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

Infinitech

D3.15 – Regulatory Compliance Tools – I

Revision Number	3.0
Task Reference	T3.6
Lead Beneficiary	ATOS
Responsible	Nuria Ituarte Aranda
Partners	AKTIF, ASSEN, ATOS, BOS, BPFI, DYN, GRAD, JSI, NBG, PI
Deliverable Type	Report (R)
Dissemination Level	Public (PU)
Due Date	2020-11-30
Delivered Date	2020-11-30
Internal Reviewers	GRAD, CTAG
Quality Assurance	CCA, INNOV
Acceptance	WP Leader Accepted and/or Coordinator Accepted
EC Project Officer	Pierre-Paul Sondag
Programme	HORIZON 2020 - ICT-11-2018
	This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no 856632

Contributing Partners

Partner Acronym	Role[1]	Author(s)[2]
ATOS	Lead beneficiary	Nuria Ituarte Aranda Darío Ruiz López
AKTIF	Contributor	Orkan Metin
ASSEN	Contributor	Ilesh Dattani
BOS	Contributor	Klaudija.Jurkosek-Seitl Sabina Podkriznik Milošević Jelena
BPFI	Contributor	Richard Walsh
DYN	Contributor	Andreas Politis Michalis Michalakoukos Anastassios Markou Christodoulos Zervas
GRAD	Contributor	Lilian Adkinson Orellana Marta Sestelo Borja Pintos
JSI	Contributor	Maja Skrjanc Mitja Jermol
NBG	Contributor	Syllignakis Manolis
PI	Contributor	Massimiliano Aschi Giusseppe Avigliano Marco Avallone Annalisa Ceccarelli

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

[2] Can be left void

Revision History

Version	Date	Partner(s)	Description
0.1	2020-06-02	Atos	ToC Version
0.2	2020-08-12	Atos	Provided contributions on table of section 4.1 with pilot, privacy issues and solutions and on section 4.3 for some pilots
0.3	2020-09-30	GRAD, PI, DYN, Atos GRAD	Adding section 3.3 with the summary of Security, Privacy and Data protection technologies in the project Added section 3.1 Data governance mechanisms
0.4	2020-10-01	GRAD, BPFI, BOS, JSI, PI, DYN	Provided contributions for pilots on sections 4.1 and 4.3
0.5	2020-10-05	Atos	Added executive summary and conclussions and section 4.2
0.6	2020-11-09	Atos, BOS, GFT	Confidential documents (pilots 7 and 8) prepared
0.7	2020-11-10	Atos, BOS, JSI	Added missing issues asked to pilot leaders
1.0	2020-11-13	Atos	First Version for Internal Review
2.0	2020-11-18	Atos	Version for Quality Assurance sent to INNOV
2.1	2020-11-26	Atos	Version for Quality Assurance sent to CCA
3.0	2020-11-27	Atos	Version for Submission

Executive Summary

This deliverable analyses regulatory compliance throughout the INFINITECH project and specifically in every pilot. It starts the analysis in a general way, considering regulations for the financial sector and the available technologies in INFINITECH. Moreover, there is an important analysis for every pilot in several aspects:

- The regulations that they should fulfil.
- Data governance mechanisms.
- The privacy, security and data protection issues.
- The technologies they use and how they comply with the regulations.
- The solutions that are provided in the pilots to comply with the regulations.

The following table shows all the INFINITECH pilots and the most important findings for every pilot, reached in this deliverable, which are the applicable regulations, the compliance solution and the technologies for Regulatory Compliance used. Each pilot's full names is in Table 2

Table 1: Regulations, Solutions, and Technologies per pilot.

Pilot	Applicable Regulations	Compliance Solution	Technologies for Regulatory Compliance
#1	None: GDPR is not applicable because the system does not ever access the data about the customers and the only persons involved are notaries, who are considered as legal persons.	Compliance based on the use of synthetic and aggregated data for both production case and the pilots.	Not Applicable because no regulation applies
#2	BASEL IV and MIFID II	Compliance based on the use of synthetic and aggregated data for the pilots.	IAM and Audit logs (MIFID II)
#3	GDPR (Need for authentication and authorization)	Compliance based on the use of synthetic and aggregated data for the pilots.	IAM and Cryptography
#4	MIFID II and GDPR	Compliance based on INFINITECH and legacy technologies to provide secure access to the personal portfolio internally by allowing secure onboarding authentication to the investor through DUOS	IAM DUOS: Digital onboarding Authentication
#5b	MIFID II and GDPR	Compliance based on the use of synthetic and aggregated data for the pilots.	IAM and Cryptography (GDPR) Audit logs (MIFID II)
#6	GDPR	Compliance based on INFINITECH and legacy technologies: Need to either secure the data or anonymize them. Use of Icarus platform to anonymize the data	Anonymization

#7	This pilot is confidential. Thus, in case of specific need to know, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us .		
#8	This pilot is confidential. Thus, in case of specific need to know, please, contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us .		
#9	Production: GDPR and AMLD4 Pilot: none	Compliance based on the use of synthetic and aggregated data for the pilots.	IAM and Cryptography (GDPR)
#10	Production: GDPR. Pilot: none	Compliance based on the use of synthetic and aggregated data for the pilots.	IAM and Cryptography (GDPR)
#11	GDPR	Compliance based on INFINITECH technologies: security framework (IAM and Consent Management) and anonymization.	Anonymization tool IAM, Consent management
#12	GDPR	Compliance based on INFINITECH technologies: The pilot will incorporate a Security framework that will provide IAM and authentication capabilities. Moreover, a regulatory compliance tool that anonymize personal data will be applied.	Anonymization tool, Access control
#13	GDPR for GPS position	Compliance based on INFINITECH technologies: consent management. The data are pseudonymized and the GPS position is anonymized.	IAM Consent management Pseudonymization Anonymization
#14	GDPR for location data and purpose of use of the data	Compliance based on INFINITECH technologies: Security framework from AGA will provide IAM, Consent Management and anonymization features	TSL IAM

Regulatory Compliance in the pilots is supported in two complementary ways:

- First, through the technologies that they are using. In some cases these technologies comprise complete solutions that have considered the applicable regulations and hence address regulatory compliance themselves.
- Second, through the INFINITECH regulatory compliance tools. In this case INFINITECH provides tools to help solving privacy and/or security issues. This second case is in-line with one of the main objectives of task T3.6 "Regulatory Compliance Tools", which is to provide regulatory compliance tools. In-line with this objective, this deliverable provides a general definition of the requirements for a regulatory compliance service, by using the DPO (Data Protection Orchestrator) tool provided by Atos.

The following regulations have been considered as part of this deliverable:

- GDPR for INFINITECH pilot systems that deal with personal data.
- MIFID II for financial consultancy services.
- PSD2 for online payment platforms.
- AMLD4 for fighting against money laundering and terrorism.

The main types of technologies that help supporting compliance to these regulations include:

- For GDPR Compliance:
    - Anonymization.
    - Pseudonymization.
    - Privacy dashboards
    - Strong authentication and authorization mechanisms
    - Encryption of data
    - Data Protector Orchestrator.
- For MIFID II Compliance:
    - Auditing logs
    - phone call recording
    - email logs
    - strong authentication, preferably multi-factor, and authorization mechanisms
- For PSD2 Compliance:
    - Strong multi-factor authentication
    - SIEM (Security Information Event Management) systems

The general philosophy of the project towards regulatory compliance tools is to use the ones already in use by the users when they are already available and already comply with applicable regulations and provide new tools when the mentioned tools are insufficient for current regulations. The summary of INFINITECH Security, Privacy and Data Protection technologies are in Table 4 and is also described in more detail in deliverable D2.5 "Specifications of INFINITECH Technologies - I" [4].

However, due to the limited resources of the project to develop new features, the project has followed a 'minimum viable product or service' strategy with three main points to ensure compliance with existing regulations:

- Provide tools for the most important features lacking in scenarios, which have been found to be:
    - anonymization tools
    - pseudonymization tools
- Ensure that the pilots will use only simulated data when some regulatory compliance tools are still pending.
- As the use of simulated data is acceptable for pilots but not for real life, provide a tool for adding regulatory tools in real life. This tool is the Data Protector Orchestrator, which allows adding new regulatory tools to the existing ones without breaking the overall workflow of the system.

Table of Contents

List of Figures

List of Tables

Abbreviations/Acronyms

Abbreviation	Definition
AgI	Agricultural Insurance
AML IV	Anti-money Laundering
BFM	Business Financial Management
DPO	Data Protection Orchestrator
DUOS	Digital User Onboarding Services
eID	electronic IDentification
eIDAS	electronic IDentification, Authentication and trust Services
EO	Earth Observatory
FATF	Financial Action Task Force (FATF)
GDPR	General Data Protection Regulation
IAM	Identity and Access Management
MDM	Mobile Device Management
MiFID	Markets in Financial Instruments Directive
MiFIR	Markets in Financial Instruments and Amending Regulation
NDA	Non-Disclosure Agreement
NIS	Network and Information Systems
OES	Operators of Essential Services
PAN	Primary Account Number
Paas	Platform as a Service
PCI DSS	Payment Card Industry Data Security Standard
PEP	Politically Exposed Person
PET	Privacy Enhancing Technology
PIA	Privacy Impact Assessment
PSD2	Payment Service Directive 2
PSP	Payment Service Provider
PSU	Payment Service User
RA	Reference Architecture
P2PP	Peer-to-Peer Payment
QTSP	Qualified Trust Service Provider
RTS	Regulatory Technical Standard

SA	Supervisory Authority
SCA	Strong Customer Authentication
SECaaS	Security-as-a- Service
SEPA	Single European Payments Area
SME	Small and Medium-Sized Enterprises
SIEM	Security Information Event Management
SSL	Secure Sockets Layer
TI	Threat Intelligence
TRA	Transaction Risk Analysis
3DS	Three-Domain Secure

1 Introduction

The current deliverable is the first one of a series of three deliverables that aim to define and develop regulatory compliance tools. This first one analyses the need of being regulatory compliance in INFINITECH developments, going in depth on them. The developments of the pilots are analysed assessing if they are regulatory compliance and identifying the need of regulatory compliance tools and preparing the field for the development of these tools.

1.1 Objective of the Deliverable

The main objective of this deliverable and of the subsequent deliverables that task "T3.6 Regulatory Compliance Tools" will produce is to ensure that all the solutions created in INFINITECH project are regulatory compliant, while providing relevant regulatory compliance tools that will boost this compliance.

This goal encompasses the following specific objectives:

- **Study in detail the regulations applicable to all the developments** included in the INFINITECH project. INFINITECH project is providing solutions for several pilots with different business objectives, as specified by the end-users of the project (i.e. financial organizations, banks, FinTechs). All the developments in the INFINITECH project are fully focused on the pilot deployments, which target the development of real-life systems that must be regulatory compliant. This deliverable study the solutions provided for every pilot and analyses the regulations that should be applicable for them (based on the previous studies of WP2).

- **Study the technologies that relate to regulatory compliance,** which the project is developing and makes available for use in the INFINITECH pilots. As part of WP2 "Vision and Specifications for Autonomous, Intelligent and Personalized Services" of the project, an initial collection of available technologies has been developed and documented in the scope of INFINITECH deliverable D2.7 "Security and Regulatory Compliance Specifications - I" [1]. In this deliverable the technologies are analysed and the ones related to regulatory compliance are outlined. The latter include the set of technologies that can be used to ensure the regulatory compliance of the INFINITECH solutions, notably technologies related to security, data protection and privacy.

- **Mapping the regulations with the technologies**. This is one of the most important goals of the task. It concerns the study of security and privacy issues that may arise in every pilot and the subsequent search for applicable regulations. Accordingly, a solution for regulatory compliance in the light of the INFINITECH technologies/pilots is sought. Many INFINITECH pilot are producing turn-key solutions that address regulatory compliance issues. In such cases, the role of WP3 "BigData/IoT Data Management and Governance for SHARP Services" is to analyse the pilot solution in order to verify its regulatory compliance.

- **To produce regulatory compliance tools** as needed for the project. This deliverable presents a preliminary overview of regulatory compliance tools for INFINITECH, which will be developed during the 26 months of the task T3.6 "Regulatory Compliance Tools". At the end of this task and in the subsequent deliverables of task T3.6, the project will produce regulatory compliance tools in-line with the needs of the INFINITECH pilots. This deliverable describes an initial design for a general regulatory compliance tool that could be used by, and adapted for, all the pilots in the project. This solution uses the Data Protection Orchestrator (DPO) provided by Atos. The DPO embeds and automate the assurance of security and privacy by design and by default in complex business flows. The DPO tool is described in this deliverable.

However, there are some pilots which, regardless of whether they make use of other regulatory tools, are conceived for providing some functionalities to support some regulations, and therefore each tool (and/or collection of tools) can be considered as a regulatory tool on its own or at least to be their own regulatory tool, too. The pilots that fall under this description are:

- Pilot 3, which aims to provide a full consent management system that allows customers of banks to grant permission to other banks to access their data for specific purposes. As this consent management is one of the obligations resulting from GDPR, pilot 3 may be considered as a regulatory compliance tool, or at least to incorporate its own one;
- Pilots 7 and 8, but as they are confidential, these internal details are located in distinct confidential annexes

1.2 Insights from other Tasks and Deliverables

This deliverable is fully cross-related to other tasks and deliverables in the project.

Let’s analyse first the inputs for this deliverable that are the regulations and technologies and the data governance mechanisms:

- INFINITECH-D2.7 “Security and Regulatory Compliance Specifications-I” [1]. This deliverable is the first version of a total of two deliverables that aim to provide the outcome of task T2.4, whose goal is the specification of the standards and regulations of the INFINITECH project. In this version selects regulations of the INFINITECH project related to the pilots’ use cases. GDPR is given high importance in BigData and analytics scenarios in INFINITECH’s sharp services. Also key regulations such as PSD II, MiFiD II and 4AML are considered for the Financial Sector with respect to the INFINITECH pilot scenarios. D2.7 is providing the main regulations to consider in the pilots to solve the privacy issues that arise in the pilots.

- INFINITECH-D2.5 “Specification of INFINITECH Technologies – I” [4]. This deliverable collects the tools and technologies currently available and in development by the technology partners of INFINITECH. The deliverable also contains specifications of the components, detailing the APIs, functionalities and specifications of the implementation technologies (e.g., BigData/IoT platforms, AI/ML toolkits, HPC infrastructures) that will be used to realize them. The current deliverable INFINITECH-D3.15 has assessed all these technologies to extract the privacy and security technologies that could be considered to participate in the creation of regulatory compliance tools and also the components have been analysed to assess if they are regulatory-compliant themselves.

- INFINITECH-D3.12 “Data governance framework and tools – I” [3]. This deliverable includes a review of the state of the art of the most common data governance mechanisms, including the following technologies: anonymization, pseudonymization, authentication against eIDAS infrastructure and digital mobile onboarding system. It also presents a preliminary design of the tools related with the mentioned technologies. This deliverable is one of the most important and practical inputs for INFINITECH-D3.15, given that the tools developed here will be direct components that will be called in regulatory compliance tools.

- INFINITECH-D2.13 “Reference Architecture-I” [2]. It presents the initial version of the INFINITECH-RA. This initial version is the starting point for the technological developments of the project, followed by the design and initial integration of the use cases. This RA will be used in INFINITECH-D3.15 to analyse the pilots and also to see the integration of regulatory compliance tools or on the other hand, to see where is the pilot implementing regulatory compliance itself.

The outputs of this deliverable will be used by various other WPs. In practice, the most important output from this task will be regulatory compliance tools that will be integrated directly in the INFINITECH pilots in WP7 "Large-Scale Pilots of SHARP Financial and Insurance Services".

1.3 Structure

The structure of this deliverable is directly associated with the objectives described in section 1.1 as shown in the figure below.

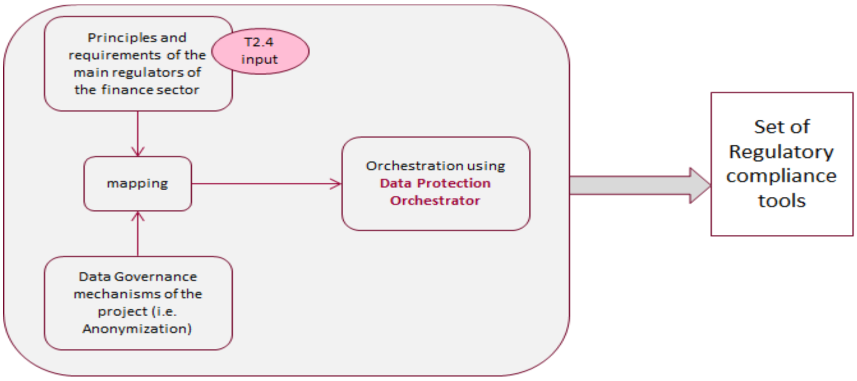
Principles and requirements of the main regulators of the finance sector
T2.4 input
mapping
Data Governance mechanisms of the project (i.e. Anonymization)
Orchestration using Data Protection Orchestrator
Set of Regulatory compliance tools

Figure 1: Regulatory compliance tools: structure

- Section 2 considers the Regulation of the financial sector based on the work done in INFINITECH-D2.7 [1] and studies the regulations applicable to every pilot.
- Section 3 collects the technologies for security, privacy and data protection based on the work performed in INFINITECH-D2.5 [4].
- Section 4 maps the regulations with the pilots, analysing the privacy and security issues and trying to give a solution through a regulatory compliance tool.

Since that this deliverable analyses in detail all the pilots, and the pilot names will be used recursively, the short names of the pilots will be used (in the format "Pilot #1" for example). The following table shows the mapping of short names versus long names of the INFINITECH pilots and will be used throughout this deliverable to know the long names of the pilots. Note that in this deliverable we do not address Pilot #15 since this pilot has recently been added to the INFINITECH project and there isn't enough information about it to perform an analysis yet.

Table 2: Map of INFINITECH Pilots

Pilot short name	Pilot long name
Pilot #1	Invoices Processing Platform for a more Sustainable Banking Industry
Pilot #2	Real time risk assessment in Investment Banking
Pilot #3	Collaborative Customer-centric Data Analytics for Financial Services
Pilot #4	Personalised Portfolio Management – Mechanism for AI based Portfolio Construction
Pilot #5b	Business Financial Management (BFM) tools delivering a Smart Business Advise

Pilot #6	Personalized and Intelligent Investment Portfolio Management for Retail Customer
Pilot #7	Avoiding Financial Crime
Pilot #8	Platform for Anti Money Laundering Supervision (PAMLS)
Pilot #9	Analysing Blockchain Transaction Graphs for Fraudulent Activities
Pilot #10	Real-time cybersecurity analytics on financial transactions’ data
Pilot #11	Personalized insurance products based on IoT connected vehicles
Pilot #12	Real World Data for novel Insurance products
Pilot #13	Configurable and Personalized Insurance Products for SMEs
Pilot #14	Big Data and IoT for the Agricultural Insurance Industry
Pilot #15	Intelligent Market Manipulation Detection: Elevating and Unveiling Hidden Patterns

2 Applicable Regulations in the Financial Sector and the INFINITECH Pilots

2.1 Main Regulations of financial sector

INFINITECH project is studying the security and Regulatory Specifications in WP2 "Vision and Specifications for Autonomous, Intelligent and Personalized Services", the main results are collected in INFINITECH D2.7 "Security and Regulatory Compliance Specifications I" [1].

2.1.1 The General Data Protection Regulation (GDPR)

GDPR is a regulatory framework aimed at providing the means by which citizens can have control over their personal data. Organizations are required to make sure that:

- personal data is gathered legally with the appropriate consent and declarations
- data collected is not misused or exploited for purposes other than for which it was collected
- rights of the data owner(s) are respected in line with the controls as set out in the regulation

The regulation relates specifically from the perspective of the project to the processing of 'personal data' meaning:
"*any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."* [17]
Processing within this context means "*any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movement. "* [18]

GDPR Technology Impact

Data collection is directly impacted by GDPR. Fintechs need to have the procedures, policies and mechanisms in place to demonstrate that their technologies and services are compliant with the regulation. This means maintaining integrity and making sure they have valid and appropriate consent for the customer data they hold, share, use and process.
Compliance breeches can result in financial penalties ranging between 2 and 4% of global turnover depending on the severity and impact of the breech. The use of technologies like blockchain requires considerable oversight to make sure that the way it is used and deployed still facilitates a core principle of GDPR, being the 'right to be forgotten'.
Other important aspects that are important and that should be considered is the need to understand data flows within systems and applications as these days most financial applications do not sit in silos – client data passes to and from multiple systems.
Finally pseudonymization rules are critical to make sure that data access is only ever on the basis of the 'need-to-know' principles. [19]

2.1.2 The Market in Financial Instruments Directive II (MIFID II)

MIFID II is aimed at all financial instruments and covers services including: advice, brokerage, dealing, storage and financial analysis. The directive seeks to harmonise oversight and governance across the industry within all member states.
It has introduced more stringent reporting requirements and tests to ensure transparency and to crack down on the use of 'dark pools' (mechanisms by which trading takes place without identities of individuals or

organizations involved being revealed). The amount of trading that can be done using 'dark pools' is as a result now limited to 8% of the overall trading in any 12 month period.
Algorithms used for automated trading under the directive now need to be registered with the applicable regulator(s), rigorously tested, and circuit breakers have to be in place.

MIFID II also seeks to increase transparency around the cost of services. In so doing, limitations are placed on the payments made to investment firms or advisors by third parties for services that they have or are providing to clients in return. Charges for different services can no longer be packaged up into a single fee. More detailed reporting will be required from brokers on the trades they carry out. Storage of all communications is recommended thereby providing a clear audit trail of all activities. [20]

MIFID II Technology Impact [21]

- **Data storage, aggregation, and analytical requirements:** All information related to trades must be retained. A data retention and archiving strategy is required to support the likely volumes of data and information.
- **Integration between disparate applications:** 'Integration of applications with trading platforms so that key data can flow through becomes a key requirement resulting from MIFID II. API-based integration is seen as the most efficient approach in most cases.
- **Enhanced and transparent client portal:** In order to provide the appropriate protection to investors, it is a requirement to maintain comprehensive client classification and client data inventories.
- **Mobile Device Management (MDM) strategy:** This relates to the need to maintain a record of all telephone calls and electronic communications. The MDM strategy should ensure that the appropriate technology is in place to facilitate this and also restrict the use of mediums of communication such as social media where communications are encrypted thereby making compliance difficult.

Specific MIFID II security requirements

Alongside data retention, security and integrity of data is also critical and could pose a challenge. Appropriate mechanisms to support access control including the use of multi-factor authentication should be put in place.

Monitoring and audit trails of data throughout the data's operational lifecycle is also critical to maintain integrity. Regular audits should be put in place in order to make sure all controls, policies and procedures are being followed.

2.1.3 Payment Services Directive 2 (PSD2)

PSD2 [22] seeks to strengthen the regulations governing online payments within the EU. At its core is the aim of developing a more integrated and seamless payments approach across all the member states of the EU.
A key requirement of the directive is the use of Strong Customer Authentication (SCA). The aim of this is to improve the levels of security around payments and to reduce fraud.
It follows on from PSD1 (adopted in 2007) and as was the case with PSD1:

- Opens up the payments market to new entrants which until now had been limited to those institutions with a banking license.
- Transparency has been increased over services offered and the resulting fees that will be incurred. This also covers both maximum execution times and exchange rates. Development of the Single Euro Payments Area (SEPA) to facilitate the execution of payments has been accelerated as a result of the directive.

The regulation has a strong impact on the whole financial services ecosystem and all the infrastructure behind payments. Furthermore, it has an impact on all businesses who are making use of payment data to better serve their customers.

Security Requirements are introduced for both the initiation and processing of all electronic payments alongside the continued need to protect data belonging to customers (including specific financial data)
Third-Party Providers also fall under the remit of PSD2. This includes all providers who have the right to access and/or aggregate accounts and provide payment services.
In essence, PSD2 is designed to give all customers access to their own data and to encourage greater levels of innovation and resulting competition by encouraging the incumbent banks to engage in secure customer data exchange with other third parties. It should open up the market for organizations in other verticals to access data with their customers' prior consent as long as the appropriate controls are in place.

PSD II Security Requirements/Guidelines

PSD2 maintains the need for Strong Customer Authentication (SCA), **Secured Communication**, **Risk Management**, and **Transaction Risk Analysis** (TRA)
In the endeavor of increasing protection for the consumer, the directive makes it a requirement for banks to implement **multi-factor authentication** for all transactions performed by any channel.
This requirement means making use of two of the following three features:

- Knowledge: information that the customer should only know such as a password, pin code, or a personal ID Number.
- Possession: an artifact that only that the customer has such as a smart card or a mobile handset.
- Inherence: the relation of an attribute to its subject, e.g. a fingerprint.

The elements used need to be mutually independent so that a breach of one cannot inherently mean a breach of any of the others.
Payment service providers (PSPs), as a result of the directive, will need to establish a framework with the required mitigation measures and controls to effectively manage all operational and security risks with respect to services they provide that formally come under the remit of the directive. These measures should include at least:

- The maintenance of an inventory of all business functions, roles and processes thereby making it possible to map functions, roles and processes and all the interdependencies that result from that where they specifically relate to operational and security risks
- Maintain an inventory of information assets, other infrastructure and the interconnection with other systems (both internal and external) so that all assets critical to the core business functions are effectively managed minimizing disruption
- Regularly monitor threats and vulnerabilities to assets critical to the effective delivery of business functions and processes

2.1.4 4th Anti money Laundering (AMLD4)

This directive [23] is designed to ensure that accountability of companies is increased with respect to any connections that can be attributable to money laundering and/or terrorist financing. Failure to comply can bring about both sanctions and reputational damage. Whilst the directive is targeted essentially at banks and other financial institutions, all corporations need to have the appropriate measures and controls in place to maintain compliance.
Key provisions under the directive include:

Key Provisions

- Wider Net – AMLD4 catches more business types than AMLD3. All of the following come under the scope of the directive: gambling institutions, real estate letting companies, individuals or companies making single transactions over €10,000, virtual exchange currency platforms as well as foreign and domestic Politically Exposed Persons (PEPs).

- An increased importance attached to risk requiring companies to be more aware of all risk factors when assessing business activities and functions. Here ten key risks are shown:
    a. Check before engaging in business dealings that owners or People of Significant Control (PSC) are not PEPs (Politically Exposed Persons).
    b. Make more rigorous checks when dealing with business sectors where there is excessive cash. Such industries are known to be key targets of money launderers.
    c. Make more rigorous checks when there is a connection with or when dealing with high-risk sectors. This covers sectors like construction, pharmaceuticals, arms, extractive industries and public procurement.
    d. Check reputable credible media sources for any allegations of "criminality of terrorism".
    e. Investigation into any potential risk of dealings with an organization or an individual who may have a record of frozen assets
    f. Business ownership and control should be clear – if there are any suspicions, complexities or lack of transparency in the structure of ownership then a thorough investigation should be instigated. Proposed business partners should be a "legal person".
    g. Doubts over identity of a beneficial owner should be further investigated
    h. Income and other sources of wealth for any potential partners should be clearly traceable back to their origin.
    i. Increased scrutiny and investigation should be undertaken when dealing with companies who have, or are suspected of having, ties to countries that appear on the sanctions list.
- Companies operating in high-risk countries should be placed under increased scrutiny and they should undergo a more rigorous set of checks. The Financial Action Task Force (FATF) releases a list three times a year which details the qualifying countries. Disclosure – jurisdictions are increasingly encouraged to disclose organizations who continue to break the rules and/or break regulatory frameworks.
- AMLD4 imposes a need for identification, closer scrutiny and increased monitoring of people who are beneficial owners of companies. This can be determined by their shareholding in the business or if they are a PSC. With this in mind, a national register linked to other national registers should be kept so that anyone who needs to can see and access the required information

Compliance professionals as a result need to be able to decide what risk a company that they are working with poses before they go on to investigate their beneficial owners. National registers not being up to date is no excuse for non-compliance.

2.1.5 Basel IV

Basel IV has a specific focus on operational risk management within the context of capital savings. Whilst it is not a direct stipulation, there is a recommendation that supervisory authorities should still maintain a strong focus on making operational risk improvements through strengthening operational resilience. This entails making sure that critical functions within banks are able to continue to operate through any periods of recovery or stress. This incorporates effective management of cyber risks and requires business continuity plans that are regularly reviewed and tested.

2.2 Main Applicable Regulations in INFINITECH Pilot Systems

INFINITECH project is studying the security and Regulatory Specifications in WP2, the main results are collected in section 4 of INFINITECH D2.7 "Security and Regulatory Compliance Specifications I" [1]. The following table collects the regulations that needs to be met for every pilot.

Table 3: Main regulations in INFINITECH pilots

Pilot	Regulations
#1	**None:** GDPR is not applicable because the system does not ever access the data about the customers and the only persons involved are notaries, who are considered as legal persons.
#2	**T**he pilot is engaged with financial markets data rather than personal data of individuals, the GDPR will not be applicable to the service. **MIFID II**: "deals with financial analysis and maintains that it has conflict of interest declaration for each employee in place" [1] As the system only provides advice but it does not carry out any operation on its own, there will be no email, phone call or electronic operation to be recorded, nor any need for a recovery system. However, the access to sensitive financial data from the customer still makes necessary to provide authentication and access control mechanisms.
#3	**GDPR** given that this pilot **"**evaluates how customer, account and transaction data is shared and analysed between banks and FinTechs using APIs to support customer-centric data services. The pilot would rely on a wide number of personal (customer) data, whose aggregation, combination and analysis would involve several implications imposed by the GDPR" [1]. It will however not be applicable at this point as it will initially use synthetic data only that do not allow inferring physical persons.
#4	**GDPR** applies since that original data will come from real clients and will be anonymized. As such, original data will be subject to GDPR and more explicitly to consent and purpose constraints of GDPR, making it mandatory to obtain informed consent of the clients to use them for this purpose. Once the data are anonymized, GDPR will not be applicable.
#5b	MIFID II and GDPR The pilot will use synthetic data, so the regulations don't apply for the pilot
#6	**GDPR** applies since the pilot would process a large number of personal data and create customer profiles (in the case that real data would be used). This could be considered as a high- risk activity considering data protection. In order to solve this issue, the pilot will anonymize personal data. Therefore, the GDPR does currently not apply **MIFID II** applies because this pilot involves making financial recommendations to real customers, even if they anonymized. Such recommendation would be subject to the legal obligations of electronic recording
#7	This pilot is confidential. Thus, in case of specific interest ("need to know"), please contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us . **GDPR** applies **MIFID II** applies **4ML** applies
#8	This pilot is confidential. Thus, in case of specific interest ("need to know"), please contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us . **GDPR** applies **MIFID II** applies **AMLD4** applies

#9	**GDPR** applies because it will collect data from financial transactions, which identify the persons behind them **MIFID II** does not apply because this use case does not provide financial but security consultancy **AMLD4** applies, because this tool may lead to the discovery of illicit transactions, subject to be informed to the authorities
#10	**GDPR** applies given that it uses data from financial transactions. The pilot will only use synthetic data that does not originate from individual persons. Therefore, the GDPR does not currently apply to the pilot.
#11	The data used in the pilot are e.g. location data, speed, acceleration forces which are considered sensitive thus will be handled under the restrictions of **GDPR**.
#12	**GDPR** since the pilot collects data such as vital signs, physical activity and subjective data. Accordingly, these types of data will fall under the GDPR.
#13	**GDPR** applies as long as the content from social media includes the identification of people. Apart from that, the Pilot #13 will use only data on legal persons and entities, which do not fall under the scope of GDPR.
#14	**GDPR**: There will be two different datasets for the pilot. The first one will be the dataset created by the insurance companies for the scope of the pilot implementation and evaluation purpose. This dataset will contain personal data. The second dataset will be the anonymized dataset. It will be provided to the service providers and will be used in the development, calibration and validation of the services that will be implemented in the pilot. Even though the dataset is anonymized, the data will be considered as personal data and will be handled under the restrictions of GDPR.

3 Technologies for security, privacy and data protection

3.1 Data governance mechanisms

According to Gartner IT Glossary, Data governance is “the specification of decision rights and an accountability framework to ensure the appropriate behaviour in the valuation, creation, consumption and control of data and analytics” [5] . That is, it is an essential axis that guarantees data security and establishes which is the route to follow in the management of information within all companies, and particularly in the Fintech ones.

Based on the above, within the task “T3.5 Data Governance Mechanisms” of the INFINITECH project some data governance building blocks are being developed. Specifically, the project will provide:

- A **pseudonymization tool** that supports pseudonymization of unique identifier and generalization of numeric and time-stamp fields.
- A **tool for anonymized data** that determines automatically the best anonymization configuration for each application.
- A **solution for authenticating** citizens and/or businesses **against the eIDAS infrastructure**, providing a cross-border strong authentication mechanism based on eIDs.
- A **mobile digital user onboarding services** with virtual eID derived from government-issued documents (ePassport or eID card).

Pseudonymization “is a security technique for replacing sensitive data with realistic fictional data that cannot be attributed to a specific individual without additional information which, according to GDPR, is to be kept separately and subject to technical and organisation measures to ensure non-attribution to an identified or identifiable person” [6]. In order to ensure privacy, there exist several pseudonymization methods that can be applied to data to either pseudonymize direct identifiers or scramble the data. Counter and random number generator (RNG), cryptographic hash function, symmetrically encrypted identifiers or data masking are examples of generating the pseudonyms out of raw identifiers [6]. The **pseudonymization tool that is being developed within the project** will support pseudonymization of unique identifier and generalization of numeric and time-stamp fields. It will work in batch mode by using a REST API and a configuration file provided by the user including input data attributes, the corresponding level of pseudonymization and the required type of generalization for numeric and time stamp data will be necessary.

Data anonymization (or de-identification) tries to handle personal data in order to irreversibly prevent identification [7]. This means that the information that can be used to link the data back to an individual is removed or transformed in such a way that the remaining data cannot be used to breach users’ privacy. The problem is that applying data anonymization correctly is a challenging task and it is necessary to assess its success, typically by measuring an individual’s risk of re-identification [8], [9]. Most of the anonymization techniques are based on two main methods: randomization and generalization methods. Noise addition [10], permutation techniques [15] or differential privacy [16] are examples of the first option, among others. Regarding the generalization of the information contained in the dataset, *k*-anonymity [13], *l*-diversity [14], *t*-closeness [15] are other alternatives to anonymize data. One of the solutions that INFINITECH will provide is a **tool for data anonymization** that determines automatically the best anonymization configuration for each application. Different anonymization algorithms will be applied to avoid the appearances of data combinations that could lead to a possible re-identification of the data subjects. Additionally, the remaining risk after the dataset anonymization and the quality of the data after anonymization can be measured using a set of privacy and utility metrics.

eIDAS (electronic Identity And trust Services) Regulation is defined in EU 910/2014 [16]. It establishes: (i) a legal framework for EU Digital Single Market in secure and cross-border transactions, (ii) a trust model for mutual recognition using e-identification means, and (iii) an e-ID and electronic Trust Services such as electronic signatures, electronic seals, time stamping, electronic registered delivery service and website authentication [25]. This regulation contributes to secure cross-border electronic transactions and central building blocks of the Digital Single Market. Within the project, a **solution for authenticating citizens and/or businesses against the eIDAS infrastructure** will be developed, named SPeIDI (Service Provider for eIDas Integration). It will provide a cross-border strong authentication mechanism based on eIDs and will support authentication for citizens, compatibility with eIDAS Network, use of eID issued by European National authorities according to the EU eID schemas, strong cross-border authentication using more secure credentials and UI easing usability and privacy.

Digital user onboarding is the process of enrolling new users ensuring that they can access all the services and products contracted in a remote and secure way. The procedure evolves two technical challenges: authentication and authorization or access control. The former is intended to check the identity of the customer while the latter checks what functionalities they are entitled to access. Virtual remote electronic identification (eID) allows this type of onboarding. INFINITECH onboarding system will be an adaptation of the DUOS (Digital User Onboarding System) developed in [24]. It allows for remote user registration using eID or electronic password and provides multi-factor authentication combining images of the face of the user with the certificates stored in the e-ID or passport.

A description in detail of the work that is being carried out in the task 3.5 can be checked on the deliverable "D3.12 - Data Governance Framework and Tools - I".

3.2 Overview of INFINITECH Security, Privacy and Data protection Technologies

The following table collects all the technologies available in INFINITECH project included the ones described in section 3.1 provided by the Data governance mechanisms deliverable. Some contents of this table have been taken from [4] which explains all the technologies available in INFINITECH.

Table 4: INFINITECH Technologies for regulatory compliance

Name tool / platform	Company	Relevance and applications for regulatory compliance
Service Provider eIDAS Integration (SPEiDI)	Atos	This technology enables the connection of online services with eIDAS infrastructure, permiting European eID use. It allows customer authentication against the pan European eID infrastructure allowing cross-border transactions. This connector requires: ● The use of eID schemas under eIDAS https://ec.europa.eu/cefdigital/wiki/display/EIDCOMMUNITY/Overview+of+pre-notified+and+notified+eID+schemes+under+eIDAS ● The use of eIDs valid for countries that notified their participation in eIDAS network This technology could be potentially used in any other INFINITECH tool, to identify users by their national digital identity. It enables cross-border usage scenario of INFINITECH tools, including privacy-preserving scenarios. Examples of possible uses could be:

		● Tracking user access to personal or sensitive data (data treatment); using digital identities for identifying users/operator, clearly define responsibilities in case of abuses or disputes. ● Digital identity: management of explicit opt-in consent from users, such as the right to request data from companies, and the right to have your data deleted Using correct identification of users by means of digital identities, increases the value provided by single INFINITECH tools managing privacy rights
Data Protection Orchestrator (DPO)	Atos	It allows embedding and automating tools for assessing security and privacy by design and by default in business flows, these being heterogeneous and complex. It orchestrates various privacy and security management functions (such as access control, encryption and anonymization). It is needed for the Swagger specification of the components (PETs) that will be called by DPO via REST The business flows must be studied and developed to perform communication with the components
Digital User Onboarding System (DUOS)	Atos	This solution allows dealing with virtual identities in a mobile device. It allows using eID or passport for remote user registration. This solution uses eIDs issued by European National authorities according to the EU eID schemas: eID cards and Passports In order to integrate DUOS, it is necessary to adapt and customize it for a user's context-of-need (e.g., Bank application) that requires user authentication This technology could be used in INFINITECH to implement "anonymous" user on boarding. The user can be securely identified by eID or e-Passport without revealing any detail about his/her identity.
Botakis Chatbot Development Network	CP	"A tool for rapid development of chatbots applications, which will be used for the development of chatbots, features in the INFINITECH pilots. Enhancements expecting to be achieved for Botakis Chatbot Platform, based on INFINITECH pilots (i.e. notably the GFT and NBG led pilots): - Built-in dialogs that utilize and are integrated with existing NLP frameworks (open or proprietary) provided by partners or every interested party - Powerful dialog system with dialogs that are isolated and composable. - Built-in prompts for simple things like Yes/No, strings, numbers, enumerations. " [4] As part of the available chatbot functionality, it will be possible to include GDPR Consent and manage requests from people exercising: 1.The Right to Be Informed 2.The Right of Access 3.The Right to Rectification 4.The Right to Erasure 5.The Right to Restrict Processing

		6.The Right to Data Portability 7.The Right to Object in the framework of the INFINITECH pilots. Regarding the ability to provide info regarding the 7 points described above, we expect that the relative responses will be included in the questions that the chatbot will cover, so we expect ny relative material to be included as part of the GDPR consent that the Pilots users will have to provide, before accessing the application.
Crowdpolicy Open (innovation) banking solution	CP	"Crowdpolicy Open (innovation) banking platform is a set of predefined and customisable banking web services and data models integrated with our own API Manager that supports access control, monitoring and authentication. This solution puts the bank (or any monetary financial institution) in control of the third-party partner relation. "[4] Crowdpolicy Open (innovation) banking platform mainly covers the requirements for Open Banking APIs as part of the PSD2 Directive, that has several modules that also are API based.
Anonymization Tool	GRAD	The anonymization tool is based on a risk-based approach that modifies data in order to preserve privacy. The tool includes different anonymization algorithms and it will determine automatically which of them (generalization, randomization, deletion, etc.) should be applied in order to preserve the maximum level of privacy for the data. "It also includes a set of privacy and utility metrics that allow to measure the risk that remains after anonymizing the dataset, and the impact of the anonymization process on the quality of the data. The component requires two inputs: the data that has to be anonymized and a configuration file that defines the structure of the data, its location and the privacy requirements. " [4] The anonymization tool is intended to be used in two modes, batch or streaming. In the case of using it in batch mode, the output of the component (anonymized data) is stored in a database. The location of the database has to be known beforehand (through the configuration file that is taken as an input). If the streaming mode is used, the output will be the queue of the service.
Blockchain-enabled Consent Management System	UBI, IBM, INNOV	The blockchain-enabled Consent Management System offers a decentralised and robust consent management mechanism that enables the sharing of the customer's consent to exchange and utilise their customer data across different banking institutions. The solution enables the financial institutions to effectively manage and share their customer's consents in a transparent and unambiguous manner. It is capable of storing the consents and their complete update history with complete consents' versioning in a secure and trusted manner. The integrity of customer data processing consents and their immutable versioning control are protected by the blockchain infrastructure [27]. To achieve this, the solution exploits the key characteristics of blockchain technology to overcome the underlying challenges of trust improvement,

	capitalising on its decentralised nature and immutability due to the impossibility of ledger falsification. The usage of blockchain enables extensibility, scalability, confidentiality, flexibility and resilience to attacks or misuse, guaranteeing the integrity, transparency and trustworthiness of the underlying data. The complete documentation of the described solution is available in deliverable D4.7 “Permissioned Blockchain for Finance and Insurance – I” of WP4 [27].

4 Mapping regulations with technologies

4.1 Pilots, regulatory issues and solutions

The following table contains a brief summary about all the aspects regarding security, privacy and data protection for every pilot of INFINITECH project. Specifically, the aspects summarised include the issues, the regulations to fulfil, the available technologies in the project and the solutions that should be applied in every pilot.

The table below has the following columns:

- *Pilot*: number of the pilot, the long name for each pilot is in Table 2.
- *Security and Privacy Issues*: a short description of the issues of the pilot.
- *Regulations*: the regulations that will be applicable in order to avoid the security and privacy issues. The regulations considered for INFINITECH are GDPR, PSD II, MiFiD II and 4AML.
- *Technologies*: the technologies for privacy/security that will be used in the pilot and will help to solve the issues.
- *Solution for the pilot*: there are two cases:
    1. The pilot provides a solution internally: in this case this field explains the solution adopted internally in the project
    2. The pilot requires a regulatory tool to solve the issues: in this case, this field explains the desired solution for the pilot

Table 5: INFINITECH pilots, privacy issues and solutions

Pilot	Security and privacy issues	Regulations	Technologies	Solutions
#1	This pilot aims to extract information automatically from notary invoices. It extracts from the invoices tables with amounts and their values from the invoices. The system does not ever access the data about the customers. Since the notary is a legal person there are no privacy issues.	Considered regulations are Not Applicable, including GDPR, because the only persons referred to in the pilot are notaries, which are considered legal persons instead of natural persons. The only data appearing in the invoices is the name of the notaries and VAT number, which is, in turn, the legal name of the company. No data from clients' mortgages appear	Not Applicable	No solution is applicable because no regulation applies

		in the invoices at all. The “notary invoice processing solution” has also been submitted to an Operational and Technological Risk Committee that has assessed data used and has concluded that data is not subject to GDPR. This information has also been communicated to the Bank of Spain.		
#2	“The pilot provides risk-assessment analytics on the fly for bank traders, risk managers and sales negotiators based upon Value-at-Risk and Expected Shortfall procedures. It estimates market risks and pre-trade risks thus facilitating decision making processes for traders.” [1] Since the pilot is engaged with financial markets data rather than personal data of individuals, there are no privacy issues.	Production: GDPR, 4AML, BASEL IV and MIFID II Pilot: none since the pilot does not consider real customers PSDII does not apply because the functionalities of the pilot do not deal with transactions	IAM and Cryptography (GDPR) Audit logs (MIFID II)	GDPR: the pilot does not use sensitive data. (the data used is “market data” which is proprietary, but not confidential and trade date which is generated synthetically). MiFiDII: the pilot does not involve any transparency issues and the pilot does not consider real customers 4AML: the pilot does not consider real customers
#3	Platform for sharing financial data	Production: GDPR: Need for authentication and authorization Pilot: Not Applicable	The use of personal data in the demonstrator would lead to applying all GDPR	Not Applicable: The pilot will work only with synthetic data

#4	The main goal of this pilot is to explore the possibilities of AI Based Portfolio construction for Wealth Management. The investor requires access to his/her personal portfolio in a secure way.	Production; MIFID II and GDPR Pilot: Partial GDPR for strong authentication	DUOS: Digital onboarding Authentication	Pilot #4 solves the secure access to the personal portfolio internally by allowing secure onboarding authentication to the investor providing him/her with access from his/her mobile phone to the bank services through DUOS
#5b	Business and financial consulting	Production: MIFID II and GDPR Pilot: None due to using synthetic data	Pilot: None due to using synthetic data	This pilot will use only synthetic data to avoid being subject to MIFID II and GDPR
#6	Personalized and intelligent investment portfolio management for Retail Customer	GDPR	need to either secure the data or anonymize them	Use of Icarus Platform to anonymize the data
#7	This pilot is confidential. Thus, in case of specific interest ("need to know"), please contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us .			
#8	This pilot is confidential. Thus, in case of specific interest ("need to know"), please contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us .			
#9	Real-Time Cybersecurity analytics on financial transactions' data	GDPR for real-life production, none for pilot.	None	Solutions: the pilot will only use synthetic data that does not originate from individual persons, but is created by a machine. Therefore, there are no privacy issues.
#10	Real time security analytics for financial data	GDPR on production mode.	Financial transactions include the people involved in them, making them become personal data	Use of synthetic data to circumvent GDPR
#11	Driver characteristics data and GPS position of the user will be	GDPR	Anonymization tool,	The pilot solves this internally by means of the security

	collected and analysed in the AI INFINITECH platform		IAM and consent management	framework (IAM and Consent management) provided by ATOS however a regulatory compliance tool that anonymize the location data will be applied.
#12	Physical activity of the user will be collected and analysed in the INFINITECH platform	GDPR	Anonymization tool, Access control	The pilot solves this internally by means of the access control framework, that ensures that only the specific user can consult the data. Moreover, a regulatory compliance tool that anonymize personal data will be applied.
#13	The main goal of Pilot #13 is to develop an insurance product configuration platform for SMEs, which will leverage large amounts of digital data in order to compute the insurance offering. An automation of the subscription process will help the insurance company to reduce costs. No security or privacy issues were found. Privacy Issues: the pilot will only use info from legal persons and it never will use personal data, so there are no privacy issues.	Considered regulations are Not Applicable.	Since the subjects to be analyzed are legal persons there are no privacy issues.	It collects driver characteristics data such as GPS position Before collecting data, there will be a manual phase of asking for the consent of the user The data are pseudonymized and the GPS position is anonymized. It will implement Role Based Access Control (RBAC)
#14	The main goal of Pilot #14 is configurable and personalized insurance products for SMEs and Agro-Insurance.	Considered regulations are: - GDPR,	The communication with the services will be done through secure and encrypted	Data anonymization: The data will be provided to the processor

	The pilot will evaluate the risks for insurance companies and offer more personalized products. The data collected for this purpose will be stored in the handler's database and no access will be allowed by third parties. Moreover, Confidentiality Agreements will be signed by the involved parties, clarifying the roles (handler, processor, owner) and how the data will be handled during the project.		connection, using the SSL protocol and the access to the data will be done only after authentication through credentials and authorisation of the user. Moreover, the server will be enhanced with firewall and the access will be authorised only to the services and to the SSL protocol and the access will be allowed only to the IP of the service providers company.	anonymized by the insurance companies since no personal data are required in order to deploy the pilot. Each field (the geospatial information along with the aforementioned details) can be accompanied by a specific id (e.g. 1, 2, 3, etc.) and no personal data can be provided. The pilot solves this internally by means of the security framework provided by AGA however a regulatory compliance tool that anonymizes the location data will be applied.

4.2 General definition of an INFINITECH Regulatory Compliance Tool

In INFINITECH, a regulatory compliance tool is a piece of software that provides the additional functionalities needed to comply with the regulations applicable to each use case, as defined in D2.7 "Security and Regulatory Compliance Specifications - I".

Due to the limited resources available in the project for developing regulatory compliance tools, its general philosophy is to use preferably those ones that are already being used by the users in their normal operations, and develop new ones only for those functionalities that are not already covered by their existing tools.

4.3 Solutions for regulatory compliance in the pilots

This section addresses every pilot and shows a summary of all them, explaining the following items:

- The security and privacy issues
- The initial solution needed, explaining how the pilot will solve the issues; in some cases the solution comes from a regulatory tool that will be prepared in this task 3.6
- The architecture for the pilot showing the components of the solution
- Results of the pilot applying the solution

Most of the pilots overview subsection have been written based on [2] which details the INFINITECH RA for every pilot

4.3.1 Pilot #1: Solution for regulatory compliance

Pilot overview

Pilot #1 aims at developing, integrating and deploying a data-intensive system to extract information from notarised invoices, in order to:

- "Establish the sustainability index of each notary based on the number of physical copies that are issued.
- Provide financial institutions with the information (properly indexed) about the documents that are finally generated by notarial services required by the bank.
- Promote notarial services from those with the higher sustainability score.

Pilot #1 will extract data from 32.300 real invoice documents from 3.000 different notaries extracted from Bankia systems. " [2]

Security and privacy issues and requirements

Pilot #1 processes invoices from notaries, which are considered legal persons, thus no personal data is processed.

The only data appearing in the invoices is the name of each notary and their VAT number. No data from clients' mortgages appear in the invoices.

The "notary invoice processing solution" has also been submitted to an Operational and Technological Risk Committee that has assessed the data used and has concluded

Solution

Not Applicable

Architecture

As this pilot will include no regulatory compliance tools, its architecture does not vary from its original one:

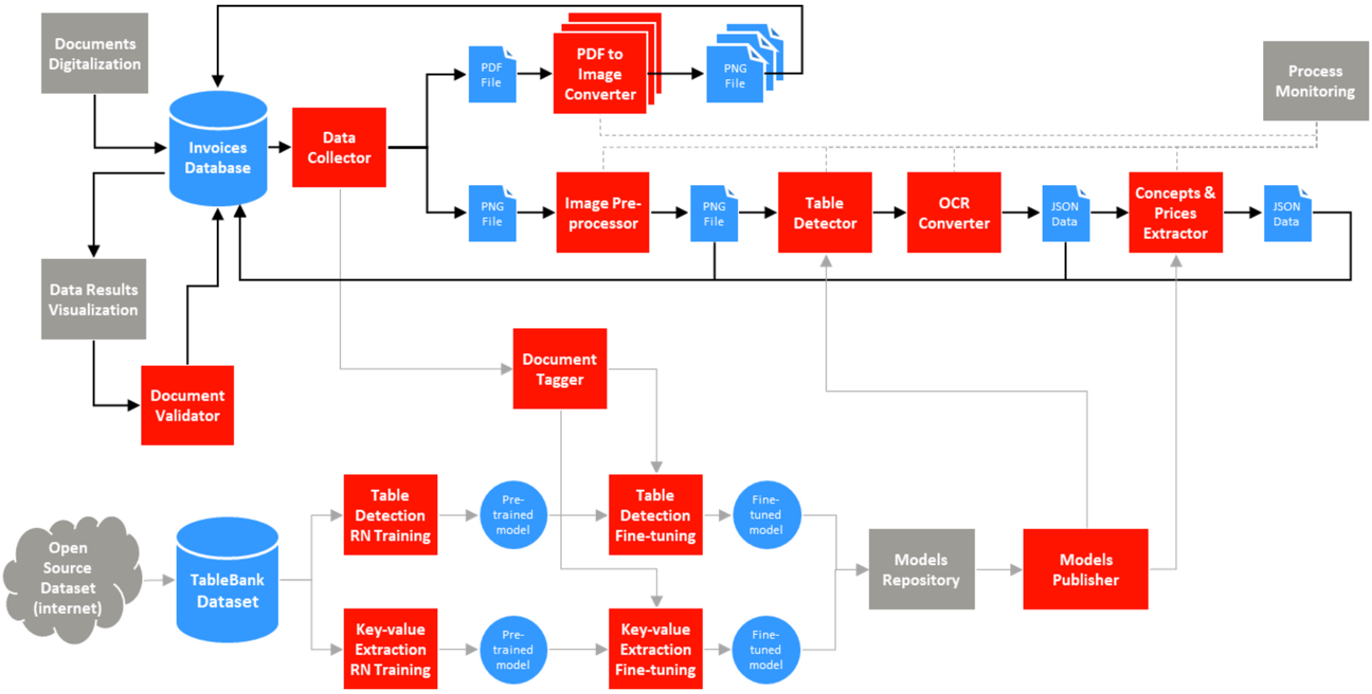

Figure 2: Pilot 1 INFINITECH RA

Expected results

Not Applicable

4.3.2 Pilot #2: Solution for regulatory compliance

Pilot overview

The high-level aim of pilot 2 is to provide "bank traders real-time information about financial assets they may wish to trade, ultimately enabling improved decision making and hence profit margins for their customers. Currently, trading information and future predictions are updated infrequently (once a day), meaning that traders are unable to exploit rapidly changing market conditions" [26]. Pilot 2 should solve this issue by providing a solution that can provide aggregate market data, trends and predicted risk/yield that updates in real-time.

Security and privacy issues and requirements

Security issues relate to the authenticity and availability of the input data. Privacy issues may arise from utilization of trade data, consisting of time and price of the trade in connection with an account number. The account number is sensitive information, but necessary to distinguish between different portfolios.

Solution

For protecting account numbers, the pilot will replace real account numbers by placeholders.

Architecture

As this pilot will include no regulatory compliance tools, its architecture does not vary from its original one:

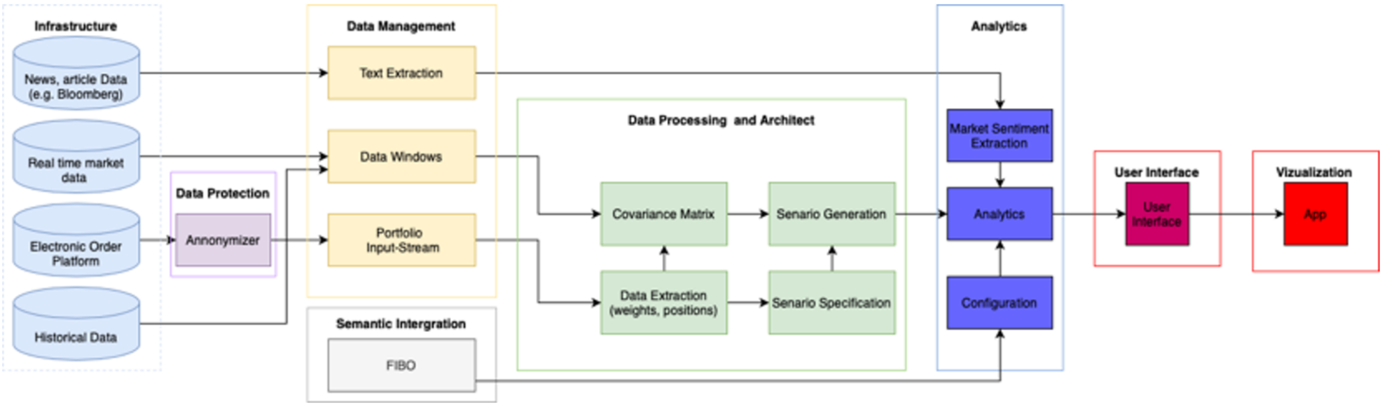

Figure 3: Pilot 2 INFINITECH RA

Expected results

Not Applicable

4.3.3 Pilot #3: Solution for Regulatory compliance

Pilot overview

Permission based service, for personal and business customers, who want to share financial information, including identity and contact details, securely and digitally, similar to message sharing. Sharing data held by financial institutions today is not easy and currently involves copying documentation and statements, emailing pdfs, or sharing login credentials. Sharing your data easily and securely makes life easier and enables taking advantage of new services. It is safe because the financial organisations you already trust to take care of the money will help to share financial information securely. Customers manage all their sharing permissions through easy to use and transparent consent management dashboard.

Security and privacy issues and requirements

In real life, as data would be linked to a real person, they would be personal data and their sharing would imply that both the application sharing the data and the one receiving it would become data processors. Additionally, sharing personal data needs managing the consent for the treatment of those data, which in this pilot is bypassed by using only synthetic data from non-real customers

Architecture including the solution

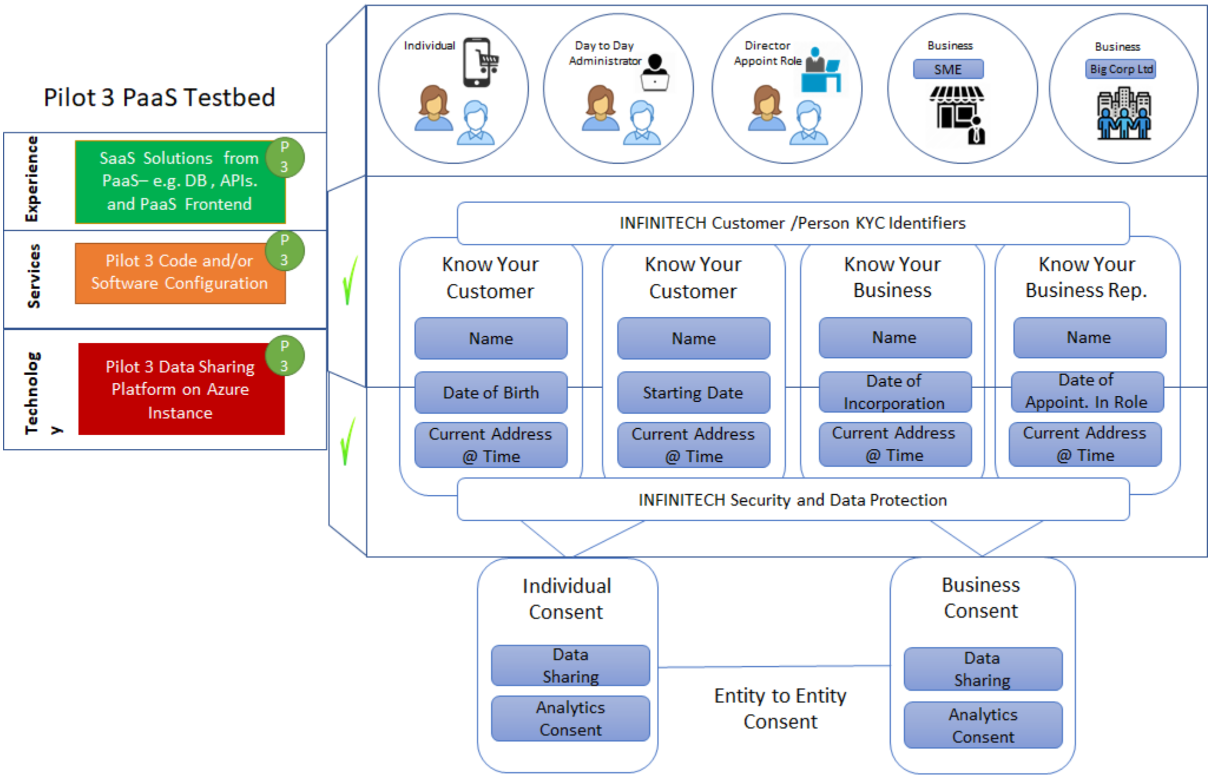

Figure 4: High level architecture diagram for Pilot 3

Expected results

Not Applicable

4.3.4 Pilot #4 - Personalised Portfolio Management – Mechanism for AI-based Portfolio Construction: Solution for regulatory compliance

Pilot overview

The goal of this pilot is to explore the possibilities of AI-based Portfolio construction for Wealth Management in general regardless which amount is to be invested. This allows Portfolio construction and optimization for all the customers, not only for the ones with more wealth.

In this pilot there will be a first phase of customer onboarding in which the investor will authenticate in his/her mobile phone and will access the bank application. The bank application could offer several services such as uploading relevant personal portfolios or starting a portfolio optimization process. The investor will select the fitness factors and constraints or preferences to perform the portfolio construction, basing themselves on the client's risk profile and his/her preferences.

Some of the data to be used by this pilot will be Customer Transactions Data, Financial Market Price Data or Financial Market Asset Master Data. "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." [2]

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.

Security and privacy issues and requirements

There are two different parts in the pilot:

- Customer authentication: the client must be authenticated in a secure way to get the results of the pilot
- AI Based Portfolio construction and optimization for Wealth Management: the data source is 300+ publicly available newsfeeds which are public by definition. No previously collected Data will be used, nor any private data are gathered

Solution

For customer authentication, the pilot is adopting a solution using DUOS (Digital User Onboarding System), a solution for dealing with virtual identities in a mobile device. This solution comes from Atos and it is described on section 3.4 of INFINITECH D3.12 “Data Governance Framework and Tools – I” [3].

Architecture including the solution

The following diagram shows the architecture of pilot #4 (it has been taken from section 6.4.5 of deliverable INFINITECH D2.13 “Reference Architecture – I” [2] including DUOS solution for digital onboarding, that will help the user to authenticate and start using the services that the pilot is providing for his/her portfolio. DUOS is an Authentication method inside Cross-cutting services inside INFINITECH-RA.

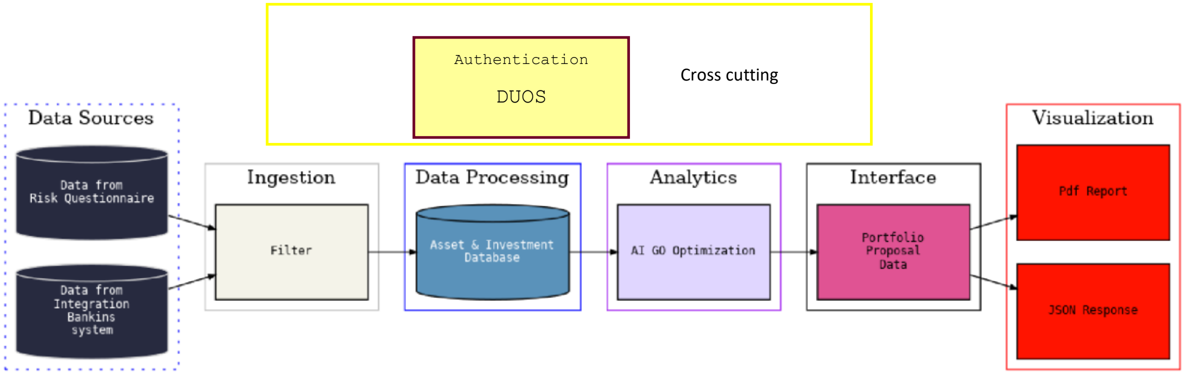

Figure 5: INFINITECH Pilot #4 Pipeline in-line with the INFINITECH-RA including compliance solution

Expected results

The solution for user authentication using DUOS will cover the secure authentication needed for the user to request his/her portfolio.

4.3.5 Pilot #5b: Solution for regulatory compliance

Pilot overview

The pilot aims to assist SME clients of 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.” [2]

Security and privacy issues and requirements

The main security and privacy issues related with this pilot comes from monitoring of transactions and the issuing of invoices, as they will contain information about the physical persons involved in them.

One of the main issues here is whether the data being processed could contain information that can identify the clients and their providers, specially data related to invoices.

Additionally, the main goal of this pilot is to produce recommendations for the clients, which makes this pilot subject to MIFID II and its obligation of recording every conversation with the client, including any recommendation sent to him

Solution

The data supplied to the system will be pseudonymized by BOC before providing them to UPRC for their processing. As they are different institutions, it will be impossible for UPRC to identify the identity of the involved persons, which in practice leads to those data being anonymized for UPRC.

To comply with the need to record the recommendations to the user, the application will produce logs whenever a recommendation is made.

Architecture

The figure below shows how the pseudonymisation module fits in the system. Note how all data travelling from one module to other passes through the pseudonymization

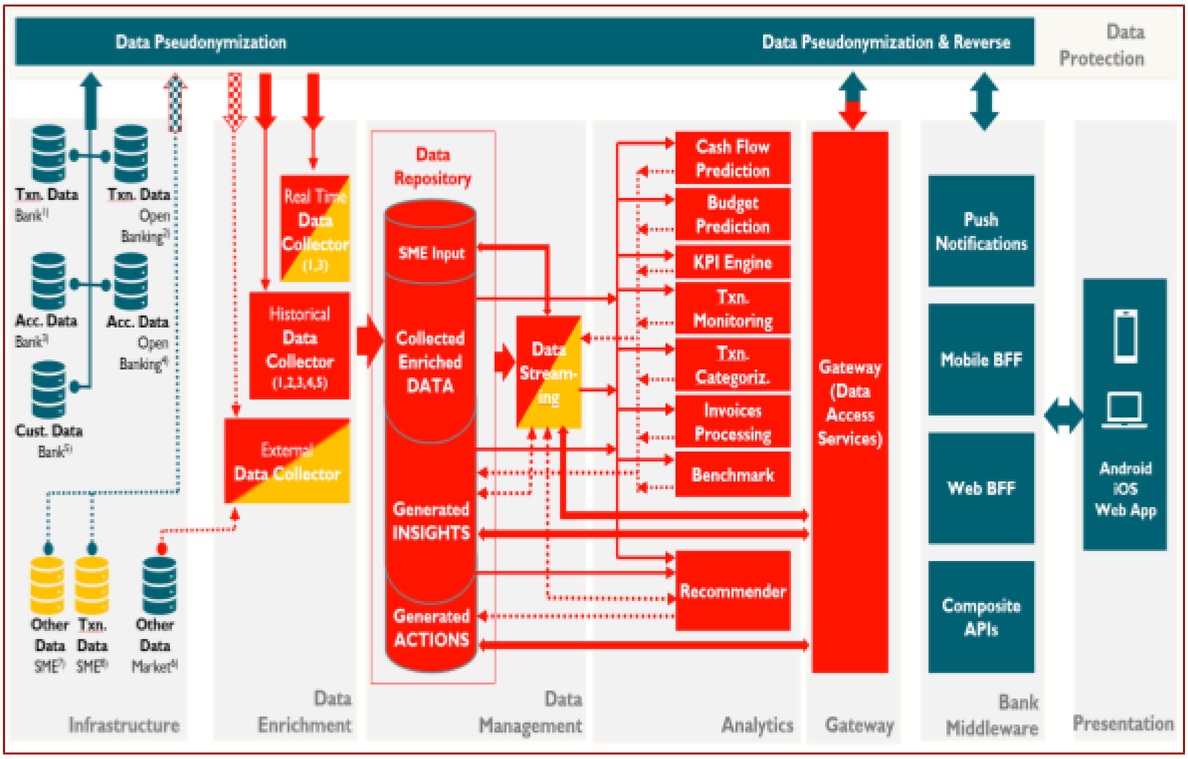

Figure 6: INFINITECH Pilot #5b RA

Expected results

Useful and operational-grade financial advice to BoC customers.

4.3.6 Pilot #6: Solution for regulatory compliance

Pilot overview

This pilot aims to leverage large customer datasets and large volumes of customer-related alternative data sources (e.g., social media, news feeds, etc) in order to explore the user benefits of the process of providing more targeted, automated, effective, investment recommendations to retail customers.

.

Security and privacy issues and requirements

The main privacy issues related with this pilot comes from processing data from customers and creating profiles.

Solution

The customers personal data are anonymized in order to avoid their identification.

Architecture

The figure below shows how the anonymization engine fits in the system. Note how all data travelling from one module to another passes through the anonymization.

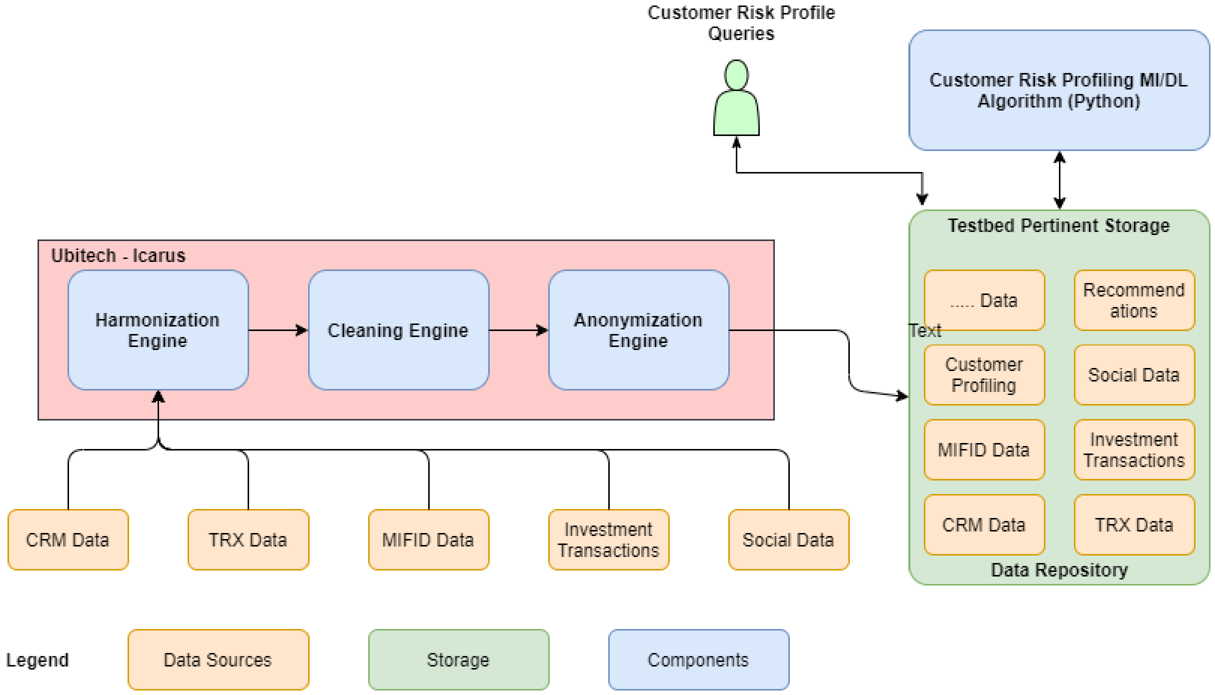

Figure 7: INFINITECH Pilot #6 RA

Expected results

Useful and operational-grade financial advice to NBG customers.

4.3.7 Pilot #7: Solution for regulatory compliance

This pilot is confidential. Its description and solutions are delivered as a confidential separate document.

4.3.8 Pilot #8: Solution for regulatory compliance

This pilot is confidential. Its description and solutions are delivered as a confidential separate document.

4.3.9 Pilot #9: Solution for regulatory compliance

Pilot overview

“Blockchain crypto currencies and tokenized assets that are obtained fraudulently can go through various transfers” [2] on the blockchain and end up as stable coins (e.g. USD, EUR, TRY tokens) in different jurisdictions. P9 PoC will demonstrate (i) parallel transaction graph construction and graph traversal-based analysis on an HPC cluster and (ii) its user interface that provides transaction graph visualization.

Security and privacy issues and requirements

Financial transactions include the identity of involved persons, who might potentially be physical ones, implying the need for compliance with GDPR. Additionally, should any fraudulent transaction be identified, it would have to be reported to national authorities.

Solution

The pilot will not use any data that allows identifying persons, making GDPR not applicable to it, and removing the need for regulatory compliance tools

Architecture

As this pilot will include no regulatory compliance tools, its architecture does not vary from its original one:

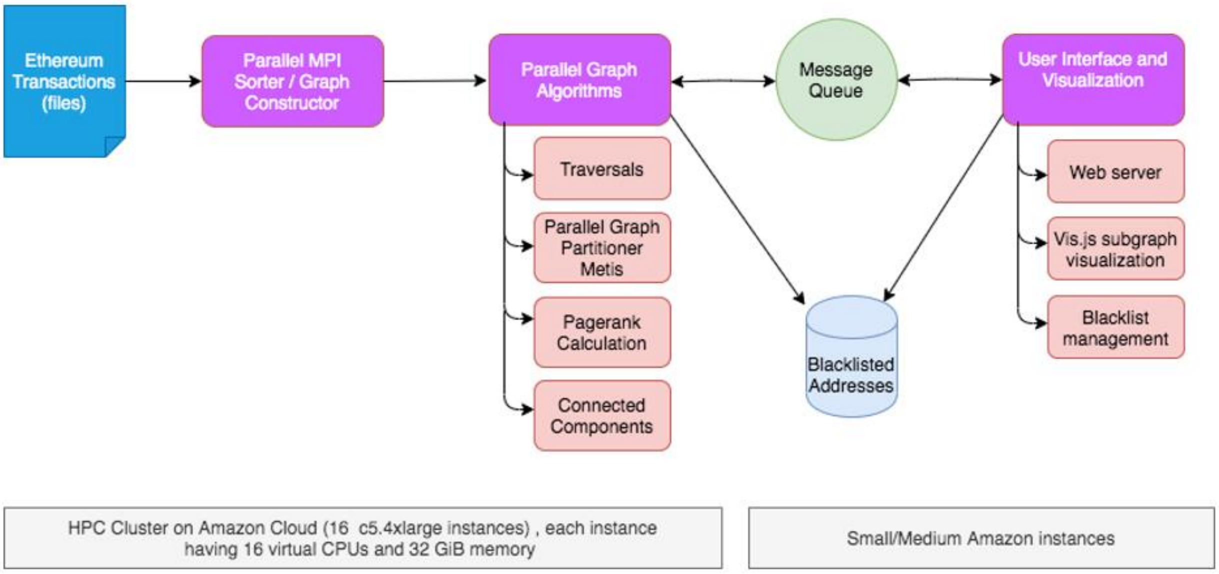
Ethereum Transactions (files)
Parallel MPI Sorter / Graph Constructor
Parallel Graph Algorithms
Traversals
Parallel Graph Partitioner Metis
Pagerank Calculation
Connected Components
Message Queue
Blacklisted Addresses
User Interface and Visualization
Web server
Vis.js subgraph visualization
Blacklist management
HPC Cluster on Amazon Cloud (16 c5.4xlarge instances) , each instance having 16 virtual CPUs and 32 GiB memory
Small/Medium Amazon instances

Figure 8: INFINITECH Pilot #9 RA

Expected results

Proof of concept of identified fraudulent activity, based on synthetic data.

4.3.10 Pilot #10: Solution for regulatory compliance

Pilot overview

Pilot #10 tries to significantly Improve the detection rate of malicious events (i.e. frauds attempts) and enable the identification of security-related anomalies while they are occurring by the analysis in real-time of the financial transactions of a home and mobile banking system.

Security and privacy issues and requirements

Financial transactions include the identity of involved persons, who might potentially be physical ones, implying the need for compliance with GDPR. Additionally, should any fraudulent transaction be identified, it would have to be reported to national authorities.

Solution

The pilot will use only synthetic data, making GDPR not applicable to it, and removing the need for regulatory compliance tools and reporting to national authorities.

Architecture

As this pilot will include no regulatory compliance tools, its architecture does not vary from its original one.

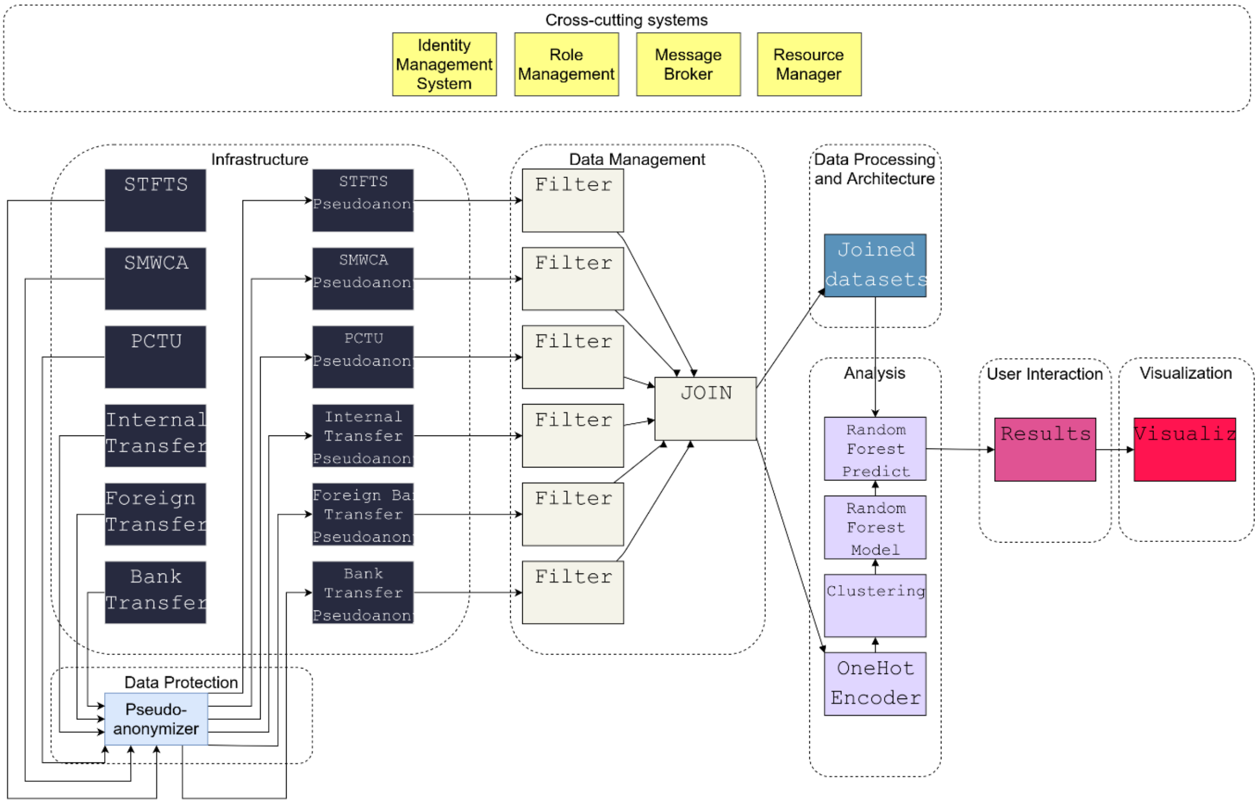

Figure 9: Pilot #10 INFINITECH RA

Expected results

Proof of concept of identified fraudulent activity, based on synthetic data.

4.3.11 Pilot #11: Solution for regulatory compliance

Pilot overview

The aim of this pilot is to improve the analysis, definition and assignment of risk profiles in car insurance, by using the information collected from connected vehicles and applying Artificial Intelligence technologies. The pilot will develop a “Pay as you drive” service to adapt insurance costs to drivers’ classifications; and a “Fraud detection” service that exploits driving profiles to identify possible drivers and helps insurance companies with this issue.

With this purpose, the tool *driver profile collection* (CTAG) will take driving characteristics data such as technical vehicle information (speed, acceleration, breaks, etc.) and location (GPS position) which can be considered as sensitive data. These collected datasets, grouped in “routes”, are necessary to define, train and test the AI models that the pilot is providing.

Security and privacy issues and requirements

The pilot has two security and privacy issues, on the one hand, the risk of unauthorized access to modules of the platform and on the other hand, the possible use of sensitive data for training models.

Solution

For preventing unauthorized access to the modules, the pilot is adopting a security framework provided by ATOS that implements and provides OAuth 2.0 based authentication and identification mechanisms. According to the data collection, the driver will answer to an “ask for consent” for the data treatment. Some of the collected data will be pseudonymized by the driver profile collection tool, that is, a user identifier is associated with the user, and other personal data, such as GPS position, will be anonymized to protect user privacy. Technical vehicle and location data will be stored and classified in the AI INFINITECH P#11 platform, also with the data coming from route simulations. The pilot will offer, as the main outcomes, two services (“Pay as you Drive” and “Fraud Detection”) to be consumed by the insurance company, using their own driving datasets. Thus, P#11 AI platform will neither store linked real driver data used to classify a real driver (data from the insurance company) nor share data used to define and train AI models.

Architecture including the solution

The figure below shows the high-level architecture for this pilot, where the security framework and the anonymization tool (anonymizer) can be found.

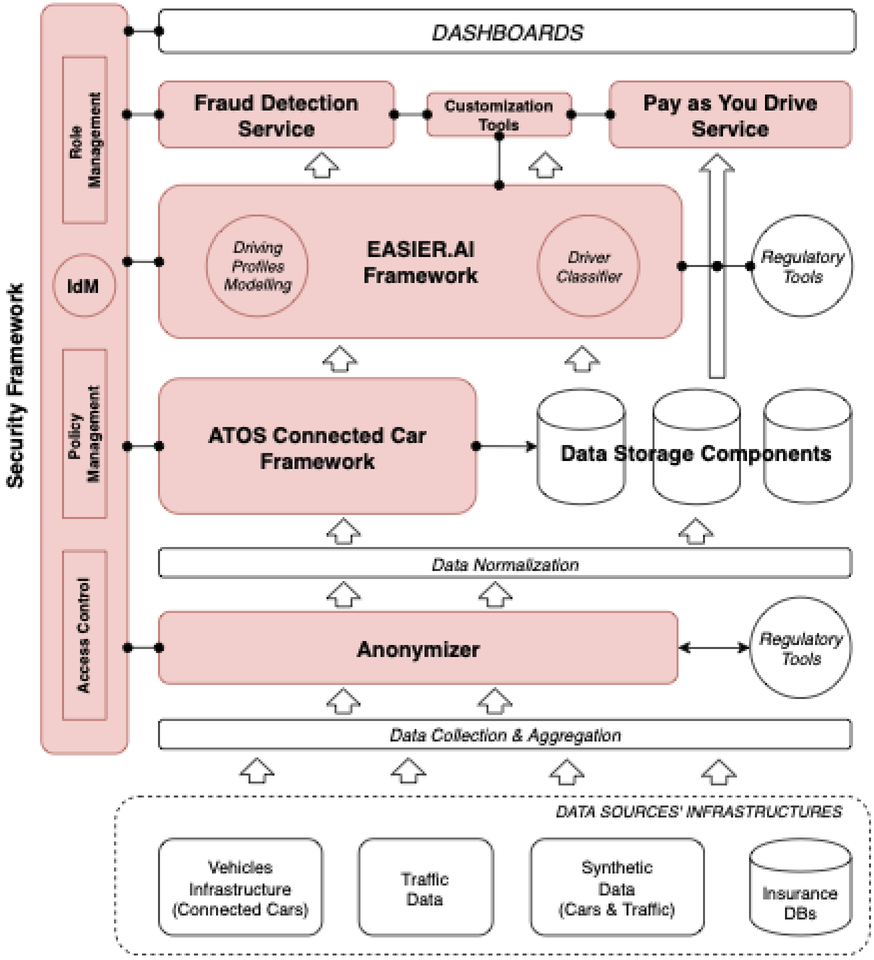

Figure 10: High level architecture diagram for Pilot 11

Expected results

Not Applicable

4.3.12 Pilot #12: Solution for regulatory compliance

Pilot overview

The aim of this pilot is to improve the analysis, definition and assignment of risk profiles in health insurance, by using the information collected from IoT devices and questionnaires, and applying ML technologies. The pilot will develop two distinct services, one performing risk assessment and another one for fraudulent behaviour detection.

To this end, the Healthentia app will collect data from pilot users by means of different activity trackers (Fitbit devices, Android phone sensors and Apple Health Kit) and questionnaires, from psychological to social and environmental aspects, as well as synthetic data (simulated lifestyle) will be collected in the context of the pilot. Thus, the collected data from the devices are sensitive because they are related to physical activity and mood of users.

Once collected, the data will be stored in the INFINITECH platform and will be used to train models in order to obtain a score for each user. The health insurance companies will use this score to adapt the price of their customers' premium.

Security and privacy issues and requirements

There exist two different security and privacy issues:

- The user authentication
- The use of personal data for training the ML models

Solution

As a solution for regulatory compliance, the Healthentia application's users will each sign a consent form in order to collect and use their data. There will be developed for this project a regulatory compliance tool to call an anonymization tool developed by GRAD; this regulatory tool will protect the user's privacy. The regulatory tool is based on the DPO developed by Atos that orchestrates the calls to different security or privacy tools; in this case it is only required to call Anonymization.

According to the user authentication, the pilot solves this internally by means of the access control framework.

Architecture including the solution

The figure below shows the high-level architecture for this pilot where the access control framework and the anonymization tool can be found.

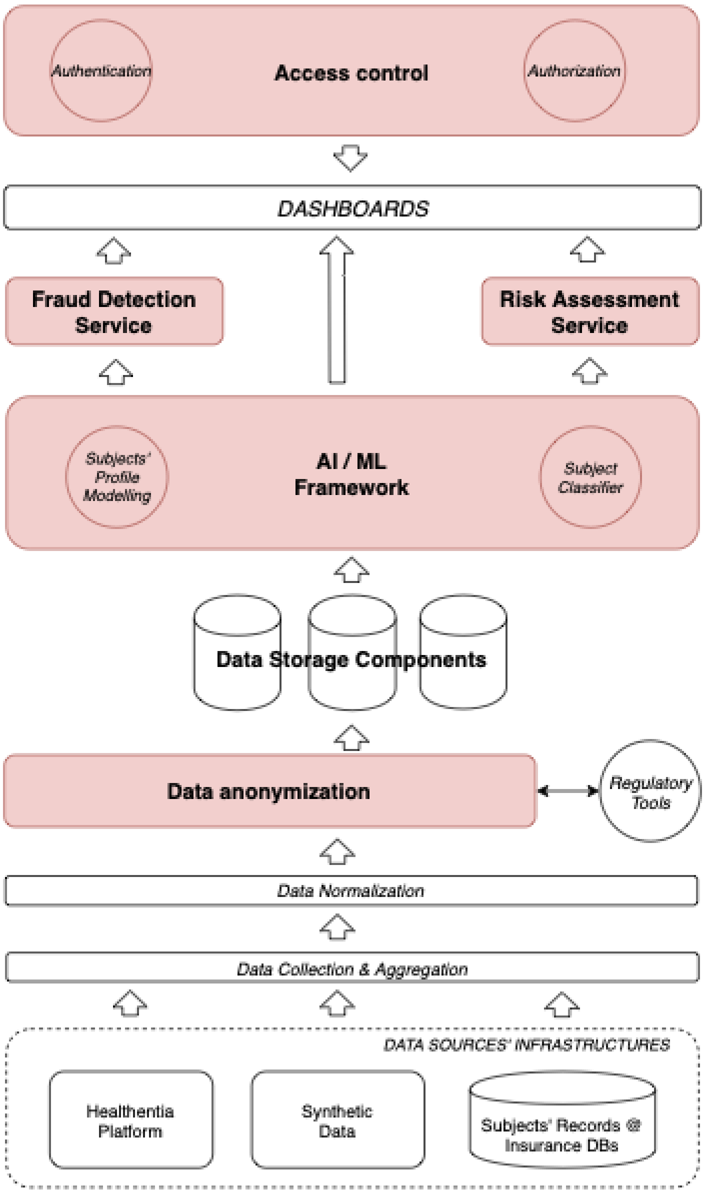

Figure 11: INFINITECH Pilot #12 RA

Expected results

Not Applicable

4.3.13 Pilot #13: Solution for regulatory compliance

Pilot overview

"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.

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

The companies (enterprises) will access our platform through a registration process and subsequent validation by assigning a package covering a 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 or API REST. The companies that use our service will have a limited number 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 targeted 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).” [2]

Security and privacy issues and requirements

There exist two different security issues and a management issue:

- The user authentication
- API rest access and encryption
- Rol management

Solution

As a solution for regulatory compliance, the Healthentia application users will sign a consent form in order to permit the collection and use of their data. In any case, the collected data will be anonymized by means of the anonymization tool developed by GRAD for protecting the user's privacy. According to the user authentication, the pilot solves this internally by means of the access control framework.

Architecture including the solution

The figure below shows the high-level architecture for this pilot, plus where the security access and data interchange can be found

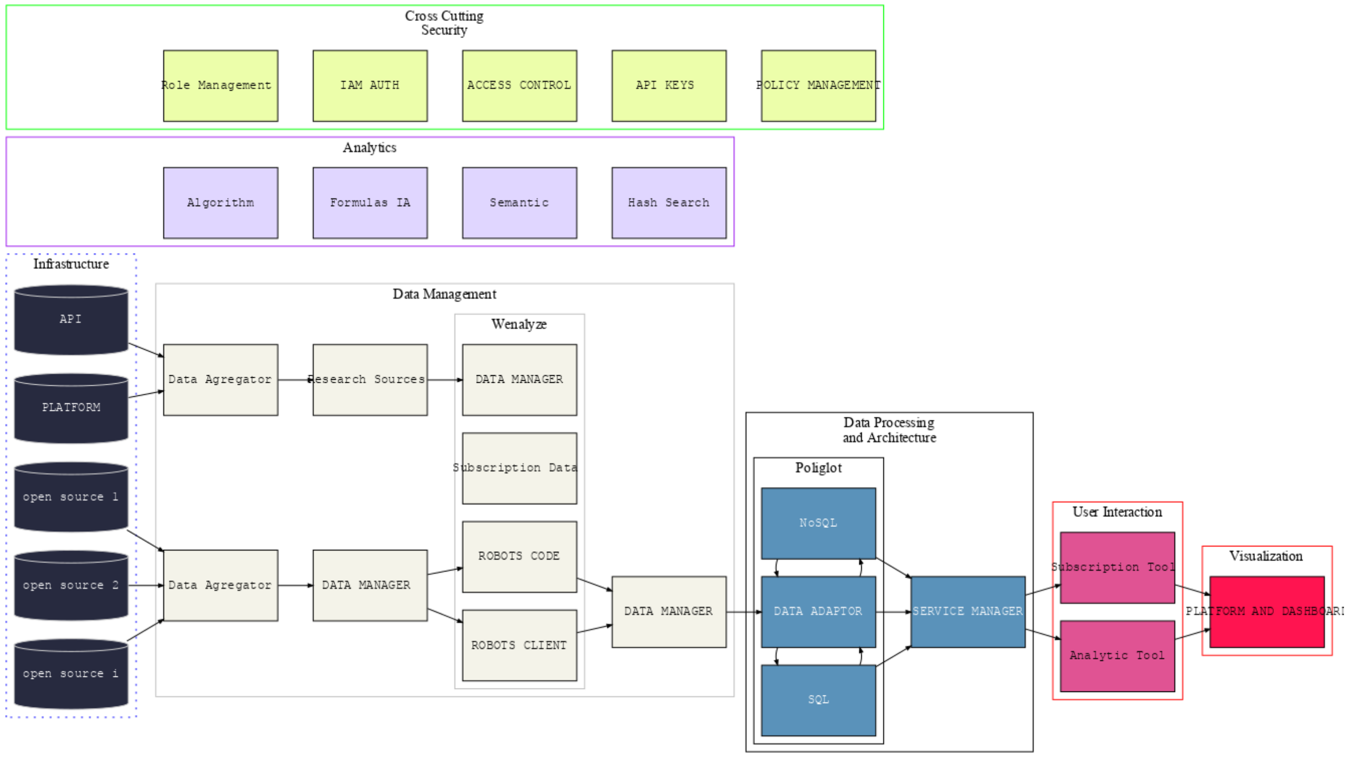

Figure 12: INFINITECH Pilot #13 RA

Expected results

Not Applicable

4.3.14 Pilot #14: Solution for regulatory compliance

Pilot overview

"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." [2]

Security and privacy issues and requirements

The use of data from banking customers might make this pilot subject to GDPR.

Solution

All data relative to banking customers will be anonymized.

Architecture

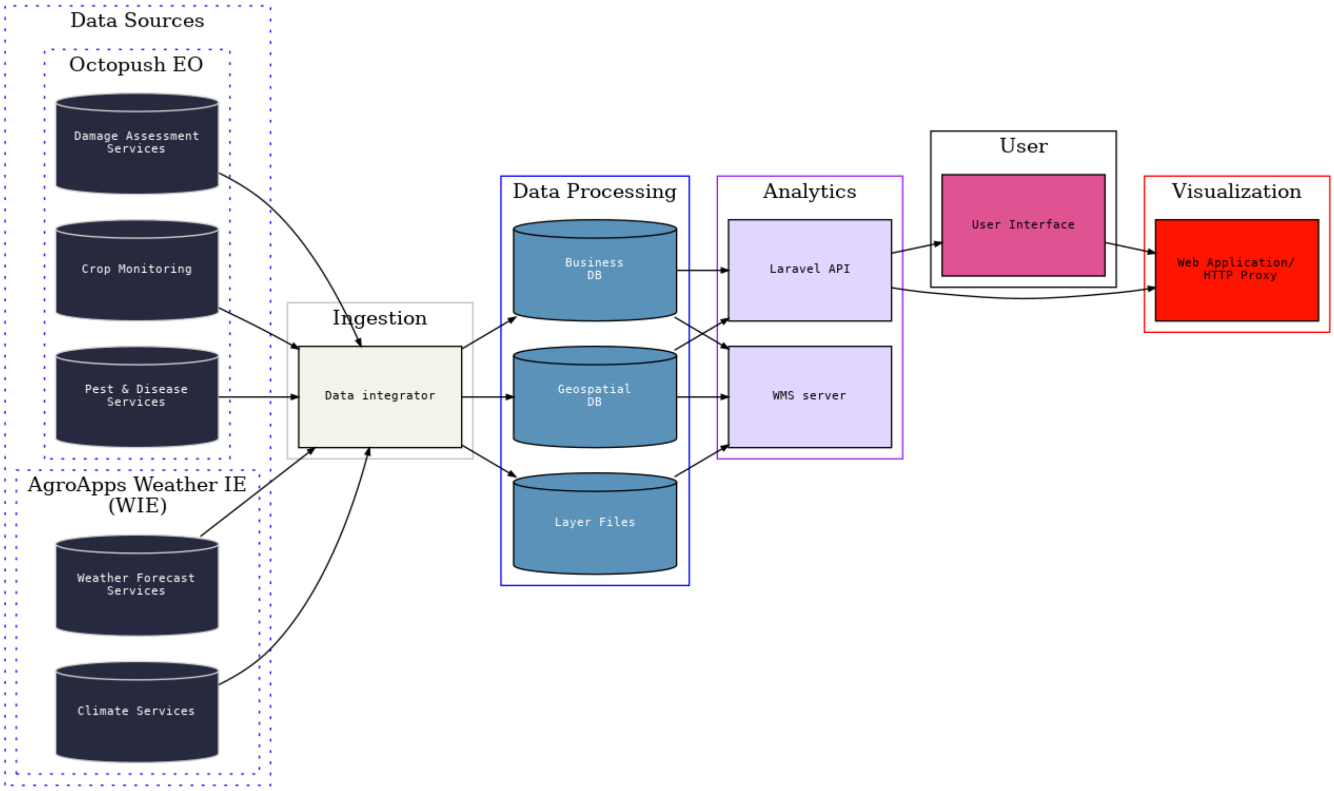

Figure 13: INFINITECH Pilot #14 RA

Expected results

Proof of concept of identified fraudulent activity, based on synthetic data.

4.4 Conclusions on regulations and technologies in INFINITECH

The analysis above of all the pilots in INFINITEC allowed us to extract details of the needs for security and privacy. Table 6 provides a summary and conclusions of mapping between regulations and technologies. It shows how the technologies listed in Table 4 can be deployed to address the different regulations applicable to INFINITECH.

Table 6: INFINITECH Technologies applicable to regulations in INFINITECH

Regulation	Need	Technology applied
GDPR	Consent management	Botakis chatbot Development Network (CP), Privacy dashboards, CMS for storing digitized documents, Blockchain-enabled Consent Management System
	anonymized data	Anonymization tool (GRAD), ICARUS (external)

	pseudonymised Data	Pseudonymization tool (JSI)
MIFID II	Recording and auditing system	ad-hoc logging implementations
PSD II	regulations for online payment services	CrowdPolicy Ooen banking solution (CP), SIEM,
AMLD4	Inclucion on local databases of PEPs	ad-hoc solutions for each country
General	Authentication	specific solutions from pilot partners, CrowdPolicy Open Solution
		DUOS for Digital User Onbording (ATOS)
	Authorization	specific solutions from pilot partners, CrowdPolicy Open Solution, IAM
	Privacy services orchestration	Data Protection Orchestrator (ATOS)

5 Conclusions

The present deliverable INFINITECH-D3.15 and the next one scheduled, INFINITECH-D3.16 “Regulatory Compliance Tools - II” are devoted to assessing regulatory compliance in INFINITECH and the tools needed to ensure it. This deliverable has performed an analysis that is fundamental to ensuring that all the pilots comply with relevant regulations. There have been analysed the regulations for the financial sector and in particular the regulations for every pilot. There have been assessed all the technologies that the partners are bringing to INFINITECH to find possible technologies that could help to give solutions for regulatory compliance. Also, we assessed the technologies that facilitate regulatory compliance. Moreover, we identified possible privacy and security issues for each pilot and studied possible solutions. In some pilots, a key value was that they are providing solutions that directly ensure regulatory compliance. In other cases, solutions must comply with regulations. In these last cases, a general preliminary solution is proposed in this deliverable by using the DPO (Data Protection Orchestrator from Atos); that will be complemented and developed in deliverables D3.16 and D3.17 of task T3.6.

The next deliverable INFINITECH-D3.16 “Regulatory Compliance Tools - II” will produce a definition of a general preliminary prototype based on DPO (Data Protection Orchestrator), and also will further describe and reflect on the applications of the technologies proposed in every pilot to comply with the regulations.

The last deliverable of this series INFINITECH-D3.17 “Regulatory Compliance Tools - III” will describe the prototype implementation of regulatory compliance tool based on DPO.

The generic regulatory compliance tool proposed for INFINITECH are based on the DPO (Data Protection Orchestrator) that is capable of orchestrate technologies for preserving privacy, data protection and security. This tool would allow to provide possible solutions for the pilots, helping to comply with the regulation.

Appendix A: Literature

[1] INFINITECH consortium, “INFINITECH D2.7 – Security and Regulatory Compliance Specifications I”, 2020

[2] INFINITECH consortium, “INFINITECH D2.13 – Reference Architecture – I”, 2020

[3] INFINITECH consortium, “INFINITECH D3.12 – Data Governance Framework and Tools – I”, 2020

[4] INFINITECH consortium, “INFINITECH D2.5 – Specifications of INFINITECH Technologies – I”, 2020

[5] I.T.Gartner, “Glossary (2020),‘Data governance,’” *URL: https://www.gartner.com/en/information-technology/glossary/data-governance*.

[6] “Pseudonymization” Imperva. https://www.imperva.com/learn/data-security/pseudonymization/.

[7] “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).

[8] A. Narayanan and V. Shmatikov, “De-anonymizing Social Networks,” in *2009 30th IEEE Symposium on Security and Privacy*, May 2009, pp. 173–187.

[9] 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.

[10] K. Mivule, “Utilizing Noise Addition for Data Privacy, an Overview,” *arXiv [cs.CR]*, Sep. 16, 2013.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] “EUR-Lex - 32014R0910 - EN - EUR-Lex.” https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.257.01.0073.01.ENG (accessed Sep. 14, 2020).

[17] GDPR – IT Governance, https://www.itgovernance.co.uk/data-privacy/gdpr-overview/gdpr-faq/gdpr-scope (accessed Nov 2020)

[18] GDPR scope – IT Governance, https://www.itgovernance.co.uk/data-privacy/gdpr-overview/gdpr-faq/gdpr-scope (accessed Nov 2020)

[19] Manisha Patel, "Top Five Impacts of GDPR on Financial Services" , Fintech Times, November 2017

[20] Will Kenton and Julius Mansa, "MiFID II – Laws&Regulations", Jul 2020, https://www.investopedia.com/terms/m/mifid-ii.asp (Accessed Nov 2020)

[21] 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)

[22] 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

[23] Mark Halstead, "What is the Fourth AML Directive (AML4D)-RedFlagAlert", https://www.redflagalert.com/articles/data/what-is-the-fourth-aml-directive-aml4d

[24] Martin David, Jorge Bernal, Julien Bringer, Nicolas Notario, Eduardo Gonzales - D3.1 – ARIES eID ecosystem technical design – July 2017

[25] "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

[26] INFINITECH consortium, "INFINITECH D3.6 – Data Streaming and Data at Rest Queries Integration – I", 2020

[27] INFINITECH consortium, "INFINITECH D4.7 – Permissioned Blockchain for Finance and Insurance – I", 2020