11: personalized insurance products based on IoT connected vehicles.

Finally, the data will be complemented with some synthetic/simulations of vehicles trips. Based on an opensource tool, SUMO and a custom developed adaptor to integrate and transform the data, according to expected data pipeline and data standards.

About data standards, the Smart Fleet platform layer, in charge of gathering, homogenizing, filtering, etc, is based on a FIWARE platform. Therefore, FIWARE Data Models will be used during the project. In that point, it is expected to contribute back to these standardization efforts fostered by the FIWARE Foundation. Some models would be adapted, or new ones would be created.

The pilot will make use of the following dataset:

- Simulated Urban Mobility Dataset (ATOS, ∼368 GB): Simulated Urban mobility data (mainly vehicles CAN Signals) through different scenarios (cities). Captured from SUMO tool.
- CAN Data (Historical Data) (CTAG): Data collected from vehicle's CAN Bus (20 vehicles driving 4 h/day 1 year). Historical data coming from existing deployments.
- Traffic Events (Historical data) (CTAG, ∼900 GB): Traffic events published by the city of Vigo and DGT (Historical data related to captured CAN Data).

- NMEA Data for vehicles (Historical) (CTAG, ~120 GB): Complementary location (GPS, Timestamp, speed, heading…) for Vehicles' CAN Data (Historical data related to captured CAN Data).
- CAN Signals (Live) (CTAG, ~150 GB): CAN data + Driving style info (revolutions, gear, hard breaking…) + Parking (close doors, windows…) + Maintenance.
- Traffic Events (Live) (CTAG, ~250 GB): Traffic events published by the city of Vigo and DGT.
- NMEA Data for vehicles (Live) (CTAG, ~50 GB): Complementary location (GPS, Timestamp, speed, heading…) for Vehicles' CAN Signal.
- Motor Insurance Data (DYN, ~500 MB): Data concerning motor insurance including data from the policies (duration, covers), data from vehicles (licence No, VIN etc.) and data from drivers (age, experience etc.).

Data Produced

Two main business services will be produced during the pilot's implementation. Therefore, it is not so focused on producing data, but, to provide services use. These services will be used, internally, by the insurance company. In any case, the data produced (or the results) by these services would be considered as data produced, that can be stored in a database, to feed new chains/workflows.

- Pay-As-You-Drive service:
  - Input: driver's trip info
  - Output: a value from 0 to 100 about the driver's behaviour.
- Fraud detection:
  - Input: driver's trip info
  - Output: a kind of driver.

It would be used to compare the kind of driver against an historical register. Example of usage in case of an accident: it would check if the kind of driver differs from previous days (stolen vehicle, identity theft).

Explainable Workflow

The data collected from the vehicle is transmitted to the INFINITECH Testbed, where the data is pipelined into a workflow with a set of steps. Before going to the Smart Fleet platform, data is prepared about regulation and anonymized to protect the driver's privacy. With the data prepared to be managed, the Smart Fleet Platform homogenize, filter, clean, and standardize the data (based on FIWARE Data models). Here the data is prepared as time series for real time management, or, it is stored as historical information. Looking at the AI Platform, it is expected to develop/train two different ML models for the two business services. Once the models have been implemented and these are available in the platform, these models will be trained, supported by the previous data gathering workflow. Getting the training data from the Smart Fleet Platform. It is important to clarify that the training process is not a matter of getting data, training and finish. The model will be constantly trained according to specific scheduling. The model will be always updated with the new data that constantly is generated by the connected car: (1) Data Management: data produced -> preprepared -> gathered -> streamed of store (2) (scheduling time raises) (3) ML Model training: data features extraction -> training model -> store the model The usage of the model, or inference service, or business service, it is an independent workflow. It just deploy a service that will exploit the previously trained model. These are interconnected, the first time the services are deployed with the model, this is linked to future training. When new training succeeds with more accurate models, the inference service will update the resulting model automatically.

Logical Schema

An INFINITECH-RA compliant architecture of the pilot is depicted in the following Figure 3.41:

The pilot's Reference Architecture and main data flows have been presented and detailed. This RA can be simplified considering:

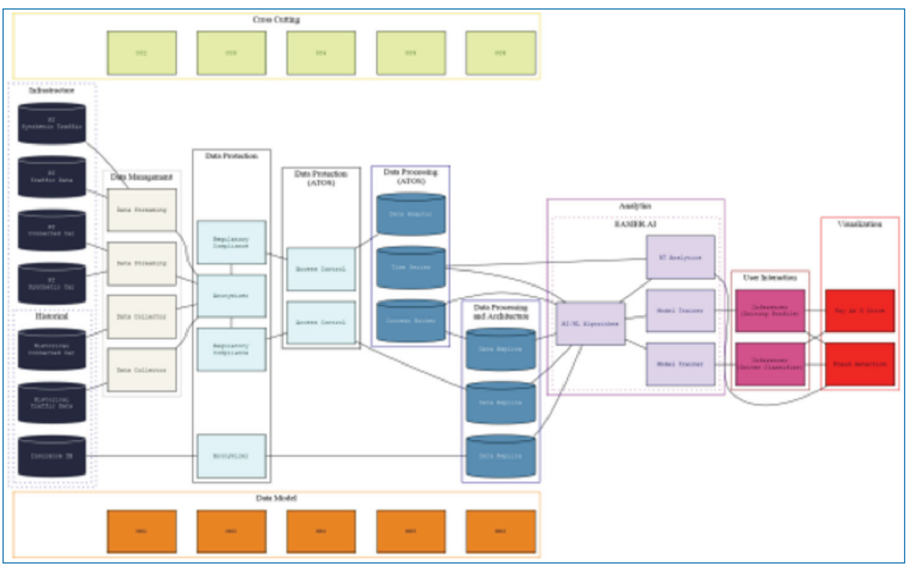


Figure 3.41. Personalized insurance products based on IoT connected vehicles pilot pipeline in-line with the IRA.

- A Data Management layer, that selects, captures and curates the data sources required to implement the pilot's functionalities. On this first stage, real connected vehicles and simulated traffic routes are the main implemented sources, assisted by weather information and traffic incidents collected for the area where the real vehicles will be driving.
- A Data Protection and Data Processing Layers in charge of homogenise and store all data collected, according specific data models (provided by FIWARE), so these are available for the analytics processes. Here are also included all the operations needed to anonymise/pseudoanonymise (as required) the captured data and protect this information from unauthorised accesses as well as data uploading from untrusted sources.
- An Analytics block, fed by the data layers, were different ML/DL technologies and visualization tools will enable data monitoring, analysis, and exploitation. Two main AI models (and inferencers) will be developed here, the Driving profiling and Driver Classifier tools (RA User Interaction), that will come up with the final services: "Pay as you Drive" and the "Fraud Detection" (in the RA's Visualization layer).

The main stakeholders for this pilot are the insurance (car) companies and their insured drivers, who will exploit the driving profiles and drivers' classifications and benefit from customised prices respectively. In Pilot #11 these stakeholders will be represented by Dynamis (DYN) that provides the user stories.

To complete this pilot: Automotive Technology Centre of Galicia (CTAG) manages the real drivers' enrolment and real connected cars, plus traffic incidences around driving areas; Atos (ATOS) provides the pilot's core platform, including the traffic simulation tool and the weather datasets. It also develops the AI models to implement the final services; and Gradient (GRAD) that implements the Anonymization Tool and takes care of all the data managed within Pilot #11 to be GDPR compliant.

Components

The main components to be used in the pilot include:

- Smart Fleet Framework (Context Broker) (Data Management in RA).
- Smart Fleet Framework (PeP Proxy) (Data Security and Privacy in RA).
- Smart Fleet Framework (Historical DB: CrateDB) (Data Source in RA).
- Smart Fleet Framework (Context DB: Mongo) (Data Source in RA).
- Smart Fleet Framework (QuantumLeap) (Data Source in RA).
- Smart Fleet Framework (Weather Injector) (Data Ingestion in RA).
- Smart Fleet Framework (IoT Agent) (Internet of Things in RA).
- Smart Fleet Framework (Grafana) (Interface in RA).

- Security Framework (IDM) (Data Security and Privacy in RA).
- Anonymiser (GRAD Anonymiser) (Data Security and Privacy in RA).
- EASIER.AI (Elasticsearch) (Data Management in RA).
- EASIER.AI (kibana) (Analytics and Machine Learning in RA).
- EASIER.AI (logstash) (Data Ingestion in RA).
- Pay as You Drive Service (Interface in RA).
- Fraud Detection Service (Interface in RA).

Conclusions – Issues and Barriers

Based on the work done so far, the main foreseen challenges would include:

- Gathering enough and relevant datasets from vehicles that allow the system to define and detect a wide enough set of profiles that cover most of the driver's population.
- Identify the proper correlations and relevant parameters from the collected datasets that better define and differentiate the profiles and so, the AI models to infer them.
- The mapping between the drivers' profiles and the context information to provide accurate risks estimations.
- The availability of real datasets (connected cars) from insured drivers to match pilots' services and exploit the results.

Pilot #12: Real World Data for Novel Insurance Products

Risk assessment is an integral part of the insurance industry, but it is usually static, done at the beginning of a contract with a client. The continuous estimation of risk factors is the aim of this pilot, an estimation based not just on medical history, but on lifestyle and behaviour, as they are continuously monitored. This allows the insurance companies to offer personalized dynamic products, where clients' premiums are calculated dynamically based on their habits. Complementary to this, a second service will help to detect possible fraud's situation. Fraud causes not fair costs to the company that would affect indirectly to the good/honest clients.

In both use cases the underline technology is based on analysing Real-World Data (RWD) of the clients. The business analysis part, which determines how healthy a client's lifestyle is, and the detection of possible frauds, will be based on ML techniques. Due to the personal data managed in the pilot, security and privacy will also play an important role (Figure 3.42).

Pilot #12 focuses on health insurance and risk analysis by developing two AI-powered services: Risk assessment, that allows the insurance company to adapt prices by classifying the client according to their lifestyle; and the Fraud Detection which helps to identify fraudulent behaviour of the clients in using the activity

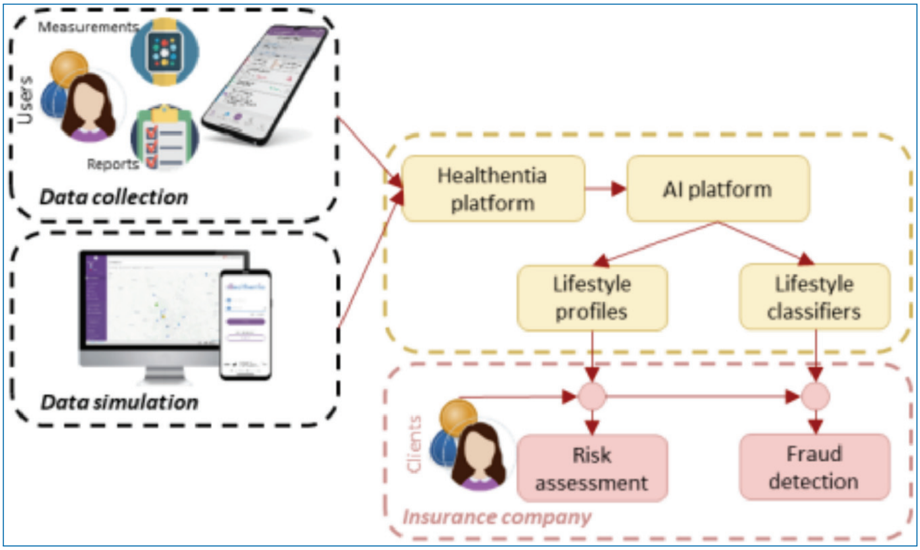


Figure 3.42. Pilot #12 real world data for novel health insurance products overview.

trackers and answering the questionnaires. These two services rely on a people modelling that requires actual data and simulated persons to train. An overview of Pilot #12 is given in the Figure.

Current health insurance services are based on medical history and very static information. The innovation of Pilot #12 lies on applying new technologies (IoT and AI) to provide more dynamic and customized services.

Technological components and Services

The High-Level Architecture presents the software components that build the Pilot's use cases (Figure 3.43). This book serieslinks the shown software components with the corresponding RA layers, providing some details about their implementation. In this sense:

- IoT infrastructures supply raw datasets required for the implementation of the final services. The pilot has already identified and linked the Healthentia platform for Real-World Data collection and the RWD Simulator, both from iSprint as the data sources. Clients' records maintained by the insurance companies are still under investigation.
- Data Collection & Aggregation and Data Normalization components (Data Management and Protection layers of the RA): Healthentia platform already handles ingestion from IoT infrastructures. Remark here the work done to implement and integrate the Real World Data (RWD) Simulator. Also,

Platform	Processors	Cores	Memory	Storage	Special Hardware
ML server	1	6 Cores / processor	32 GB	250GB	Access to GPU special AI hardware is very welcome
Application server	1	6 Cores / processor	16 GB	250GB	None
Anonymization server	1	4 Cores / processor	16 GB	500GB	None
TOTAL	**4**	**18**	**64 GB**	**1TB**	

Figure 3.43. Pilot's #12 hardware testbed.

Gradiant's Anonymizer tool analyses and anonymises (when required) the collected data before being uploaded.

- The ML component (RA Analytics layer) builds upon Scikit-Learn and Keras/TensorFlow the subjects' profiling and subjects' classification inferencers (User Interaction RA layer) that will support the two services (Visualization RA Layer).

Testbed

Pilot's #12 deployment relies on UNINOVA infrastructure, as detailed. Further details of the software/hardware first analysis and their results are summarised in Figure 3.44.

UNINOVA infrastructure it's currently being dimensioned to provide support to several clusters, so it is not still available for deployments. Pilot's #12 demonstrator is currently being designed following a docker+kubernetes approach for their deployment to facilitate possible initial deployment at a temporary server and final deployment at the UNINOVA testbed.

Other non-technical requirements

The success of pilot #12 depends on the wealth of data made available for training and inference. Data is obtained from users employing measurement devices and the will to participate in the pilot by using the measurement infrastructure, reporting symptoms, liquids and meals, and answering the questionnaires. The RWD Simulator is being built to fill in the necessary data volume, but its models depend on the observations made on the actual data being collected.

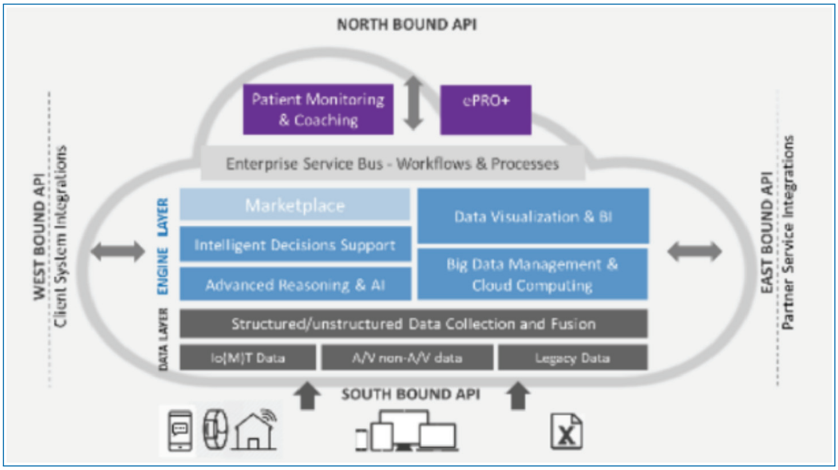


Figure 3.44. Pilot #12: real world data for novel insurance products.

Implementation of a first Proof of Concept

The primary focus of the Proof of Concept demonstrator of Pilot #12 is data collection: what to measure, what to ask for, how to collect and how to simulate. Secondary points of focus are the pilot's testbed and the risk analysis service.

Pilot 12 data collection is based on the Healthentia platform, by Innovation Sprint. Healthentia is an eClinical system that comprises mobile apps at the data source (the pilot participants and a platform for collecting the data. A postal app allows data visualization. The pilot's first goal has been to repurpose Healthentia from the clinical to the health insurance domain. To this extend, the data collected, and the questionnaires forwarded to the pilot participants have been selected and defined. Currently we collect physiological data from four possible sources, a Garmin connector, a Fitbit connector, an Apple Health Kit connector and a proprietary Android sensing service. Our questionnaires span symptoms, liquid and food consumption and the selfassessment of quality of life and health, the EQ-5D-5L questionnaire.

Data are also being provided by the RWD Simulator built for INFINITECH. The simulator accepts people's personality traits and health profiles whilst simulates their activities and the corresponding measurements and questionnaire answers. The simulator data have the exact same structure as the actual ones and are also collected by Healthentia.

Regarding the pilot's testbed, a temporary setup is being managed by Innovation Sprint using a Linux 2020LTS server at Hetzner. It is a VM with 2 vCPUs, 8 GB RAM and 80 GB storage (CX31 instance). There Ubitech's Data Capturing Tool has been configured to capture data from the Healthentia API and store it in the LeanXcale DB.

Finally, regarding the risk analysis service, classifiers have been built using the simulated data to predict if the health of a person is expected to improve or not during a week, based on the week's measurements and reports. Both Random Forest and fully connected Neural Networks classifiers have been trained, with the NN one performing slightly better, achieving 78% correct identification of the health trend.

Expected Outcomes

The pilot will produce a personalized life insurance system, with interfaces for both citizens and insurance companies.

- Dynamic individualized adaptation of coverage and pricing according to client's behaviour and automated risk calculation.
- Fraud Detection as a mechanism that analyze client's behaviour with the aim of historical data.
- Automated data privacy risk assessment and mitigation.

Figure below illustrates the Pilot #12 workflow in detail. It comprises of north, south, east and west bound APIs and different layers applied in the pilot.

Datasets

The main data source in the pilot is the RWD collected by Healthentia. Healthentia is a platform for measuring and reporting RWD. Measurements are based on sensors on smartphones or IoT wearable devices. Reports employ questionnaires that the clients periodically answer utilising the Healthentia app. A secondary source of data is the records of the clients of the health insurance companies. Finally, the data will be complemented with synthetic/simulated data.

- Healthentia Live (average 720kB per user per week): Measured physical activity (steps, floors, sleep and heart rate) and user reported data from users of
- Healthentia SaaS who have given consent Healthentia Simulated (average 720kB per user per week): Simulated physical activity and reported data
- Activity tracking datasets based on 100s of individuals/users that will be engaged in the pilot by RRD
- 100s' of Citizens' feedback datasets
- 1000s' Nutritional information datasets
- Simulated of activity datasets from 1000s of patients based on the simulation module of the Healthentia platform

Data Produced

Two main business services will be produced during the pilot's implementation. Therefore, it is not focused on producing data, but on service provision. These services will be used by the insurance company. To have these services, the ML module will be producing models which can be considered as data produced, that can be stored in a database, to feed new chains/workflows.

- Risk assessment service:
  - o Input: client's lifestyle, enumerated by long-term, short-term averages and trends of physiological parameters that have to do with activity, sleep, the heart, nutrition, hydration, body signals (blood presure, temperature), weight and symptoms (pain, fatigue, diarrhea, nausea, cough).
  - o Output: decisions on health outlook are accumulated across time, forming a health assessment ranginh from −100 to +100.
- Fraud detection:
  - o Input: client's lifestyle enumerated as above, models of all clients.
  - o Output: probability of fraud, enumerating mismatch of current behavior from past behavior of client and other clients.

Explainable Workflow

The RWD collected from the client using Healthentia and the secondary sources is transmitted to the INFINITECH Testbed, where they are aggregated together, anonymised for protection and stored. Stored data are either used to (re)train the risk and fraud assessment models. The trained models are used by the services on input data without anonymisation to provide the risk and fraud assessments. The outputs of the services are offered to the health insurance professionals via the presentation layer of the pilot, together with all collected RWD for human insights/verification. The presentation layer is the Healthentia portal app.

Logical Schema

An initial mapping of the pilot architecture to the INFINITECH-RA is depicted in the following Figure 3.45.

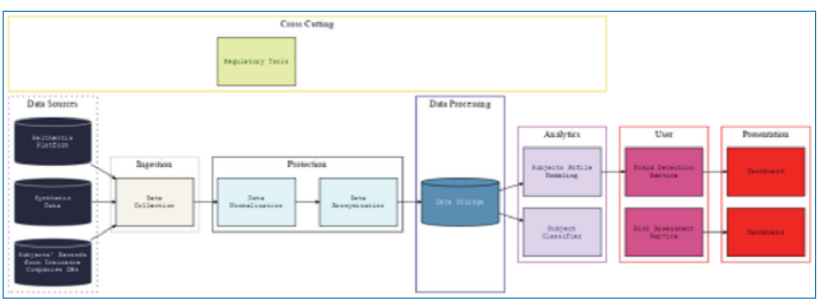


Figure 3.45. Real world data for novel health-insurance products pilot pipeline in-line with the IRA.

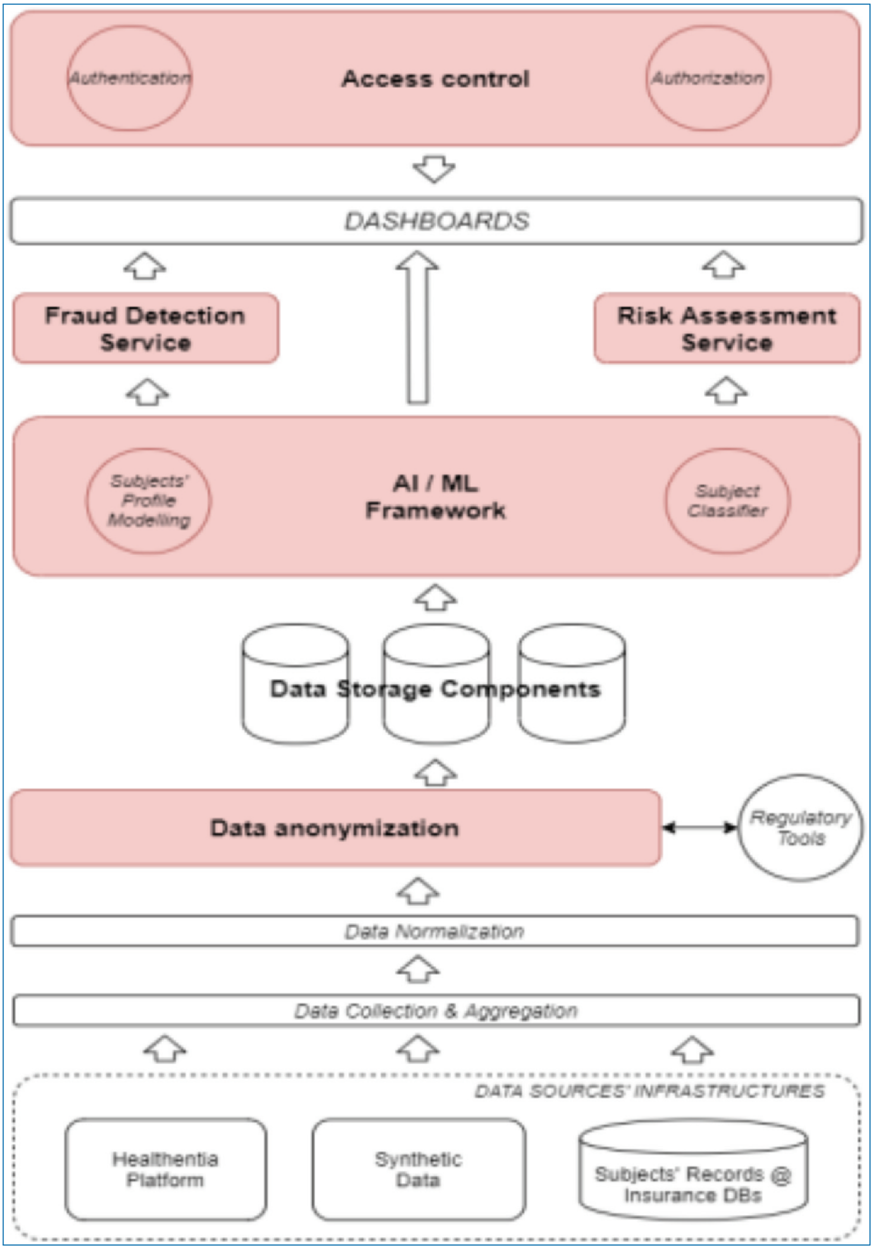


Figure 3.46. Pilot #12 high level architecture.

Pilot's Reference Architecture and main data flows have been presented (Figure 3.46). This RA can be simplified considering:

- A Data Management layer, that selects, captures and curates the data from the actual and the simulated people.

- A Data Protection and Data Processing Layers in charge of homogenise and store all collected data, so these are available for the analytics processes. Here are also included all the operations needed to anonymise/pseudoanonymise (as required) the captured data and protect this information from unauthorised accesses.
- An Analytics layer, fed by the data layers, where different ML/DL technologies and visualization tools will enable data monitoring, analysis and exploitation. Two main AI models (and inferencers) will be developed here, the subject profiling and subject classifier tools, that will come up with the final services in the RA's Visualization layer.

The main stakeholders for this pilot are the health insurance companies and their insured clients, who will exploit the subjects' profiles and subjects' classifications and benefit from customised prices respectively. In Pilot #12 these stakeholders will be represented by Dynamis (DYN) that provides the user stories. To complete this pilot: Roessingh Research and Development (RRD) manages the real subjects' enrolment; Innovation Sprint (iSprint) provides the data collection platform and the subject simulator. Singular Logic (SiLo) and Innovation Sprint (iSprint) develop the AI models to implement the final services; and Gradient (GRAD) implements the Anonymization Tool and takes care of all the data managed within Pilot #12 to be GDPR compliant.

Components

The following components will be deployed and used as part of the pilot:

- UBITECH Data Capturing Tool (Data Ingestion in RA).
- LeanXcale Database (Data Management in RA).
- Innovation Sprint's ML services (risk assessment and fraud detection) (Analytics and Machine Learning in RA).
- ATOS Regulatory tool through Data protection Orchestrator (DPO) (Data Security and Privacy in RA).
- GRAD Regulatory tool through Anonymization Component (Data Security and Privacy in RA).

Conclusions – Issues and Barriers

The PoC of Pilot 12 allowed us to implement the data collection system, addressing both, the what and the how. The low engagement of the PoC participants (about 50%) is alarming and will be addressed in the Data Sharing Acceptance and Usability Study from the data, privacy and UI/UX aspects. It is also being addressed technically by increasing the measurement options and optimising the Android sensing service. Our aim is to be gathering soon enough and relevant data from diverse users to facilitate both risk assessment and fraud detection services.

The testbed will be transferred to its permanent location at the NOVA server, but the PoC already set in motion all the collaborations necessary for its setup amongst the INFINITECH partners not members of the pilot.

The risk assessment service has been addressed at the PoC via an initial predictor of weekly variations of health. Both classifiers and regressors will be built in the coming months, the feature vector used to train them will be optimised and as a result the heart of the service will be in place. The fraud detection has not been addressed yet, and this is a concern, since today people cheat on their activity trackers just to get a badge in their favourite wellness app. This could escalate when health insurance discounts are involved.

3.4 Predictive Financial Crime and Fraud Detection Pilots

Pilot #7: Avoiding Financial Crime

The aim of this pilot is to see if we can detect Financial Crime more accurately and sooner than any existing system by using AI and advanced computational power abilities. The goal of Operation Whitetail is to explore how next generation technical solutions like Machine Learning and AI could help to create a more accurate, comprehensive and near real-time picture of suspicious behaviour in the Financial Crime remit (Anti-Money Laundering and Combat Terrorist Financing). The goal is to explore more accurate, comprehensive and near real-time pictures of suspicious behavior in Financial Crime, Fraud, in the use case of instant loans. Such loans can be requested online and are subject to fraud and crime, e.g. identity theft. Based on comprehensive data including KYC and transaction data a financial crime risk score is calculated by AI/ML algorithms. This way the instant loan can be approved or denied related to this score.

Within the pilot the following processes are addressed:

- KYC (Know Your Customer), for screening the vast amount of available data sources in near-real time, to ensure that KYC data is automatically updated to the most recent information available on the customer facilitating data quality.
- Customer risk profiling, based on feeding the transaction-based customer's behavioural profile data and KYC results leading to an advanced risk score that could provide a holistic customer risk profile and will enable the business to respond quicker to newly identified risk and changes in criminal behavior.

Therefore, the pilot plans to utilize use synthetic or anonymized data as source. Bank internal and bank external sources of KYC data shall be joined in an advanced KYC data store.

The external data sources include public sources or sources actively shared by the customer. Information from external sources will be obtained traditionally, e.g. credit reference agencies; sanctions lists.

The advanced KYC data are used to extract a customer profile. Additionally, customer transactions patterns are extracted from the customers' transaction data.

Expected Outcomes

- Better detection of suspicious customers and transactions (transaction monitoring).
- Fewer false positives.
- Near-real time update of customer's extended KYC profile.
- Near-real time update of the customer's behavioural profile.
- New, holistic risk score based on a complex risk model.
- Analyses financial crime alerts/anomalies more effective and efficiently.

Figure 3.47 illustrates the design architecture of the pilot #7 workflow.

Datasets

- Transactions and Customer attributes (anonymized).
- The pilot will use synthetic or anonymized data as source. In the bank internal data pool sources will be accessed. This data pool also includes bank internal and external KYC data and internal transactional data. These data shall be joined in an advanced KYC data source and the relevant data for the use case will be extracted from that Due to compliance rules, these data need to be treated confidential. In a 1st step use related data representing customer profiles will be extracted facilitating the development of synthesized data sets giving insight to the financial crime risk score and facilitating the development of AI/ML models.

Data Produced

The pilot will produce data giving insight to the financial crime, i.e. instant loan, risk score. This may include a risk score, customer data, transaction patterns and details. The detailed data, which will be presented, are yet to be specified depending on the advice of Financial Crime experts in the bank.

Explainable Workflow

Within the pilot the following processes are addressed: KYC (Know Your Customer), for screening the available data sources in near-real time, to ensure that KYC data is automatically updated to the most recent information available on the customer facilitating data quality. Customer risk profiling, based on feeding the transaction-based customer's behavioural profile data and KYC results leading

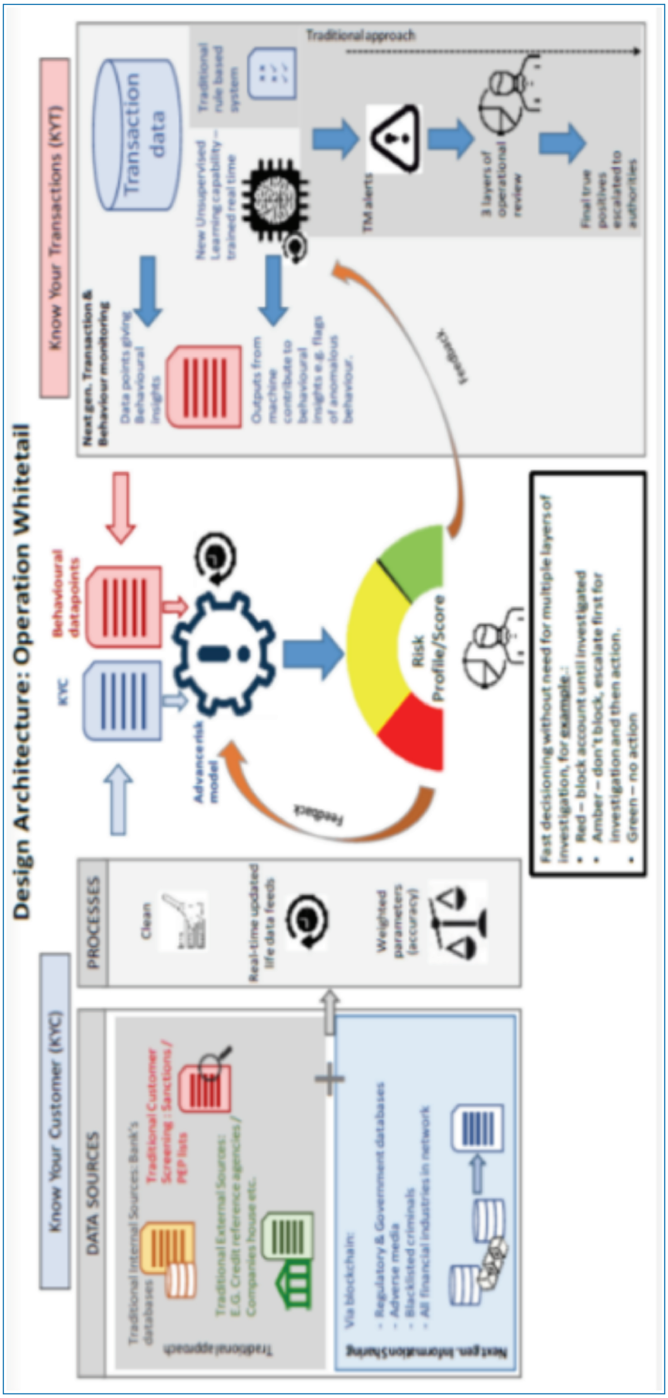


Figure 3.47 Pilot #7: avoiding financial crime workflow.

to an advanced risk score that could provide a holistic customer risk profile and will enable the business to respond quicker to newly identified risk and changes in criminal behaviour.

The workflow will produce data giving insight to the financial crime risk score. This may include a risk score, customer data, transaction patterns and details. The detailed data, which shall be produced, are yet to be specified depending on the advice of Financial Crime experts in the bank (Figure 3.48).

Logical Schema

Figure below summarizes financial crime pilot pipeline in detail. The component involved are described in the following section.

Components

Due to strict compliance and approval procedures in CXB the pilot operations are facilitated splitting the tasks in a pre- and pilot processing part. The pre-processing part may be mimicked by INFINTECH tools based on beforehand synthesized/anonymized data. However, for a smooth progress of the pilot development, a bank internal and an INFINITECH process will be considered as a first step.

A List of the main components to be deployed and used in the pilot follows:

Pre-processing – Inside the bank by bank approved tools:

- Bank Data Pool (Data Sources in the RA).
- Bank Data Pool Extraction (Data Management in the RA).
- Bank Data Pool Join (Data Management in the RA).
- Data synthesation/anonymization (Data Source in the RA).

Pilot-Processing – The synthesized/anonymized data then are used in the INFINITECH Pilot (Figure 3.49)

- Synthesized/anonymized data (Data Source in the RA).
- Data Ingestion (Ingestion in the RA).
- Data Analytics/Scoring (Analytics in the RA).
- Visualization (Presentation in the RA).

Pilot #8: Platform for Anti Money Laundering Supervision (PAMLS)

The objective of the Pilot, is to develop a Platform for anti-money laundering Supervision (PAMLS), which will improve the effectiveness of the existing supervisory activities in the area of anti-money laundering and combating terrorist financing (AML/CTF) by processing large quantity of data (Big Data) owned by the Bank of Slovenia (BOS) and other competent authorities (FIU).

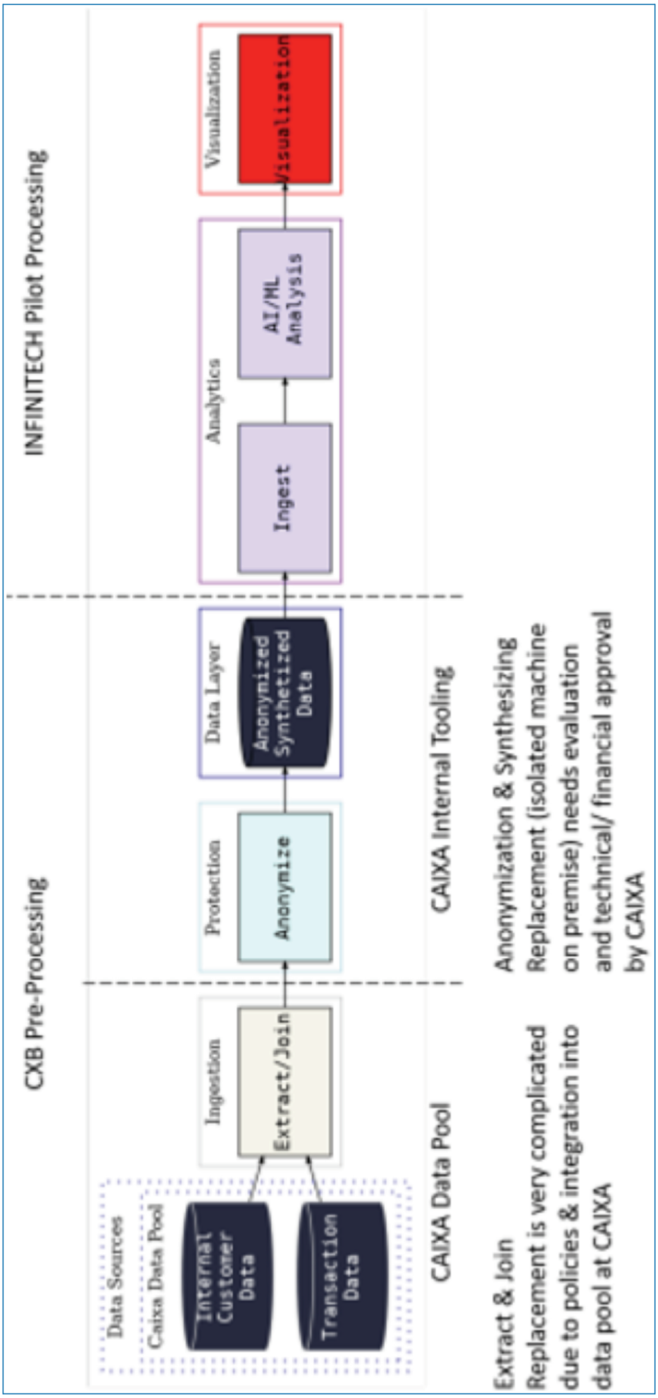


Figure 3.48 Financial crime pilot pipeline in-line with the IRA.

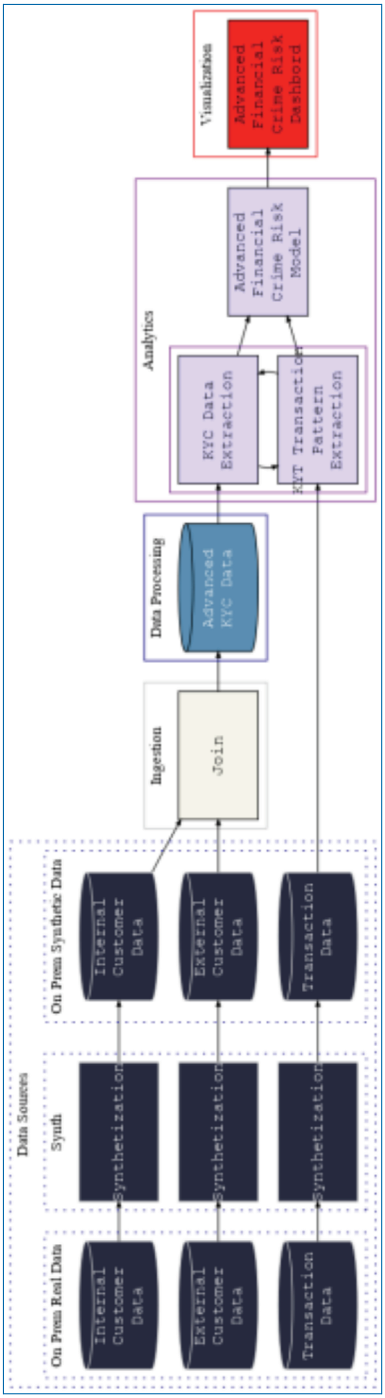


Figure 3.49 Pilot #7 RA.

Figure 3.50. Pilot #8: components of anti money laundering supervision (PAMLS).

The book series will develop a platform named PAMLS that will improve the effectiveness of the existing supervisory activities in the area of ML/TF, by (Figure 3.50):

- Automated and transparent data gathering that will include data quality control.
- Improved analysis of big data coming from wide range of different sources (e.g. payment transactions, data acquired from the FI; business register etc.).
- Improved Risk Assessment (as an ongoing and cyclical process) with automated feeds from big data analysis.
- More cost-efficient risk assessment process due to less time-consuming data gathering tasks, assessments of the FI and the financial sector and semi-automated features.
- A more effort-efficient risk assessment process, additional resources can be focused on the supervision of identified high risks.

PALMS will consist of four main business services:

- Risk assessment tool: to assess the money laundering and terrorist financing (ML/FT) risks of financial institutions (FIs) and the risk of a whole sector to support risk based supervision,
- Screening tool: for screening payment transactions, enriched with data from business register (ePRS) and transactions accounts register (eRTR), to recognize unusual patterns that could indicate typologies and risks of ML/FT at level of individual FI or the whole sector,

- Search engine: allowing supervisor to look for a specific transaction or a sample of transactions,
- Distribution channel: for secure gathering data that will feed risk assessment tool and screening tool.
- Risk Assessment tool: which provide risk assessment functionalities within PAMLS.

Technological components and Services

Following a list of components and grouped by the Reference Architecture layers for Data Analytics and User Interfaces.

Data Analytics Layer main components:

- Risk Calculation engine and Complex search services, which will be implemented specifically for Pilot8 requirements and therefore will be tailored to BOS specific:
  - Current status: 1st version developed on scrambled data
  - Next version: M27
- Anomaly detection & prediction analysis, which will provide functionalities for anomaly detection and prediction for time series data including Pattern analysis, which will provide analytical services on data graphs, including detection of complex patterns on data graphs:
  - Current status: to be developed
  - First version: M27
- Stream story is a component for the analysis of multivariate time series. It computes and visualizes a hierarchical Markov chain model which captures the qualitative behaviour of the systems' dynamics, where system is described with a group of time series.
  - Current status: to be developed
  - First version: M27

User Interfaces Layer main components:

- Risk Assessments tool – 1 st version already developed, next version M27.

Expected Outcomes

- Automated, more accurate and more dynamic detection of money laundering transactions.
- Scalable, multimodal data platform, compliant with legal and regulatory framework, measured KPIs associated with the volume of data processed, the speed of processing and the effort required for processing will be tracked.

Datasets

Relevant datasets, planned to be analyzed within PAMLS are:

- TARGET2 transactions:
  - Transactions executed by the Slovenian payment institutions within TARGET2 (TransEuropean Automated Real-time Gross Settlement Express Transfer System)
  - High value (above 50.000 EUR), urgent transactions in EUR
  - Transactions processed through BOS payment systems (responsible BOS Payment Settlement and Systems department – PPS) Confidential data.
- SEPA transactions:
  - Transactions executed by the Slovenian payment institutions within SEPA (Single Euro Payments Area)
  - Domestic and international transactions within SEPA area in EUR under 50.000EUR value
  - Transactions processed through payment systems by third party provider Confidential data.
- FIU transactions (public data):
  - Transactions related to high risk countries above 15.000 EUR reported to the Slovene
  - Financial Intelligence Unit (FIU) Public data.
- FI identification data
  - Identification information about Financial Institution (FI)
  - Aggregated statistical data on the FI inherent risk and control environment (number of clients, number of Suspicious transactions reports (STR) etc.)
  - FI reports to the BOS (reports are confidential) Confidential data.
- ePRS data
  - Slovenian Business Register (public data on legal entities) Public data.
- eRTR data
  - Slovenian Transactions Accounts Register (public data on legal entities) Public data.
- High risk country list
  - List of countries defined as high risk due to lack of or not effective AML/CTF system
  - List is managed and published by the Slovene FIU (public data) Public data. Personal data will be anonymized by the source, prior data delivery to PAMLS.

Data Produced

Ongoing risk assessment for the purpose of the Anti-Money Laundering and Combating Terrorist Financing Supervision over the FI and FI sector.

Explainable Workflow

PAMLS will use various data sources, which we can divide in to three groups:

The first group consists of transactional data process through payment service providers in Slovenia (TARGET and SEPA transactions). This group will first be enriched with ePRS and eRTR data, and then pseudo-anonymized (for end user anonymized). Before data will be stored in PAMLS internal data storage it will also be joined with High risk country List.

The second set of data sources represents public data (FIU transactions) that will also be enriched with ePRS and eRTR data and than joined with High risk country List and stored in PAMLS internal data storage.

Third group of data sources represents FI data (data on FI inherent risk and control environment), which will stored in PAMLS internal data storage after positive Data Quality Check.

After data is ingested in PAMLS platform, it needs to be preprocessed in a way, that information is properly enriched and it needs to be provided in a suitable data format (vectors, graphs). Process of feature engineering, tailored to specific goals, will follow. PAMLS will develop and test novel approaches for detecting unusual patterns of ML/TF, which could be labelled as high risk later in the process and will have an effect on final FI risk assessment. Part of the PAMLS is also Risk Calculation engine. There the risk calculation will be continuously calculated on a level of a sector or a particular FI, using predefined Risk Assessment methodology. To empower bank analysis, to develop and test novel approaches, PAMLS provides three components: Stream story, Pattern discovery & matching, Anomaly detection & prediction. With introduction of enriched graph topologies and a hierarchical Markov chain models, PAMLS will capture the qualitative behaviour of the systems' dynamics and enable analyst to discover new regularities and correlations on a larger scale. These components will enable iterative development of potentially new upgrades to existing Risk Assessment methodology and discovery of novel or additional money laundering and terrorist financing typologies. PAMLS will also provide three different user interfaces, which corresponds to 3 different use cases.

Logical Schema

The following figure illustrates an initial logical mapping of the pilot components to the layers and pipelines approach of the INFINITECH-RA (Figure 3.51).

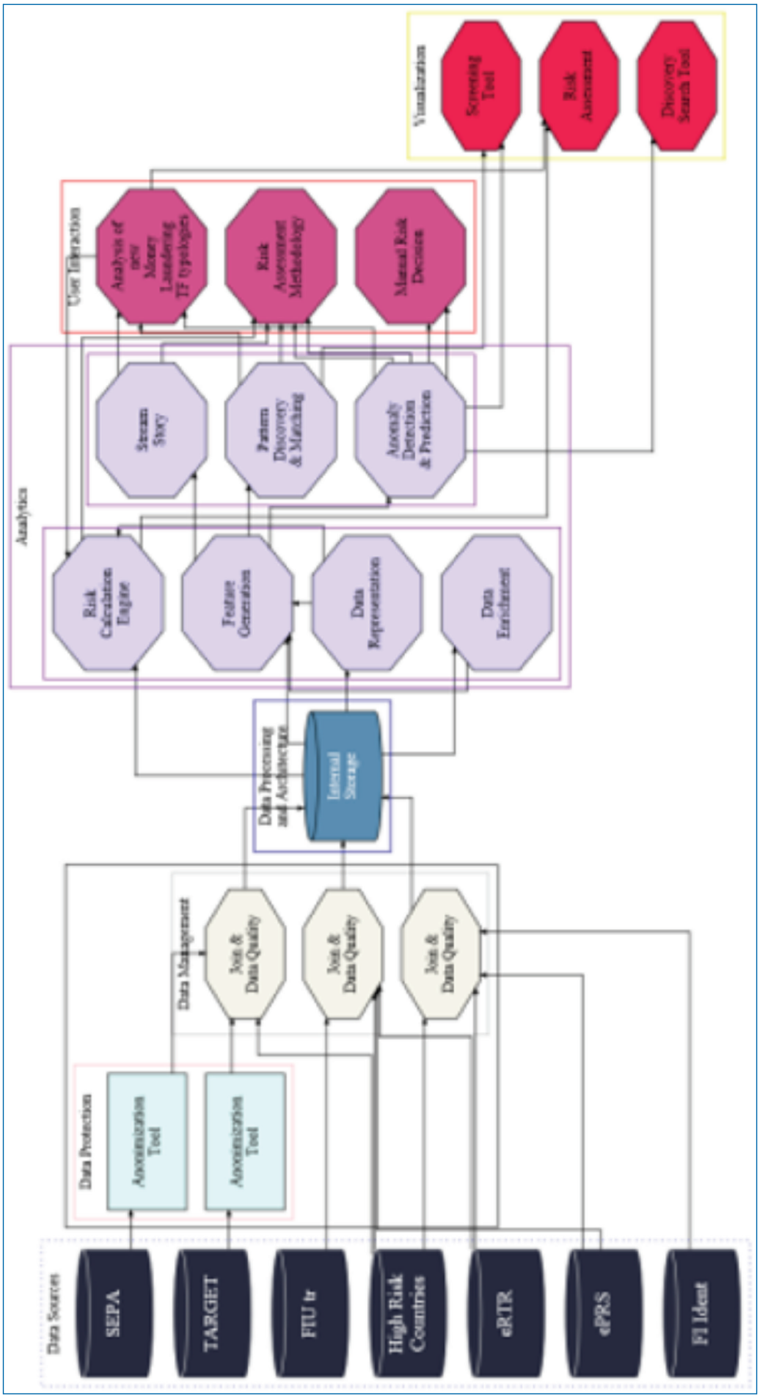


Figure 3.51 PAMLS pilot pipeline in-line with the IRA.

The components implemented for the first PoC include:

- Data sources:
  - Synthetic FI data (data about inherent risk and their control environment)
  - Implemented APIs
  - Implemented DQ 1st version
- Risk methodology – framework
- Risk engine:
  - Defined technical requirements (flexibility, scalability, DQ..)
  - 1st version implemented
  - Validation & verification of risk calculations

Additionally, the Stream Story component was used to search data for meaningful patterns, however, since it was applied on scrambles data, meaningful verification of data patterns was not possible.

Testbed

Pilot#8 will be hosted at the Testbed on the premises of the BOS, it is ready and it has already deployed the software components and data to implement the PoC. The testbed has been specified in the following manner:

Hardware Description
HP Z4 G4 WKS CPU: Intel XeonW-2125 4.0 4C
RAM: 256GB (8x32GB)
DDR4 Graphic: NVIDIA Quadro P400 2GB (3)mDP Graphics
Disk: Z Turbo Drv 1TB PCIe NVMe OPAL2 TLC SSD

Testbed is based on Windows operating system and include software:

- Libraries (QMiner & SNAP).
- External tools (candidates): (PostgreSQL/Elastic Search).
- Programming tools (C/C++ compilers (GNU or Microsoft, Python, Node.js).

Others non-technical requirements

Due to standard security measures at BOS, in order to use the Pilot#8 testbed physical presence of JSI development team at the BOS premises is required. As a consequence of the strict measures to mitigate the spread of Covid-19 and additional security measures external parties do not have granted access to the BOS premises (JSI as a partner on Pilot#8 included) (Figure 3.52). During PoC implementation the development was done on scrambled data in order to preserve data privacy. Therefore PoC implementation was done at JSI site, however validation and initial testing was done by BOS in several phases, where risk calculations were validated.

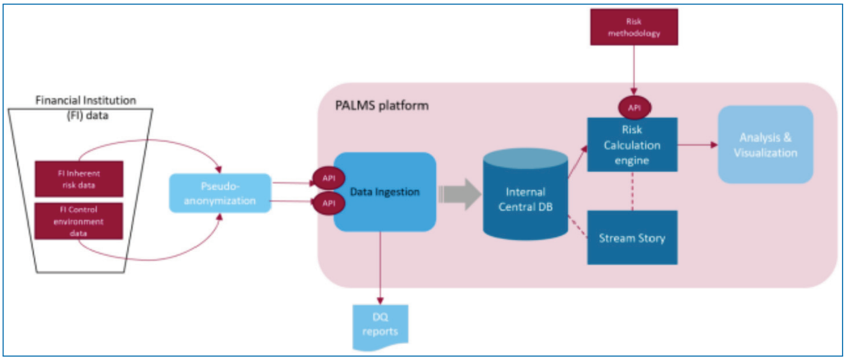


Figure 3.52. Risk assessment tool data flow.

At the next phase, PoC will be transferred to the Pilot#8 testbed at BOS site. For next development phases of Pilot#8, it is crucial that during initial development of AI components appropriate test data is available, while proper validation and testing needs to be done on real data at BOS site.

Implementation of a first Proof of Concept

In accordance with the development timeline first prototype of the Risk assessment tool was developed:

Risk Assessment Tool – PoC was implemented in agile quick cycles. It provides the following functionalities:

1. Sector Risk Assessment view allows us to review the risk of all financial institutions (FIs) based on the assessment of their inherent risk and control environment in the specific year. Based on their risk, FIs are placed in the Risk Assessment Matrix in to low – medium – medium high – or high risk. Supervisory authority will focus on those presenting higher risk. Since the final risk assessment is an evaluation of inherent risk and control environment the view also enables graphical schema of those two important elements of the risk assessment. Changes in the risks of the specific FI is also an important factor. Therefore Sector Risk Assessment view enables also historical view for selected FIs.
2. Inherent risk/Control environment view: Inherent risk (and similar for control environment) consists of different risk areas and those consist from different elements. In this view supervisor can drill down to the specific elements and compare FIs amongst each other. Also, supervisor receives information which areas or elements of the inherent risk or of the control environment present more risk for a specific FI and can therefore focus on those areas during the on-site supervision.
3. Bank Profile view enables the supervisor to select a FI for a detailed review. In the first version of PoC the view consists of FI basic information (FI ID

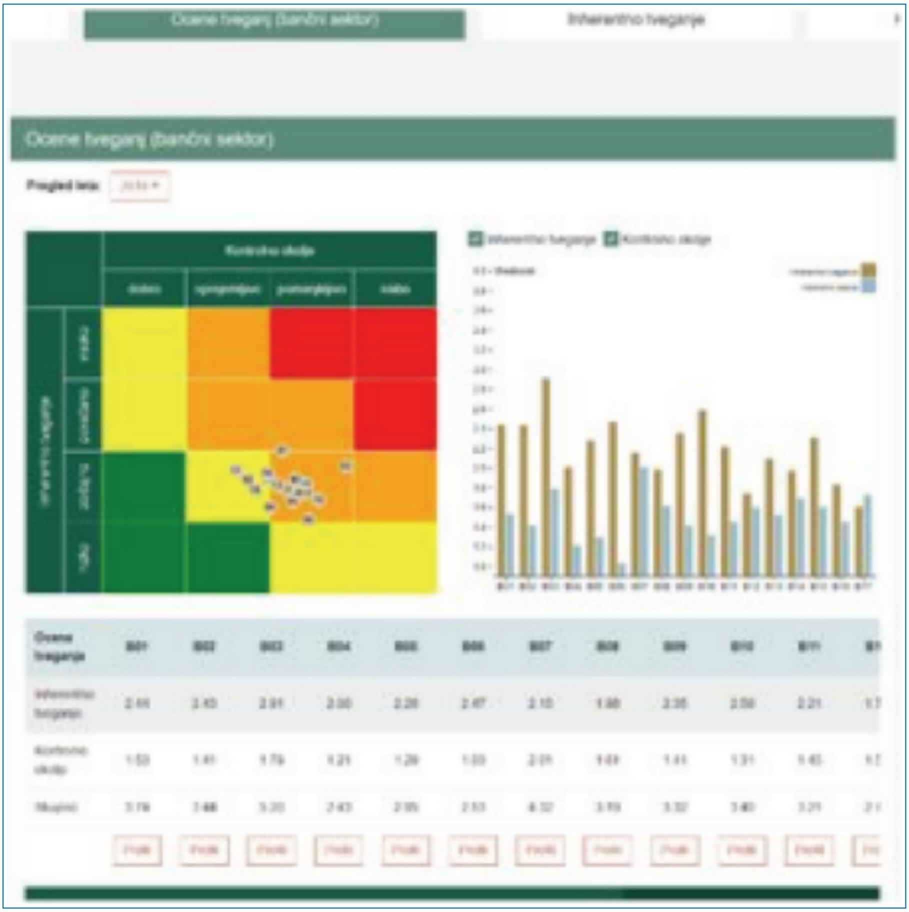


Figure 3.53. Sector risk assessment view.

Figure 3.54. Inherent risk and control environment view.

Figure 3.55. Bank profile view.

Card), graph on the FIs risk assessment changes through the year (historical view) and detailed information on FI inherent risk and control environment (Figure 3.55).

Components

The pilot will use the following components:

- Risk Calculation engine and Complex search services (Analytics in the RA)
- Anomaly detection and prediction component (Analytics in the RA): will provide functionalities for anomaly detection and prediction for time series data including Pattern analysis. The latter will provide analytical services on data graphs, including detection of complex patterns on data graphs;
- StreamStory component (Analytics in the RA): a component for the analysis of multivariate time series. It computes and visualizes a hierarchical Markov chain model which captures the qualitative behaviour of the systems' dynamics, where system is described with a group of time series;

- Pattern discovery and matching component (Analytics in the RA)
- Pseudo-anonimization tool (Data Managenent in the RA)
- PostgreSQL (Data Managenent in the RA)
- ElasticSearch (Data Managenent in the RA)
- NEO4J (Presentation in the RA)

Conclusions – Issues and Barriers

Development in Pilot#8 is going according to the plan. As PoC we provided 1st version of one of the use cases – Risk Assessment tool. It will facilitate supervision activities in terms of providing relevant data analysis on the fly. It will enable risk analyst to gather and analyse risk assessment data more efficiently and provide straightforward analysis of risk methodology on one hand and analysis of Slovenian FIs in terms of Inherent and Control risks through the years. It enables comparison of particular risk categories and provide detail insights.

Regulatory requirements used within the Pilot#8 requires additional actions that were not foreseen at the start of the project (approval by compliance and management board, anonymization requirements etc.). Although such additional actions could affect the pilot development timeline, it does not change planned development and defined use cases set.

Pilot #9: Analyzing Blockchain Transaction Graphs for Fraudulent Activities

There can be blockchain crypto currencies and tokenized assets (e.g. USD, EUR, TRY tokens) that are obtained fraudulently as a result of ransomware and theft of funds. These fraudulent assets can go through various transfers on the blockchain and enter the regulated environments in different jurisdictions. As a result, it is possible that a company may accept deposits of crypto currencies and tokens that can be traced to addresses involved in fraudulent activities.

Pilot #9 is developing a parallel and scalable transaction graph analysis system that can construct and operate on the massive Bitcoin and Ethereum blockchain transaction graph with distributed dynamic data structures on an HPC cluster. During Period 1 of the project, the pilot has implemented parallel graph algorithm based fraudulent activity analysis. In the Period 2 of the project, it has also initiated implementation of machine learning based analysis algorithms. The pilot is also providing a user interface that provides various queries and visualization of results using graph drawing package.

Pilot #9 aims to detect fraudulent activities monitoring blockchain transactions. Blockchain crypto currencies and tokenized assets that are obtained fraudulently can go through various transfers on the blockchain and end up as stable coins (e.g. USD, EUR, TRY tokens) in different jurisdictions. As a result, it is possible that a company that accepts crypto-currencies, or stable coins, is paid by stable coins that can be traced to addresses involved in fraudulent activities. Holding crypto-currencies or stable coins that originated from fraudulent or sanctioned addresses can be risky for the company. Hence, construction of the massive blockchain transaction graph and its analysis is necessary to trace and detect fraudulent addresses. Since blockchain data is constantly accumulating and will be growing at increasing rates in the future, a parallel scalable transaction graph analysis system is being developed that runs on HPC cluster and that can process the growing transaction graph without encountering performance bottlenecks.

The main innovation of the pilot lies in the applicability of HPC technologies to analyse Blockchain (huge) transaction graphs, to quickly detect possible frauds based on blacklists.

Following the main components of the pilots and the partners in charge of the development:

- Blockchain Transaction Dataset Preparation Component (developed by BOUN (Bogazici University)),
- Scalable Transaction Graph Analysis Component (developed by BOUN (Bogazici University)),
- User Interface for Blockchain Transaction Reports and Visualization Component (developed by AKTIF Bank).

The final users will be banks who need to do analysis of blockchain addresses. Developed services can also be offered as a service to companies who need to do such checks, for example, companies that accept token payments.

The first year aimed at massive blockchain dataset preparation, an HPC based cluster parallel transaction graph analysis system construction and coding of traversal-based graph algorithms. The second year will concentrate on machine learning based approaches for analysis using, in particular, the graph system developed in the first year for feature extraction.

Testbed

Figure 3.56 depicts the testbed which is currently set-up and running on the Amazon cloud. The following is the hardware and software configuration that is used for the testbed:

Hardware:

- HPC Cluster on Amazon Cloud (16 c5.4xlarge instances), each instance having 16 virtual CPUs, 32 GiB memory and 500 GB SSD storage.
- A medium Amazon instance for running message queue.

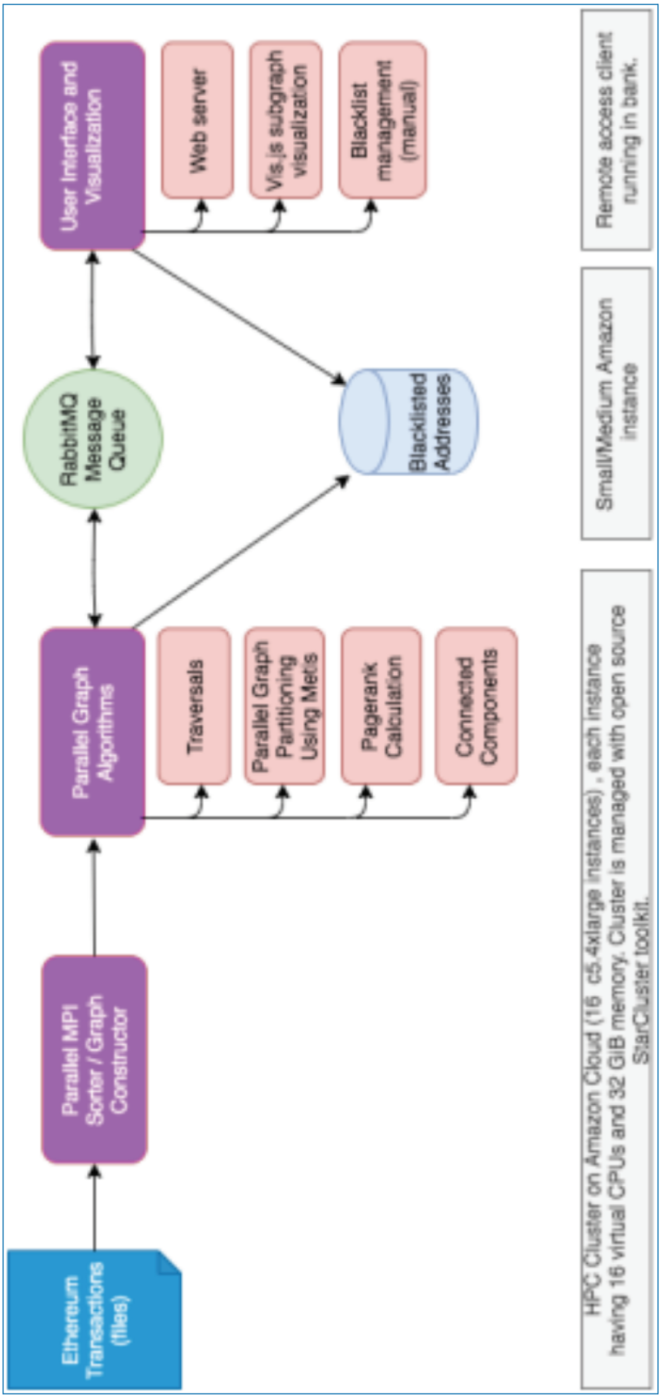


Figure 3.56 Pilot #9 architecture of the proof of concept.

Software:

- Ubuntu Linux operating system
- StarCluster HPC cluster toolkit.
- MPI message passing interface
- Rabbit MQ message queue
- Metis Parallel graph partitioner
- Vis.js open source graph visualization software for web interface.

Implementation of a first Proof of Concept

The figure below shows the architecture of the Proof of Concept system that has been implemented.

Expected Outcomes

- Software that runs on hybrid CPU/GPU cluster;
- Partitioned transaction graphs of the current blockchains easy to be managed;
- Blacklist of hacked/fraudulent account addresses on bitcoin and Ethereum that is collected from public sources on the Internet

Figure 3.56 shows the Pilot #9 workflow. The input data are Ethereum and Bitcoin blockchains and their nodes. The inputs are fed into blockchain transaction graph analysis which ultimately provides bank or business queries.

Services to be implemented according the user stories.

PoC currently offers parallel scalable blockchain transaction graph construction, parallel graph traversals that trace customer addresses to blacklisted addresses by returning the traced subgraph. Parallel Pagerank algorithm that finds important addresses is also offered as a service. The transaction graph can also be partitioned in parallel using the Metis software. These services are offered on the whole dataset graph having 633M transactions.

Components implemented, interactions and deployment

1. The following components of the project have been built as proof of concept: Blockchain Transaction Dataset Preparation Component.
2. Scalable Transaction Graph Analysis Component.
3. User Interface for Blockchain Transaction Reports and Visualization Component.

Component (1) parses Ethereum raw data to extract transactions which are saved as files. (2) constructs the distributed and partitioned graph on HPC cluster using the transaction files and performs parallel graph algorithms. (3) communicates with (2) via RabbitMQ message service, submits queries and displays returned results on web page and produces graph visualization output using the Vis.js package.

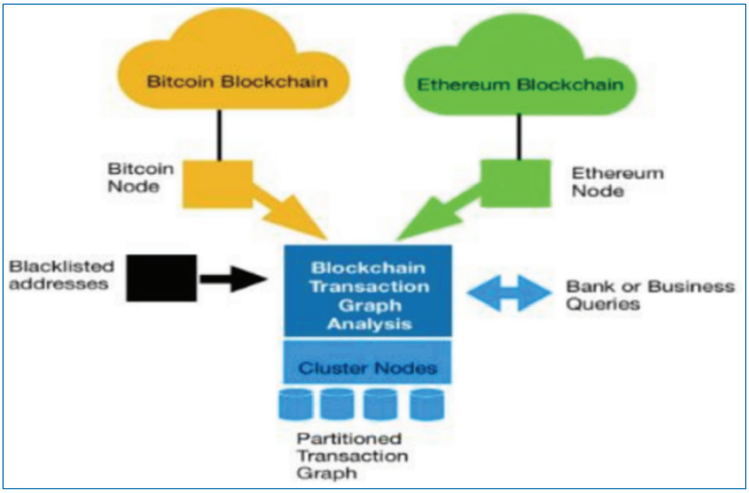


Figure 3.57. Pilot #9: Analyzing blockchain transaction graphs for fraudulent activities workflow.

Datasets

The following data sources are used:

- Public Bitcoin Blockchain Data (BOUN)

Bitcoin transfers (send transactions);

- Public Ethereum Blockchain Data (BOUN)

– Ether transfers (send transactions) and ERC20 Token Smart contract transactions (major popular tokens including stable coins like EURS, GUSD, USDT, TRYB, PAX, TUSD, QCAD, XAUT)

- Bitcoin and Ethereum Addresses Database (AKTIF)

Database of all Bitcoin (within block ranges 0-674999) and Ethereum addresses (within block ranges 0-10199999) are maintained as a database with capability to label each address with features.

Blacklisted Bitcoin and Ethereum blockchain addresses that are obtained from the Internet by manual search for published hacked/fraudulent accounts and addresses involved in ransomware activities.

Data Produced

The following data generated:

- Extracted Ethereum and major ERC20 token transaction data that is also made available at https://zenodo.org/record/4718440#.YXkLhtZBw1I. It can be downloaded by researchers and businesses;

- Paths and subgraphs that show tracing of blockchain addresses to blacklisted addresses;
- Importance values of addresses computed by running parallel Pagerank algorithm is produced as data. This rank data can be used as an important feature in the machine learning algorithms.

Explainable Workflow

Pilot #9's Blockchain Transaction Dataset Preparation Component parses raw blockchain data and extracts Bitcoin, Ethereum and major ERC20 token transactions (such as Gemini USD (GUSD), Tether USD (USDT), Tether Gold (XAUT), Statis Euro (EURS) and Turkish BiLira (TRYB)) that come from the Bitcoin and Ethereum Mainnet blockchains. After retrieving all the blocks up until now, this component is run periodically to retrieve newly generated blockchain blocks during the period.

Scalable Transaction Graph Analysis Component of the pilot takes the full bitcoin and Ethereum public transaction dataset. Graph traversal algorithms are used to analyze the data. Parallel graph traversals are used to extract features that are in the form of subgraphs. Since the transaction graph size is massive and dynamically growing, it constructs distributed and partitioned transaction graph in parallel using MPI message passing libraries in order to achieve scalability. Graph analysis service is interacted through a message queue that takes commands in YAML format. The outputs of the service are in the form of graph paths or subgraphs that show tracing of Blockchain addresses to blacklisted addresses. In the second period of the project, machine learning algorithms have been started to be developed. Bitcoin and Ethereum transaction data and blacklisted address lists as well as pageranks that are computed in parallel are used in machine learning algorithms.

Finally, the User Interface for Blockchain Transaction Reports and Visualization functional service interacts with the Scalable Transaction Graph Analysis and presents results in a web browser. When subgraphs are returned that trace customer addresses to blacklisted addresses, these subgraphs are output in vis.vj graph visualization software format for viewing in browsers. The business service is provided through a RabbitMQ message queue that takes commands in the YAML format. Visualization of transaction graph traces as well as a simple address score based on shortest path from blacklisted addresses is also provided.

Logical Schema

The following figure illustrates how the pilot architecture can be expressed in terms of the layers and the pipelines approach of the INFINITECH-RA.

Platform for data gathering (Related Reference Architecture Layers (Figure 3.58): Blockchain, Infrastructure and Data management).

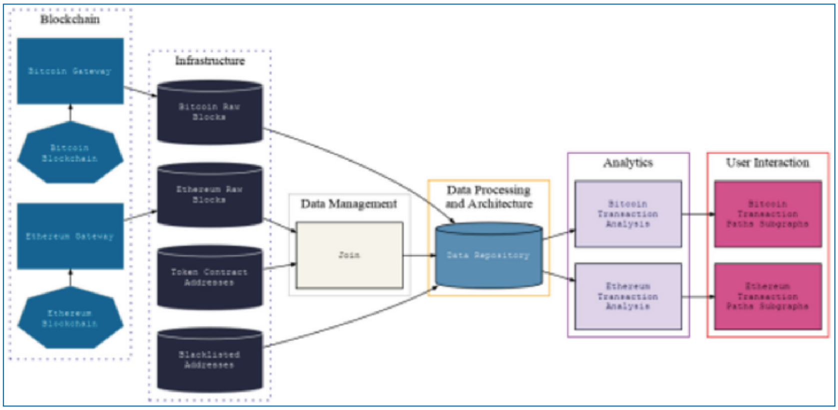


Figure 3.58. Blockchain transaction graphs analysis pilot pipeline in line-with the IRA.

Blockchain dataset component is implemented as scripts that retrieve blockchain data as raw block data and parse these to extract crypto-currency and token transaction. Sources of blockchain raw data are Cloudflare Ethereum Gateway, Google Bigtables and blockchain nodes.

Big Data management (Related Reference Architecture Layers (Figure 3.59): Data Processing and Analytics). One cannot assume that massive blockchain data will fit in one computer node. Therefore, a distributed inmemory storage on an HPC cluster is essential. Currently, Scalable Transaction Graph Analysis Component which is implemented using C/C++ and MPI message passing libraries constructs a partitioned graph in parallel and provides big data management and processing capability.

Statistics, analysis, AI (Related Reference Architecture Layers (Figure 3.59): User interface).

In order to carry out analytics and report various statistics, two types of approaches are to be utilized (i) Graph Algorithms Approach and (ii) Machine Learning Approach. For Machine learning, K-Means, Support Vector Machines, Naive Bayes, Logistic Regression, Random Forest, Artificial Neural Networks (Multilayer Perceptron) methods will be used by making use of the existing Python Scikit-learn and Pytorch machine learning software. Analytics layer, in Figure 3.21:, shows these functionalities.

Readiness, matureness, level of development (TRL level).

Currently, a proof of concept (PoC) implementation of the pilot is available. As a whole, the current level of development is at TRL3. On the other hand, industrially relevant environment for Pilot9 is defined to be an environment where the real world blockchain data is used. When carrying out our tests in Pilot9, we do use massive industrially relevant blockchain data. The eventual target TRL level is TRL7.

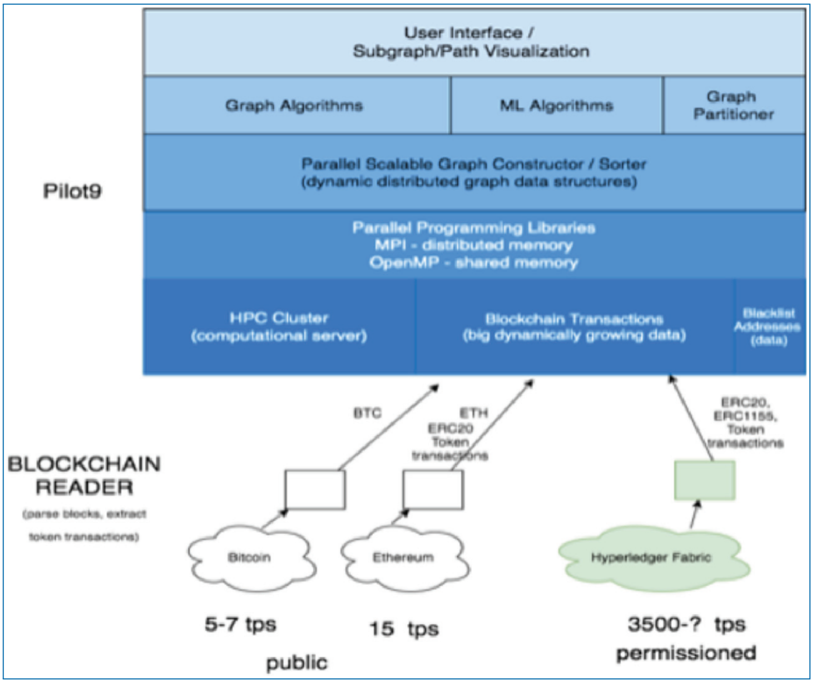


Figure 3.59. Layered architecture of the scalable blockchain transaction graph analysis system.

Data Components

The following components will be developed, deployed and used in the pilot:

- Blockchain Transaction Dataset Preparation Component (Data ingestion in RA).
- Scalable Transaction Graph Analysis Component (Data Management and Analytics in RA).
- User Interface for Blockchain Transaction Reports and Visualization Component (Interface and Analytics in RA). A database of bitcoin and ethereum addresses as well as blacklisted addresses is also managed by this component.

Conclusions – Issues and Barriers

The first year of the Pilot9 has focused on (i) collection and parsing of public massive blockchain data and (ii) design and development of a scalable parallel transaction graph system (iii) development of a simple web interface that would query the graph system and output visualizations of subgraphs returned. We concentrated mainly on Ethereum Mainnet blockchain data, because it was more challenging to deal with due to smart contract support. Code needed to be written to extract transactions from token contract calls.

Whereas implemented parallel algorithms for graph construction, Pagerank computation, tracing and extracting of subgraphs have been tested successfully, our parallel connected algorithms has an issue with it. It is working on small test graph with 1M transactions. But on the whole 633M transaction graph, it is taking too long and not possibly terminating either due to a bug or because the parallel algorithm coded is not efficient due to excessive communication. This will be fixed in the future by coding a more efficient algorithm.

Even though there exists massive public blockchain transaction data and this data can be obtained easily by writing scripts, the same cannot be said for blacklisted addresses. Publicly available blacklisted addresses had to be located through google searches by hand and extracted manually. Collection and tagging of blacklisted addresses information remain as challenging issues because often this type of data may be private and not publicly available.

For machine learning, we need data that can be used for training in our models. In particular, licitness and illicitness information about addresses are needed, but little information is available about this – just the roughly 4K Ethereum blacklisted addresses available from various sites on the web are available to start with. On the other hand, there are roughly 70 million addresses on the Ethereum Mainnet. Hence availability of illicit addresses is limited. Identities of owners of addresses are also not available. This is currently the biggest issue and barrier that we currently have. However, the fact that a parallel cluster graph analysis system has been built means that we can do fast graph queries and traversals on massive data. As a result, we plan to tackle these challenging issues and barriers, by developing graph algorithms that provide information about licitness and illicitness. For example, Pagerank algorithm can be ran to find out important addresses. These addresses are more on the side of licitness since they are addresses of popular services like exchanges that are regulated. Since exchanges verify addresses of customers, then transactions going to addresses from such services are more likely to be licit since KYC/AML checks are carried out by exchanges. Hence, the graph traversal algorithms can be used to report features related to possible licit or illicit addresses in this manner without having information about the addresses in question. These extracted features can then be used in Machine Learning algorithms.

Pilot #10: Real-time Cybersecurity Analytics on Financial Transactions' Data

Pilot #10 aims to significantly improve the detection of cases of suspected fraudulent transactions, to enable the identification of security-related anomalies while they are occurring by the analysis in real-time of the financial transactions of a home and mobile banking system (Figure 3.60). The ability to detect anomalies faster (i.e.

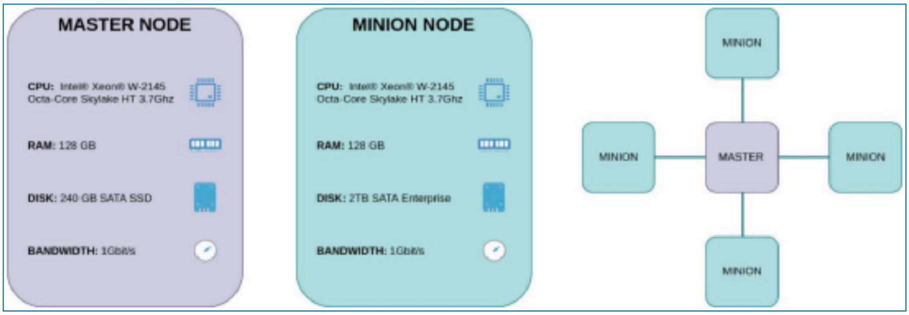


Figure 3.60. Pilot #10 testbed logic.

in real time) and to unveil potential hidden patterns of cyber-attacks are among the main innovations of the pilot.

The use case envisages a pre-processing of transaction data and model training in a batch layer (to periodically retrain the predictive model with new data) while in a stream layer, the real time fraud detection is handled based on new input transaction data.

A fraud detection system is proposed 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 analyzed 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.

Testbed

With regards to the pilot #10 design and execution, the testbed definition (that is the setting of hardware resources, like Storage, Compute and Network…) aims to consider the deployment of an instance of ALIDA asset (Figure 3.61).

The set of resources needed is described in the following picture:

The infrastructure setup consists in a single-master four-worker nodes running a as kubernetes on-premise cluster, each node has a wide enough set of allocable resources to run the testbed safely and without running into disk pressure and memory pressure issues for the expected workload. In any case, the system can be scaled both horizontally and vertically. The machines are equipped with 128 GB of RAM, 2TB of storage and one octa-core 3.7Ghz processor.

ALIDA is cloud native software, this means that can be seamlessly deployed both in an on-premise environments and on the cloud environments provisioned by the

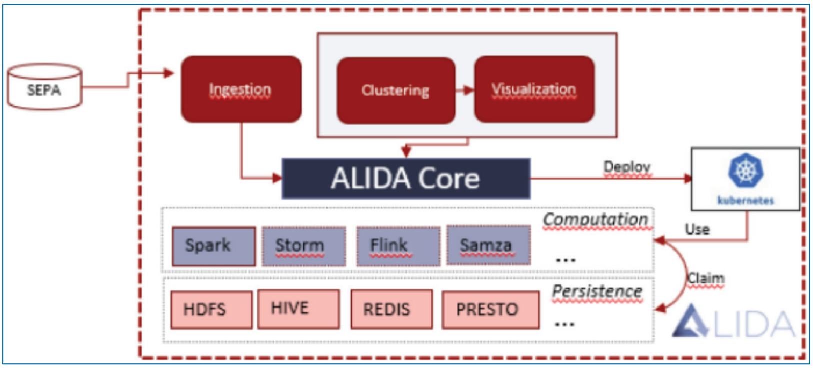


Figure 3.61. Pilot #10 PoC (October 2020).

widely known providers such as Microsoft Azure, Amazon AWS and Google Cloud Platform.

To summarize, the software requirement to get ALIDA up and running are:

- Helm and Tiller 2.14+.
- Kubernetes 1.14+.
- To enable ingresses, a valid ingress provider is required, Traefik is recommended.
- A DNS service provider is recommended to use ingresses with Traefik.
- A persistent volume provisioner support in the underlying infrastructure.

Implementation of a first Proof of Concept

Current implementation status on Pilot#10 is shown in Figure 3.62.

PI create Synthetic and Realistic data set on "Bank Transfer SEPA" transactions that are consistent with the real data present in the data operations environment. These data sets are going to be used by Pilot #10 and, more in concrete, for the first PoC. To develop the services and workflows and ALIDA instance was deployed on ENG premise. As a Preliminary step: a job to transfer synthetic data set on "Bank Transfer SEPA" transactions from an SFTP server to ALIDA HDFS, was designed and it is up and running (Figure 3.63).

With the data ready to be processed, and using ALIDA, a first Batch processing/workflow has been created. This workflow converts qualitative fields into quantitative one, train a KMeans model and makes the clustering process. The Figure 3.11 shows developed ALIDA workflow based on three steps (string-indexer, trains the data with a KMeans models and the clustering creation).

After that, the data is grouped and visualized by clusters (Figure 3.63). Here a domain expert has to label which clusters would be suspicious of fraud. After that the Stream processing would start labelling and detecting new incoming data in real time. But this part is not implemented yet.

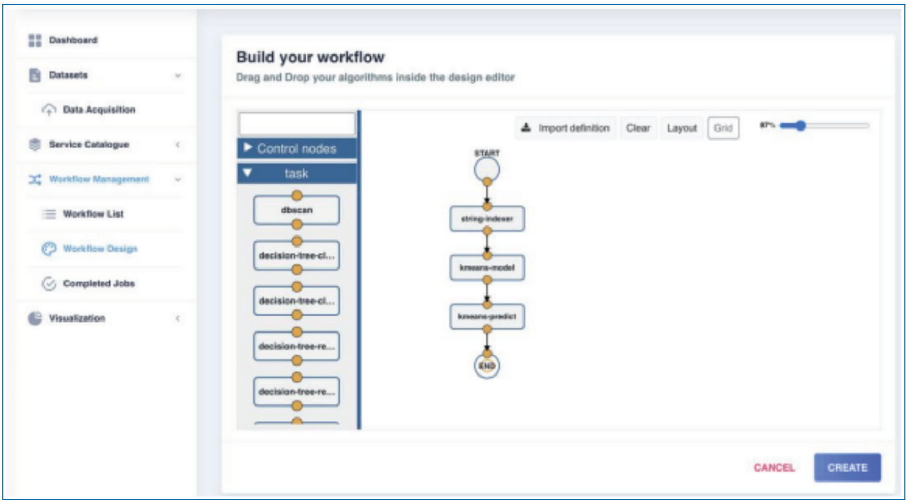


Figure 3.62. Pilot #10 clustering workflow.

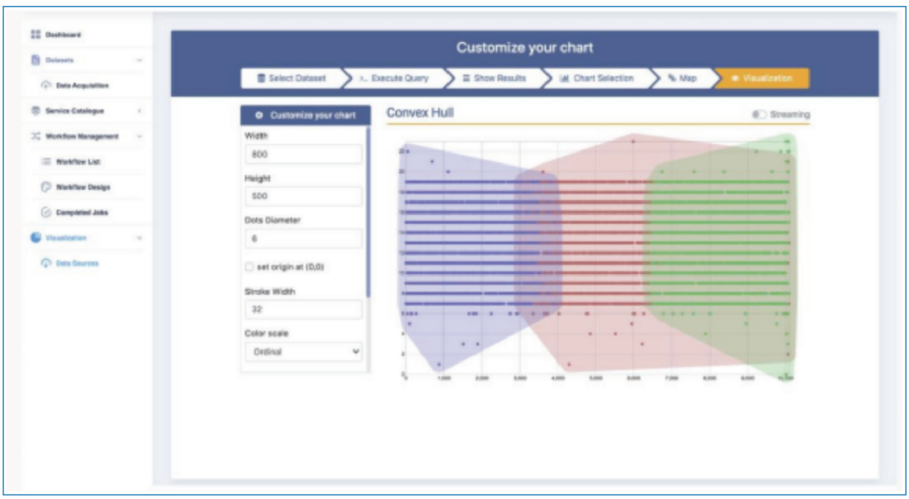


Figure 3.63. Pilot #10 clustering results.

Expected Outcomes

- Validation of specific systems, models and tools for the real-time analysis of big data.
- Collection of quantitative evidences on the performances of the solution and indications on potential further improvements and potentialities.
- Validation of a set of cyber-risk rating metrics.

Figure 3.64 shows the dashboard and the workflow of Pilot #10: Real-time cyber-security analytics on financial transactions' data.

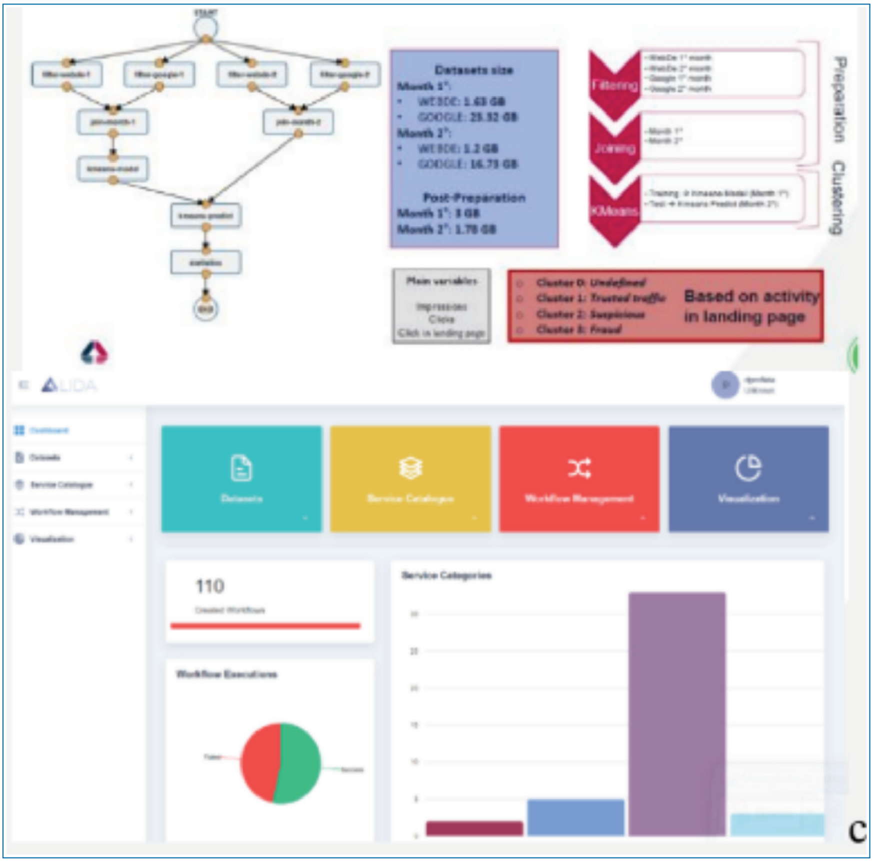


Figure 3.64. Pilot #10: real-time cybersecurity analytics on financial transactions' data.

Datasets

- Synthetic Financial flow Dataset.
- Logs for Correlation and Security Analytics.

The data sets in input of the batch workflow are related to several types of transactions: – Bank Transfer SEPA (The Single Euro Payments Area (SEPA): a payment-integration initiative of the European Union for simplification of bank transfers denominated in euro. SEPA covers predominantly normal bank transfers.

A data generator, implemented by ENG, will simulate real-time transactions (SEPA) which includes informations about the emission date, the beneficiary and the orderer accounts, the amount, the IP address of the orderer's connection and its location (futher informations on Explainable Worflow section). These data will be collected and stored on a dataset to later retrain machine learning models batch-wise; at the same time they are analyzed at real-time for fraud detection with previously trained models.

Data Produced

The data produced consist of a list of suspected fraudolent transactions with an associated probability of actually being frauds exstimated based on the models used.

Explainable Workflow

To meet the abovementioned goals, Pilot#10 envisages two layers (batch and stream layers) implementing the following ML pipelines:

- Unsupervised training (batch) of an outilier detection model (Isolation Forest) on all the collected data.
- Supervised training (batch) of a classifier on the data labeled by the domain expert user.

Real-time detection (stream) of ouliers: which consists of both data preparation services and the application of the Isolation Forest model.

Real-time detection (stream) of fraudolent transactions using the supervisely trained model.

For the first training the goal is to try to identify the outliers using the collected data over time and a method called Isolation Forest, an unsupervised teqnique that identifies anomalies isolating points in a n-dimensional space using binary trees. These points are not necessarily fraudolent transactions, but assuming that the illegitimate ones are a very small percentage of the whole dataset, it is likely that they are as well outliers; therefore it is important to collect outliers and make them available to the domain expert for further analysis. While analising them, he will also label the data, distinguishing between true positive and false positive fraudulent transactions.

The second training consists of the generation of a supervised classifier model. Since the domain expert labels a portion of the data at real-time, and those are collected in batches, we can exploit the work done so far in order to offer a second exstimation of the probability of the transaction being illegal or not. This second exstimation is going to help the system in filtering what transactions the domain expert must analyze and what are the ones he can ignore, reducing his work but at the same time tring not to reduce the reliability of the fraud detection mechanism.

At the same time a real-time analysis is needed. Before the real-time detection the data pass through a process of cleaning and filtering in order to create new features that will be more useful in the predictive model or to enhance other features, improving model performance. We are supposing that it will be needed to analyze datasets made of mixed-type data, where numeric and nominal features coexist. These data must be then elaborated: e.g., instead of dates, time intervals might be more interesting; instead of user names or IP addresses, their location might be more useful during the model's training.

The real-time detection of outliers consists on the application of the unsupervised model described before; at the same time the supervised model is used for inference and a second estimation of the probability that the transaction's data belong to a fraud is produced. The results of both predictions are analized by the domain expert which can take action and block the transaction in time, while at the same time all the data produced are collected, including the labels generated. These data are used on the next cycle of batch tranings, improving models performance over time.

For the Pilot #10 aims, ALIDA (https://home.alidalab.it/) is adopted and extended to design Big Data Analytics (BDA) services batch and stream workflows. In a nutshell, ALIDA is a micro-service based platform, developed by ENG (Engineering), for composition, deployment, execution and monitoring of workflows of BDA services; it is entirely developed with open source technologies.

ALIDA offers a catalogue of BDA services (for ingestion, preparation, analysis, visualization), implemented as Spring Boot Applications and deployable as docker images. User designs his own (stream/batch) workflow by choosing the BDA services from it, indicates which Big Data set he wants to process, launches and monitors the execution of the workflow and personalizes the results visualization by choosing from a set of available graphs. All this without worrying about having software developer skills or particular knowledge on big data technologies.

Some BDA services for preparation and machine learning, as KMeans and Random Forest modelling and prediction, are already available within the ALIDA Catalogue. Even though they need to be reviewed (and in some cases redesigned) to meet specific pilot requirements.

Concerning the remaining BDA services (especially pseudo anonymization one) the pilot will make use of the services made available within the project.

Preliminary step:

To load data sets related to several types of transactions (SEPA bank transfer, foreign bank transfers, internal transfers of funds, PCTU, SMWCA, STFTS) into the HDFS storage of the ALIDA instance, by means an ingestion job.

Batch processing, building and labelling clusters (training):

Stored data sets are properly filtered (to remove some columns and rows unneeded for the ML) and joined to get only one unlabeled data set to be used for the unsupervised machine learning.

In this phase the goal is to cluster such data, to create labeled samples to feed the supervised machine learning classifier of the next phase. Clustering process groups the data according to automatically detected similarities. These clusters/groups still need a domain expert, PI (PosteItaliane), who determine which clusters present a fraudulent behaviour and properly assign labels to such clusters.

Stream processing:

After learning the mapping, the Random Forest (RF) classifier can map new real-time unlabeled transaction data to their corresponding high-level information (i.e. label) on the basis of the model trained in the batch layer. In that way, financial fraud events can be detected while happening.

Logical Schema

An initial mapping of the pilot's components and modules to the INFINITECH-RA pipelines approach is illustrated in the following figure.

Figure 3.65 shows a logical view of the components identified for the Pilot#10 according to the mapping with the INFINITECH Reference Architecture. The list of main components to be deployed and used in the Pilot#10 includes:

- Identity Management System: It is a cross cutting system that guarantees user authentication, authorization and management. It allows or denies the access to the federated services that run within the architecture.
- Role Management: It implements and handles the roles and the privileges that can be associated to the users. It is often tightly coupled with the Identity Management System.
- Message Broker: works as an intermediary software that allows system components to communicate each other effectively, implementing a common communication protocol over message buses.
- Resource Manager: It is a lower-level software that consents to handle and use the infrastructure resources seamlessly and dynamically, according to the number of requests received per time interval.
- Pseudoanonymizer: tool to pseudonymize personal or sensitive data at source, in order to preserve privacy according to GDPR regulation.
- Filter: Filtering component to remove specific rows and columns.
- Join: Service to join two or more datasets where at least one column must be the same.
- OneHotEncoder: Service to transform categorical variables into numerical ones.
- Clustering (Kmeans): Given a set of observations (x1, x2, …, xn), where each observation is a ddimensional real vector, k-means clustering aims to partition the n observations into k ($\leq$n) sets S = {S1, S2, …, Sk} so as to minimize the within-cluster variance.

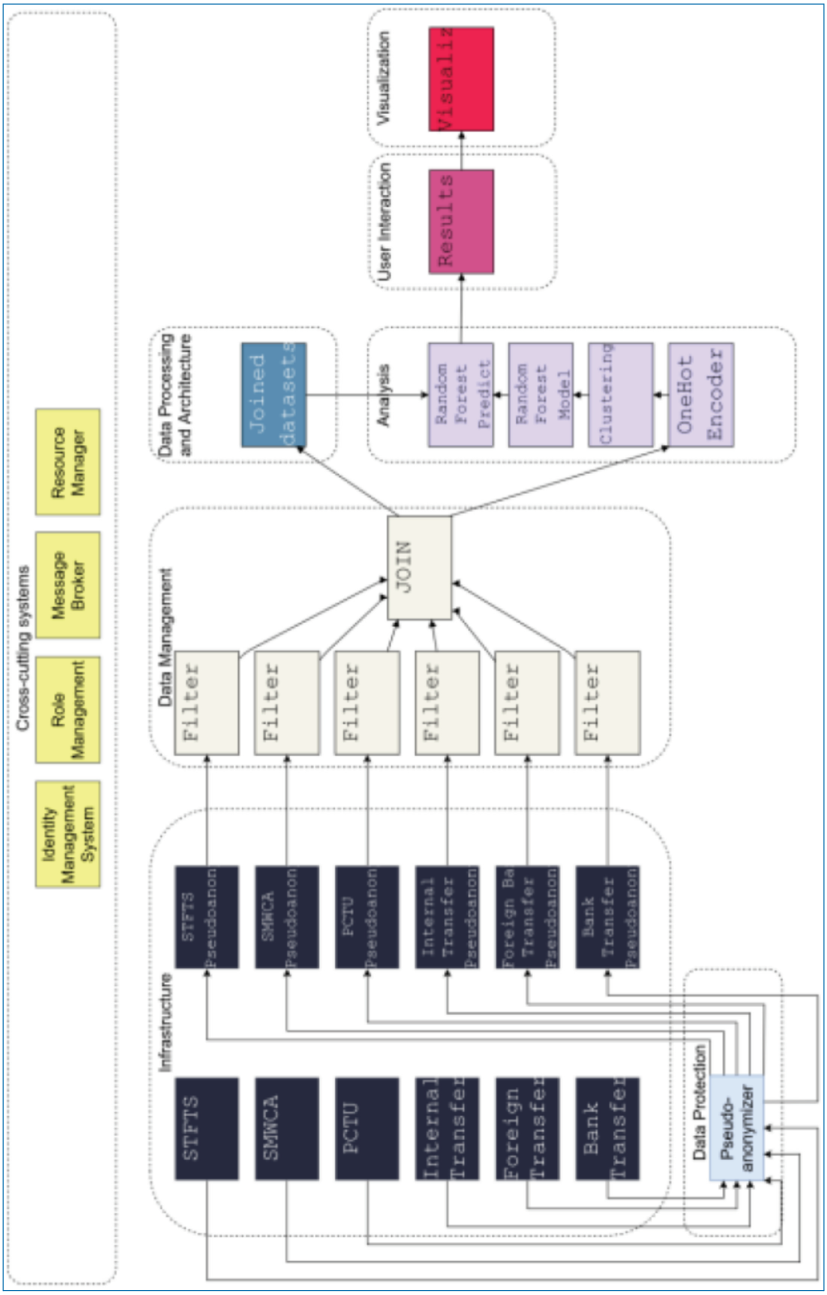


Figure 3.65 Pilot #10 reference architecture.

- Random Forest: An ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.
- Results: Storage that contains all the processed data elaborated by the workflow.
- Visualization: The service that gathers the resulting datasets to be delivered to the visualization clients.

Components

The list of main components to be deployed and used in the pilot includes:

- Filter: Filtering component to remove specific rows and columns.
- Join: Service to join two or more datasets where at least one column must be the same.
- Prelaboration: Service to transform categorical variables into numerical ones throgh different calculation.
- Outliers detection (Isolation forest): Given a set of observations (x1, x2, …, xn), where each observation is a d-dimensional real vector, isolation forest associate to each Fraudolent transaction detection: exploiting a supervised classifier algorithm (e.g., random forest classifier, neural network classifier), classify incoming data in two categores: suspected frauds or clean transactions. observation a value that expresses how much it differs from the distribution calculated on each dimension.
- Fraudolent transaction detection: exploiting a supervised classifier algorithm (e.g., random forest classifier, neural network classifier), classify incoming data in two categores: suspected frauds or clean transactions.
- Results: Storage that contains all the processed data elaborated by the workflow published to be visualized.
- Visualization: The service that gets the resulting datasets to be delivered to the visualization.

Conclusions – Issues and Barriers

Current setup clearly demonstrates some of the most relevant capabilities of the pilot:

- availability of a significant dataset for analysis
- data collection from source and preparation
- data ingestion

- AI model training
- prediction
- data visualization

Some pre-processing services are supposed to be needed in both batch and stream stages before the ML algorithms are invoked. They will be implemented once the data schema on transactions will be defined.

In order to fulfil GDPR requirements both at pilot stage and in potential production stage with real transaction data, a fully synthetic dataset will be also pseudonymized at source. Currently, synthetic data are pseudonymized at generation time, therefore data analysis will work on pseudonymized data, but we expect a pseudonymization tool will be made available in the framework of the INFINITECH project for potential production use with real data.

Expected Business Impact of Technologies adopted in this Pilot

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

This viewpoint, built on information sharing activities currently running in the banking sector, is also reinforced, and strengthened by trusted industry reports. With some surveys and reports pointing to issues, such as: "recover less than 25 percent of fraud losses", "Increase fraud typologies globally, from recent years, include identity theft and account takeover, cyber-attack, card not present fraud and authorized push paymentsscams", "6 is the average number of frauds reported per company studied", "56% asked companies investigated their worst fraud incident. many organisations are failing to respond effectively". These, and other issues in these reports, demonstrate the importance of developing new technologies and approaches, such as real time analytics, to enhance the need of fighting against cyber frauds.

Pilot #16: Data Analytics Platform to Detect Payments Anomalies Linked to Money Laundering Events

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

The pilot purpose is to preside anomalous scenarios linked to money laundering, adhering to European AML regulatory compliance policies, by notifying detected cases to the Italian Financial Intelligence Unit (FIU). The innovation potential of current pilot lies in introducing novel technologies like, machine learning, artificial intelligence, graph database to detect anomalous scenarios, which allows to automatically detect complex anomalous money-laundering scenarios.

The adoption of pilot platform will improve quality and efficiency of AML users work and, at the same time, will concur in reducing risk of unmatched scenarios related to money laundering events.

Data Sources

The following data sources will be integrated and used in the pilot:

- Cardholders transaction operations.
- Cardholders information registry.
- Merchants transaction operations.
- Merchants information registry.
- AML Anomalies Features Store.
- AML Suspicious Activities Report (SAR) practices collection.
- Master and reference data.

All above-mentioned data sources are in an anonymized format and are stored and collected into a Data Lake environment to enable agile development and processing.

The pilot will use a graph database, to model many-to-many relations that belongs to anomalous events linked to money laundering; thanks to this technology we can find out any relationship occurring between a suspicious payment events and individuals or merchants.

Data Produced

The three data outputs produced during the pilot are:

- Anomalous subjects.
- Cluster of anomalous subjects.
- Anomaly risk score for each subjects.

Decision rules developed into the graph database, based on many-to-many associations, will produce periodically (monthly or quarterly accordingly) anomalous subjects (1) or groups, clusters, of anomalous subjects (2) intercepted.

During the pilot we will develop an algorithm that update anomaly risk scores (created with machine learning supervised classification algorithm) associated to any subject: those updated score is the last data product generated (3).

Subjects can be, cardholders, legal entities, organizations, legal representatives. Any sensitive information, such as person ID, is anonymized so that it's not traceable to physical or legal person.

The output format will be a csv text or a parquet file stored into the Data lake. Those formats allow not to be bounded to any particular database technology. In a post-processing phase, we can then accordingly choose how to use and model those data: a relational database to perform SQL queries and to be the back-end of a data visulization tool, or just a plain .csv file to perform exploratory data analysis with data science frameworks.

Explainable Workflow

The data workflow considers all steps and data shapes needed to develop a ML solution to find anomalous scenarios : the collection of data (listed in the previous paragraph), the processing step to create a training, validation and test set, the training of risk score with ML algorithms and the presentation layers to communicate results.

As a first step, we collect in anonymized format historical data about Nexi clients behaviour, such as transaction payments, withdrawals, merchants information, reversals, money transfer, past SAR reported to Italian Financial Unit (IFU), into a Data Lake.

Afterwards, we apply the Transformation step of a typical Extract, Load, Transform(ELT) workflow to create the Feature Store ; that is, a dataset containing Machine Learning features for each cardholder (or organization) together with the target variable, the outcome of the ML algorithm, that in this case is binary variable, representing whether a cardholder has been notified to IFU. It is updated monthly in batch mode.

Once the Feature Store is ready, we follow these steps:

- Training machine learning model and, based to the predictions generated, we get the anomaly risk score for each cardholders (or organization).
- Create a graph database, inserting cardholders, organizations, legal representatives behaviours data (from both Data Lake and feature store) and ML based risk score.
- Perform a Personalized Page Rank algorithm to adjust risk scores, taking into account the many-to-many relationships modelled with graph data structures.
- Define rule based anomaly events detection to find customers of groups of customers (clusters) to whom AML users would pay attention.

All the steps mentioned above are then stored into the Data Lake in a file format (.csv) or compressed file like parquet, and then modelled into relational database tables or views to make those accessible to analytics users.

Finally, a visualization dashboard allows to end users (Customers Due Diligence team members) to explore and visualize outputs of the data workflow,

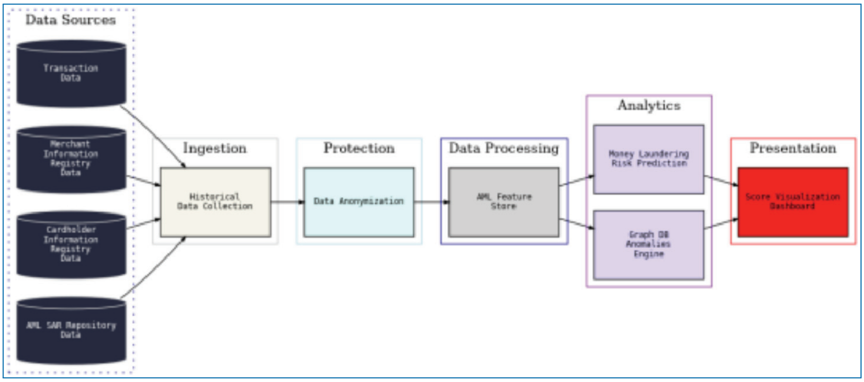


Figure 3.66. Pilot #16 pipeline in-line with the IRA.

and so can discover and analyse riskier subjects as suggested by algorithms developed (Figure 3.66).

Logical Schema

The explainable workflow of the pilot to the INFINITECH-RA layers is in the following figure.

Components

The following main components will be deployed and used in the pilot pipelines:

- BigData Management Layer to collect and process data (Data Management in RA).
- Money Laudering Risk Prediction supervised classification model and Graph database engine to adjust risk scores (Analytics and Machine Learning in RA).
- Visulaization dashboard of customers wirh higher risk of money laundering event (Visualization in RA).

3.5 Smart, Reliable and Accurate Risk and Scoring Assessment

Pilot #1: Invoices Processing Platform for a More Sustainable Banking Industry

The main objective of the pilot is to develop, integrate and deploy a data-intensive system to extract information from notary invoices, in order to: (i) Establish the sustainability index of each notary based on the number of physical copies that are issued. (ii) Provide to financial institutions the information (properly indexed) about the documents that are finally generated by notarial services required by the bank. (iii) Promote notarial services from those with the higher sustainability score.

The innovation of the pilot lies in the applicability of Artificial Intelligence technologies over scanned physical documents (notary invoices) for cost savings and increased effectiveness. Currently, many physical documents, and copies (some of the redundant), have to be managed. Each physical copy and its control cause significant costs over the period of the financial products lifetime. 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.

Following a list of partners participating in the pilot and their different roles/contributions:

- Bankia's Auditing department provides the business use case, the functional requirements, and the expert knowledge about entities to extract, business rules, alert generation and information dashboard. It also provides the cloud environment for the deployment of the storage and computation platform. It provides the expert knowledge for invoice tagging and validation of the final product.
- GFT provides the architectural design, platform implementation, algorithm design, training, validation and implementation, together with the document pre-processing. GFT carries out the development of the different components.
- Final users are Bankia's internal auditing department, where a real task will be automatized, thus obtaining a real return of investment and key performance indicators.
- In the pilot will also collaborate: FBK (Fondazione Bruno Kessler) as the data science expert, RB (ReportBrain) as solution expert with expertise in text-analytics and sentimental analysis and INSO (Insomnia Digital Innovation Hub) as business and development advisor.

Due to the modular nature and close interaction between technology and business actors, the solution has been showcased to different European and US potential customers, that have showed a keen interest. Feedback is that, with adaptations with regards to their business workflows and retraining for their documents, both of which is completely feasible, this solution will constitute a compelling technology and business case.

The technology developments have an impact on the implicit training of different actors:

- Business users that have been training in the statistical nature of the results coming from such tools, and therefore the interaction and cooperation human-AI tools.

- Business managers to assess the impact and specificity of the use of such technologies.
- Data scientists and software architects and developers that have been trained during the development of the present project.

Technological components and Services

The main technological components that will be implemented and integrated as part of this pilot are:

- Invoices and invoicing workflow database.
- Document ingestion.
- Document pre-processing: document pagination, PDF to image conversion, image normalization, OCR.
- Document entities and region-of-interest extraction: machine learning models and Natural Language Processing extractors for the identification and extraction of entities of interest: billable concepts, prices, headers, addresses, etc.
- Entity association: graph deep neural networks for the identification of related concepts: e.g. that a certain billable concept corresponds with an identified price and identified.
- Business rules engine: application of compliance business rules for the generation of alerts and reports.
- Data Tagger: for the tagging of training invoices examples.
- Document validator: for the verification of processed invoices.
- Training and inference orchestrated pipelines.
- MLOps tools: Models and data Repository, code repository.

Testbed

A cloud-based testbed will be implemented using AWS Bankia Private Cloud, with and estimated volume of data of 2TB. The test bed is already available, and the hardware to be used will depend on the task to be accomplished:

- For training and tagging AWS EC2 instance of the type g4dn.xlarge with 200 GB of disk with GPU.
- For inference, normal computing optimized instances c6g.2xlarge or the same type g4dn.xlarge, with the AWS Deep Learning AMI (Ubuntu 18.04).

Some more details about different tools and software components to be deployed follow:

- Data management: linux file system, S3, elastic search.
- Data processing: kafka, Kubeflow.

- Data analytics and AI related tools: tensorflow 1.5, sklearn, pandas, numpy, seaborn.
- Data tagging: labelme.
- Data visualization: kibana, Floent, Prometheus.

Implementation of a first Proof of Concept

This first Proof of Concept is addressed to the implementation of a two-sided machine-learning based system at scale.

From one side, to automatize the capture and extraction of the unstructured information in scanned documents using computer vision and machine learning deep neural networks. This implies to develop, integrate and deploy a data-intensive system to extract information from notary invoices to establish a sustainability index of notary services based on the number of physical copies issued, that will be used by the bank.

From the other side, to capture the business rules expressed as concept associations (e.g. invoiceable concept + related quantity + related price) that in an unstructured way are scattered along the document with high variability. Finally, automatizing the whole process at scale coupling with automated workflows for document capture and reporting in a real financial institution environment.

Expected Outcomes

In terms of technical results, the following components will be developed, integrated, deployed and operated:

- A batch processing architecture to process notary invoices.
- A computer vision system to identify and extract tables.
- An AI/Machine Learning system to extract information from tables.
- A visual console to show results and the sustainability score.

Datasets

- Real Invoices.
- Physical copies.

Data from 32.300 real invoices documents and from 3.000 different notaries extracted from Bankia systems are the source of the Pilot. Invoice documents to be digitalized in PDF format or may also arrive already digitalized from other channels (email attachments, bulk sftp, etc.). Data type will be: PDF/Image/Text. Data format will be: PDF/ PNG/ TXT. Estimated data volume will be: 2 TB. The dataset TableBank, which consists of 500.000 documents, will be used as Table Benchmark for Image based Table Detection and Recognition.

Data Produced

Digitization of contracting and invoicing processes will allow an automated analysis of the digitized documents enabling a smart and autonomous scoring of notary services. Rating notaries based on a "Sustainability Index Score" 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.

Explainable Workflow

Invoice documents will be securely storage in a data lake. The system will parallelize different jobs to pre-process, process and post-process the documents and the outcomes. For instance: Image preprocessing (cropping, adjusting brightness, contrast, etc.); converting PDF to Text; OCR; text correction. A computer Vision system will identify and extract tables from invoices that will allow extracting sensible information to establish a sustainability scoring. And using machine learning we will extract information from the identified and extracted tables. The extracted information will be displayed so it can be validated and re-introduced to the system. The AI models will be trained (offline process) with a combination of public huge datasets and specific invoices samples. Trained models will be published to the runtime processing time after an expert evaluation.

Figure 3.67 shows interactions and workflow, from high level point of view, between the main components. Invoices are automatically ingested by the system to start the processing, yielding the OCRed document, together with the extracted fields, the association between the corresponding fields, and the application of the rules. Later, results will be stored in the ElasticSearch database and finally, the summary accessible by a dashboard (Figure 3.68).

Logical Schema

The following figure illustrates the logical architecture of the pilot in-line with INFINITECH-RA constructs and approach.

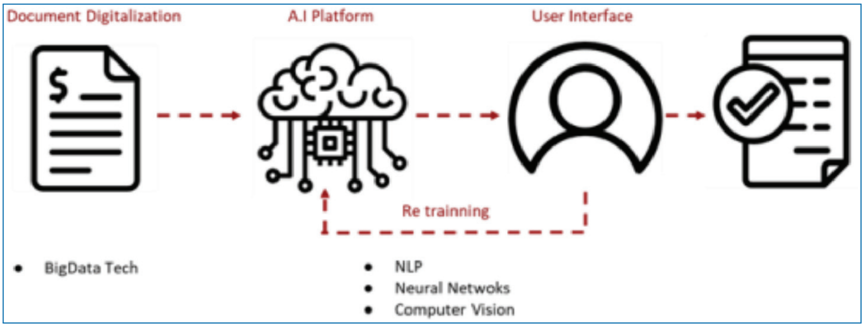


Figure 3.67. Pilot #1 main components interactions.

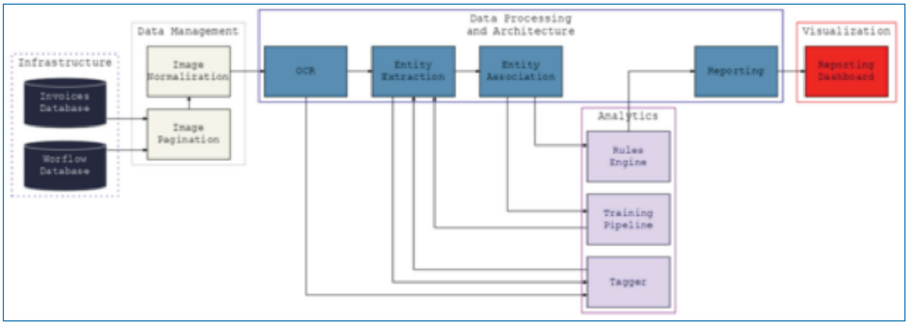


Figure 3.68. Invoices processing pilot pipeline in-line with the IRA.

Components

The main technological components that will be implemented and integrated as part of this pilot are:

- Invoices and invoicing workflow database.
- Document ingestion.
- Document pre-processing: document pagination, PDF to image conversion, image normalization, OCR.
- Document entities and region-of-interest extraction: machine learning models and Natural Language Processing extractors for the identification and extraction of entities of interest: billable concepts, prices, headers, addresses, etc.
- Entity association: graph deep neural networks for the identification of related concepts: e.g. that a certain billable concept corresponds with an identified price and identified.
- Business rules engine: application of compliance business rules for the generation of alerts and reports.
- Data Tagger: for the tagging of training invoices examples.
- Document validator: for the verification of processed invoices.
- Training and inference orchestrated pipelines.
- MLOps tools: Models and data Repository, code repository.
- Reporting business dashboards and operational databases.

Conclusions – Issues and Barriers

End-users from the auditing department have been involved in the development following an agile methodology. Their roles has been crucial in the:

(1) Validation of the information to be extracted and the definition of the ground truth for the document samples.

(2) Elicitation of the operation workflow and reporting dashboards.
(3) Definition of the business rules for the concepts' association.

Mismatch of end-user expectations and requirements with the actual project implementation has been addressed by the active involvement of the users in the bi-weekly review meetings.

At the present time, the critical path consists in the coordination and adjustment of the different elements of the pipeline, a characteristic typical for projects with intensive use of machine learning algorithms that result in the combination of many moving parts.

The main barriers like the availability of data, tagged data and expert knowledge for problem definition are mainly removed at the present time.

Pilot #2: Real Time Risk Assessment in Investment Banking

The pilot will implement a real time risk assessment and monitoring procedure for two standard risk metrics – VaR (Value-at-Risk) and ES (Expected Shortfall). Both can be applied for measuring various types of risk, above all, market risk of portfolios of assets. The pilot will implement both 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. In addition, it will implement the evaluation of what-if-scenarios allowing pre-trade analysis, i.e. estimating changes in risk measures before a new trading position is entered. Moreover, the pilot will implement a sentiment-based decision support indicator derived from financial and economic news data and social media channels. 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.

The aim of this use case is to give traders in investment banking a precise and timely indication of the risk of a given portfolio and specifically changes in risk due to market changes or changes of the portfolio. The need of such knowledge comes from operational as well as supervisory requirements that every regulated financial institute must comply with.

Risk assessment is based on a common risk metric – Value at Risk (VaR) – to be calculated and updated in real-time on both, portfolio level as well as for each individual asset. A second risk measure – Expected Shortfall (ES) – indicating not the maximum amount of a potential loss, but the expected loss with a given probability, will be derived at a later stage. The pilot will furthermore implement the evaluation of what-if-scenarios allowing pre-trade analysis, i.e. estimating changes in risk measures before a new trading position is entered. In addition, the pilot will implement a sentiment-based decision support indicator derived from financial and economic news data and social media channels.

The pilot will support institutional traders, asset managers, risk managers and wealth management experts in:

- Calculating the Value-at-Risk (VaR) of their Portfolios. Emphasis will be paid on FOREX (FX) portfolios3 , yet the system will be applicable for other types of portfolios as well.
- Evaluating what-if scenarios for alternative Portfolios based on their VaR. In practice the system will simulate alternative investment strategies and will provide relevant information to the end-users to allow them to shape their investment decisions.

The main innovations of the pilot lie in:

- The calculation of VaR at very short timescales based on the processing of high-ingestion data.
- The employment of ML-based VaR calculation techniques that will yield more accurate values and will facilitate traders in better understanding and framing the risks of their portfolios.

Technological components and Services

The components to be implemented are depicted in Figure 3.69: Pilot #2 Data Science Pipeline, which illustrates the data science pipeline for the pilot. They include:

Data ingestion component. Ensures the ingestion of real-time data in the database of the pilot (XLS database). It is destined to cope with the high ingestion rates of the real time data.

- Market Sentiment component. Extract market sentiment for specific assets of the portfolio and provides this information to the data to reinforce the accuracy of the VaR calculation and/or to provide alternative methods for VaR calculation. The component is not implemented in the early Proof of Concept that is described in this deliverable.
- VaR calculation component(s). Scientific Computing and Machine Learning components, which calculate the VaR of the portfolio based on different methods (e.g., historic method, variance-covariance, monte carlo simulation). They harness data from the LXS dataset.
- End-User Dashboard component. Provides user friendly visualization of the VaR parameters for different portfolios owned by the user.
- Semantic Interoperability component. Provides an interface for access to FIBO data, while supporting their parsing. The semantic annotation and the structuring of the data according to FIBO that is performed and hence the relevant description is beyond the scope of this deliverable.

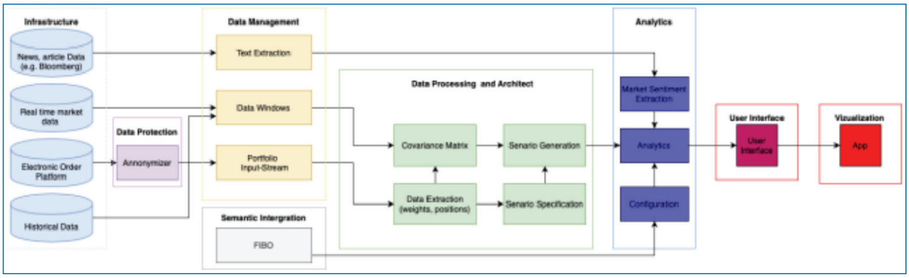


Figure 3.69. Pilot #2 data science pipeline.

The pilot pipeline conforms to the INFINITECH-RA specification i.e. the various blocks are structured according to the modules and layers of the INFINITECH-RA. Likewise, the deployment of the components adheres to the guidelines of the INFINITECH reference testbed.

The implemented pilot's deployment diagram for the first proof of concept can be seen in Figure 3.70. More specifically:

- The main elements of this deployment are:
  - LXS Database (Docker container) containing Historical Ticker data.
  - predict_var (Dockerized python scripts) for time series pre-processing and VaR prediction
  - visualize_var (Dockerized Flask Web application) to visualize FOREX assets' historical statistics, VaR predictions and perform What if Analysis.
- New ticker data (test set) are injected from a csv file to the predict_var docker using kafka in between. In the next version LXS DB will be used instead of a csv file.
- predict_var docker reads once historical data from LXS DB to be used as a training set for VaR calculation. As new ticker data is created (from the test set) the training set is updated in predict_var docker.
- The predicted results are written back to the LXS DB.
- visualize_var read predictions from LXS DB to update dashboards dynamically.

Other non-technical requirements

At later stages of the pilot experiments with more data will be carried out, based on access to data from other trading platforms (e.g., Forex platforms that provide APIs for different assets). Likewise, the estimations for the open source datasets to be used are subject to revision.

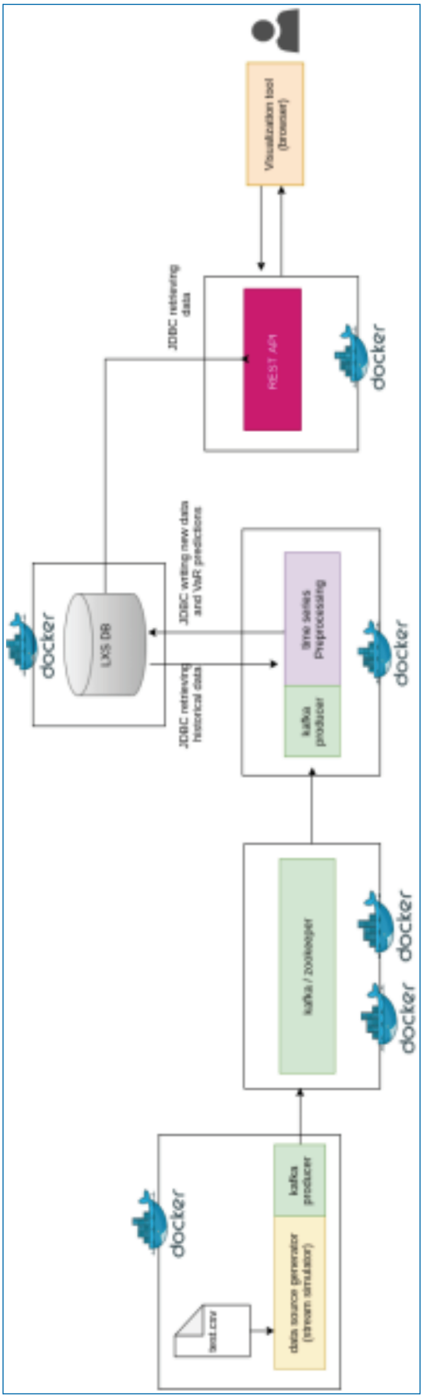


Figure 3.70 Pilot #2 deployment diagram.

Implementation of a first Proof of Concept

The Proof-of-Concept implementation comprises the following components in-line with the pilot data science pipeline (Figure 3.71):

- Data Ingestion Component implemented within LXS database.
- Scientific computing components in Python that calculate the Value-at-Risk (VaR) using three different methods, namely:
  - The Historical Method: This is probably the simplest VaR calculation method. It relies on significant volumes of historical market data (e.g., typically one trading year data for conventional assets and much more than

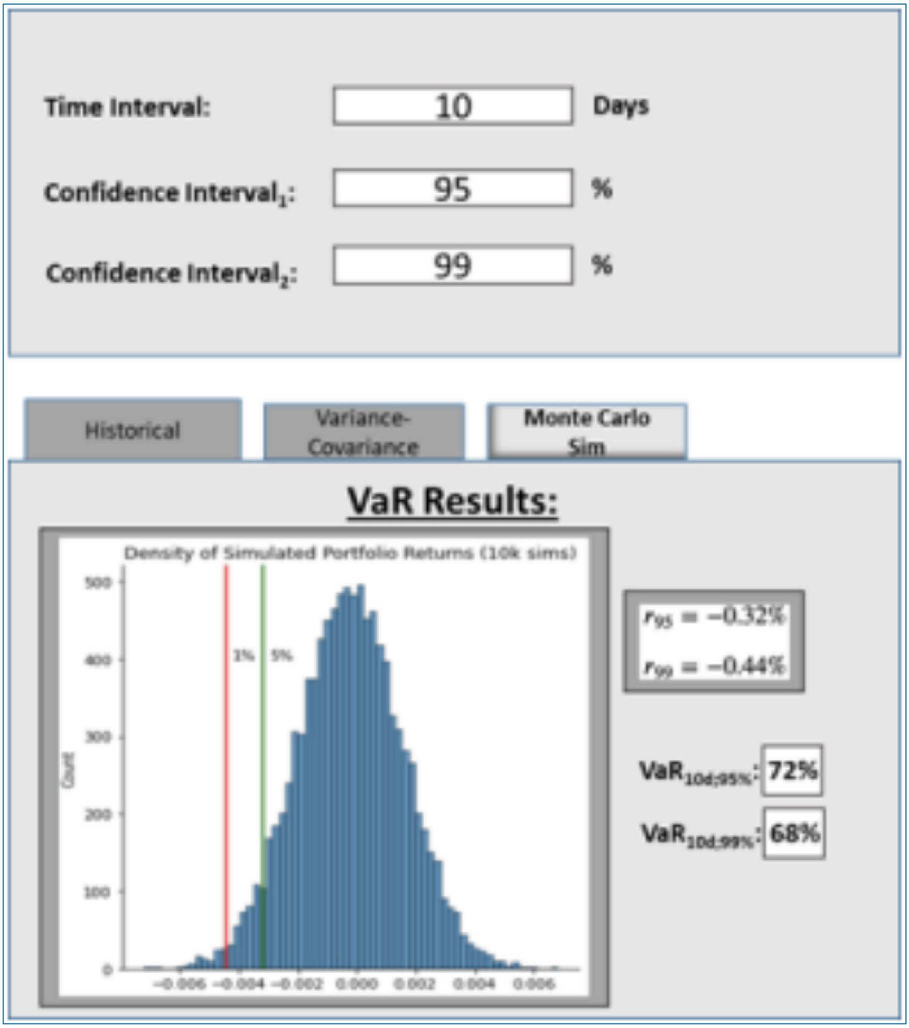


Figure 3.71. Pilot #2 Dashboard for parameter configuration and visualization.

that for hedge funds) to calculate the price changes for all the assets of the portfolio. Accordingly, it calculates the value of the portfolio for each one of the price changes i.e. the value of the portfolio is simulated many times in-line with the number of price changes in the historic data (e.g., approx. 250–260 times for one trading year). These simulated/estimated values for the portfolio can be sorted and used to form a distribution. Then the VaR at a given confidence level (e.g., 99%) is computed as the mean of the simulated values minus the lowest values (e.g., 1% lowest value for the 99% case) in the series of simulated portfolio values.

- The Variance-Covariance Method: This is also called parametric method. It assumes that returns follow a normal distribution, which is a simplistic yet acceptable assumption during normal market conditions. Given this assumption two parameters can be computed i.e. an expected return and a standard deviation for the portfolio. In case of a portfolio with many assets, the standard deviation should consider the correlation in the price changes of the different assets. The latter requires the computation of the covariance matrix of the various assets (i.e. the correlation coefficient of the assets). Based on the mean and the variance of the portfolio its value distribution is calculated and the value at the 95% or 99% confidence interval is produced. The method works quite well when there is a large sample size for the assets of the portfolio, as well as when the distributions of the asset prices are known.
- The Monte Carlo Method: This method develops randomly scenarios for the future price of the portfolio based on some non-linear pricing models. Accordingly, it creates the distribution of these future prices and takes their losses at the target confidence interval. The method is more reliable when dealing with complex portfolios and complicated risk factors. Its advantage compared to the first two methods is that it is not restricted to scenarios seen in the past, but may also consider scenarios more extreme than those contained in the historical data due to its random component and thus is expected to be more realistic.

- The visualization dashboard, which displays VaR Charts for each one of the three methods and two confidence intervals (95%, 99%)5 . A snapshot of the charts of the dashboard is depicted in Figure 3.71.

A comparative visualization view of all three methods with different parametrization and their development over time on a daily basis as depicted in Figure 3.72.

The Proof of concept leverages reduced versions of the "Trade Data" and "Tick Data" datasets i.e.:

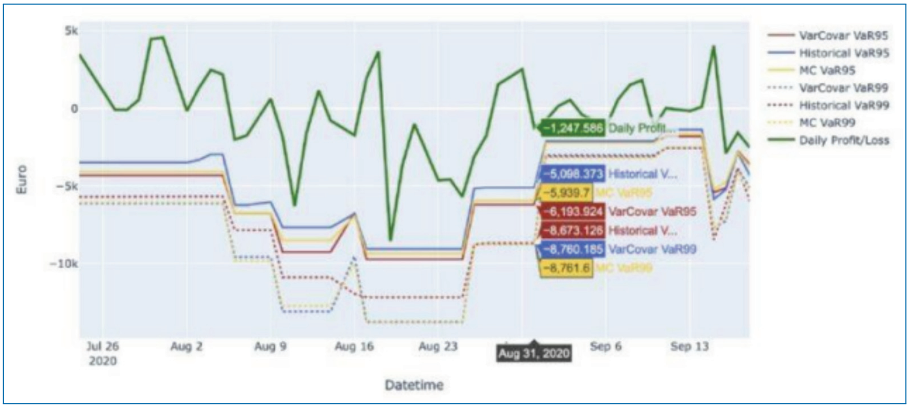


Figure 3.72. Pilot #2 charts for three VaR calculation method.

- The Trades comprise 3 popular FX assets. The scale-up of the Proof-of-Concept will support more complex portfolios.
- The Tick Data comprises historic data about the corresponding Forex assets in the period March 2020 – October 2020. It is considered a sufficient dataset for the Proof-of-Concept and the validation of the various methods. However, the scale-up of the pilot will use more data and will experiment with different historic windows.

Expected Outcomes

The pilot will implement a real time risk assessment and monitoring procedure for three standard risk metrics (Figure 3.73):

- VaR (Value-at-Risk).
- ES (Expected Shortfall).
- Pre-trade analysis.

Datasets

Data will be extracted from several data sources: real-time market data, historical market data, synthetic electronic order platform (trades data), and financial news/article data. The pilot will leverage FOREX (FX) data provided by the JRC Platform and other trading platforms via Forex APIs. The data will include:

- Trade Data (i.e. data with the assets' positions) of the user that will be used to define the portfolio(s) of the user and their VaR/ES;
- Tick Data (i.e. Historical market data) that will be used in the different methods for VaR calculations including standard methods such as Monte Carlo simulations, VarianceCovariance, Historical Simulation and a novel one based on deep neural networks, the socalled DeepVaR.

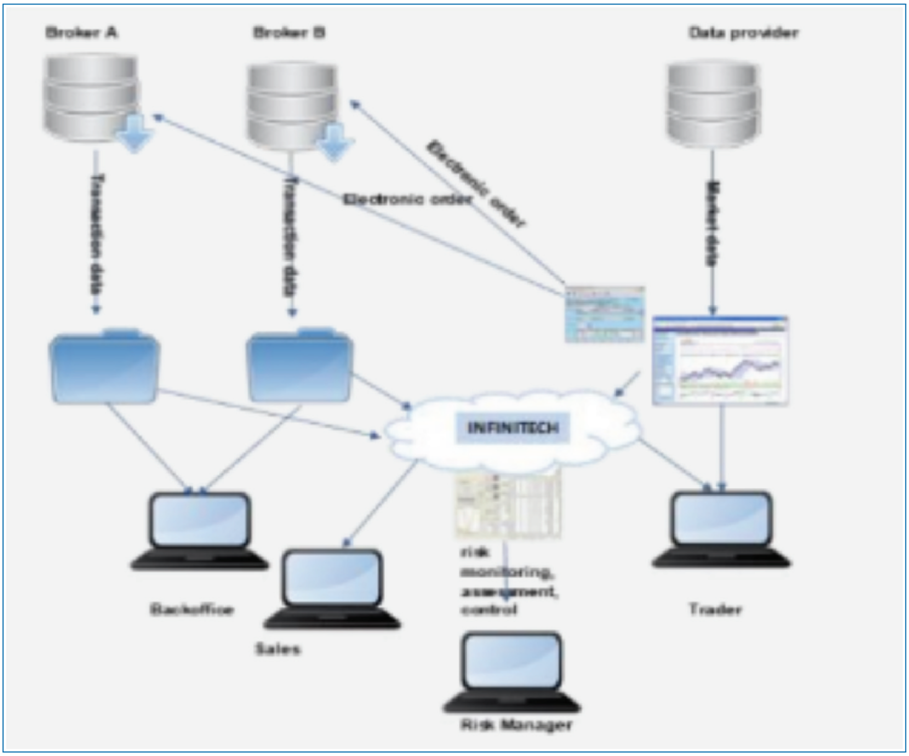


Figure 3.73. Pilot #2: real-time risk assessment in investment banking workflow.

- Alternative data (e.g., data from news feed) that will be used for market sentiment analysis based on NLP (Natural Language Processing Techniques). Such data will be obtained from Open API's (e.g., Google News API, Twitter API and Interactive Brokers API).

Trade Data and Tick Data will contain information such as: the name of the instrument in FOREX trading (ex. GBPUSD for the exchange of GBP to USD), Timestamp that denotes when the trading took place, the Quantity and the Closing Price.

The main data computed and produced include the VaR (Value-at-Risk) and ES (Expected Shortfall) estimations. In addition, the injected real time data are both processed and saved as historical market data as it is (ticker data) and processed (i.e., aggregated market data in frequencies of 1 min, 5 min, 1 hour, 1 day). Moreover, the pilot's sentiment-based decision support indicator derived from financial and economic news data and social media channels will produce a sentiment score (positive, neutral, negative) for each article/description coming from the news feed.

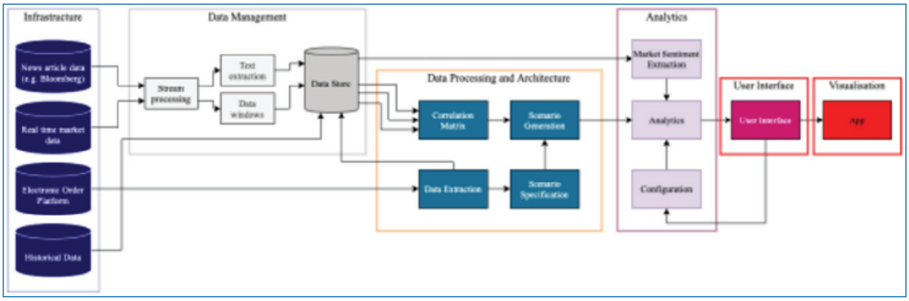


Figure 3.74. Real-time risk assessment pilot pipeline in-line with the IRA.

Explainable Workflow

Data from the real-time market database and the news feed databases is injected into the Data Management layer through a stream processing component which is capable of handling large volumes of data that feature very high ingestion. The real-time data is initially concatenated with the historical data and then is appropriately transformed using a data windows component (i.e., the Online Aggregates Component), creating segments of time series. Data from the electronic order platform are managed using a data extractor. These data will also serve as input for both the correlation matrix and the scenario specifications components. The processed market data (historical and real-time) will then feed the correlation matrix component together with the processed data from the electronic order platform database. The correlation matrix processes and calculates the ingested data, merging the different data sources. The output will then serve as input, together with the scenario specifications component, for the scenario generation, the basis for the Monte Carlo simulation. The processed data will then go into the Analytics component where VaR/ES estimation takes place.

On the other side, data from the news article database are processed using the text extraction component and then market sentiment extraction one. Therefore, sentiment and behavioural analysis will be performed, serving as well as input for the Analytics component (Figure 3.74).

The analytics component will perform calculations on the data from the above-described flows and from the inputs of the configurator. The latter involves interaction of the user, in order to configure specifications for the scenario generation.

The results are depicted in the User Interface which is responsible not only to visualize the VaR/ES predictions but also to perform pre-trade analysis leveraging the developed risk assessment models.

Logical Schema

Components

The workflow leverages the following components:

- BigData Management Layer i.e. INFINISTORE and Online Aggregates (Data Management in RA).
- Custom Injection Simulator (Data Ingestion in RA).
- Kafka (namespace Cross Cutting in RA).
- Zookeeper (Cross Cutting in RA).
- AI model for VaR prediction (Analytics and Machine Learning in RA).
- UI Risk Assessment based on VaR (Interface in RA).
- Sentiment Analysis for financial news (Analytics and Machine Learning in RA).

Conclusions – Issues and Barriers

The implementation progresses smoothly in terms of its BigData and data analytics parts. Nevertheless, the datasets used are still quite limited. Furthermore, the prototype of the market sentiment component is not available. Likewise, the NOVA testbed is not fully operational. These are two of the main risks that have to monitored and cleared in the coming months i.e. within the period M13-M18, so as to ensure that the full-scale implementation is on track.

This pilot is enhanced with semantic interoperability features/functionalities, which were not originally foreseen. Specifically, the pilot systems will support inputs (e.g., Trades Data) in FIBO (Financial Industry Business Ontology) 7 semantic format, to support VaR calculation in cases of portfolios for large investors (e.g., large investment banks, institutional investors) that might hold assets/trades across multiple platforms. In this case, the VaR of a portfolio might have to be calculated based on data from multiple platforms that produce data in different semantics and formats. FIBO will ensure the semantic integration and semantic interoperability of these streams/trades towards facilitating VaR calculation for large portfolios. The integration of a semantic interoperability module in the pilot system is considered as a highlight for the pilot.

DOI: 10.1561/9781638282310.ch4

Chapter 4

INFINITECH Conclusions

4.1 Conclusions

This Book Series provides an overview of FinTech services and applications developed in particular pilot locations across Europe and beyond i.e. Israel, Turkey, etc. The aim is to specify different aspects of each large-scale pilot: readiness; development; and validation of different services and components. In so doing, validation becomes the core element in this process as the main objectives of INFINITECH is to test innovative (IoT, BigData, AI, ML, Blockchain and more) technologies towards improving business services in the Financial and Insurance sector. We report on the readiness of the various pilot sites to test the INFINITECH innovative AI, IoT and BigData technologies into the testbeds/sandoxes that are developed during the project, while validating their ability to improve the business processes of end-user organizations (i.e. financial organizations, banks, and FinTech firms).

In the second chapter we describe innovative technologies for financial sector. We explain work package 4 of INFINITECH project in detail. Work package 4, which is Interoperable data exchange and semantic interoperability, focuses on establishing the foundation for common, shared meaning across the several data sources and message and event feeds within the INFINITECH platform while facilitating the technical implementation of the INFINITECH principles. It comprises of six tasks which are described thoroughly in this chapter. We further present background and related works and concepts and definitions.

Furthermore, we describe 16 pilots organized in 5 categories in detail. More specifically, for each pilot we describe the overall objective of each pilot, technological components and services, testbed, implementation of a first Proof of concept, expected outcomes, datasets, data produced, explainable workflow, logical schema, components, and conclusions – issues and barriers. During the first year of the project, pilot focused on use cases definition, requirements identification, reference architecture, and corresponding deliverables. Great effort from all pilots covering requests coming from different partners and workpackages; working as a whole. Communications have been crucial to organise and progress in a proper way. The effort of project partners led into a Proof of Concept (PoC) for all pilots (with the exceptions explained at the introduction) that summarizes developments and achievements. This PoC also refined the targets of each pilot, whilst helped them to identify new requirements and envision possible constraints and issues. This way, every pilot can work on an improved and more fruitful outcomes within its cluster. The following pilot provides an overview of the status of the various pilots and illustrates that most pilots have managed to implement an initial proof-of-concept and demonstrator:

Pilot Theme	Implementation Status (As of November 2020)
Invoices Processing Platform for a more Sustainable Banking Industry	• Stakeholders Mobilized. • Architecture Finalized. • Partial Implementation of PoC.
Real-time risk assessment in Investment Banking	• Stakeholders Mobilized. • Architecture Finalized. • Implementation of Initial Integrated PoC.
Open Inter-Banking Pilot	• Pilot Defined and Planned. • Stakeholders Engagement Planned.
Collaborative Customer-centric Data Analytics for Financial Services	• Pilot still at Specification Stage due to late inclusion of key stakeholders in the project.
Personalised Portfolio Management ("Why Private Banking cannot be for everyone?")	• Stakeholders Mobilized. • Architecture Finalized. • Implementation of Initial Integrated PoC.
Business Financial Management (BFM) tools delivering a Smart Business Advise	• Stakeholders Mobilized. • Architecture Finalized. • Implementation of Initial Integrated PoC.
Personalized Closed-Loop Investment Portfolio Management for Retail Customers	• Stakeholders Mobilized. • Architecture Finalized. • Implementation of Initial Integrated PoC.
Avoiding Financial Crimes	• Stakeholders Mobilized. • Architecture Finalized. • Implementation of Initial Integrated PoC.
Platform for AML supervision	• Stakeholders Mobilized. • Architecture Finalized. • Implementation of Initial Integrated PoC.

Figure 4.1. Continued

Real-time cybersecurity analytics on financial transactions' data	• Stakeholders Mobilized. • Architecture Finalized. • Implementation of Initial Integrated PoC.
Analysing Blockchain Transaction Graphs for Fraudulent Activities	• Stakeholders Mobilized. • Architecture Finalized. • Implementation of Initial Integrated PoC.
Personalized insurance products based on IoT connected vehicles	• Stakeholders Mobilized. • Architecture Finalized. • Implementation of Initial Integrated PoC.
Real World Data for Novel Health-Insurance products	• Stakeholders Mobilized. • Architecture Finalized. • Implementation of Initial Integrated PoC.
Alternative and automated insurance risk selection and insurance product recommendation for SME's	• Stakeholders Mobilized. • Architecture Finalized. • Implementation of Initial Integrated PoC.
Big Data and IoT for the Agricultural Insurance Insurance Products Industry	• Stakeholders Mobilized. • Architecture Finalized. • Implementation of Initial Integrated PoC.

Figure 4.1. High level overview of pilots' implementation status.

The successful implementation of Proof-of-Concepts for most of the INFINITECH pilots provides evidence of progress and readiness for the pilots, while at the same time manifesting the collaborative efforts and the synergies between the INFINITECH partners.

So far, pilot's development has been running pretty in parallel, because of time restrictions, with the technological WPs. Not showing, in general, the technological match between pilots needs and INFINITECH provided technologies. This also happened because of INFINITECH technologies have been involved in a first definition process. The work done to contribute to this report and putting together all PoCs helped to break these silos between pilots and share technologies and architecture components. Pilots results will be able to show, use and integrate the results provided by technological INFINITECH workpackages. This will happen with the INFINITECH technologies more defined (deliverables in WP3, WP4 and WP5), and the testbeds and sandboxes ready to support this integration.

WP6 testbeds and sandboxes will unify the way that pilots are prepared, from an infrastructure and deployment perspective, to set the base for similar use cases, or stakeholders with similar needs, to try INFINITECH technologies. Reference Architecture settles the bases for multi-layer architecture, with different components that can be plugged and combined. This approach has been followed by all pilots and it will finish with the all the corresponding testbeds deployment. This way, Kubernetes and Docker's orchestration framework will demonstrate these multi-layer-plugable approach defended by INFINITECH. Pilots already started to work (some fully prepared) with Docker containers for a smooth transition and implementation of sandboxes.

This first phase could be summarized with pilots focusing on data collection and preparation and deployment of first set of components. Data capture, filtering, homogenization (following the Reference Architecture) is already there, and it is starting to work. In the coming phase, these components will be complemented with the results of INFINITECH technologies and services (W3-WP5) and the creation of sandboxes (WP6) that fill finalize a common way of working. While finishing this deliverable, a testing infrastructure is been put in place. Pilots will have an infrastructure to manage their software components, CI/CD tools for deployment and an environment to create/use blueprints for their architectures. These blueprints will help replicability of similar scenarios and needs, e.g. how to get/inject data through a data pipeline into a LeanXcale database for later analysis.

Cluster 1 comprehends different pilots linked mainly by services and risk assessment purposes. The overall development and deployment of the pilots is proceeding as planned. Generally, the requirements and development phase for the first two pilots is at an advanced state, on one hand, having already implemented the PoC, deployed on a onpremise cloud-based testbed, on the other hand, having already implemented the PoC that will be deployed on the shared testbed hosted by NOVA. Instead, Pilot#15 will be deployed in the testbed blueprint: indeed Cluster 1 provides a comprehensive view of pilots' deployment by exploiting the three different "typologies" of testbeds established by the INFINITECH project. Overall, the development and training of machine learning applications is proceeding, enhancing the innovative components of the pilots. To conclude, Pilot#1 already planned to perform relevant activities in stakeholders' engagement, demonstrating the matureness and readiness of such pilot, whereas Pilot#2 is a clear example of how INFINITECH technologies can be exploited to develop a trading-based risk assessment use case.

Cluster 2 of Pilots that are related on Personalized Retail and Investment Banking Services, based on the progress until now, they are progressing following their initial plans (except Pilot #3 that is in the process of redefining its scope). The majority are in the process of building the ground for each pilot, which includes mainly the AI power tools that will be used as basis for the final deployment. Most of the pilots either established or in the process of testbed deployment and now based on the relative blueprint definition will start working towards to INFINITECH way of deployment. Even though the main activities already reported are mainly focus on the technical site, the actual goal for each pilot focus on providing technologies that will improve the financial health of individuals and SMEs, either through better and personalized investment propositions or better financial management tools.

At cluster 3 level, we can summarize the current status of implementation of pilots at a general good and promising point. Each of these pilots implemented a PoC initial prototype, provided initial bunch of data and a testbed installation to implement the first serviced and develop the need its technology. First data analytics components, risk calculation engines, complex search services, and user interfaces (for risk assessments), have been developed. Financial Information synthetic data are currently used in research environment at JSI. Blockchain technologies are started demonstration to provide more secured and trusted transactions systems, facing the difficulties of a so high computational demand derived from these technologies: transaction dataset preparation, huge (scalable) transaction graph analysis and visualization tools. A case apart is represented by Pilot #7, because of a change in pilot partners, and the subsequent need for updated specifications, as well as for the redefinition of the pilot according to the new partners.

Cluster 4 is focused on different insurance services customisation, by exploiting real world data collected from users through different AI powered technologies that evaluate the insured client's behaviour and his/her associated potential risks. This first stage on the cluster pilots' development analysed the different available data sources, identifying which are relevant for the use cases to be played and built all the mechanisms needed to gather, curate, and homogenise these identified datasets. In parallel, the infrastructure to collect, store and classify the information has been defined and implemented, so, aligned with INFINITECH development and deployment guidelines, the second stage, which will design, build, and run all the novel ML models. These ML/DL models will be based on cutting-edge AI technologies and will be specifically created to solve the particularities of each scenario. In turn, they will be the key component on the final new services to be offered to insurance companies and insured clients.

Cluster 5 is focused on customized and configurable insurance products based on non-traditional data sources and not obtained directly from the insured subjects. The objective is to obtain a better determination of the insured risk, the insured enterprises and agricultural sector. On the one hand, to offer a more adjusted and personalised insurance and on the other hand to speed up the payment of the compensation. The technologies that will be used are based on Machine Learning and AI on large amounts of data obtained from sources both in text format and in satellite images. The process is composed into three phases, the determination of the relevant sources to provide data to the models and their homogenization for processing. The second is the management of the data within the reference architecture established in the lines of INFINITECH. Finally, ML/DL models will be based on cutting-edge AI technologies and will be specifically created to solve the particularities of each scenario.

Finally, most pilots are now focused on technological developments and not so focused on Business Processes and Stakeholders Involvements:

- Business Process Change and Innovation: What is the system changing in the business? How things are done today and how they will be done after INFINITECH?
- Stakeholders' Involvement: Who is involved from the business side? Who are the end-users and how they are involved in the pilots? Are there stakeholders' workshops planned to evaluate the pilot systems? How many participants are expected when they will be scheduled? Do we need to train some users to use the system?

DOI: 10.1561/9781638282310.ch5

Chapter 5

References

5.1 Infinitech Public Deliverables

D1.3 Data Management Plan
D1.5 Data Management Plan – II

D2.1 User Stories and Stakeholders' Requirements – I
D2.2 User Stories and Stakeholders' Requirements – II
D2.3 Reference Scenarios and Use Cases – I
D2.4 Reference Scenarios and Use Cases – II
D2.5 Specifications of INFINITECH Technologies – I
D2.6 Specifications of INFINITECH Technologies – II
D2.7 Security and Regulatory Compliance Specifications – I
D2.8 Security and Regulatory Compliance Specifications – II
D2.9 Initial Specification of Testbeds, Data Assets and APIs – I
D2.10 Initial Specification of Testbeds, Data Assets and APIs – II
D2.11 Data Model Specifications – I
D2.12 Data Model Specifications – II
D2.13 INFINITECH Reference Architecture – I
D2.14 INFINITECH Reference Architecture – II
D2.15 INFINITECH Reference Architecture – III

D3.1 Hybrid Transactional/Analytics Processing for Finance and Insurance Applications – I
D3.2 Hybrid Transactional/Analytics Processing for Finance and Insurance Applications – II
D3.3 Hybrid Transactional/Analytics Processing for Finance and Insurance Applications – III
D3.4 Integrated (Polyglot) Persistence – I
D3.5 Integrated (Polyglot) Persistence – II
D3.6 Data Streaming and Data at Rest Queries Integration – I
D3.7 Data Streaming and Data at Rest Queries Integration – II
D3.8 Data Streaming and Data at Rest Queries Integration – III
D3.9 Automatic Parallelization of Data Streams and Intelligent Pipelining – I
D3.10 Automatic Parallelization of Data Streams and Intelligent Pipelining – II
D3.11 Automatic Parallelization of Data Streams and Intelligent Pipelining – III
D3.12 Data Governance Framework and Tools – I
D3.13 Data Governance Framework and Tools – II
D3.14 Data Governance Framework and Tools – III
D3.15 Regulatory Compliance Tools – I
D3.16 Regulatory Compliance Tools – II
D3.17 Regulatory Compliance Tools – III
D3.18 Integrated (Polyglot) Persistence – III

D4.1 Semantic Models and Ontologies – I
D4.2 Semantic Models and Ontologies – II
D4.3 Semantic Models and Ontologies – III
D4.4 Semantic Streams Analytics Engine – I
D4.5 Semantic Streams Analytics Engine – II
D4.6 Semantic Streams Analytics Engine – III
D4.7 Permissioned Blockchain for Finance and Insurance – I
D4.8 Permissioned Blockchain for Finance and Insurance – II
D4.9 Permissioned Blockchain for Finance and Insurance – III
D4.10 Blockchain Tokenization and Smart Contracts – I
D4.11 Blockchain Tokenization and Smart Contracts – II
D4.12 Blockchain Tokenization and Smart Contracts – III
D4.13 Encrypted Data Querying and Personal Data Market – I
D4.14 Encrypted Data Querying and Personal Data Market – II
D4.15 Encrypted Data Querying and Personal Data Market – III

D4.16 Visualization Frond-End for Aggregated Information – I
D4.17 Visualization Frond-End for Aggregated Information – II

D5.1 Library of Parallelized Incremental Analytics – I
D5.2 Library of Parallelized Incremental Analytics – II
D5.3 Library of Parallelized Incremental Analytics – III
D5.4 Framework for Declarative and Configurable Analytics – I
D5.5 Framework for Declarative and Configurable Analytics – II
D5.6 Framework for Declarative and Configurable Analytics – III
D5.7 Library of ML/DL Algorithms – I
D5.8 Library of ML/DL Algorithms – II
D5.10 Data Management Workbench and Open APIs – I
D5.11 Data Management Workbench and Open APIs – II
D5.12 Data Management Workbench and Open APIs – III
D5.13 Datasets for Algorithms Training & Evaluation – I
D5.14 Datasets for Algorithms Training & Evaluation – II

D6.1 Testbeds Status and Upgrades – I
D6.2 Testbeds Status and Upgrades – II
D6.3 Testbeds Status and Upgrades – III
D6.4 Tools and Techniques for Tailored Sandboxes and Management of Datasets – I
D6.5 Tools and Techniques for Tailored Sandboxes and Management of Datasets – II
D6.6 Tools and Techniques for Tailored Sandboxes and Management of Datasets – III
D6.7 Sandboxes in Incumbent Testbeds – I
D6.8 Sandboxes in Incumbent Testbeds – II
D6.9 Sandboxes in Incumbent Testbeds – III
D6.10 Sandboxes for FinTech/InsuranceTech Innovators – I
D6.11 Sandboxes for FinTech/InsuranceTech Innovators – II
D6.12 Sandboxes for FinTech/InsuranceTech Innovators – III
D6.13 Testbeds Support and Certification Services – I
D6.14 Testbeds Support and Certification Services – II
D6.15 Testbeds Support and Certification Services – III

D7.1 Report on Pilot Sites Preparation – I
D7.2 Report on Pilot Sites Preparation – II
D7.18 Pilot Sites Synergies and Collaboration – I
D7.19 Pilot Sites Synergies and Collaboration – II

D7.20 Pilots' Evaluation and Stakeholders' Feedback – I
D7.21 Pilots' Evaluation and Stakeholders' Feedback – II
D7.22 Pilots' Evaluation and Stakeholders' Feedback – III

D8.1 Market Platform and VDIH Specifications – I
D8.2 Market Platform and VDIH Specifications – II
D8.3 BigData and AI Solutions Marketplace – I
D8.4 BigData and AI Solutions Marketplace – II
D8.5 IoT and Blockchain Solutions Marketplace – I
D8.6 IoT and Blockchain Solutions Marketplace – II
D8.7 Virtualized Digital Innovation Hub – I
D8.8 Virtualized Digital Innovation Hub – II
D8.9 Third-Party FinTech and InsuranceTech Solutions – I
D8.10 Third-Party FinTech and InsuranceTech Solutions – II
D8.11 Ecosystem Support Services and Third Party Solutions – I
D8.12 Ecosystem Support Services and Third Party Solutions – II

D9.1 Detailed Dissemination/Communication Plan
D9.2 Dissemination and Communication Activities – I
D9.3 Dissemination and Communication Activities – II
D9.4 Dissemination and Communication Activities – III
D9.5 Dissemination and Communication Activities – IV
D9.6 Contributions to Standards, Associations and Clusters – I
D9.7 Contributions to Standards, Associations and Clusters – II
D9.8 Contributions to Standards, Associations and Clusters – III
D9.9 Community Building Report – I
D9.10 Community Building Report – II
D9.11 Community Building Report – III
D9.12 Exploitation and Sustainability Plan – I

5.2 References

Accenture – Technology Vision for Insurance 2021 https://www.accenture.com/us-en/insights/insurance/technology-vision-insurance-2021.

A. Narayanan and V. Shmatikov, "De-anonymizing Social Networks," in 2009 30th IEEE Symposium on Security and Privacy, May 2009, pp. 173–187.

A. Machanavajjhala, D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber, "Privacy: Theory meets Practice on the Map," 2008 IEEE 24th International Conference on Data Engineering. 2008, doi: 10.1109/icde.2008.4497436.

A. Crespo, "PowerPoint presentation: eIDAS-Compliant Cross-Border Authentication – Alberto Crespo." 2018.

"ARIES – ReliAble euRopean Identity EcoSystem." [Online]. Available: https://www.aries-project.eu/. [Accessed: 07-Jul-2021].

A-System Requirements. ICAO – Security and facilitation. Retrieved June 24, 2021, from https://www.icao.int/Security/FAL/PKD/BVRT/Pages/System-Requirements.aspx.

"Article 29 Working Party. Opinion 05/2014 on Anonymisation Techniques (2014)." https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf (accessed Sep. 14, 2020).

A. Narayanan and V. Shmatikov, "De-anonymizing Social Networks," in 2009 30th IEEE Symposium on Security and Privacy, May 2009, pp. 173–187.

A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," ACM Trans. Knowl. Discov. Data, vol. 1, no. 1, p. 3–es, 2007.

"Amazon Kinesis.", https://aws.amazon.com/kinesis/.

Acerbi, C., & Tasche, D. (2002). Expected Shortfall: a natural coherent alternative to Value at Risk. Economic notes, 31(2), 379–388. Economic notes.

Allen, D. M. (2012). The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction. Technometrics, 16(1), 125–127. doi/abs/10.1080/00401706.1974.10489157.

Alpaydin, E. (2014). Introduction to Machine Learning (3rd ed.). MIT Press.

Altman, N.S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46(3), 175–185.

A. W. Service, "Amazon Kubernetes Service," [Online]. Available: https://aws.amazon.com/it/quickstart/architecture/amazon-eks/.

A. C. P. C. J. T. E. B. a. R. S. W. Radomski, "ERC-1155 multi token standard," 2015. [Online]. Available: https://eips.ethereum.org/EIPS/eip-1155. [Accessed 20 November 2021].

Alberto Crespo García, Nicolás Notario McDonnell, Carmela Troncoso et al., Privacy- and Security-by-Design Methodology Handbook, December 2015, https://fr.slideshare.net/richard.claassens/priparemethodologyhandbookfinalfeb242016 and https://ipen.trialog.com/images/ipen/a/a1/PRIPARE_Methodology_Handbook_Final_Feb_24_2016.pdf (accessed in November 2020).

ARIES consortium, "D4.1 ARIES prototype instantiation." 2018, [Online]. Available: https://www.ariesproject.eu/sites/default/files/aries/public/contentfiles/deliverables/D4.1%20ARIES%20prototype%20instantiation.pdf.

A. Polyviou, P. Velanas, and J. Soldatos, "Blockchain Technology: Financial Sector Applications Beyond Cryptocurrencies," Proceedings, vol. 28, no. 1, p. 7, 2019.

A. Camerra, J. Shieh, T. Palpanas, T. Rakthanmanon, and E. J. Keogh. Beyond one billion time series: indexing and mining very large time series collections with iSAX2+. Knowledge and Information Systems (KAIS), 39(1):123–151, 2014.

A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," ACM Trans. Knowl. Discov. Data, vol. 1, no. 1, p. 3–es, 2007.

Alberto Crespo García, Nicolás Notario McDonnell, Carmela Troncoso et al., Privacy- and Security-by-Design Methodology Handbook, December 2015, https://fr.slideshare.net/richard.claassens/priparemethodologyhandbookfinalfeb242016 and https://ipen.trialog.com/images/ipen/a/a1/PRIPARE_Methodology_Handbook_Final_Feb_24_2016.pdf (accessed in November 2020).

AWS, "Amazon EKS," Amazon, 2020. [Online]. Available: https://aws.amazon.com/eks/?nc1=h_ls.

A. Z. R. a. B. S. Dandekar, "Comparative evaluation of synthetic data generation methods," in ACM Conference (Deep Learning Security Workshop), 2017.

A. Palavalli, D. Karri, and S. Pasupuleti, "Semantic internet of things," in 2016 IEEE Tenth International Conference on Semantic Computing (ICSC), 2016, pp. 91–95.

A. Narayanan and V. Shmatikov, "Robust De-anonymization of Large Sparse Datasets," in 2008 IEEE Symposium on Security and Privacy (sp 2008), May 2008, pp. 111–125.

Alexandrov, A., Benidis, K., Bohlke-Schneider, M., Flunkert, V., Gasthaus, J., Januschowski, T., Maddix, D.C., Rangapuram, S., Salinas, D., Schulz, J., Stella, L., Caner Turkmen, A., and Wang, Y, (2019). GluonTS: probabilistic time series models in Python. arXiv preprint arXiv:1906.05264.

"Apache Calcite", https://calcite.apache.org/.

"Article 29 Working Party. Opinion 05/2014 on Anonymisation Techniques (2014)." https://ec.europa.eu/justice/article-29/documentation/opinionrecommendation/files/2014/wp216_en.pdf (accessed Sep. 14, 2020).

AWS, "What is aws," Amazon, 2020. [Online]. Available: https://aws.amazon.com/it/what-is-aws/.

Apache.org, "Maven," [Online]. Available: https://maven.apache.org/.

"Apache Tomcat," [Online]. Available: http://tomcat.apache.org/. [Accessed 15 October 2020].

Abouzeid, K. Badja-Pawlikowski, D. Abadi, A. Silberschatz, A. Rasin, "HadoopDB: an architectural hybrid of MapReduce and DBMS technologies for analytical workloads", PVLDB, vol. 2, pp. 922–933 (2009).

Apache Impala, http://impala.apache.org/.

Apache Drill – Schema-free SQL Query Engine for Hadoop, NoSQL and Cloud Storage, https://drill.apache.org/.

A. De Nicola and M. Missikoff, "A lightweight methodology for rapid ontology engineering," Communications of the ACM, vol. 59, no. 3, pp. 79–86, 2016.

A. Martidis, I. Tomasic, and P. Funk, "Deliverable 5.3: CREATE Interoperability." Aug. 25, 2014.

Adar, E., and Huberman, B. (2001). A market for secrets. First Monday, 6(8).

A. Mueen, Y. Zhu, M. Yeh, K. Kamgar, K. Viswanathan, C. Gupta, and E. Keogh. The fastest similarity search algorithm for time series subsequences under euclidean distance, August 2017. http://www.cs.unm.edu/~mueen/FastestSimilaritySearch.html.

Alexis Hope, Catherine D'Ignazio, Josephine Hoy, Rebecca Michelson, Jennifer Roberts, Kate Krontiris, and Ethan Zuckerman. (2019). Hackathons as Participatory Design: Iterating Feminist Utopias. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, 61.

Arnab Nandi and Meris Mandernach. (2016). Hackathons as an informal learning platform. In Proceedings of the 47th ACM Technical Symposium on Computing Science Education. ACM, 346–351.

Angarita, M. A. M., & Nolte, A. (2020). What do we know about hackathon outcomes and how to support them?–A systematic literature review. In International Conference on Collaboration Technologies and Social Computing (pp. 50–64). Springer, Cham.

Anti-money laundering and countering the financing of terrorism legislative package. (2021). European Commission. https://ec.europa.eu/info/publications/210720-anti-money-laundering-countering-financing-terrorism_en (Accessed: January 2022).

AIOTI |The Alliance for the Internet of Things Innovation. Retrieved January, 2022 from https://aioti.eu/https://aioti.eu/.

https://cordis.europa.eu/project/id/951972.

Apache SparkTM – Unified Analytics Engine for Big Data. (2019). Apache.org. https://spark.apache.org/ (Accessed: January 2022).

Apache Kafka. https://kafka.apache.org/ (Accessed: January 2022).

Apache Flink: Stateful Computations over Data Streams. (2014). Apache.org. https://flink.apache.org/ (Accessed: January 2022).

"Apache Storm.", http://storm.apache.org/.

Alexander Nolte, Ei Pa Pa Pe-Than, Anna Filippova, Christian Bird, and Herbsleb James D. Scallen, Steve. (2018). You Hacked and Now What? – Exploring Outcomes of a Corporate Hackathon. Proceedings of the ACM on Human-Computer Interaction 2, CSCW (2018), 129:1–129:23.

Anne-Laure Mention (2019) The Future of Fintech, Research-Technology Management, 62:4, 59–63, DOI: 10.1080/08956308.2019.1613123.

Arasu A. et al. (2016) STREAM: The Stanford Data Stream Management System. In: Garofalakis M., Gehrke J., Rastogi R. (eds) Data Stream Management. Data-Centric Systems and Applications. Springer, Berlin, Heidelberg.

A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," IEEE Pervasive Comput., vol. 2, no. 1, pp. 46–55, Jan. 2003.

A. Simitsis, K. Wilkinson, M. Castellanos, U. Dayal, "Optimizing analytic data flows for multiple execution engines", in ACM SIGMOD, 2012, pp. 829–840.

A. Tomasic, L. Raschid, P. Valduriez, "Scaling access to heterogeneous data sources with DISCO", IEEE Trans. On Knowledge and Data Engineering, vol. 10, pp. 808–823, 1998.

B. Gedik and L. Liu, "Location Privacy in Mobile Systems: A Personalized Anonymization Model," 25th IEEE International Conference on Distributed Computing Systems (ICDCS'05). doi: 10.1109/icdcs.2005.48.

Baik Hoh and M. Gruteser, "Protecting Location Privacy Through Path Confusion," in First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SECURECOMM'05), Sep. 2005, pp. 194–205.

Baik Hoh, M. Gruteser, Hui Xiong, and A. Alrabady, "Enhancing Security and Privacy in Traffic-Monitoring Systems," IEEE Pervasive Comput., vol. 5, no. 4, pp. 38–46, Oct. 2006.

"Beating financial crime: Commission overhauls anti-money laundering and countering the financing of terrorism rules. Press Release. Website of the European Union." 2021. [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_3690 [Accessed July-2021].

Baptista, A. I., & Silva, A. P. (2018). FINANCIAL TECHNOLOGIES EFFECT ON FINANCIAL SERVICES FROM AN OPEN INNOVATION PERSPECTIVE. Theory and Applications in Game Theory, 179.

B Campbell C Mortimore, "Security Assertion Markup Language (SAML) 2.0 Profile for OAuth 2.0 Client Authentication and Authorization Grants." 2015, [Online]. Available: https://tools.ietf.org/html/rfc7522.

Brown, R. G. (1959). Statistical forecasting for inventory control. McGraw/Hill.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.

Becker and Whisler, 1967 https://econpapers.repec.org/article/ucpjnlbus/v_3a40_3ay_3a1967_3ap_3a462.html.

Busilytics – Digital Transformation https://www.busilytics.com/digital-transformation/.

"Balisage Paper: RESTful Service Description Language (RSDL)," [Online]. Available: http://www.balisage.net/Proceedings/vol10/html/Robie01/BalisageVol10-Robie01.html. [Accessed 24 03 2021].

BOOST4.0, D2.5 – BOOST 4.0 Reference Architecture Specification v1, Available at: https://cordis.europa.eu/project/id/780732/results [Accessed: 15-May-2020].

Big Data Value Association. BVA SRIA—European big data value strategic research and innovation agenda. 2017. http://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf. Accessed 18 Feb 2020.

Boid D, Chang W. NIST Big Data Interoperability Framework: Volume 6, RA Version 2. NIST Big Data Program. 2018. https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-6r1.pdf.

Bons, R.W.H., Alt, R., Lee, H.G. et al. Banking in the Internet and mobile era. Electron Markets 22, 197–202 (2012). https://doi.org/10.1007/s12525-012-0110-6.

B. Kolev, P. Valduriez, C. Bondiombouy, R. Jiménez-Peris, R. Pau, J. Pereira, "CloudMd-sQL: querying heterogeneous cloud data stores with a common language", Distributed and Parallel Databases, vol. 34, pp. 463–503. Springer (2015).

Botts, M., Percivall, G., Reed, C. and Davidson. J. OGC Sensor Web Enablement: Overview and High Level Architecture. Technical report, OGC, December 2007.

Barros, R., & Santos, S. (2018). A large-scale comparison of concept drift detectors. Information Sciences, 451–452, 348–370.

Bhattacharya, A. (2020) Effective Approaches for Time Series Anomaly Detection. towardsdatascience.com.

Bergstra, J., & Bengio, Y. (2012) Random Search for Hyper-Parameter Optimization. Journal of Machine Learning Research, 13(1), 281–305.

Bifet, A., & Gavaldà, R. (2007). Learning from time-changing data with adaptive windowing. In Proceedings of the Seventh SIAM International Conference on Data Mining, SDM '07, SIAM, pp. 443–448.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. Springer Link. doi.org/10.1023/A:1010933404324.

Buchanan, B. G. (2019). Artificial Intelligence in finance. The Alan Turing Institute, 00(00), 3–49. 10.5281/zenodo.2626454.

"Blockchain basics: Hyperledger Fabric," IBM Developer, 2020. [Online]. Available: https://developer.ibm.com/technologies/blockchain/articles/blockchain-basics-hyperledgerfabric/. [Accessed 29 August 2020].

Bishop, C.M. (2006). Pattern recognition. Machine learning 128, no. 9.

Breiman, L. (2001). Random Forests. Machine Learning, 45 (1), pp. 5–32.

Berg, Larry K., et al. "Evaluation of a modified scheme for shallow convection: Implementation of CuP and case studies." Monthly Weather Review 141.1 (2013): 134–147.

Briscoe G, Mulligan C, (2014). "Digital innovation: the hackathon phenomenon" Creative works London/QMUL, London. Available at: http://www.creativeworkslondon.org.uk/wpcontent/uploads/2013/11/Digital-Innovation-TheHackathon-Phenomenon1.pdf.

Bard Rosell, Shiven Kumar, and John Shepherd. (2014). Unleashing innovation through internal hackathons. In Innovations in Technology Conference (InnoTek), 2014 IEEE. IEEE, 1–8.

BDVA Task Force 6: Technical. https://www.bdva.eu/task-force-6 (Accessed: January 2022).

B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, Estimating the Support of a High-Dimensional Distribution. Neural Comput. 13, 7, 1443–1471, 2001.

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

"CEF Digital Home." https://ec.europa.eu/cefdigital/wiki/cefdigital/wiki/display/CEFDIGITAL/CEF+Digital+Home (accessed Sep. 14, 2020).

C. Gómez, "eID under eIDAS Building trust in a digital society – DG CONNECT-European Commission." [Online]. Available: http://st.fbk.eu/sites/st.fbk.eu/files/20180316_eidas_oauth_security_workshop.pdf.

C. Dwork, A. Roth, and Others, "The algorithmic foundations of differential privacy," Found. Trends Theor. Comput. Sci., vol. 9, no. 3–4, pp. 211–407, 2014.

C. Bettini, X. S. Wang, and S. Jajodia, "Protecting Privacy Against Location-Based Personal Identification," in Secure Data Management, 2005, pp. 185–199.

C. A. Ardagna, M. Cremonini, E. Damiani, S. D. C. di Vimercati, and P. Samarati, "Location Privacy Protection Through Obfuscation-Based Techniques," Data and Applications Security XXI. pp. 47–60, 2007, doi: 10.1007/978-3-540-73538-0_4.

Çetintemel U. et al. (2016) The Aurora and Borealis Stream Processing Engines. In: Garofalakis M., Gehrke J., Rastogi R. (eds) Data Stream Management. Data-Centric Systems and Applications. Springer, Berlin.

Chira Barua et al., "The last pit stop? Time for bold late-cycle moves – McKinsey Global Banking Annual Review", McKinsey & Co, 2019.

Compliance Maturity Level https://www.leancompliance.ca/post/capabilities-maturity-model-for-compliance*. (https://www.leancompliance.ca/post/capabilities-maturity-model-for-compliance).

"Consul," [Online]. Available: https://www.consul.io/. [Accessed 10 October 2020].

C. Bondiombouy, B. Kolev, O. Levchenko, P. Valduriez, "Multistore big data integration with CloudMdsQL", Transactions on Large-Scale Data and Knowledge-Centered Systems (TLDKS), pp. 48–74. Springer (2016).

C. Iso and others, "ISO/IEC 2382-1: 1993 Information Technology-Vocabulary-Part 1: Fundamental terms," 1993.

Compton, M., Barnaghi, P., Bermudez, L., Castro, R. G., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K. Kelsey, W. D., Phuoc, D. L., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheath, A. and Taylor, K. The SSN Ontology of the Semantic Sensor Networks Incubator Group. Journal of Web Semantics: Science, Services and Agents on the World Wide Web, ISSN 1570-8268, Elsevier, 2012.

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. ICML '06: Proceedings of the 23rd International Conference on Machine learning, 161–168. ACM Digital Library. doi.org/10.1145/1143844.1143865.

Chawla, N. V., Herrera, F., Garcia, S., & Fernandez, A. (2018) SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. Journal of Artificial Intelligence Research, 61, 863–905.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. Springer. doi.org/10.1007/BF00994018.

Cristianini, N., & Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press. doi.org/10.1017/CBO9780511801389.

"Consent Receipt Specification – Kantara Initiative," Kantara Initiative, 2020. [Online]. Available: https://kantarainitiative.org/download/7902/. [Accessed 1 September 2020].

Cranor C.D., Johnson T., Spatscheck O. (2016) Stream Processing Techniques for Network Management. In: Garofalakis M., Gehrke J., Rastogi R. (eds)

Data Stream Management. Data-Centric Systems and Applications. Springer, Berlin, Heidelberg C.

C. Dwork, A. Roth, and Others, "The algorithmic foundations of differential privacy," Found. Trends Theor. Comput. Sci., vol. 9, no. 3–4, pp. 211–407, 2014.

C. Faloutsos, M. Ranganathan, and Y. Manolopoulos. Fast subsequence matching in timeseries databases. In Proceedings of the International Conference on Management of Data (SIGMOD), pages 419–429, 1994.

CNIL, Infographic on DPIA methodology, retrieved on Mai 13th 2020 via https://www.cnil.fr/sites/default/files/atoms/files/171019_fiche_risque_en_cmjk.pdf.

Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, 2018, pp. 591–594. [MIRAB12] Microsoft Industry Reference Architecture for Banking (MIRA-B), Microsoft Corporation Whitepaper, May 2012.

"Credit Card Fraud Detection", 2016. [Online]. Available: https://www.kaggle.com/mlg-ulb/creditcardfraud

«Cloning,» [Online]. Available: https://sdorsett.github.io/post/2018-12-24-using-terraform-toclone-a-virtual-machine-on-vsphere/).

«CNCF,» Linux Foundation, [Online]. Available: https://www.cncf.io/.

Christoersen, P.F.: Evaluating interval forecasts. International economic review pp. 841–862 (1998).

Chesbrough, H. (2019). Open innovation results: Going beyond the hype and getting down to business. Oxford University Press.

Chainalysis Team Indirect Exposure: Why you need to look beyond direct counterparties to understand cryptocurrency address risk, https://blog.chainalysis.com/reports/cryptocurrency-risk-blockchain-analysisindirect-exposure, November 2020.

Chesbrough, H., and Bogers, M. (2014) Explicating open innovation: Clarifying an emerging paradigm for understanding innovation, In H. Chesbrough, W. Vanhaverbeke, & J. West (Eds.), New Frontiers in Open Innovation: 3–28. Oxford: Oxford University Press.

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL Towards better implementation of the EU's anti-money laundering and countering the financing of terrorism frameworkCOM/2019/360 final https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX%3A52019DC0360 (Accessed: January 2022).

"Circuit Breaker Pattern," 2020. [Online]. Available: https://docs.microsoft.com/enus/azure/architecture/patterns/circuit-breaker. [Accessed 10 October 2020].

"Communication in a microservice architecture," [Online]. Available: https://docs.microsoft.com/enus/dotnet/architecture/microservices/architect-micros ervice-container-applications/communication-inmicroservice-architecture. [Accessed 24 03 2021].

Contentcal.io what-is-a-good-social-media-engagement (https://www.contentcal.io/blog/what-is-a-good-social-media-engagement-rate/#:~:text=A%20good%20engagement%20rate%20for%20Linkedin%20is%20about%202%25%2C%20however,type%20of%20content%20you%20share.)

Context Information Management (CIM); NGSI-LD API Disclaimer. (n.d.). https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.05.01_60/gs_CIM009v010501p.pdf (Accessed: January 2022).

Cilium. [Online]. Available: https://cilium.io/.

Data Protection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019, https://www.scl.org/news/10384-data-protection-privacy-and-electronic-communicationsamendments-etc-eu-exit-regulations-2019 [accessed March 2022].

D. G. Murray, F. McSherry, R. Isaacs, M. Isard, P. Barham, and M. Abadi, "Naiad: A Timely Dataflow System," in Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles, New York, NY, USA, 2013, pp. 439–455.

D. J. Patterson, L. Liao, D. Fox, and H. Kautz, "Inferring High-Level Behavior from Low-Level Sensors," in UbiComp 2003: Ubiquitous Computing, 2003, pp. 73–89.

D. Taibi, V. Lenarduzzi and C. Pahl, "Architectural patterns for microservices: a systematic mapping study," 2018.

Docker, "What container," 2020. [Online]. Available: https://www.docker.com/resources/what-container.

D. G. Neil MacDonald, "12 things to get right for successful devsecops," Gartner, December 2019. [Online]. Available: https://www.gartner.com/en/documents/3978490/12-things-to-get-right-forsuccessful-devsecops.

Devlin, J., Chang, M.-W., Lee K., Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), pp. 4171–4186.

D. J. Abadi et al., "The Design of the Borealis Stream Processing Engine," p. 13.

Dealroom.co – The state of European Insurtech 2021 https://dealroom.co/blog/the-state-of-european-insurtech-2021.

D. Yaga, P. Mell, N. Roby and K. Scarfone, Blockchain technology overview, National Institute of Standards and Technology, 2018.

D. Deuber, B. Magri and S. A. K. Thyagarajan, "Redactable Blockchain in the Permissionless Setting," 2019 IEEE Symposium on Security and Privacy (SP), pp. 124–138, 2019.

De Montjoye, Y.-A., Shmueli, E., Wang, S.S., and Pentland, A. (2014). OpenPDS: Protecting the privacy of metadata through SafeAnswers. PloS One, 9 (7), e98790.

Dubey, A., and Pentland, A. (2020a). Differentially-private federated linear bandits. In Proceedings of the Conference on Neural Information Processing Systems (NeurIPS 2020).

Dubey, A., and Pentland, A. (2020b). Kernel methods for cooperative multi-agent contextual bandits. In Proceedings of the 37th International Conference on Machine Learning (ICML), 119:2740–2750.

Dubey, A., and Pentland, A. (2020c). Private and byzantine-proof cooperative decision-making. In Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2020), 357–365.

D. S. E. E. a. N. S. W. Entriken, "EIP-721: ERC-721 non-fungible token standard," 2018. [Online]. Available: https://eips.ethereum.org/EIPS/eip-721. [Accessed 15 November 2021].

Deloitte, "Open banking – Privacy at the epicentre," Deloitte, 2018.

De Marchi, S., Schaback, R., Wendland, H. (2005). Near-optimal data-independent point locations for radial basis function interpolation, Adv. Comput. Math. 3, 23, 317–330.

D. DeWitt, A. Halverson, R. Nehme, S. Shankar, J. Aguilar-Saborit, A. Avanes, M. Flasza, J. Gramling, "Split query processing in Polybase", in ACM SIGMOD, pp. 1255–1266 (2013).

Dr. Ann Cavoukian, former Information and Privacy Commissioner of Ontario, Canada; for PbD see https://iapp.org/resources/article/privacy-by-design-the-7-foundational-principles/ (accessed in November 2020).

DBeaver, SQL UI client, https://dbeaver.io/.

Dimitrios Salampasis and Anne-Laure Mention (2021) Transformation Dynamics in FinTech: An Open Innovation Ecosystem Outlook. https://doi.org/10.1142/12355.

Dong-Hee Shin, Hongbum Kim, Junseok Hwang, Standardization revisited: A critical literature review on standards and innovation, Computer Standards & Interfaces, Volume 38, 2015, Pages 152–157, ISSN 0920-5489, https://doi.org/10.1016/j.csi.2014.09.002.

Directive (EU) 2015/849 of the European Parliament and of the Council of 20 May 2015 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, amending Regulation (EU) No 648/2012 of the European Parliament and of the Council, and repealing Directive 2005/60/EC of the European Parliament and of the Council and Commission Directive 2006/70/EC (Text with EEA relevance) 4th AML

Directive-4AMLD https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32015L0849 (Accessed: January 2022).

Directive (EU) 2018/843 of the European Parliament and of the Council of 30 May 2018 amending Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, and amending Directives 2009/138/EC and 2013/36/EU (Text with EEA relevance) Directive (EU) 2018/843 (Accessed: January 2022).

Debezium https://debezium.io/ (Accessed: January 2022).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

"DevTest and DevOps for microservices," Microsoft, 2020. [Online]. Available: https://docs.microsoft.com/en-us/azure/architecture/solution-ideas/articles/dev-test-microservice.

De Filippi, P., and McCarthy, S. (2012). Cloud computing: Centralization and data sovereignty- European Journal of Law and Technology, 3(2), available at SSRN: https://ssrn.com/abstract=2167372.

D. Kyriazis, et al, "BigDataStack: A Holistic Data-Driven Stack for Big Data Applications and Operations", BigData Congress, San Francisco, CA, USA, 2018: 237–241.

Document Object Identifier Accessible here: http://www.doi.org/handbook_2000/DOIHandbook-v4-4.1.pdf.

Exxact Corporation. (2020). PyTorch vs TensorFlow in 2020: What You Should Know About These Frameworks. Exxact. https://blog.exxactcorp.com/pytorch-vs-tensorflow-in-2020-what-you-should-know-about-these-frameworks/.

"EUR-Lex – 32014R0910 – EN – EUR-Lex." https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=uriserv:OJ.L_.2014.257.01.0073.01.ENG (accessed Sep. 14, 2020).

EU General Data Protection Regulation (EU-GDPR), https://www.privacyregulation.eu/en/index.htm [accessed March 2022].

European Commission. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Da. https://eur-lex.europa.eu/eli/reg/2016/679/oj.

European Commission, "Technology readiness levels (TRL" [Online]. Available: https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf.

EUR-Lex – 32014R0910 – EN – EUR-Lex.” https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=uriserv:OJ.L_.2014.257.01.0073.01.ENG (accessed Sep. 14, 2020).

European Commission (2020) On Artificial Intelligence – A European approach to excellence and trust (White paper), COM(2020) 65 final, available at https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

European Banking Authority (2020), EBA report on Big Data and Advanced Analytics, EBA/REP/2020/01, available at https://eba.europa.eu/sites/default/documents/files/document_library//Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf.

E. Union, “Official Journal of the European Union,” 2016. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN. [Accessed 05 November 2021].

E. &. Y. G. Limited., “How banks can balance GDPR and PSD2,” 2019. [Online]. Available: https://www.ey.com/en_lu/banking-capital-markets/how-banks-can-balance-gdpr-and-psd2. [Accessed 14 November 2021].

EY Global Wealth Research Report. EYG no. 003263-21Gbl. (2021) “Where will wealth take clients next?” Source: https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/wealth-and-asset-management/ey-2021-global-wealth-research-report.pdf.

EY – Global Consumer Banking Survey 2016 – EY – Global. (2016). http://www.ey.com/gl/en/industries/financial-services/banking---capital-markets/ey-global-consumer-banking-survey-2016.

Ernst & Young. 2017. EY FinTech Adoption Index: Fintech Services Poised for Mainstream Adoption in the US With 1 in 3 Digitally Active Consumers Using Fintech. Press release, June 28. https://www.ey.com/us/en/newsroom/news-releases/news-ey-fintech-adoptionindex.

Emily Porter, Chris Bopp, Elizabeth Gerber, and Amy Voida. (2017). Reappropriating Hackathons: The Production Work of the CHI4Good Day of Service. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, 810–814.

Erman, C., (2017). Financial technologies effect on financial services from an open innovation perspective.

ETSI GS CIM 009 V1.5.1 (2021-11) Context Information Management (CIM); NGSI-LD API https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.05.01_60/gs_CIM009v010501p.pdf.

Ethics guidelines for trustworthy AI. (2019, April 7). European Commission. Retrieved January 2022, from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

European Commission. (2020). White Paper on Artificial Intelligence: a European approach to excellence and trust (White Paper No. COM(2020) 65 final). European Commission. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approachexcellence-and-trust_en.

EBA report identifies key challenges in the roll out of Big Data and Advanced Analytics |European Banking Authority (europa.eu) https://www.eba.europa.eu/eba-report-identifies-key-challenges-roll-out-big-data-and-advanced-analytics (Accessed: January 2022).

Ethereum Magicians forum https://ethereum-magicians.org/ (Accessed: January 2022).

Ethereum Trust Alliance Https://Ethtrust.Org/ (Accessed: January 2022).

Ethereum Improvement Proposals, https://eips.ethereum.org/ (Accessed: January 2022).

European Blockchain Services Infrastructure (EBSI), https://ec.europa.eu/digital-single-market/en/european-blockchain-services-infrastructure (Accessed: January 2022).

EBSI Documentation, Ledger Protocols, https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/EBSI (Accessed: January 2022).

Einpresswire – Fintech market boosted by rising demand for digitization in organizations https://www.einnews.com/pr_news/549826454/fintech-market-boosted-by-rising-demand-for-digitization-in-organizations.

ESCP Business school repository https://www-emerald-com.revproxy.escpeurope.eu/insight/content/doi/10.1108/MF-02-2017-0028/full/pdf?title=electronic-finance-recent-developments.

European Commission (2020) On Artificial Intelligence – A European approach to excellence and trust (White paper), COM(2020) 65 final, available at https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligencefeb2020_en.pdf.

European Banking Authority (2020), EBA report on Big Data and Advanced Analytics, EBA/REP/2020/01, available at https://eba.europa.eu/sites/default/documents/files/document_library//Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf.

ESCP Business school repository https://web-a-ebscohost-com.revproxy.escpeurope.eu/ehost/pdfviewer/pdfviewer?vid=4&sid=a9e463e0-f2fd-42c2-bfad-5444579e0422%40sdc-v-sessmgr03.

Emerald Insight – Detecting money laundering transactions with machine learning https://www.emerald.com/insight/content/doi/10.1108/JMLC-07-2019-0055/full/html.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the econometric society, 987–1007.

"ETSI NGSI-LD specification," [Online]. Available: https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.01.01_60/gs_CIM009v010101p.pdf.

EY – Four ways COVID-19 is reshaping consumer banking behavior https://www.ey.com/en_kr/banking-capital-markets/four-ways-covid-19-is-reshaping-consumer-banking-behavior.

Electronic Privacy Information Center, Privacy and Human Rights 2005: An International Survey of Privacy Laws and Developments. 2007.

European Union Agency for Cybersecurity, "Pseudonymisation Data protection GDPR Pseudonymisation techniques and best practices," 2019, doi: 10.2824/247711.

ETSI, "ETSI TS 186 020 V3.1.1 – Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IMS-based IPTV interoperability test specification," Technical, 2011.

EICTA, "EICTA INTEROPERABILITY WHITE PAPER." 2004, [Online]. Available: http://www.agoria.be/www.wsc/webextra/prg/nwAttach?vWebSessionID=8284&vUserID=999999&appl=Agoriav2&enewsdetid=73029&attach=DescFile10.720108001095774198.pdf.

"EU data protection rules," European Commission, 2020. [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/eu-data-protection-rules_en. [Accessed 27 August 2020].

EPCglobal: Reader Protocol Standard, Version 1.1, 3 Ratified Standard, 4 June 21, 2006.

EPCglobal: The Application Level Events (ALE) Specification, Version 1.1.1 Part I: Core Specification, EPCglobal Ratified Standard, 13 March 2009.

EPCglobal: The EPCglobal Architecture Framework, EPCglobal Final Version 1.2 Approved 10 September 2007.

EPCglobal: EPC Information Services (EPCIS) Version 1.0.1 Specification Approved September 21, 2007.

EPCglobal: Reader Management Standard 1.0.1, 3 May 31, 2007.

"EU data protection rules," European Commission, 2020. [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/eu-data-protection-rules_en. [Accessed 27 August 2020].

"EIP-1155: ERC-1155 Multi Token Standard," Ethereum Improvement Proposals, 2020. [Online]. Available: https://eips.ethereum.org/EIPS/eip-1155. [Accessed 10 September 2020].

E. J. Keogh, K. Chakrabarti, M. J. Pazzani, and S. Mehrotra. Dimensionality reduction for fast similarity search in large time series databases. Knowledge and Information Systems (KAIS), 3(3):263–286, 2001.

"EIP-20: ERC-20 Token Standard," 2020. [Online]. Available: https://eips.ethereum.org/EIPS/eip20. [Accessed 1 September 2020].

E. B. Association, "B2B Data Sharing: Digital Consent Management as a Driver for Data Opportunities," Euro Banking Association, 2018.

E. Commission, "European Commission," 2015. [Online]. Available: https://ec.europa.eu/info/law/payment-services-psd-2-directive-eu-2015-2366_en. [Accessed 10 November 2021].

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 226–231.

ETSI GS CIM 009 V1.5.1 (2021-11) Context Information Management (CIM); NGSI-LD API https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.05.01_60/gs_CIM009v010501p.pdf.

Elliott, T. (2019). The State of the Octoverse: machine learning. GitHub. Retrieved 11 10, 2020, from https://github.blog/2019-01-24-the-state-of-the-octoverse-machine-learning/.

Edwards, L., & Veale, M. (2017). Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For. Duke Law & Technology Review, 16(18), 1–67. SSRN. dx.doi.org/10.2139/ssrn.2972855.

ECML-PKDD 2018 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, http://www.ecmlpkdd2018.org/wp-content/uploads/2018/09/567.pdf (http://www.ecmlpkdd2018.org/wp-content/uploads/2018/09/567.pdf).

«EC,» [Online]. Available: https://aws.amazon.com/ec2/instance-types/.

F. Vogelsteller and V. Buterin, "ERC-20 token standard," 2015. [Online]. Available: https://eips.ethereum.org/EIPS/eip-20 [Accessed 1-October-2020].

FAFT, Updated Guidance for a Risk-Based Approach to Virtual Assets and Virtual Asset Service Providers, https://www.fatf-gafi.org/publications/fatfrecommendations/documents/guidance-rba-virtual-assets-2021.html [accessed March 2022].

Feizi, Andisheh & Wong, Chui Yin. (2012). Usability of user interface styles for learning a graphical software application. 1089–1094. 10.1109/ICCISci.2012.6297188.

Final Report: Guidelines on the security measures for operational and security risks of payment services under Directive (EU) 2015/2366 (PSD2), European Banking Authority, Dec 2017.

Ferrari Dacrema, M., Cremonesi, P., & Jannach, D. (2019). Are we really making much progress? A worrying analysis of recent neural recommendation approaches. RecSys '19: Proceedings of the 13th ACM Conference on Recommender Systems, 101–109. ACM Digital Library. https://doi.org/10.1145/3298689.3347058.

Filice, S., Castellucci, G., Da San Martino, G., Moschitti, A., Croce, D., & Basili, R. (2018). KELP: a Kernel-based Learning Platform. Journal of Machine Learning Research 18(191):1–5.

Financial Stability Board (2017), Artificial intelligence and machine learning in financial services, www.fsb.org/wp-content/uploads/P011117.pdf.

Fine, S., Singer, Y., & Tishby, N. (1998). The hierarchical hidden Markov model: Analysis and applications. Machine Learning, 32, 41–62.

Fionda, V., & Pirró, G. (2019). triple2Vec: Learning Triple Embeddings from Knowledge Graphs. arxiv Preprint. arXiv. https://arxiv.org/abs/1905.11691.

Fortuna, B., Rupnik, J., Brank, J., Fortuna, C., Jovanoski, V., Karlovcec, M., Kazic, B., Kenda, K., Leban, G., Muhic, A., Novak, B., Novlian, J., Papler, M., Rei, L., Sovdat, B., Stopar, L., Grobelnik, M., & Mladenic, D. (2014). QMiner: Data Analytics Platform for Processing Streams of Structured and Unstructured Data. In Proceedings of the Software Engineering for Machine Learning Workshop at Neural Information Processing Systems (NIPS 2014).

Fasnacht, Daniel. (2018) "Open innovation ecosystems." Open Innovation Ecosystems. Springer, Cham, 2018. 131–172.

Fasnacht, Daniel. (2020) "The Ecosystem Strategy: Disruptive Business Model Innovation." Zeitschrift Führung und Organisation (zfo) 89.3: 168–173.

Florea, I. C., Vochin, O. A., Ciachir, L., & NAGEL-PICIORUS, C. P. (2017). Competition for Innovation in the Financial Software Industry-A Research on Hackathons. In BASIG INTERNATIONAL CONFERENCE: NEW TRENDS IN SUSTAINABLE BUSINESS AND CONSUMPTION (Vol. 2017, pp. 234–242).

Feldmann, A., & Teuteberg, F. (2020). Success factors for hackathons: German banks collaborate to tame the economic crisis. Journal of Business Strategy.

Flores, M. et al. (2018). How Can Hackathons Accelerate Corporate Innovation? In: Moon, I. et al. (eds.) Advances in Production Management Systems. Production Management for Data Driven, Intelligent, Collaborative, and Sustainable Manufacturing. pp. 167–175 Springer International Publishing, Cham.

Fasnacht, Daniel. (2019). "Open Innovation Ecosystem: The Winner Takes It All." In: Open Innovation Ecosystems. Management for Professionals. Springer, Cham. DOI 10 (2019): 978-3.

FINSEC Reference Architecture II – Available at https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ce3a941d&appId=PPGMS.

FBK, " D5.9 – Library of ML/DL Algorithms – III," p. 12.

FATF. (2019). Guidance for a risk-based approach, virtual assets and virtual asset service providers, FATF, 2019, Paris. https://www.fatf-gafi.org/publications/fatfrecommendations/documents/guidance-rba-virtual-assets.html.

"FIWARE," [Online]. Available: https://www.fiware.org/developers/.

F. Montesi and J. Weber, "Circuit breakers, discovery, and API gateways in microservices," 2016.

"FIWARE Vehicle Data model," [Online]. Available: https://fiwaredatamodels.readthedocs.io/en/latest/Transportation/Vehicle/Vehicle/doc/spec/index.html.

"FIWARE Alert Data Model," [Online]. Available: https://fiwaredatamodels.readthedocs.io/en/latest/Alert/doc/spec/index.html.

"FIWARE Weather Observed Data model," [Online]. Available: https://fiwaredatamodels.readthedocs.io/en/latest/Weather/WeatherObserved/doc/spec/index.html.

F. Khan, N. Akhtar and M. A. Qadeer, "RFID Enhancement in Road Traffic Analysis by Augmenting Reciever with TelegraphCQ," 2009 Second International Workshop on Knowledge Discovery and Data Mining, Moscow, 2009, pp. 331–334.

F. Moutinho, L. Paiva, J. Köpke, and P. Maló, "Extended Semantic Annotations for Generating Translators in the Arrowhead Framework," IEEE Transactions on Industrial Informatics, vol. 14, no. 6, pp. 2760–2769, Jun. 2018, doi: 10.1109/TII.2017.2780887.

F. V. a. V. Buterin, "ERC-20 token standard," 2015. [Online]. Available: https://eips.ethereum.org/EIPS/eip-20. [Accessed 19 November 2021].

Farrugia, S., Ellul, J., & Azzopardi, G. (2020). Detection of illicit accounts over the Ethereum blockchain. Expert systems with applications 150: 113318.

FATF. Guidance for a risk based approach, virtual assets and virtual asset service providers, FATF, paris. https://www.fatf-gafi.org/publications/fatfrecommendations/documents/guidance-rba-virtual-assets.html, 2019.

Frisendal, T. (2012). Design thinking for business analysis. In Design Thinking Business Analysis (pp. 15–24). Springer, Berlin, Heidelberg.

FIWARE Foundation, "Smart Data Models," https://github.com/smart-data-models (Accessed: January 2022).

FIWARE Foundation, "Smart Data Models," https://smartdatamodels.org/ (Accessed: January 2022).

FIWARE Foundation, "Smart Data Models," https://github.com/smart-data-models/dataModel.Transportation (Accessed: January 2022).

FIWARE Foundation, "Smart Data Models," https://github.com/smart-data-models/dataModel.Transportation/tree/master/Vehicle (Accessed: January 2022).

FIWARE Foundation, "Smart Data Models," https://github.com/smart-data-models/dataModel.Weather/tree/master/WeatherObserved (Accessed: January 2022).

FIWARE Foundation, "Smart Data Models," https://github.com/smart-data-models/dataModel.Transportation/tree/master/Road (Accessed: January 2022).

FIWARE Foundation, "Smart Data Models," https://github.com/smart-data-models/dataModel.Transportation/tree/master/RoadSegment (Accessed: January 2022).

FIWARE Foundation, "Smart Data Models," https://github.com/smart-data-models/dataModel.Alert/tree/master/Alert (Accessed: January 2022).

FIWARE Foundation, "Smart Data Models," https://github.com/smart-data-models, 2021. https://github.com/smart-data-models/dataModel.Transportation/tree/master/RoadAccident (Accessed: January 2022).

FIWARE Foundation, "Smart Data Models," https://github.com/smart-data-models/dataModel.Transportation/tree/master/RestrictedTrafficArea (Accessed: January 2022).

"FlinkCEP.", https://ci.apache.org/projects/flink/flink-docs-release-1.11/dev/libs/cep.html.

Gareth, J., Witten, D., Hastie, T., & Robert Tibshirani. (2013). An Introduction to Statistical Learning. Springer-Verlag New York. 10.1007/978-1-4614-7138-7.

Goldberger, J., Hinton, G., Roweis, S., & Salakhutdinov, R. R. (2004). Neighbourhood Components Analysis. Advances in neural information processing systems, 17(1), 513–520. https://proceedings.neurips.cc/paper/2004/file/42fe880812925e520249e808937738d2-Paper.pdf.

GDPR – IT Governance, https://www.itgovernance.co.uk/data-privacy/gdpr-overview/gdpr-faq/gdpr-scope (accessed Nov 2020).

GDPR scope – IT Governance, https://www.itgovernance.co.uk/data-privacy/gdpr-overview/gdprfaq/gdpr-scope (accessed Nov 2020).

G. Wondracek, T. Holz, E. Kirda, and C. Kruegel, "A Practical Attack to De-anonymize Social Network Users," in 2010 IEEE Symposium on Security and Privacy, May 2010, pp. 223–238.

G. Zhong and U. Hengartner, "A distributed k-anonymity protocol for location privacy," 2009 IEEE International Conference on Pervasive Computing and Communications. 2009, doi: 10.1109/percom.2009.4912774.

G. Wondracek, T. Holz, E. Kirda, and C. Kruegel, "A Practical Attack to De-anonymize Social Network Users," in 2010 IEEE Symposium on Security and Privacy, May 2010, pp. 223–238.

Growth Share Matrix https://fourweekmba.com/bcg-matrix/.

Gartner, "DevOps Gartener Glossary," Gartner, 2020. [Online]. Available: https://www.gartner.com/en/information-technology/glossary/devops.

Git, "Git," Software Freedom Conservancy, 2020. [Online]. Available: https://git-scm.com/about.

GitLab, "Gitlab docs," GitLab, 2020. [Online]. Available: https://docs.gitlab.com/charts/installation/deployment.html.

Google, "MLOps: Continuous delivery and automation pipelines in machine learning," Google, 2019. [Online]. Available: https://cloud.google.com/solutions/machine-learning/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning.

Google, "Kubernetes-Kubectl," [Online]. Available: https://kubernetes.io/docs/tasks/tools/install-kubectl/.

Google, "Containers," Google, 2018. [Online]. Available: https://cloud.google.com/containers.

G. A. Stevens, B. Magri, D. Venturi and E. Andrade, "Redactable Blockchain – or – Rewriting History in Bitcoin and Friends," 2017 IEEE European Symposium on Security and Privacy (EuroS&P), pp. 111–126, 2017.

GDPR Recital (1).

Gaur, N., Desrosiers, L., Ramakrishna, V., Novotny, P., Baset, S.A., and O'Dowd, A. (2018). Hands-on Blockchain with Hyperledger: Building decentralized applications with Hyperledger Fabric and Composer. Packt Publishing.

GitLab homepage, GitLab, [Online]. Available: https://about.gitlab.com/.

Guidelines FIWARE data models https://fiware-datamodels.readthedocs.io/en/latest/guidelines/index.html.

GENE Ontology – bioinformatics initiative Accessible here: http://www.geneontology.org.

G. D. Orio, P. Maló, J. Barata, M. Albano, and L. L. Ferreira, "Towards a Framework for Interoperable and Interconnected CPS-populated Systems for Proactive Maintenance," in 2018 IEEE 16th.

G. Wood, "Ethereum: A Secure Decentralised Generalised Transaction Ledger, Ethereum Project Yellow Paper", 2018. [Online] Available at: https://ethereum.github.io/yellowpaper/paper.pdf. [Accessed: 20- October-2020].

Grafana, [Online]. Available: https://grafana.com/.

H. Hacigümüs, J. Sankaranarayanan, J. Tatemura, J. LeFevre, N. Polyzotis, "Odyssey: a multistore system for evolutionary analytics", PVLDB, vol. 6, pp. 1180–1181 (2013).

H. S. Pinto, S. Staab, and C. Tempich, "DILIGENT: Towards a fine-grained methodology for DIstributed, Loosely-controlled and evolvInG Engineering of oNTologies," in Proceedings of the 16th European Conference on Artificial Intelligence, 2004, pp. 393–397.

Heitmann, B., Kinsella, S., Hayes, C. and Decker, S. Implementing Semantic Web Applications: Reference Architecture and Challenges. In International Workshop on Semantic Web enabled Software Engineering, collocated with the 8th International Semantic Web Conference (ISWC2009), 2009.

HELM, "Quickstart," HELM, 2020. [Online]. Available: https://helm.sh/docs/intro/quickstart/.

Henson, C. A., Pschorr, J. K., Sheth, A. P. and Thirunarayan, K. SemSOS: Semantic sensor observation service. Collaborative Technologies and Systems, International Symposium on, 0:44–53, 2009.

H. van der Veer and A. Wiles, "Achieving technical interoperability," European Telecommunications Standards Institute, 2008.

He, H. (2019, 10 10). The State of Machine Learning Frameworks in 2019. The Gradient. https://thegradient.pub/state-of-ml-frameworks-2019-pytorch-dominates-research-tensorflow-dominates-industry/.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T.-S. (2017). Neural Collaborative Filtering. WWW '17: Proceedings of the 26th International Conference on World Wide Web, 173–182. ACM Digital Library. https://doi.org/10.1145/3038912.3052569.

Ho, T. K. (1995). Random decision forests (Proceedings of 3rd International Conference on Document Analysis and Recognition ed.). IEEE. doi.org/10.1109/ICDAR.1995.598994.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8). ACM Digital Library. doi.org/10.1162/neco.1997.9.8.1735.

Holton, G. A. (2014). Value-at-Risk: Theory and Practice (2nd ed.). London: Academic Press.

Hardjono, T., and Pentland, A. (2017). Open algorithms for identity federation. arXiv:1705.10880.

Hardjono, T., and Pentland, A. (2019). Data Cooperatives: Towards a foundation for decentralized personal data management. arXiv: 1905.08819.

Hu, V., Ferraiolo, D., Kuhn, D., Friedman, A., Lang, A., Cogdell, M.M., Schnitzer, A., Sandlin, K., Miller, R., and Scarfone, K. (2014). Attribute-Based Access Control Guide to Attribute Based Access Control (ABAC) Definition and Considerations. National Institute of Standards and Technology, NIST SP 800-162.

H. Xu, M. Chen, Y. Zhou, B. Du and L. Pan, "A Novel Comprehensive Quality Index QoX and the Corresponding Context-aware System Framework," 2018 IEEE 4th International Conference on Computer and Communications (ICCC), Chengdu, China, 2018, pp. 2415–2419.

Hu, V., Kuhn, D., and Ferraiolo, D. (2015). Attribute-based access control. Comput. J. 48, 864–866. doi: 10.1109/MC.2015.33.

Harvard Business Review – You need an innovation strategy https://www.homeworkforyou.com/static_media/uploadedfiles/You%20Need%20an%20Innovation%20Strategy.pdf.

Heikki Topi and Allen Tucker. (2014). Computing Handbook, Third Edition: Information Systems and Information Technology. CRC Press.

High-Level Expert Group on Artificial Intelligence. (2018, June 14). European Commission. Retrieved January 2022, from https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligenc.

Huxley – The most in-demand skills for Fintech https://www.huxley.com/en-ae/blog/2020/04/the-most-in-demand-skills-for-fintech-covid/.

"Hyperledger Fabric – Hyperledger," 2020. [Online]. Available: https://www.hyperledger.org/use/fabric. [Accessed 5 September 2020].

Hong, Song-You, Yign Noh, and Jimy Dudhia. "A new vertical diffusion package with an explicit treatment of entrainment processes." Monthly weather review 134.9 (2006): 2318–2341.

H. Surendra and H. Mohan, "A review of synthetic data generation methods for privacy preserving data publishing," International Journal of Scientific & Technology Research, vol. 6, no. 3, pp. 95–101, 2017.

HAproxy, "HAproxy," HAproxy, [Online]. Available: http://www.haproxy.org/#desc.

Hashicorp, "Introduction to Terraform," Hashicorp, [Online]. Available: terraform.io/intro/index.html.

HELM, [Online]. Available: https://helm.sh/.

HCL, [Online]. Available: https://www.terraform.io/docs/configuration-0-11/syntax.html.

Harbor homepage, Linux Foundation, [Online]. Available: https://goharbor.io/.

Hubble, [Online]. Available: https://docs.cilium.io/en/v1.9/intro/#intro.

https://www.redflagalert.com/articles/data/what-is-the-fourth-aml-directive-aml4d (accessed in December 2020).

https://www.loc.gov/law/help/artificial-intelligence/regulation-artificial-intelligence.pdf (accessed in December 2020).

https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework (accessed in December 2020).

https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf (accessed in December 2020).

https://docs.docker.com/get-started/overview/, Docker. [Online]. [Accessed 24 03 2021].

https://www.oreilly.com/library/view/software-architecture-with/9781786468529/ch08s04.html.

http://www.dossier-andreas.net/software_architecture/index.html.

http://www.se.rit.edu/~emad/teaching/slides/CISC322_06_ArchitectureStyles_sep20.pdf.

https://sarasanalytics.com/blog/data-pipeline-architecture.

https://www.astera.com/type/blog/data-pipeline-architecture.

https://airflow.apache.org/docs/apache-airflow/stable/index.html.

https://luigi.readthedocs.io/en/stable.

https://www.knime.com/.

https://c3.ai/products/c3-ai-ex-machina/.

https://pipeline.loni.usc.edu/learn/user-guide/building-a-workflow/.

https://streampipes.apache.org/.

https://advisera.com/27001academy/what-is-iso-27001/, page 1 (accessed in December 2020).

https://fcicyber.com/top-5-ways-the-financial-services-industry-can-leverage-nist-for-cybersecuritycompliance/ (accessed in December 2020).

https://financialservices.mazars.com/gdpr-psd2-issues-fintechs/ (accessed in December 2020).

https://www.thalesgroup.com/en/markets/digital-identity-and-security/banking-payment/digitalbanking/psd2 (accessed in December 2020).

https://eba.europa.eu/sites/default/documents/files/documents/10180/2060117/d53bf08f-990b-47ba-b36f-15c985064d47/Final%20report%20on%20EBA%20Guidelines%20on%20the%20security%20measures%20for%20operational%20and%20security%20risks%20under%20PSD2%20(EBA-GL-2017-17).pdf (accessed in December 2020).

https://www.thalesgroup.com/en/markets/digital-identity-and-security/banking-payment/digitalbanking/psd2 (accessed in December 2020).

https://www.investopedia.com/terms/m/mifid-ii.asp (accessed in December 2020).

https://www.redflagalert.com/articles/data/what-is-the-fourth-aml-directive-aml4d (accessed in December 2020).

https://www.loc.gov/law/help/artificial-intelligence/regulation-artificial-intelligence.pdf (accessed in December 2020).

https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework (accessed in December 2020).

https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf (accessed in December 2020).

https://www.fiware.org/developers/smart-data-models/.

https://fiware-orion.readthedocs.io/en/master/.

https://fiware.github.io/specifications/ngsiv2/stable/.
https://quantumleap.readthedocs.io/en/latest/.
https://app.swaggerhub.com/apis/smartsdk/ngsi-tsdb/0.1.
https://konghq.com/kong/.
https://sumo.dlr.de/docs/index.html.
https://opendata.aemet.es/.
https://cgiarcsi.community/data/srtm-90m-digital-elevation-database-v4-1/.
https://soilgrids.org/#!/?layer=TAXNWRB_250m&vector=1.
https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs.
http://crowdpolicy.com/.
https://www.standardsplusinnovation.eu/, retrieved 12/12/2020.
https://databox.com/new-vs-returning-visitors.
https://ec.europa.eu/info/publications/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics-0_en.
https://www.eventbrite.com/blog/asset/ultimate-way-reduce-no-shows-free-events/.
https://ec.europa.eu/digital-single-market/en/news/ict-verticals-and-horizontals-blockchain-standardisatio.
https://marketinginsidergroup.com/content-marketing/why-social-media-is-important-for-business-marketing/.
https://www.webfx.com/internet-marketing/why-is-social-media-so-important.
https://www.campaignmonitor.com/resources/guides/why-email/.
https://www.campaignmonitor.com/blog/email-marketing/8-epic-email-newsletters-you-can-learn-from/.
https://v12data.com/blog/top-strategies-to-personalize-your-email-campaigns/&sa=D&source=editors&ust=1623772876236000&usg=AOvVaw3QVzfE7H7Qh5tpWHBVMOum.
https://thenextscoop.com/company-host-hackathon/.
https://technology.finra.org/articles/why-enterprises-should-host-hackathons.html.
https://www.livechat.com/success/company-can-benefit-from-hackathons/.
https://digitalagencynetwork.com/webinar-marketing-strategy-how-to-promote-your-upcoming-webinar/.
https://www.meraevents.com/blog/benefits-of-attending-workshops-for-professionals-businessowners.
https://yellowdoorcollective.com/blog/3-reasons-why-hosting-a-workshop-will-add-value-toyour-business/.
https://marketplace.infinitech-h2020.eu/workshops/stakeholders-workshops-series-1.

https://digital-strategy.ec.europa.eu/en/activities/edihs.

https://advisera.com/27001academy/what-is-iso-27001/, page 1 (accessed in December 2020).

https://fcicyber.com/top-5-ways-the-financial-services-industry-can-leverage-nist-for-cybersecurity-compliance/ (accessed in December 2020).

https://financialservices.mazars.com/gdpr-psd2-issues-fintechs/ (accessed in December 2020).

https://www.thalesgroup.com/en/markets/digital-identity-and-security/banking-payment/digital-banking/psd2 (accessed in December 2020).

https://eba.europa.eu/sites/default/documents/files/documents/10180/2060117/d53bf08f-990b-47bab36f-15c985064d47/Final%20report%20on%20EBA%20Guidelines%20on%20the%20security%20measures%20for%20operational%20and%20security%20risks%20under%20PSD2%20(EBA-GL-2017-17).pdf (accessed in December 2020).

https://www.thalesgroup.com/en/markets/digital-identity-and-security/banking-payment/digital-banking/psd2 (accessed in December 2020).

https://www.investopedia.com/terms/m/mifid-ii.asp (accessed in December 2020).

https://rancher.com/docs/rancher/v2.5/en/installation/requirements/#hardware-requirem.

https://www.abe-eba.eu/thought-leadership-innovation/open-banking-working-group/management-summary-artificial-intelligence-in-the-era-of-open-banking/.

International Organization for Standardization, "ISO/IEC 29115:2013 Information technology – Security techniques – Entity authentication assurance framework." 2013, [Online]. Available: https://www.iso.org/standard/45138.html.

IBM – Security, Policy and Compliance, https://www.ibm.com/cloud/garage/architectures/securityArchitecture/security-policy-governance-risk-compliance [accessed March 2022].

ICAO, "ICAO Document 9303 – Machine Readable Travel Documents." [Online]. Available: https://www.icao.int/publications/pages/publication.aspx?docnum=9303. [Accessed: 07-Jul-2021].

I.T. Gartner, "Glossary (2020),'Data governance,' " URL: https://www.gartner.com/en/information-technology/glossary/data-governance.

Innovation Project Portfolio Management: A Meta-Analysis, June 2012, International Journal of Product Development: 77–94. https://www.researchgate.net/publication/259557845_Innovation_Project_Portfolio_Management_A_Meta-Analysis.

Innovation management techniques and tools: a review from theory and practice, Antonio Hidalgo and Jose Albors Department of Business Administration.

Universidad Politecnica de Madrid, c/Jose Gutierrez https://oa.upm.es/2406/1/INVE_MEM_2008_55568.pdf.

IBM, "Containerization," [Online]. Available: https://www.ibm.com/cloud/learn/containerization. [Accessed 24 3 2021].

IBM, "Containers," Google, 2018. [Online]. Available: https://www.ibm.com/cloud/learn/containerization. [Accessed 24 03 2021].

Intellectual Property Rights https://www.researchgate.net/figure/Intellectual-PropertyReadiness-Level-Part-of-TRL-Hasenauer-et-al-Managing_fig4_313063121.

IBM Institute for Business Value, Moving to a token-driven economy Enabling the digitization of real-world assets, 2018. [Online] Available at: https://www.ibm.com/downloads/cas/YMRKPOJ8 [Accessed 15-October-2020].

IBM digital transformation https://www.ibm.com/thought-leadership/institute-business-value/report/digital-transformation#.

Istio, "What is Istio," Istio, 2020. [Online]. Available: https://istio.io/latest/docs/concepts/what-is-istio/.

International Conference on Industrial Informatics (INDIN), Jul. 2018, pp. 146–151, doi: 10.1109/INDIN.2018.8472041.

"Introduction — hyperledger-fabricdocs master documentation," Hyperledger, 2020. [Online]. Available: https://hyperledger-fabric.readthedocs.io/en/release-2.2/whatis.html. [Accessed 03 September 2020].

I. Karagiannis, K. Mavrogiannis, J. Soldatos, D. Drakoulis, E. Troiano and A. Polyviou, "Blockchain Based Sharing of Security Information for Critical Infrastructures of the Finance Sector," Computer Security Lecture Notes in Computer Science, Computer Security. IOSEC 2019, MSTEC 2019, FINSEC 2019, vol. 11981, pp. 226–241, 2020.

Iacono, Michael J., et al. "Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models." Journal of Geophysical Research: Atmospheres113.D13 (2008).

IEEE. (2019). ETHICALLY ALIGNED DESIGN A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems First Edition. Retrieved January 2022, from https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf.

IEEE Blockchain Standards, https://blockchain.ieee.org/standards (Accessed: January 2022).

IEEE Blockchain Standards, https://blockchain.ieee.org/standards (Accessed: January 2022).

IEEE Big Data Governance and Metadata Management (BDGMM) https://standards.ieee.org/industry-connections/BDGMM-index.html Accessed: January 2022).

I. Cambridge Semantics, "Leveraging Semantic and Graph Technology to Tame the Enterprise Data Storm." combridgesemantics.com, Jun. 2019, Available Online: https://blog.cambridgesemantics.com/.

International Organization for Standardization. https://www.iso.org/home.html (Accessed: January 2022).

ISO – ISO/IEC JTC 1 – Information Technology. ISO. Retrieved January, 2022 from https://www.iso.org/isoiec-jtc-1.html.

Information security, cybersecurity and privacy protection. JTC 1. Retrieved January, 2022 from https://jtc1info.org/technology/subcommittees/information-security-cybersecurity-privacy-protection/.

Imerman, M. B., & Fabozzi, F. J. (2020). A Conceptual Framework for FinTech Innovation. Available at SSRN 3543810.

ISO/IEC Guide 59:2019(en): ISO and IEC recommended practices for standardization by national bodies.

i-SCOOP – Digital transformation strategy: the bridges to build https://www.i-scoop.eu/digitaltransformation/digital-transformation-strategy/.

IETF – Uniform Resource Names Accessible here: http://tools.ietf.org/html/rfc2141 [W3C-RDF] World Wide Web Consortium – Resource Description Framework, Accessible here: http://www.w3.org/TR/rdf-syntax-grammar/.

J. Dumortier, "REGULATION (EU) NO 910/2014 ON ELECTRONIC IDENTIFICATION AND TRUST SERVICES FOR ELECTRONIC TRANSACTIONS IN THE INTERNAL MARKET (EIDAS REGULATION)," in EU Regulation of E-Commerce, Edward Elgar Publishing, 2017.

J. Froehlich and J. Krumm, "Route prediction from trip observations, Soc," Automot. Eng. Spec. Publ, vol. 2193, p. 53, 2008.

J. Krumm and E. Horvitz, "Predestination: Inferring Destinations from Partial Trajectories," in UbiComp 2006: Ubiquitous Computing, 2006, pp. 243–260.

J. Duggan, A. J. Elmore, M. Stonebraker, M. Balazinska, B. Howe, J. Kepner, S. Madden, D. Maier, T. Mattson, S. Zdonik, "The BigDAWG polystore system", SIGMOD Record, vol. 44, no. 2, pp. 11–16, 2015.

J. LeFevre, J. Sankaranarayanan, H. Hacıgümüs, J. Tatemura, N. Polyzotis, M. Carey, "MISO: souping up big data query processing with a multistore system", in ACM SIGMOD, pp. 1591–1602 (2014).

J. Wang, T. Baker, M. Balazinska, D. Halperin, B. Haynes, B. Howe, D. Hutchison, S. Jain, R. Maas, P. Mehta, D. Moritz, B. Myers, J. Ortiz, D. Suciu, A. Whitaker, S. Xu, "The Myria big data management and analytics system and cloud service", in Conference on Innovative Data Systems Research (CIDR) (2017).

J. Krumm, "Inference Attacks on Location Tracks," in Pervasive Computing, 2007, pp. 127–143.

J. Köpke, "Annotation paths for matching xml-schemas," Data & Knowledge Engineering, vol. 122, pp. 25–54, 2019.

J. Köpke and J. Eder, "Semantic annotation of XML-schema for document transformations," in OTM Confederated International Conferences "On the Move to Meaningful Internet Systems", 2010, pp. 219–228.

J. Dafflon, J. Baylina, and T. Shababi, "ERC-777 token standard," 2015. [Online]. Available: https://eips.ethereum.org/EIPS/eip-777 [Accessed 10-October-2020].

Jolliffe, I.T. (2002). Principal Component Analysis, second edition, New York: Springer-Verlag New York, Inc.

Jiménez, Pedro A., et al. "A revised scheme for the WRF surface layer formulation." Monthly Weather Review 140.3 (2012): 898–918.

«Jenkins homepage,» [Online]. Available: https://www.jenkins.io/.

J. Wang, Y. Luo, Y. Zhao, and J. Le, "A Survey on Privacy Preserving Data Mining," in 2009 First International Workshop on Database Technology and Applications, Apr. 2009, pp. 111–114.

Jacobs, I. and Walsh, N. Architecture of the World Wide Web, Volume One, World Wide Web Consortium, Recommendation REC-webarch-20041215, 2004.

J.Y. Lee, "A decentralized token economy: How blockchain and cryptocurrency can revolutionize business", Business Horizons (2019). [Online]. Available at: www.sciencedirect.com [Accessed: 10- October-2020].

Jenkins, "Jenkins," [Online]. Available: https://jenkins.io/.

Joining forces for blockchain standardisation https://europa.eu/newsroom/events/joining-forces-blockchain-standardisation_en (Accessed: January 2022).

James Moar: "Three Trends Accelerating the Growth of Digital Identity", Juniper Research, July 2019.

Jeff Galvin et al., "Synergy and disruption: Ten trends shaping Fintech", McKinsey & CO, December 2018.

Keycloak – Authorization Services Guide, https://www.keycloak.org/docs/latest/authorization_services/#authorization-services, [Accessed: March 2022].

KIE API, https://docs.jboss.org/drools/release/6.3.0.Final/kie-api-javadoc/overview-summary.html, [accessed March 2022].

K. K. Petros Kavassalis, "D4.3 Operational and Technical Documentation of SP (ATHEX, Hellenic Post) integration (production)." 2018, [Online]. Available: http://www.lepsproject.eu/sites/default/files/leps/public/contentfiles/deliverables/LEPS%20D4.3%20Operational%20and%20Technical%20Documentation%20of%20SP%20integration.pdf.

K. Mivule, "Utilizing Noise Addition for Data Privacy, an Overview," arXiv [cs.CR], Sep. 16, 2013.

KPMG – Insurtech 10: Trends for 2019 https://assets.kpmg/content/dam/kpmg/xx/pdf/2019/03/insurtech-trends-2019.pdf.

K. Mivule, "Utilizing Noise Addition for Data Privacy, an Overview," arXiv [cs.CR], Sep. 16, 2013.

Kubernetes, "What is kubernetes?," Kubernetes, [Online]. Available: https://kubernetes.io/docs/concepts/overview/what-is-kubernetes/. [Accessed 24 03 2021].

K. Philippe, "Architectural Blueprints – The "4+1" View Model of Software Architecture", IEEE Software, vol. 12, pp. 42–50, November 1995.

Kubernetes, "What is kubernetes?," Kubernetes, 5 August 2020. [Online]. Available: https://kubernetes.io/docs/concepts/overview/what-is-kubernetes/.

Kubernetes – Namespaces, [Online]. Available: https://kubernetes.io/docs/concepts/overview/workingwith-objects/namespaces/.

kfp kubeflow sdk, [Online]. Available: https://www.kubeflow.org/docs/components/pipelines/sdk/sdk-overview/.

kfp.dsl package, [Online]. Available: https://kubeflowpipelines.readthedocs.io/en/latest/source/kfp.dsl.html.

K. K. Framework. [Online]. Available: https://www.kubeflow.org/docs/external-add-ons/kserve/kserve/.

Kubernetes, [Online]. Available: https://kubernetes.io/. [Accessed 31 05 2021].

Kubeflow, "About Kubeflow," Kubeflow, 2020. [Online]. Available: https://www.kubeflow.org/docs/about/kubeflow/.

K. Awada, M. Eltabakh, C. Tang, M. Al-Kateb, S. Nair, G. Au, "Cost Estimation Across Heterogeneous SQL-Based Big Data Infrastructures in Teradata IntelliSphere", in EDBT, pp. 534–545 (2020).

Kılıç, B., Özturan, C., & Şen, A. (2021). Finding convex subgraphs in Blockchain transaction graphs, submitted to BCCA 2021 conference (under review).

K. I. P. L. D. V. C. J. P. K. A. Bhaskaran, "Double-blind consent-driven data sharing on blockchain.," in 2018 IEEE International Conference on Cloud Engineering (IC2E), 2018.

Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R.G.L., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P.B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konecný, J., Korolova, A, Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Raykova, M., Qi, H., Ramage, D., Raskar, R., Song, D., Song, W., Stich, S.U., Sun, Z., Suresh, A.T., Tramér, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F.X., Yu, H., and Zhao, S. (2019). Toward trustworthy AI development: Mechanisms for supporting verifiable claims. arXiv:1912.04977.

K. Tuononen, "The impact of PSD2 directive on the financial services industry.," 2019.

KPMG, "KPMG," 2021. [Online]. Available: https://home.kpmg/ph/en/home/insights/2019/07/openbanking-opens-opportunities-for-greater-value.html. [Accessed 15 November 2021].

Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R.G.L., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P.B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konecný, J., Korolova, A, Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Raykova, M., Qi, H., Ramage, D., Raskar, R., Song, D., Song, W., Stich, S.U., Sun, Z., Suresh, A.T., Tramér, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F.X., Yu, H., and Zhao, S. (2021). Advances and open problems in federated learning. Foundations and trends in machine learning, 14 (1–2), pp. 1–210.

KPMG, "Open banking opens opportunities for greater customer," KPMG, 2021.

Kruchten, Philippe (1995, November). Architectural Blueprints – The "4+1" View Model of Software Architecture. IEEE Software 12 (6), pp. 42–50.

Komulainen, H., Makkonen, H. Customer experience in omni-channel banking services. J Financ Serv Mark 23, 190–199 (2018). https://doi.org/10.1057/s41264-018-0057-6.

Kumar, N., Singh, A., Handa, A., & Shukla, S.K. (2020). Detecting malicious accounts on the Ethereum blockchain with supervised learning. In Proceedings of International Symposium on Cyber Security Cryptography and Machine Learning, pp. 94–109. Springer, Cham.

K. Chan and A. W. Fu. E cient time series matching by wavelets. In Proceedings of the International Conference on Data Engineering (ICDE), pages 126–133. IEEE Computer Society, 1999.

K. W. Ong, Y. Papakonstantinou, and R. Vernoux, "The SQL++ semi-structured data model and query language: a capabilities survey of SQL-on-Hadoop, NoSQL and NewSQL databases", CoRR, abs/1405.3631, 2014.

Kubernetes IP management https://kubernetes.io/docs/concepts/services-networking/service/#choosing-your-own-ip-address.

Kubernetes. [Online] https://kubernetes.io.

Kubernetes IP management [Online]. Available: https://kubernetes.io/docs/concepts/servicesnetworking/service/#choosing-your-own-ip-address.

Konstantinos Chatzikokolakis, Ehab Elsalamouny, Catuscia Palamidessi. Efficient Utility Improve-ment for Location Privacy. Proceedings on Privacy Enhancing Technologies, De Gruyter Open, 2017, 2017(4), pp. 308–328. 10.1515/popets-2017-0051.hal-01422842v2.

Kathleen Brandenburg. Design as Business Strategy: Key Principles of Successful Design Thinking https://iacollaborative.com/perspective/design-as-business-strategy-key-principles-of-successful-design-thinking/.

L. Haas, D. Kossmann, E. Wimmers, J. Yang. Optimizing Queries across Diverse Data Sources. Int. Conf. on Very Large Databases (VLDB), pp. 276–285 (1997).

Lundberg, S.M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), pp. 4768–4777.

L. Neumeyer, B. Robbins, A. Nair, and A. Kesari, “S4: Distributed Stream Computing Platform,” in 2010 IEEE International Conference on Data Mining Workshops, 2010, pp. 170–177.

Longerstaey, J., & Spencer, M. (1996). Riskmetrics — technical document. Morgan Guaranty Trust Company of New York: 51, 54.

LeewayHertz – How can blockchain simplify KYC and AML processes? https://www.leewayhertz.com/blockchain-in-aml/.

L. Encrypt, “Let’s Encrypt,” [Online]. Available: https://letsencrypt.org/.

Le-Phuoc, D., Dao-Tran, M. Parreira, J. X. and Hauswirth, M. A Native and Adaptive Approach for Unified Processing of Linked Streams and Linked Data. Proceedings of the 10th International Conference on The Semantic Web (ISWC’11), Springer, 2011.

Le-Phuoc, D., Nguyen Mau, H., Parreira, J. X. and Hauswirth, M.. The Linked Sensor Middleware – Connecting the Real World and the Semantic Web. Proceedings of the 10th International Conference on The Semantic Web (ISWC’11), Springer, 2011.

Le-Phuoc, D. and Hauswirth, M. Linked open data in sensor data mashups. Proceedings of the 2nd International Workshop on Semantic Sensor Networks (SSN09) in conjunction with ISWC 2009.

Liu, F.T., Ting, K.M., & Zhou, Z.-H. (2008). Isolation Forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422.

Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y. P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., & Wiessner, E. (2018). Microscopic Traffic Simulation using SUMO. 21st International Conference on Intelligent Transportation Systems (ITSC), 2575–2582. IEEE Xplore. /doi.org/10.1109/ITSC.2018.8569938.

L. Oliveira, L. Zavolokina, I. Bauer, and G. Schwabe, "To Token or not to Token: Tools for Understanding Blockchain Tokens", Thirty Ninth International Conference on Information Systems, San Francisco 2018.

L. Monso, "Asset tokenization: What is it and why does it matter?", 2019. [Online]. Available at: https://medium.com/@Metaco/asset-tokenization-what-is-it-and-why-does-it-matter-3f6892273dfe [Accessed: 20-October-2020].

L. Monso, "Asset tokenization: Benefits and challenges ahead", 2019 [Online]. Available at: https://medium.com/@Metaco/asset-tokenization-benefits-and-challenges-ahead-475d737eba57 [Accessed: 20-October-2020].

Long G., Tan Y., Jiang J., Zhang C. (2020) Federated Learning for Open Banking. In: Yang Q., Fan L., Yu H. (eds) Federated Learning. Lecture Notes in Computer Science, vol 12500. Springer, Cham. https://doi.org/10.1007/978-3-030-63076-8_17.

Landgren, P., Srivastava, V., and Leonard, N.E. (2016). On distributed cooperative decisionmaking in multi armed bandits. In Proceedings of the IEEE European Control Conference (ECC), 243–248.

"Leveraging Semantic and Graph Technology to Tame the Enterprise Data Storm." https://info.cambridgesemantics.com/hubfs/whitepapers/Perfect%20Storm%20Fin%20Svcs%20Whitepaper.pdf (accessed Mar. 26, 2022).

Laudet, D., (2017). Infographic: Worldwide Hackathon Figures in 2016, Available at: http://agency.bemyapp.com/insights/infographics-hackathon-figures-in-2016.html.

Lafourcade, P., & Lombard-Platet, M. (2020). About blockchain interoperability. Information Processing Letters, 161, 105976.

Links, Skip. "High-Resolution Land Use and Land Cover Mapping".

M. Armbrust, R. Xin, C. Lian, Y. Huai, D. Liu, J. Bradley, X. Meng, T. Kaftan, M. Frank-lin, A. Ghodsi, M. Zaharia, "Spark SQL: relational data processing in Spark", in ACM SIGMOD, 2015, pp. 1383–1394.

Manisha Patel, "Top Five Impacts of GDPR on Financial Services" , Fintech Times, November 2017.

Mark Halstead, "What is the Fourth AML Directive (AML4D)-RedFlagAlert", https://www.redflagalert.com/articles/data/what-is-the-fourth-aml-directive-aml4d.

Martin David, Jorge Bernal, Julien Bringer, Nicolas Notario, Eduardo Gonzales – D3.1 – ARIES eID ecosystem technical design – July 2017.

Marcus Brandenburger, Eduarda Freire, "Witdom Project, D4.2 – Final specification of an end-to-end secure architecture", August 2016. [Online]. Available: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ac7baf1a&appId=PPGMS [Accessed 10-July-2021].

M. Gruteser and B. Hoh, "On the Anonymity of Periodic Location Samples," in Security in Pervasive Computing, 2005, pp. 179–192.

M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability," Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security – CCS '13. 2013, doi: 10.1145/2508859.2516735.

M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability," Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security – CCS '13. 2013, doi: 10.1145/2508859.2516735.

M. Duckham and L. Kulik, "A Formal Model of Obfuscation and Negotiation for Location Privacy," Lecture Notes in Computer Science. pp. 152–170, 2005, doi: 10.1007/11428572_10.

M. Duckham and L. Kulik, "Location privacy and location-aware computing," Dynamic & mobile GIS: investigating change in, 2006, [Online]. Available: https://www.academia.edu/download/44974954/Location_privacy_and_locationaware_comp20160421-7519-143glj.pdf.

Miguel E. Andrés and Nicolás Emilio Bordenabe and Konstantinos Chatzikokolakis and Catuscia Palamidessi (2012). Geo-Indistinguishability: Differential Privacy for Location-Based Systems. CoRR, abs/1212.1984.

M. Gruteser and D. Grunwald, "Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking," Proceedings of the 1st international conference on Mobile systems, applications and services – MobiSys '03. 2003, doi: 10.1145/1066116.1189037.

M. Hirzel et al., "IBM Streams Processing Language: Analyzing Big Data in motion," in IBM Journal of Research and Development, vol. 57, no. 3/4, pp. 7:1–7:11, May-July 2013.

Miklos Dietz et al., "Cutting through the noise around financial technology", McKinsey & Co, February 2016.

Michael Lamer: "The Future of FINTECH ~ The new Standard", Juniper Research, May 2019.

More Than Digital – Innovation – Definition, Innovation Types And Meaning https://morethandigital.info/en/innovation-definition-innovation-types-and-meaning/.

MIT Sloan Management Review https://sloanreview.mit.edu/projects/strategy-drives-digital-transformation/.

Meng, Z., McCreadie, R., Macdonald, C., Ounis, I., Liu, S., Wu, Y.,… & Zhang, Q. (2020, September). Beta-rec: Build, evaluate and tune automated recommender systems. In Fourteenth ACM conference on recommender systems (pp. 588–590).

"Microservices a definition of this new architectural term," [Online]. Available: https://martinfowler.com/articles/microservices.html#footnote-etymology. [Accessed 24 03 2021].

"Microservices.io," [Online]. Available: https://microservices.io/. [Accessed 20 October 2020].

M. Dudjak and G. Martinović, "An API-first methodology for designing a microservice-based Backend as a Service platform.," Information Technology and Control, vol. 49, no. 2, pp. 206–223, 2020.

M. J. T. D.-M. J. B. S. A. M. N. T. M. I. … & D.-L.-H.-V. E. Naeem, "Trends and Future Perspective Challenges in Big Data. In Advances in Intelligent Data Analysis and Applications," in Springer, Singapore, 2021.

Medium – How to: Business Model Canvas explained https://medium.com/seed-digital/how-to-business-model-canvas-explained-ad3676b6fe4a.

Medium – Introduction to Lean Canvas https://medium.com/@steve_mullen/an-introduction-to-lean-canvas-5c17c469d3e0.

M. Noura, M. Atiquzzaman, and M. Gaedke, "Interoperability in internet of things: Taxonomies and open challenges," Mobile Networks and Applications, vol. 24, no. 3, pp. 796–809, 2019.

M. Elkhodr, S. Shahrestani, and H. Cheung, "The Internet of Things: vision & challenges," in IEEE 2013 Tencon-Spring, 2013, pp. 218–222.

M. Fernández-López, A. Gómez-Pérez, and N. Juristo, "Methontology: from ontological art towards ontological engineering," 1997.

M. Niranjanamurthy, B. N. Nithya and S. Jagannatha, "Analysis of Blockchain technology: pros, cons and SWOT," Cluster Computing, vol. 22, no. S6, pp. 14743–14757, 2018.

M. Casey, J. Crane, G. Gensler, S. Johnson, N. Narula and C. A. Wyplosz, The impact of blockchain technology on finance, Geneva: International Center for Monetary and Banking Studies (ICMB), 2018.

Muselli, M. (2006). Springer Berlin Heidelberg. In Neural Nets (pp. 23–30). Springer Berlin Heidelberg. doi.org/10.1007/11731177_4.

M. di Angelo and G. Salzer, "Tokens, Types, and Standards: Identification and Utilization in Ethereum" IEEE International Conference on Decentralized Applications and Infrastructures (DAPPS), 2020.

Mun, M., Hao, S., Mishra, N., Shilton, K., Burke, J., Estrin, D., Hansen, M., and Govindan, R. (2010). Personal data vaults: A locus of control for personal data streams. In Proceedings of the 6th International Conference Co-NEXT '10, 1–12.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., Aguera y Arcas, B. (2017). CommunicationEfficient Learning of Deep Networks from Decentralized Data, AISTAT.

Miguel Lara and Kate Lockwood. (2016). Hackathons as Community-Based Learning: a Case Study. TechTrends 60, 5 (2016), 486–495.

M. Jordan. (2012) Planning a hackfest. Open Data Learning Summit.

Mueller, J. (2017). FinTech: Considerations on How to Enable a 21st Century Financial Services Ecosystem. Viewpoints], Milken Institute.

Mention, A. L. (2021). The age of FinTech: Implications for research, policy and practice. The Journal of FinTech, 1(01), 2050002.

mlflow.org, "An open source platform for the machine learning lifecycle," [Online]. Available: https://mlflow.org/.

M. Westerkamp, F. Victor, and A. Küpper, "Tracing manufacturing processes using blockchain-based token compositions", Digital Communications and Networks, 2019.

MetaCX for Healthcare. [Online]. Available: https://metacx.com/healthcare.

Marko Komssi, Danielle Pichlis, Mikko Raatikainen, Klas Kindström, and Janne Järvinen. (2015). What are Hackathons for? IEEE Software 32, 5 (2015), 60–67.

«MetalLB,» [Online]. Available: https://metallb.org/.

Mahendraratnam, N., Sorenson, C., Richardson, E., Daniel, G.W., Buelt, L., Westrich, K., Qian, J., Campbell, H., McClellan, M., and Dubois, R.W. (2019). Value-based arrangements may be more prevalent than assumed. The American Journal of Managed Care, 25(2), pp. 70–76.

N. Li, T. Li, and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," in 2007 IEEE 23rd International Conference on Data Engineering, Apr. 2007, pp. 106–115.

Netguru – Top 7 Technology Trends in Fintech in 2021 https://www.netguru.com/blog/top-7-technology-trends-in-fintech-in-2021.

"NGSI specifications," [Online]. Available: http://fiware.github.io/specifications/ngsiv2/stable/.

N. Gaur, L. Desrosiers, V. Ramakrishna, P. Novotny, SA. Baset, and A O'Dowd, Hands-On Blockchain with Hyperledger, 2018.

NCEP. (2015). NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. U.S.: Weather Service/NOAA/U.S.

Niu, Guo-Yue, et al. "The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local scale measurements." Journal of Geophysical Research: Atmospheres 116.D12 (2011).

Next-gen Technology transformation in Financial Services https://www.mckinsey.com/~/media/mckinsey/industries/financial%20services/our%20insights/nex t-gen%20technology%20transformation%20in%20financial%20services/next-gen-technology-transformation-in-financial-services.pdf.

N. Guarino, Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6–8, Trento, Italy, vol. 46. IOS press, 1998.

N. Kapsoulis, A. Psychas, G. Palaiokrassas, A. Marinakis, A. Litke and T. Varvarigou, "Know Your Customer (KYC) Implementation with Smart Contracts on a Privacy-Oriented Decentralized Architecture," Future Internet, vol. 12, no. 2, p. 41, 2020.

National Institute of Standards and technologies https://www.nist.gov/ (Accessed: January 2022).

Nick Taylor and Loraine Clarke. (2018). Everybody's Hacking: Participation and the Mainstreaming of Hackathons. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, 172.

Nolte, A., Chounta, I. A., & Herbsleb, J. D. (2020). What Happens to All These Hackathon Projects? Identifying Factors to Promote Hackathon Project Continuation. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2), 1–26.

Osterwalder, A. P. (2015). Value Proposition Design. New York: Wiley. Retrieved from Osterwalder, A., Pigneur, Y., Bernarda, G., Smith, A., & Papadakos, T. (2015). Value Proposition Design. New York: Wiley.

Osterwalder, A., & Pigneur, Y. (2013). Business Model Generation. Hoboken, NJ: Wiley.

Oehmichen, A., Jain, S., Gadotti, A., and de Montjoye, Y.-A. (2019). OPAL: High performance platform for large-scale privacy-preserving location data analytics. In Proceedings of BigData 2019: 1332–1342.

Osterwalder, A., Pigneur, Y. et al. (2010). 'Business Model Generation', self published.

Osterwalder, Alexander; Pigneur, Yves. (2013). Business Model Generation. Hoboken, NJ: Wiley.

OECD (2019), "Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)", OECD Digital Economy Papers, No. 291, OECD Publishing, Paris, https://doi.org/10.1787/d62f618a-en.

Oehmichen, A., Jain, S., Gadotti, A., and de Montjoye, Y.-A. (2019). OPAL: High performance platform for large-scale privacy-preserving location data analytics. In Proceedings of BigData 2019: pp. 1332–1342.

"Open API specification," 2020. [Online]. Available: https://swagger.io/specification/#:~:text=The%20OpenAPI%20Specification%20(OAS)%20defines,or

%20through%20network%20traffic%20inspection. [Accessed 20 October 2020].

"OData (Open Data Protocol)," [Online]. Available: https://www.odata.org/. [Accessed 24 03 2021].

"OpenAPI Specification," [Online]. Available: http://spec.openapis.org/oas/v3.0.3. [Accessed 24 03 2021].

[Online]. Available: Lead Beneficiary, Contributor, Internal Reviewer, Quality Assurance.

[Online]. Available: Can be left void.

[Online]. Available: https://en.wikipedia.org/wiki/Bastion_host.

OpenLDAP, "OpenLDAP," [Online]. Available: https://www.openldap.org/.

«Oralce Virtual Box,» [Online]. Available: https://www.virtualbox.org/.

«Open-LDAP,» Openldap Foundation, [Online]. Available: https://www.openldap.org/.

OECD (2019), "Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)", OECD Digital Economy Papers, No. 291, OECD Publishing, Paris, https://doi.org/10.1787/d62f618a-en.

"OpenAPI Specification," [Online]. Available: http://spec.openapis.org/oas/v3.0.3. [Accessed 24 03 2021].

"OpenAPI Specification v2.0," [Online]. Available: https://swagger.io/specification/v2/. [Accessed 24 03 2021].

OECD, Gennaio 2020. "The Impact of Big Data and Artificial Intelligence (AI) in the Insurance Sector". Page 10–19. Retrieved January 2022, from: https://www.oecd.org/finance/TheImpact-Big-Data-AI-Insurance-Sector.pdf.

OXPER (2019). Building FinTech Ecosystems: Emerging Trends & Policy Implications. Insights from the 5th Annual Oxford Entrepreneurship Policy Roundtable (OXEPR). Available at: https://www.sbs.ox.ac.uk/sites/default/files/2020-09/OXEPR%202019%20%28Building%20FinTech%20Ecosystems%29.pdf.

Philippe Bracke, Anupam Datta, Carsten Jung and Shayak Sen, "Machine learning explainability in finance: an application to default risk analysis", Bank of England, Staff Working Paper No. 816, August 2019.

P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," 1998, [Online]. Available: http://epic.org/privacy/reidentification/Samarati_Sweeney_paper.pdf.

Patricia Johnson, WhiteSource, "MiFID II Reforms and Their Impact on Technology and Security", Feb 2018, https://resources.whitesourcesoftware.com/blog

-whitesource/mifid-ii-reforms-and-their-impact-on-technology-and-security (accessed Nov 2020).

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., and Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. In Proceedings of Advances in Neural Information Processing Systems 31 (NeurIPS 2018).

Park, Hoyong, Eric Hsiao, and Andy Piper. "Continuous query language (CQL) debugger in complex event processing (CEP)." U.S. Patent No. 9,329,975. 3 May 2016.

P. Carbone, A. Katsifodimos, S. Ewen, V. Markl, S. Haridi, and K. Tzoumas, "Apache FlinkTM: Stream and Batch Processing in a Single Engine," p. 12.

P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," 1998, [Online]. Available: http://epic.org/privacy/reidentification/Samarati_Sweeney_paper.pdf.

Pnevmatikakis, A., Kanavos, S., Matikas, G., Kostopoulou, K., Cesario, A., & Kyriazakos, S. (2021). Risk assessment for personalized health insurance based on real-world data. Risks. 9(3):46. https://doi.org/10.3390/risks9030046.

P. Treleaven, R. Gendal Brown and D. Yang, "Blockchain Technology in Finance," Computer, vol. 50, no. 9, pp. 14–17, 2017.

Pelleg, D., & Moore, A. (1999). Accelerating exact k-means algorithms with geometric reasoning. KDD '99: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 5(1), 277–281. ACM Digital Library. doi.org/10.1145/312129.312248.

Pentland, A. (2012). Society's Nervous System: Building Effective Government, Energy, and Public Health Systems. IEEE Computer, 45(1): 31–38.

Pinkas, B., and Lindell, Y. (2009). A proof of security of Yao's for two-party computation. J Cryptol 22, 161–188.

Pentland, A. (2014) Saving Big Data from itself. Scientific American, 65–68.

Perentis, C., Vescovi, M., Leonardi, C., Moiso, C., Musolesi, M., Pianesi, F., and Lepri, B. (2017). Anonymous or not? Understanding the factors affecting personal mobile data disclosure. ACM Trans. Internet Techn. 17(2): 13:1–13:19.

Pinkas, B., and Lindell, Y. (2009). A proof of security of Yao's for two-party computation. J Cryptol 22, 161–188.

Poursafaei, F., Hamad, G.B., & Zilic, Z. (2020). Detecting malicious Ethereum entities via application of machine learning classification. In Proceedings of the 2020 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS). IEEE.

P2957 – Standard for a Reference Architecture for Big Data Governance and Metadata Management. Standards.ieee.org. Retrieved January, 2022 from https://standards.ieee.org/project/2957.html.

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on information accompanying transfers of funds and certain crypto-assets (recast) COM/2021/422 final https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2021:422:FIN (Accessed: January 2022).

Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the mechanisms to be put in place by the Member States for the prevention of the use of the financial system for the purposes of money laundering or terrorist financing and repealing Directive (EU) 2015/849 COM/2021/423 final. EUR-Lex – 52021PC0423 – EN – EUR-Lex (europa.eu) (Accessed: January 2022).

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing COM/2021/420 final (Accessed: January 2022).

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL establishing the Authority for Anti-Money Laundering and Countering the Financing of Terrorism and amending Regulations (EU) No 1093/2010, (EU) 1094/2010, (EU) 1095/2010 COM/2021/421 final https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0421 (Accessed: January 2022).

Pantelis Angelidis, Leslie Berman, Maria de la Luz Casas-Perez, Leo Anthony Celi, George E Dafoulas, Alon Dagan, Braiam Escobar, Diego M Lopez, Julieta Noguez, Juan Sebastian OsorioValencia, et al. (2016). The hackathon model to spur innovation around global mHealth. Journal of medical engineering & technology 40, 7–8, 392–399.

Pe-Than, E.P.P. et al. (2018): Designing Corporate Hackathons With a Purpose: The Future of Software Development. IEEE Softw. 36, 1.

"Pseudonymization," Imperva. https://www.imperva.com/learn/data-security/pseudonymization/.

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-globalfintech-report-17.3.17-final.pdf.

P. M. N. Maló, "Hub-and-spoke Interoperability: an out of the skies approach for large-scale data interoperability," 2013, Accessed: Jan. 18, 2016. [Online]. Available: http://run.unl.pt/handle/10362/11397.

P. Andrews, I. Zaihrayeu, and J. Pane, "A classification of semantic annotation systems," Semantic Web, vol. 3, no. 3, pp. 223–248, 2012.

Proposal for a Regulation of the European parliament and of the council on Markets in Crypto-assets, and amending Directive (EU) 2019/1937, 2020, https://eu r-lex.europa.eu/legal-content/EN/ALL/?uri=COM%3A2020%3A0593%3AFIN.

P2957 – Standard for a Reference Architecture for Big Data Governance and Metadata Management. (n.d.). Standards.ieee.org https://standards.ieee.org/project/2957.html.

Policy and investment recommendations for trustworthy Artificial Intelligence. (2019, June 25). European Commission. Retrieved January 2022, from https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence.

Pidorycheva, I. Y. (2020). Innovation ecosystem in contemporary economic researches. Economy of Industry, (2 (90), 54–92.

Pascal Lafourcade, Marius Lombard-Platet, About blockchain interoperability, Information Processing Letters, Volume 161, 2020.

PReparing Industry to Privacy-by-design by supporting its Application in REsearch, Project Description, https://cordis.europa.eu/project/id/610613 (accessed in November 2020).

Presto – Distributed Query Engine for Big Data, https://prestodb.io/.

Priest, A. Na, M., Niedzwiadek, H. and Davidson, J. Sensor observation service. Technical Report OGC 06-009r6, October 2007.

Packer. [Online]. Available: https://www.packer.io/.

Q. Zhang, N. Koudas, D. Srivastava, and T. Yu, "Aggregate Query Answering on Anonymized Tables," in 2007 IEEE 23rd International Conference on Data Engineering, Apr. 2007, pp. 116–125.

Recommendations on outsourcing to cloud service providers, https://www.eba.europa.eu/sites/default/documents/files/documents/10180/2170121/5fa5cdde-3219-4e95-946d-0c0d05494362/Final%20draft%20Recommendations%20on%20Cloud%20Outsourcing%20%28EBA-Rec-2017-03%29.pdf?retry=1, [accessed March 2022].

Recital 19 EU GDPR, https://www.privacy-regulation.eu/en/recital-19-GDPR.htm, [accessed March 2022].

R. Chen, G. Acs, and C. Castelluccia, "Differentially private sequential data publication via variablelength n-grams," Proceedings of the 2012 ACM conference on Computer and communications security – CCS '12. 2012, doi: 10.1145/2382196.2382263.

R. Dewri, "Local Differential Perturbations: Location Privacy under Approximate Knowledge Attackers," IEEE Transactions on Mobile Computing, vol. 12, no. 12. pp. 2360–2372, 2013, doi: 10.1109/tmc.2012.208.

R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar, "Preserving User Location Privacy in Mobile Data Management Infrastructures," Privacy Enhancing Technologies. pp. 393–412, 2006, doi: 10.1007/11957454_23.

R. Castro Fernandez, M. Migliavacca, E. Kalyvianaki, and P. Pietzuch, "Making State Explicit for Imperative Big Data Processing," presented at the USENIX ATC'14: 2014 USENIX Annual Technical Conference, Philadelphia, USA, 2014.

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN.

Rancher, "Rancher overview," SUSE, [Online]. Available: https://rancher.com/docs/rancher/v2.5/en/overview/architecture/.

"RAML Version 1.0: RESTful API Modeling Language," [Online]. Available: https://github.com/raml-org/raml-spec/blob/master/versions/raml-10/raml-10.md/. [Accessed 24 03 2021].

R. M. M. M. M. A. Irakli Nadareishvili, Microservice Architecture: Aligning Principles, Practices, and Culture, O'Reilly Media, Inc., 2016.

Risks, an online journal, an international, scholarly, peer-reviewed, open access journal for research and studies on insurance and financial risk management. Risks is published monthly online by MDPI.

R. Campos-Rebelo, F. Moutinho, L. Paiva, and P. Maló, "Annotation Rules for XML Schemas with Grouped Semantic Annotations," in IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society, 2019, vol. 1, pp. 5469–5474.

Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2009). Bpr: Bayesian personalized ranking from implicit feedback. UAI '09: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, 452–461. ACM Digital Library. doi: 10.5555/1795114.1795167.

Rendle, S., Krichene, W., Zhang, L., & Anderson, J. (2020). Neural Collaborative Filtering vs. Matrix Factorization Revisited. RecSys '20: Fourteenth ACM Conference on Recommender Systems, 240–248. ACM DigitalLibrary. https://doi.org/10.1145/3383313.3412488.

Rendle, S., Zhang, L., & Koren, Y. (2019). On the difficulty of evaluating baselines: A study on recommender systems. arxiv Preprint, Information Retrieval. arXiv. https://arxiv.org/abs/1905.01395.

Rokach, L., & Maimon, O. (2005). Clustering methods. Data mining and knowledge discovery handbook. Springer US, 321–352.

Ruta, M., Noia, T. Di, Scioscia, F., Di Sciascio E. Semantic-enhanced EPCglobal Radio-Frequency IDentification. SWAP 2007.

RKE. [Online]. Available: https://rancher.com/docs/rke/latest/en/.

Rancher. [Online]. Available: https://rancher.com/.

R. S. M. S. W. M. &. S. R. Norvill, "Blockchain for the Simplification and Automation of KYC Result Sharing," in 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), 2019.

Rasmussen, C. E., Williams, C. K. I. (2006). Gaussian Processes for Machine Learning, The MIT Press.

R. Agrawal, C. Faloutsos, and A. N. Swami. E cient similarity search in sequence databases. In Proceedings of the International Conference on Foundations of Data Organization and Algorithms (FODO), pages 69–84. Springer-Verlag, 1993.

R. Cole, D. Shasha, and X. Zhao. Fast window correlations over uncooperative time series. In Proceedings of the International Conference on Knowledge Discovery and Data Mining (SIGKDD), pages 743–749. ACM, 2005.

Ruby. [Online]. Available: https://www.ruby-lang.org/en/.

Ribeiro, Marco & Singh, Sameer & Guestrin, Carlos. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. 1135–1144. 10.1145/2939672.2939778.

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN.

Rahimi, A., and Recht, B. (2008). Random features for large-scale kernel machines. In Proceedings of Advances in Neural Information Processing Systems 20.

"Random Forest Regressor", 2021 [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html#sklearn.ensemble.RandomForestRegressor.

R. J. Hyndman and G. Athanasopoulos. Forecasting: Principles and Practice. https://otexts.com/fpp2.

Richards, D., Rebeschini, P., and Rosasco L. (2020). Decentralised learning with random features and distributed gradient descent. In Proceedings of the 37th International Conference on Machine Learning (ICML), 119: 8105–8115.

Rissanen, E. (2013). Extensible Access Control Markup Language (XACML) Version 3.0. OASIS Standard. Available online at: http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html.

S.-S. Ho and S. Ruan, "Differential privacy for location pattern mining," Proceedings of the 4th ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS – SPRINGL '11. 2011, doi: 10.1145/2071880.2071884.

Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting 36(3), 1181–1191.

Spark Streaming | Apache Spark., https://spark.apache.org/streaming/.

Stratosphere?» Next Generation Big Data Analytics Platform. http://stratosphere.eu/.

Samza., http://samza.apache.org/.

S. Schneider, H. Andrade, B. Gedik, A. Biem, and K. L. Wu, "Elastic scaling of data parallel operators in stream processing," in 2009 IEEE International Symposium on Parallel Distributed Processing, 2009, pp. 1–12.

Saft – How is the IoT world shaping up in 2021 and what trends will influence the future of IoT? https://www.saftbatteries.com/energizing-iot/how-iot-world-shaping-2021-and-what-trends-will-influence-future-iot-infographic.

Solera – Accelerating digital transformation for the future of automotive claims https://get.solera.com/SOLGLOBAL2020-04MTR1219Digital-Transformation-Webinar_05-Recording-Download.html.

S. Nexus, "Sonatype Nexus OSS," [Online]. Available: https://www.sonatype.com/nexus-repository-oss.

Slack, "Slack features," Microsoft, [Online]. Available: https://slack.com/intl/en-it/features.

Soldatos, John et al. "Big Data and Artificial Intelligence in Digital Finance: Increasing Personalization and Trust in Digital Finance Using Big Data and AI." Springer International Publishing AG, 2022.

Sonarqube, "Sonarqube architecture," SonarSource SA, 2020. [Online]. Available: https://docs.sonarqube.org/latest/architecture/architecture-integration/.

Spring Boot Framework, [Online]. Available: https://spring.io/projects/spring-boot. [Accessed 20 October 2020].

Spring Cloud Circuit Breaker, [Online]. Available: https://spring.io/projects/spring-cloudcircuitbreaker. [Accessed 21 October 2020].

Salesforce – What Is Digital Transformation? https://www.salesforce.com/products/platform/what-is-digital-transformation/.

Stack – Online guide to digital business transformation https://www.stackstudio.digital/onlineguide-to-digital-business-transformation/.

Shin, D. H., Kim, H., & Hwang, J. (2015). Standardization revisited: A critical literature review on standards and innovation. Computer Standards & Interfaces, 38, 152–157.

Self-Registration, 2020. [Online]. Available: https://microservices.io/patterns/self-registration.html. [Accessed 10 October 2020].

S. Assefa, D. Dervovic, M. Mahfouz, T. Balch, P. Reddy and M. Veloso, "Generating synthetic data in finance: opportunities, challenges and pitfalls," JPMorgan Chase & Co, 2020.

Springer (2021) "Digital Finance" Source: https://www.springer.com/journal/42521.

Source: https://www.fatf-gafi.org/publications/fatfrecommendations/documents/guidance-rba-virtual-assets2021.html.

Source: https://docs.google.com/spreadsheets/d/1_LghI9zgA5wGArjd1q6VeP7xWJzw1gx/edit?usp=sharing&ouid=115760979640495528344&rtpof=true&sd=true.

Su, Y. S., Zheng, Z. X., & Chen, J. (2018). A multi-platform collaboration innovation ecosystem: the case of China. Management Decision.

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

Salehi, A., Aberer, K. «GSN, Quick and Simple Sensor Network Deployment», European conference on Wireless Sensor Networks (EWSN), Netherlands, 2007.

Scherp, A., Franz, T. Saatho, S. Staab. F–a Model of Events Based on the Foundational Ontology DOLCE+DnS Ultralight. In: International Conference on Knowledge Capturing (K-CAP), Redondo Beach, CA, USA., 2009.

Sharpe, W.F. (1994). The Sharpe Ratio. The Journal of Portfolio Management 21(1), 49–58.

Sheth, A. Henson, C., Sahoo. S. Semantic Sensor Web. IEEE Internet Computing 12 (4), 2008.

"Semantic Annotations for WSDL and XML Schema." https://www.w3.org/TR/sawsdl/ (accessed Jul. 22, 2020).

S. Peroni, "A simplified agile methodology for ontology development," in OWL: Experiences and Directions–Reasoner Evaluation, Springer, 2016, pp. 55–69.

S. A. McIlraith, T. C. Son, and H. Zeng, "Semantic web services," IEEE intelligent systems, vol. 16, no. 2, pp. 46–53, 2001.

S. Minukhin, V. Fedko and D. Sitnikov, "SQL-On-Hadoop Systems: Evaluting Performance of Polybase for Big Data Processing," 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, 2018, pp. 591–594. [MIRAB12] Microsoft Industry Reference Architecture for Banking (MIRA-B), Microsoft Corporation Whitepaper, May 2012.

Shearer C., The CRISP-DM model: the new blueprint for data mining, J Data Warehousing (2000); 5:13—22. [Simitsis12] A. Simitsis, K. Wilkinson, M. Castellanos, U. Dayal, "Optimizing analytic data flows for multiple execution engines", in ACM SIGMOD, 2012, pp. 829–840.

Singh, A. (2019). Anomaly detection in the Ethereum network. A thesis for the degree of Master of Technology/Indian Institute of Technology Kanpur.

Sra, S., & Dhillon, I. (2005). Generalized Nonnegative Matrix Approximations with Bregman Divergences. Advances in Neural Information Processing Systems, 18(1), 283–290.

Stopar, L., Skraba, P., Grobelnik, M., & Mladenic, D. (2019). StreamStory: Exploring multivariate time series on multiple scales. IEEE Transactions on Visualization and Computer Graphics, 25(4), pp. 1788–1802.

Strumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. Knowledge and Information Systems, 41(3), 647–665. Springer Link. doi.org/10.1007/s10115-013-0679-x.

Supporting Material – Archived Groups – Kantara Initiative, Kantarainitiative.org, 2020. [Online]. Available: https://kantarainitiative.org/confluence/display/archive/1+-+Supporting+Material. [Accessed 2 September 2020].

Sagi, O., and Rokach, O. L. (2018). Ensemble learning: A survey, WIREs Data Mining and Knowledge Discovery.

S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system”, 2008. [Online] Available at: https://bitcoin.org/bitcoin.pdf [Accessed 19-October-2020].

Sen, J. (2013). Homomorphic encryption: Theory and applications. In J. Sen (ed.) Theory and practice of cryptography and network security protocols and technologies. INTECH Publishers.

Shamir, A. (1979). How to share a secret. Communications of the ACM, 22 (11): 612–613.

Staiano, J., Oliver, N., Lepri, B., de Oliveira, R., Caraviello, M., and Sebe, N. (2014). Money walks: A human-centric study on the economics of personal mobile data. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 583–594.

Staiano, J., Zyskind, G., Lepri, B., Oliver, N., and Pentland, A. (2019). The rise of decentralized personal data markets. In T. Hardjono, D. Shrier, and A. Pentland (eds.) Trust :: Data. A new framework for identity and data sharing. MIT Press.

Squirrel, SQL UI client, http://squirrel-sql.sourceforge.net/.

SPARQL Query Language for RDF Implementation Accessible here: http://www.w3.org/TR/rdf-sparql-query/.

The Financial Instrument Global Identifier https://www.openfigi.com/about/figi (Accessed: January 2022).

The Object Management Group OMG.org, an international, non-profit standards organization, founded in 1989.

The Financial Industry Business Ontology https://spec.edmcouncil.org/fibo/ (Accessed: January 2022).

The Financial Industry Business Ontology https://spec.edmcouncil.org/fibo/ (Accessed: January 2022).

The European project for Standardized Transparent Representations in order to Extend Legal Accessibility (Estrella, IST-2004-027655) http://www.estrellaproject.org.

The European project for Standardized Transparent Representations in order to Extend Legal Accessibility (Estrella, IST-2004-027655) http://www.estrellaproject.org.

"The EU regulation on electronic identification and certification services", 2016, https://www.dpc.bg/p/doc/dpc-dpco-the-eu-regulation-on-electronic-identification-and-certification-services-paving-the-way-forward-towards-more-secure-internet-transactions-07-04-2016-637.pdf, accessed Nov 2020.

"Tokenization Market by Component, Application Area (Payment Security, Application Area, and Compliance Management), Tokenization Technique (API-based and Gateway-based), Deployment Mode, Organization Size, Vertical, and Region – Global Forecast to 2023", 2018. MarketsandMarkets Research Report, [Online]. Available at: https://www.marketsandmarkets.com/MarketReports/tokenization-market76652221.html?gclid=Cj0KCQjw2or8BRCNARIsAC_ppyY4t6bLXgBQG1pregsjn3hKtxG8goxXR6RgjRXxlP1H7MY2IXUSqGsaAjNAEALw_wcB [Accessed: 20-October-2020].

Text Retrieval conference https://trec.nist.gov/ (Accessed: January 2022).

The Power MBA – The Lean Startup Methodology Cycle: 4 Steps to Risk-Free Success https://www.thepowermba.com/en/blog/lean-startup-methodology.

The Viima Blog – Innovation Management https://www.viima.com/blog/innovationmanagement?hs_amp=true.

The Viima Blog – Innovation Management – Processes https://www.viima.com/blog/innovation-management-processes?hs_amp=true.

The EC blockchain strategy in a nutshell, https://ec.europa.eu/digital-single-market/en/blockchaintechnologies (Accessed: January 2022).

The Enterprisers project – What is digital transformation https://enterprisersproject.com/sites/default/files/what_is_digital_transformation_2020.pdf.

Testing two-factor authentication, Michelangelo van Dam, 05/08/2017. https://www.in2it.be/2017/08/testing-two-factor-authentication/.

The KTH Innovation Readiness Level https://kthinnovationreadinesslevel.com/take-a-tour/.

Team Readiness Capability Assessment Model https://teamreadiness.com/?page_id=122.

T. Yuanyuan, T. Zou, F. Özcan, R. Gonscalves, H. Pirahesh, "Joins for hybrid warehouses: exploiting massive parallelism in hadoop and enterprise data warehouses", in EDBT/ICDT Conf., pp. 373–384 (2015).

T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," Scientific american, vol. 284, no. 5, pp. 34–43, 2001.

Tapanainen, T. (2020). Toward Fintech Adoption Framework for Developing Countries-A Literature Review based on the Stakeholder Perspective. Journal of Information Technology Applications and Management, 27(5), 1–22.

Tsiatsis, V., Gluhak, A., Bauge, T., Montagut, F., Bernat, J., Bauer, M., Villalonga, C., Barnaghi, P.M., Krco, S. The SENSEI Real World Internet Architecture. Future Internet Assembly, IOS Press, 2010.

Team, C. (2020). Indirect exposure: Why you need to look beyond direct counterparties to understand cryptocurrency address risk. Nov. 2020, https://blog.chainalysis.com/reports/cryptocurrency-risk-blockchain-analysis-indirect-exposure.

T. Cameron McKenzie, "open API (public API)," [Online]. Available: https://searchapparchitecture.techtarget.com/definition/open-API-public-API. [Accessed 24 03 2021].

T. Slimani, "Semantic annotation: The mainstay of semantic web," arXiv preprint arXiv:1312.4794, 2013.

T. M. Siebel, Digital transformation: survive and thrive in an era of mass extinction. RosettaBooks, 2019.

Thompson, Gregory, et al. "Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization." Monthly Weather Review 136.12 (2008): 5095–5115.

«Terraform,» [Online]. Available: https://www.terraform.io/.

«Terraform homepage,» HashiCorp, [Online]. Available: https://www.terraform.io/docs/language/files/index.html.

T. Özsu, P. Valduriez, Principles of Distributed Database Systems, 4th ed. Springer, 700 pages (2020).

Takemori, S., and Sato, M. (2020). Approximation methods for kernelized bandits, arXiv 2010.12167.

Tamer Özsu, Patrick Valduriez. Principles of Distributed Database Systems, 4th Edition, Springer, 2020.

UN AHEG (2020), First version of a draft text of a recommendation on the ethics of Artificial Intelligence, SHS/BIO/AHEG-AI/2020/4, limited distribution.

Understanding JSON Schema, [Online]. Available: http://json-schema.org/understanding-json-schema/. [Accessed 24 03 2021].

Unified Medical Language System Accessible here: http://www.nlm.nih.gov/research/umls/index.html.

UN AHEG (2020), First version of a draft text of a recommendation on the ethics of Artificial Intelligence, SHS/BIO/AHEG-AI/2020/4, limited distribution.

Upbit hacking incident announcement, Nov. 2019, https://upbit.com/service_center/notice?id=1085.

Verbatim text from https://resources.whitesourcesoftware.com/legal/mifid-ii-reforms-and-their-impact-on-technology-and-security (accessed in December 2020).

V. Ayala-Rivera, P. McDonagh, T. Cerqueus and L. Murphy, "Synthetic Data Generation using Benerator Tool," University College Dublin, 2013.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. Advances in Neural Information Processing Systems, 30, pp. 5998–6008.

V. Buterin, "A next-generation smart contract and decentralized application platform, Ethereum Project White Paper", [Online] Available at: https://github.com/ethereum/wiki/wiki/White-Paper [Accessed: 20-October-2020].

Vescovi, M., Lepri, B., Perentis, C., Moiso, C., and Leonardi, C. (2014). My data store: Toward user awareness and control on personal data. In Proceedings of UbiComp Adjunct: 179–182.

Verbatim text from https://resources.whitesourcesoftware.com/legal/mifid-ii-reforms-and-their-impact-on-technology-and-security (accessed in December 2020).

Vagrant. [Online]. Available: https://www.vagrantup.com/.

VMware. [Online]. Available: https://www.vmware.com/.

v. ESXi. [Online]. Available: https://www.vmware.com/products/esxi-and-esx.html.

«VMware Workstation,» [Online]. Available: https://www.vmware.com/products/workstationpro.html.

VMware. [Online]. Available: https://www.vmware.com/.

V. Buterin, Chain interoperability, https://www.r3.com/wp-content/uploads/2017/06/chain_interoperability_r3.pdf (Accessed: January 2022).

V. Gadepally, P. Chen, J. Duggan, A. J. Elmore, B. Haynes, J. Kepner, S. Madden, T. Mattson, M. Stonebraker, "The BigDawg polystore system and architecture", in IEEE High Performance Extreme Computing Conference (HPEC), 2016, pp. 1–6.

V. Gulisano, R. Jiménez-Peris, M. Patiño-Martínez, C. Soriente and P. Valduriez, "StreamCloud: An Elastic and Scalable Data Streaming System," in IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 12, pp. 2351–2365, Dec. 2012.

WSO2 CEP., https://wso2.com/products/complex-event-processor/.

WITDOM Project Documentation https://bit.ly/3cEA75e.

Web Services Description Language (WSDL) 1.1, [Online]. Available: https://www.w3.org/TR/2001/NOTE-wsdl-20010315. [Accessed 24 03 2021].

Web Application Description Language, [Online]. Available: https://www.w3.org/Submission/wadl/. [Accessed 24 03 2021].

WBCINNO Project – Methodology for Innovation Management, University of Kragujevac, Prof. Dr. Slobodan Arsenijević http://www.wbc-inno.kg.ac.rs/pub/download/13953889933298_wbc_inno_academic_methodology_for_innovation_management_eng.pdf.

WBC Inno – Methodology for Innovation Management http://www.wbc-inno.kg.ac.rs/pub/download/13953889933298_wbc_inno_academic_methodology_for_innovation_management_eng.pdf.

Wikipedia, "Bastion Host," [Online]. Available: https://en.wikipedia.org/wiki/Bastion_host.

World Wide Web Consortium – Linked Open Data, Accessible here: http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData.

World Wide Web Consortium – Resource Description Framework Schema Accessible here: http://www.w3.org/TR/rdf-schema.

World Wide Web Consortium – Turtle Serialisation Specification Accessible here: http://www.w3.org/TeamSubmission/turtle/.

World Wide Web Consortium – N-Triples format specification Accessible here: http://www.w3.org/TR/rdf-testcases/#ntriples.

World Wide Web Consortium – Ontology Web language Accessible here: http://www.w3.org/TR/owl-ref.

World Wide Web Consortium – Resource Description Framework in Attributes Accessible here: http://www.w3.org/TR/xhtml-rdfa-primer/.

Wang, X., He, X., Wang, M., Feng, F., & Chua, T.-S. (2019). Neural Graph Collaborative Filtering. SIGIR'19: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 165–174. ACM digital Library. https://doi.org/10.1145/3331184.3331267.

W. Entriken, D. Shirley, E. Evans, and N. Sachs, "ERC-721 non-fungible token standard," 2018. [Online]. Available: https://eips.ethereum.org/EIPS/eip-721. [Accessed 1-October-2020].

Want, R., Pering, T., Danneels, G., Kumar, M., Sundar, M., and Light, J. (2002). The personal server: Changing the way we think about ubiquitous computing. In Proceedings of 4th International Conference on Ubiquitous Computing, 194–209.

What is a lambda architecture? https://databricks.com/glossary/lambda-architecture.

World Economic Forum Companion to the Model AI Governance Framework—Implementation and Self-Assessment for Organizations (Infocomm Media Development Authority of Singapore, 2020) Retrieved January 2022 from, https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGIsago.pdf.

Will Kenton and Julius Mansa, "MiFID II – Laws&Regulations", Jul 2020, https://www.investopedia.com/terms/m/mifid-ii.asp (Accessed Nov 2020).

Wedge R., Kanter J. M., Veeramachaneni K., Rubio S. M., and Perez S. I., (2018). Solving the false positives problem in fraud prediction using automated feature engineering. In Proceedings of the.

W. Radomski, A. Cooke, P. Castonguay, J. Therien, E. Binet, and R. Sandford, "ERC-1155 multi token standard," 2015. [Online]. Available: https://eips.ethereum.org/EIPS/eip-1155 [Accessed 10-October-2020].

W. Radomski, A. Cooke, P. Castonguay, J. Therien, E. Binet, and R. Sandford, "ERC-1155 multi token standard," 2015. [Online]. Available: https://eips.ethereum.org/EIPS/eip-1155 [Accessed 10-July-2021].

Y. Liao, M. Lezoche, H. Panetto, and N. Boudjlida, "Semantic annotation model definition for systems interoperability," in OTM Confederated International Conferences "On the Move to Meaningful Internet Systems", 2011, pp. 61–70.

Y. Song, X. Lu, S. Nobari, S. Bressan, and P. Karras, "On the privacy and utility of anonymized social networks," International Journal of Adaptive, Resilient and Autonomic Systems (IJARAS), vol. 4, no. 2, pp. 1–34, 2013.

Y.-A. de Montjoye, Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the Crowd: The privacy bounds of human mobility," Scientific Reports, vol. 3, no. 1. 2013, doi: 10.1038/srep01376.

Yagoubi, D.-E., Akbarinia, R., Kolev, B., Levchenko, O., Masseglia, F., Valduriez, P., Shasha, D., 2018. ParCorr: Efficient Parallel Methods to Identify Similar Time Series Pairs across Sliding Windows. Data Mining and Knowledge Discovery, vol. 32(5), pp. 1481–1507. Springer.

Yang, L., Bagdasaryan, E., Gruenstein, J., Hsieh, C.-K., & Estrin, D. (2018). Openrec: A modular framework for extensible and adaptable recommendation algorithms. WSDM '18: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 664–672. ACM Digital Library. https://doi.org/10.1145/3159652.3159681.

Z. Minpeng, R. Tore, "Querying combined cloud-based and relational databases", in Int. Conf. on Cloud and Service Computing (CSC), 2011, pp. 330–335.

Zyskind, G., Nathan, O., and Pentland, A. (2015). Decentralizing privacy: Using blockchain to protect personal data. In Proceedings of IEEE Symposium on Security and Privacy Workshops: 18.

Zeng, S., Tay, Y., Yao, L., Wu, B., & Sun, A. (2019). Deeprec: An open-source toolkit for deep learning based recommendation. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, 6581–6583. dblp computer science bibliography. https://doi.org/10.24963/ijcai.2019/963.

Zyskind, G., Nathan, O., and Pentland, A. (2015). Decentralizing privacy: Using blockchain to protect personal data. In Proceedings of IEEE Symposium on Security and Privacy Workshops: 180–184.

"3rd-Party Registration," 2020. [Online]. Available: https://microservices.io/patterns/3rd-party-registration.html. [Accessed 15 October 2020].