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

Infinitech

D3.17 – Regulatory Compliance Tools III

Revision Number	**3.0**
Task Reference	T3.6
Lead Beneficiary	ATOS
Responsible	Nuria Ituarte Aranda (ATOS)
Partners	Participating partners in Task according to DOA
Deliverable Type	Dem
Dissemination Level	PU
Due Date	2022-03-31
Delivered Date	2022-04-01
Internal Reviewers	IBM, NOVA
Quality Assurance	INNOV
Acceptance	WP Leader Accepted and/or Coordinator Accepted
EC Project Officer	Beatrice Plazzotta
Programme	HORIZON 2020 - ICT-11-2018
	This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no 856632

Contributing Partners

Partner Acronym	Role[1]	Author(s)[2]
ATOS	Lead beneficiary	Nuria Ituarte Aranda Ignacio Elicegui Carmen Perea
AKTIF	Contributor	Orkan Metin Ömer Bora Zeybek
ASSEN	Contributor	Ilesh Dattani
BOS	Contributor	Klaudija.Jurkosek-Seitl Sabina Podkriznik Milošević Jelena
DYN	Contributor	Andreas Politis
GRAD	Contributor	Inés Ortega Fernández Lilian Adkinson Orellana
JSI	Contributor	Maja Skrjanc Mitja Jermol
NBG	Contributor	Syllignakis Manolis Eleni Perdikouri Georgia Prokopaki
PI	Contributor	Massimiliano Aschi Giusseppe Avigliano
BPFI	Contributor	Phil Atherton
CXB	Contributor	Mario Maawad
Wenalize	Contributor	Carlos Albo Portero
JRC	Contributor	Konstantina Tripodi
iSPRINT	Contributor	Aristodemos Pnevmatikakis
AGROAPPS	Contributor	Grigoris Mygdakos
ABILAB	Contributor	Barbara Cacciamani

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

[2] Can be left void

IBM	Internal Review	Fabiana Fournier
NOVA	Internal Review	Guilherme Brito
INNOV	Internal Review & QA	John Soldatos
GFT	Internal Review & QA Contributor	Ernesto Troiano Vittorio Monferrino Marina Cugurra

Revision History

Version	Date	Partner(s)	Description
0.1	2021-11-10	Atos	ToC Version
0.9	2022-02-20	Atos, All	Version with partners contributions
2.5	2022-03-16	Atos	First Version for Internal Review
2.8	2022-03-24	Atos	Version for Quality Assurance
3.0	2022-03-31	Atos	Version for Submission

1 Executive Summary

Important note

This is the last deliverable of Regulatory Compliance Tools series. The first deliverable is D3.15 – Regulatory Compliance Tools – I [1], the second version is deliverable D3.16 – Regulatory Compliance Tools – II[2] which provides an extension of the regulations that applied to Pilot #15 that was incorporated in INFINITECH, an update of technologies for Regulatory Compliance and an extension of the Regulatory Compliance Tools for every pilot.

This last deliverable D3.17, compared with the previous iteration D3. 16 [2], extends and improves D3.16 to provide the complete version of Regulatory Compliance Tools in INFINITECH. D3.17 updates the regulations applicable to INFINITECH pilots, the solutions of Regulatory Compliance Tools for each pilot and an analysis of the Regulatory Compliance for both, the pilot case and the real case. An assessment and recommendation for every pilot are also provided. And finally the complete description of the General INFINITECH Regulatory Compliance Tool based on DPO.

Extending D3.16 into D3.17 in these ways has been preferred to the creation of an amendment as it gives complete information in the document, facilitating the reader's understanding.

This deliverable completes the analysis of regulatory compliance throughout the INFINITECH project and in every pilot. The analysis started with "Regulatory Compliance Tools – I" [1] and was updated in "Regulatory Compliance Tools – II" [2]. The document starts with the regulations for the financial sector and the available technologies in INFINITECH. It provides the final analysis for every pilot participating in the INFINITECH project and considers the pilots assessment and evaluation.

The aspects analysed for every pilot are the following:

- The regulations that they should comply with.
- 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 solutions that could be provided in case of production case
- The assessment for the pilot
- The recommendations for regulatory compliance for every pilot

The general approach of the project towards regulatory compliance tools is to use the ones already in use by the applications and solutions if they are already available and compliant with applicable regulations, and to provide new tools when the tools already in use are insufficient for current regulations.

However, due to the limited resources of the project to develop new features especially in advance of current regulations, 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
AI/ML	Artificial Intelligence and Machine Learning
AML IV	Anti-money Laundering
API	Application Programming Interface
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
HPC	High Performance Computing
IAM	Identity and Access Management
IoT	Internet of Things
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
NLP	Natural language processing
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
P2PP	Peer-to-Peer Payment
QTSP	Qualified Trust Service Provider
RTS	Regulatory Technical Standard
RA	Reference Architecture
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 third and last one of a series of three deliverables that aim to define and develop regulatory compliance tools. The first one, D3.15 – Regulatory Compliance Tools – I [1] analysed the need in INFINITECH to comply with regulations. The pilots were analysed, assessing if they are regulatory-compliant, identifying the need of regulatory-compliance tools and preparing the field for the development of these tools.

The second one D3.16 – Regulatory Compliance Tools – II [2] was an update of [1], adding information from Pilot #15 that was incorporated in this period and with more detailed information of the regulatory compliance solutions applied in every pilot. This second iteration also provided a first approach of the definition of the INFINITECH Regulatory Compliance Tool based on the Data Protection Orchestrator (DPO), describing this component, its architecture, technical design and interfaces and integration with Anonymization from partner Gradiant.

This third and last iteration of Regulatory Compliance Tools is an update to the second one completing the work regarding regulatory compliance in INFINITECH, considering the complete pilots set now completed with the new recently incorporated Pilot #16. It provides the applicable regulations and technologies for security and privacy in INFINITECH and provides the solutions of Regulatory Compliance in pilots, both for the implementation of each pilot case, and for the description of the solution for the production environment. The solutions for the pilots are assessed (considering stakeholders evaluations form D7.21 [15]) and some recommendations for regulatory compliance are given accordingly. The deliverable also provides the final definition, interfaces, architecture and integration of the General INFINITECH Regulatory Compliance Tool based on the DPO.

1.1 Updates from the previous version (D3.16)

This last deliverable D3.17, compared with the previous iteration D3. 16 [2], extends and improves D3.16 to provide the complete version of Regulatory Compliance Tools in INFINITECH. Specifically, Section 2 regarding the regulations applicable to INFINITECH pilots has been updated on the correspondent table from D3.16 with the regulations that apply to Pilot #16, since these have been incorporated in INFINITECH after the release of D3.16. Section 4 regarding the Solutions of Regulatory Compliance Tools for every pilot has been extended based on the advanced status of the pilots at this stage. The Section 4 provides an added value by analysing the Regulatory Compliance for both, the pilot case and the real case, also a summary of the assessment of the solution coming from D7.21 [15]. These analyses allow to finally provide recommendations for Regulatory compliance for every pilot. Finally, Section 5 provides the complete description of the General INFINITECH Regulatory Compliance Tool based on DPO.

1.2 Objective of the Deliverable

The main objective of this final deliverable of the task “T3.6 Regulatory Compliance Tools” is to ensure that all the solutions created in INFINITECH project are regulatory-compliant, while providing the relevant regulatory compliance tools that will boost this compliance.

This goal encompasses the following specific objectives:

- **Review of main regulations applicable to INFINITECH**. INFINITECH project contributes (stated in D3.15 [1]) solutions for several pilots with different business objectives, as specified by the end-users of the project (i.e. financial organizations, banks, insurance companies, and 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 updates the study of the solutions provided for every pilot and also updates the analysis of the regulations that should be applicable to them (based on the previous studies of WP2 in deliverable D2.8 "Security and Regulatory Compliance Specifications – II" [4]).

- **Review of technologies for security, privacy and data protection** (stated in D3.15 [1])**.** The INFINITECH project is developing these technologies and making them 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 deliverables D2.5 and its update D2.6 "Specifications of INFINITECH Technologies – II" [7]. In T3.6 "Regulatory Compliance Tools" deliverables, the technologies are analysed, and the ones related to regulatory compliance are outlined. The present deliverable includes the complete 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 in INFINITECH pilots**. This was performed mainly during the first iteration of this task. It encompasses 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 under the INFINITECH technologies/pilots scope is sought. Many INFINITECH pilots 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 a general regulatory compliance tool** as needed for the project. This deliverable presents the final definition of the General Regulatory Compliance Tool for INFINITECH. This General tool enables the production of regulatory compliance tools in-line with the needs of the INFINITECH pilots. This solution could be adapted for all the pilots in the project. This solution uses the Data Protection Orchestrator (DPO) provided by Atos. The DPO embeds and automates the assurance of security and privacy by design and by default in complex business flows.
- **To examine all the Pilots solutions for regulatory compliance.** Compliance assessment methodologies in INFINITECH are summarized. All the pilots are reviewed showing the solution for regulatory compliance for the pilot and production case, the stakeholder's evaluation performed in the pilot's workshops is collected and recommendations for the pilots are provided.

Table 1 shows the most important findings for every INFINITECH pilot, the applicable regulations, the compliance solution and the technologies for Regulatory Compliance used. Each pilot's full name can be found in Since this deliverable relates to all the pilots, and their names are used intensively, the pilot's short names are used (in the format "Pilot #1" for example) instead of the long names. Table 2 shows the mapping of short names versus long names of the INFINITECH pilots.

Table 2. Note that Pilot #1 and #5a have been discontinued in the project and therefore are not shown in the table

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

Pilot	Applicable Regulations	Compliance Solution	Technologies for Regulatory Compliance
#2	BASEL IV (Changes to the global capital requirements the Basel Committee agreed to in 2017) and MIFID II (Markets in financial instruments directive 2)	Compliance based on the use of synthetic and aggregated data for the pilots.	IAM and Audit logs (MIFID II)

#3	GDPR (General Data Protection Regulation (EU) 2016/679) 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 (Digital User Onboarding System)	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	GDPR, MIFID II and AMLD4 (the Fourth Anti-Money Laundering Directive (EU) 2015/849)	Compliance based on the use of Synthetic data or anonymization	Anonymization
#8	GDPR and AMLD4	Compliance based on performing pseudonymization of personal data about the individuals and confidential information on legal entities within the transactions prior to data delivery to PAMLS	Pseudonimization
#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	Anonymization tool, General regulatory

		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.	Compliance Tool based on DPO (Data Protection Orchestrator) 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
#15	None: GDPR is not applicable because the service assessment application analyses a subset of process operating documents for classification purposes, without ever accessing the data about the customers.	Compliance based on the analysis and process of a subset of data not related to personal customer data.	Not Applicable because no regulation applies
#16	GDPR and AMLD4	Compliance based on the anonymization of the used data	Anonymization

As stated in D3.15 [1] 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 directly 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 general regulatory compliance tools. In-line with this objective, this deliverable provides the final definition or the General INFINITECH Regulatory Compliance Tool, by using the DPO (Data Protection Orchestrator) tool provided by Atos.

As stated in D3.15 [1], 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 blocking funding for terrorism.

The main types of technologies that help support compliance with 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

1.3 Insights from other Tasks and Deliverables

As stated in the previous iterations of this deliverable [1] and [2], the current one is fully cross-related to other tasks and deliverables in the project.

Let's analyse first the inputs that are in the previous iteration of this deliverable, the regulations and technologies and the data governance mechanisms:

- **INFINITECH D3.16 "Regulatory Compliance Tools – II** "[2]. The current deliverable starts from the content of deliverable D3.16 [2]; from that base, its coverage has been increased and improved by providing updates on regulations and technologies and also perform a detailed and complete analysis of all the pilots from the regulatory compliance point of view, analysing the solutions of the pilot and real case and detailing the descriptions of the regulatory compliance tools of the pilots that are providing them. Also, it describes the final general definition and development of INFINITECH Regulatory Compliance Tool based on DPO.

- **INFINITECH-D2.8 "Security and Regulatory Compliance Specifications** – **II"** [4]. It was considered for INFINITECH-D2.16 [2]. This deliverable is the last version 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. It 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. D2.8 provides the main regulations to consider in the pilots to solve the privacy issues that arise in the pilots.

- **INFINITECH-D2.6 "Specification of INFINITECH Technologies – II"** [7]. This deliverable collects the tools and technologies available and under 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-D3.17 collects the final set of privacy and security technologies that could be considered by participants in the creation of regulatory compliance tools and also describing the components that are providing regulatory-compliance themselves.

- **INFINITECH-D3.14 "Data governance framework and tools – III"** [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 and digital mobile onboarding system. It also presents the final description and development of the tools related with the mentioned technologies. This deliverable is one of the main and most practical inputs for INFINITECH-D3.17, given that the tools developed here will be direct components that will be called by regulatory compliance tools.

- **INFINITECH-D2.14 "Reference Architecture** – **II"** [2]. It presents the second version of the INFINITECH-RA. This version presents the design and initial integration of the use cases. This RA will be used in INFINITECH-D3.17 to analyse the pilots and also to ensure the integration of regulatory compliance tools or to describe where is the pilot implementing regulatory compliance itself.

- **INFINITECH D9.15 "Business Models and Innovation Management - I"** [13]. This deliverable has been considered for the assessment methodology to apply for the pilots combined with D2.1 [19] providing the results in D9.16 [14]

- **INFINITECH D7.21 "Pilots' Evaluation and Stakeholders' Feedback - II"** [15]. This deliverable collects stakeholders' feedback for all pilots mainly from the workshops This feedback has been collected in the present deliverable in order to be considered in the solution assessment.

The outputs of this deliverable and the previous ones of the series are used by other WPs. In practice, the most important output from this task will be the general regulatory compliance tool that will be use in one of the pilots in WP7 "Large-Scale Pilots of SHARP Financial and Insurance Services " and the review of all the solutions provided by the pilots for regulatory compliance.

1.4 Structure

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

Section 2 provides a review of main applicable regulations of the financial sector in pilots based on the work done in INFINITECH-D2.8 [4].

Section 3 reviews the technologies for security, privacy and data protection based on the work performed in INFINITECH-D2.6 [7] and provides an update of the work done in D3.15 [1].

Section 4 aims to show for every pilot the solutions that have been applied in INFINITECH project both for the pilot case and also the solutions that should be applied in case that they would work in real life in a production environment. It also show the assessment results for every pilot and recommendation for assessing the compliance.

Section 5 provides the final definition of a general INFINITECH Regulatory Compliance tool based on DPO. It details the DPO component by supplying the complete description and providing its architecture and interfaces. It also specifies the integration with the Anonymization tool from Gradiant.

Since this deliverable relates to all the pilots, and their names are used intensively, the pilot's short names are used (in the format "Pilot #1" for example) instead of the long names. Table 2 shows the mapping of short names versus long names of the INFINITECH pilots.

Table 2: Map of INFINITECH Pilots

Pilot short name	Pilot long name
Pilot #2	Real time risk assessment in Investment Banking

Pilot #3	Collaborative Customer-centric Data Analytics for Financial Services
Pilot #4	Personalized Portfolio Management (“Why Private Banking cannot be for everyone?”)
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 Health-Insurance products
Pilot #13	Alternative/automated insurance risk selection - product recommendation for SME
Pilot #14	Big Data and IoT for the Agricultural Insurance Industry
Pilot #15	Open Inter-banking Pilot
Pilot #16	Data Analytics Platform to detect payments anomalies linked to money laundering events

Note that Pilot #1 and Pilot #5a became discontinued and therefore omitted from the table.

2 Review of main applicable regulations in INFINITECH pilot

The following table summarizes the main regulations applicable to each of the INFINITECH Pilots. It is an update of the previous version, in D3.16 Regulatory Compliance Tools – II [2]. This update takes into account the findings documented in D2.8 “Security and Regulatory Compliance Specifications II” [4], in particular the requirements for each pilot (and related test beds, sandboxes and applied technologies). It considers the requirements resulting from state-of-the-art security and privacy standards (e.g. ISO 27000, NIST Cyber Security Framework), regulatory frameworks/directives such as MIFID II, PSD II and requirements resulting from AI regulations. Regarding AMLD4 directive, the European Commission has just (July 2021) presented a package of legislative proposals to strengthen the EU's anti-money laundering and countering terrorism financing (AML/CFT) rules. The aim of this package is to improve the detection of suspicious transactions and activities and includes a sixth Directive on AML/CFT (“AMLD6”), replacing the existing Directive 2015 (AMLD4) [11]. The applicability of regulations is further addressed per each pilot in Section 4.

The main change in the table below (taken from D3.16 [2]) is the introduction of Pilot #16 “Data Analytics Platform to detect payments anomalies linked to money laundering events”, which was described in D2.19 “Reference Scenarios and Use Cases – Version II” [26].

Table 3: Main regulations in INFINITECH pilots

Pilot	Regulations
#2	The 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 it 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 data on 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 are anonymized. Such recommendation would be subject to the legal obligations of electronic recording.
#7	This pilot uses confidential data. 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 uses confidential data. Thus, in case of specific interest ("need to know"), please contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us . **GDPR** 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, which must 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.
#15	**None:** GDPR is not applicable because the service assessment application analyses a subset of process operating documents for classification purposes, without ever giving access to the data about the customers.

#16	**GDPR:** applies as data will be utilised that includes information regarding cardholders, merchants, organizations, and digital payment authorizations and transactions. The pilot will use data that is anonymized however it can still be construed as personal in the cases where the information relates to an individual and GDPR will still apply in that situation **AMLD4:** this applies in the limited context where any findings that suggest the potential of money laundering activity must be reported to relevant regulators based on the local jurisdiction

3 Review of technologies for security, privacy and data protection

Table 4 taken from [1] collects a review of the technologies for security, privacy and data protection available in INFINITECH project. Contents of this table have been taken from [2], [4] and [7] which explains all the technologies available in INFINITECH. The table has been taken from [2], no changes have been produced in INFINITECH technologies descriptions.

These are described in the table 4 below.

Table 4: INFINITECH Technologies for regulatory compliance

Name tool / platform	Company	Relevance and applications for regulatory compliance
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 requires the Swagger specification of the components (Privacy Enhancing Technologies-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 (notably the GFT- and NBG-led pilots): • Built-in dialogs that utilize and are integrated with existing NLP (Natural language processing) 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." [9] 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 With regard to the ability for providing information regarding the 7 points described above, we expect that the relative responses will be included in the questions that the chatbot will cover, as well as any relative material as part of the GDPR consent that the Pilot 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. “[7] 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. Enhancements in INFINITECH project are: • technology scale-up is to democratise the use and exploitation of open banking APIs even for users with no development skills, building FinTech software development kits. • implement a complete programmable framework to integrate different services and APIs using protocols, thus providing similar user experience as zapier, “yahoo pipes” and “IFTTT”. The main objective is to provide a graphical user interface for building data and FinTech services mashups that aggregate open banking APIs, open available data sets and rules, and also for creating Web based apps from various sources, and publishing those apps.[7]
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 and deletion,) 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. [9] The anonymization tool is intended to be used in two execution 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 [10]. 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 D4.9 "Permissioned Blockchain for Finance and Insurance – III" in WP4 [10].
Pseudo-anonymization Tool	JSI	The tool developed within INFINITECH will be used for pseudo-anonymization of financial transactions' data, but the service itself will be general enough to handle various types of inputs. Typical data fields that need to be pseudo-anonymized in transactional data are for example: names, company names, bank identifications, IBAN numbers, but also amounts, timestamps and textual data (e.g., comments and transaction descriptions). Details on Pseudo-anonymization tool are described in D3.13 "Data Governance Frameworks and Tools II" [3].
OpenSource AI/ML frameworks	FTS	These frameworks facilitate the development of AI/ML based tools, which shall be applied to Financial Crime and Fraud, e.g. on so called Instant Loans. Today a number of open source tools for AI/ML development are available where the AI/ML community keeps on actively and dynamically progressing these technologies, thus providing the basis for solution development while facilitating more specific solution for a wide range of business problems as encountered in INFINITECH. These open source tools provide the foundation for development towards off-the-shelf modules being part of the INFINITECH RA. The solution of Pilot #7 will comprise extraction of customer and transactional features as well as an advanced scoring model indicating the risk of a fraudulent instant loan. The frameworks will be mainly applied in the solutions of Pilot#7.
Data Layer - REST API	GFT	A Data Layer to support Security Data Model with REST API based on a not relational database (MongoDB). Supports heterogeneous sources. Developed upon FLASK-Python3 framework and dockerized to be deployed on

		Kubernetes infrastructure. It will be applied in Pilot#15. For more information, refer to https://gitlab.infinitech-h2020.eu/datamanagement/infinistore.
Data Check-In Mechanism	UBI	A sophisticated data check-in mechanism that enables the preparation and uploading of the data provider's (public or confidential) datasets in the cloud platform that is one of the results of the ICARUS H2020 project. The data check-in mechanism is deployed on the premises of the data provider as a stand-alone desktop application and receives as input a list of data check-in jobs that incorporate a set of instructions with all the actions that will be performed on a specific dataset, residing on the local storage of running operating system, in order to enable the data preparation and uploading of new datasets in a secure manner. Internally, the mechanism handles the orchestration and execution of the designed instructions with the use of incorporated (micro) services for: a) data mapping of data source entities to the designed common data schema, b) data cleaning operations on the data source entities, c) anonymization operations on the data source entities and d) encryption of the data source entities. This list of (micro) services is expandable based on the needs of each platform. The data check-in mechanism is offered in the form of a local client for all OS (Mac, Linux, Windows) and is designed and developed using the latest technologies for desktop apps with the aim to offer end-to-end security on the data preparation and data upload tasks. The specific technology served as the basis for the design and implementation of the INFINITECH Data Collection component.
Analytics Library	UPRC	In the scope of the ATMOSPHERE (Adaptive, Trustworthy, Manageable, Orchestrated, Secure Privacy-assuring Hybrid, Ecosystem for REsilient Cloud Computing) project, the UPRC team, focused on the delivery of the library of services, which can be utilized as a baseline for the INFINITECH library. (WP5) It is based on the idea of an Update of the library to enable the inclusion of metadata relevant to security and privacy constraints of the INFINITECH algorithms, and make them available through the library. The library has been incorporated in the INFINITECH market platform and will support the metadata structures to describe the algorithms through the respective descriptors of the marketplace assets. For ore information refer to: https://www.atmosphere-eubrazil.eu/

3.1 Mapping regulations and technologies in INFINITECH

Table 5 taken from [1] provides a summary and conclusions of the mapping between regulations and technologies. It shows how the technologies listed in Table 3 can used to address the different regulations applicable to INFINITECH.

Table 5: 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)

4 Solutions of Regulatory Compliance tools in pilots and solutions assessment

This section provides a complete view of the solutions for regulatory compliance in all the INFINITECH pilots.

The section aims to show:

- Compliance assessment. In INFINITECH, there are two proposed ways to assess/evaluate every pilot:
    - Compliance assessment methodology. It's described in 4.1.1 and encompasses three areas of interest: Regulation and directive compliance, Internal Compliance framework and intellectual property rights. This methodology comes from D9.15 [13] and will be applied in D9.16[14]
    - Stakeholders' evaluation. It is described in 4.1.2, it takes stakeholders' feedback considering monitored KPIs for every pilot, risks and critical problems reported. This evaluation comes from D7.21 [15]. The final score values are ranged between 1 to 5. The results of this evaluation are given for every pilot in the subsection "assessment solution"
- For every pilot the following information is provided:
    - The pilots' solutions that have been applied in INFINITECH project regarding regulatory compliance
    - The pilots' solutions for regulatory compliance that would be adopted in case they would work in real life in a production environment
    - Assessment results for every pilot coming from Stakeholders' evaluation from D7.21 [15].
    - Recommendations for assessing the compliance in the pilot using the most suitable way. In this section recommendations for fulfilling compliance in the pilot to fill the gaps with the regulations (in case) are given.

4.1 Compliance Assessment

Pilots Compliance Assessment is being performed in INFINITECH project through two main ways. The first one is described in 4.1.1 as "Compliance assessment methodology" defining a process to follow in order to conclude with an assessment value, and the second one is provided in 4.1.2 coming mainly from stakeholder's feedback

In the present deliverable, the provided assessment result for every pilot comes from 4.1.2 which considers D7.21 [15] that collects the Pilot's evaluation and stakeholders' feedback. Every pilot has been evaluated by stakeholders assessing their satisfaction with respect to the INFINITECH Pentagon`s Innovation dimensions (the dimensions are customer, market, resources, compliance and technology). The important value for the present deliverable is the compliance value which has been considered in the assessment solution and recommendations of every pilot.

4.1.1 Compliance assessment methodology

This section presents the methodology to validate compliance assessmentbased on section 5 of D9.15 [13]and also D1.2 [19] regarding risk management. This methodology will be used in D9.16 [14] to perform the validation

As stated in D9.15[15], the solution must comply with the current regulations it applies to, carefully shaping and modifying details to also exploit them. Moreover, the proposed solution has to be compliant to internal processes, IP limitations and the interactions with the environment.

Compliance assessment encompasses three areas of interest: Regulation and directive compliance, Internal Compliance framework and intellectual property rights, which are addressed in the following subsections.

The application of this assessment methodology to every pilot results in the compliance value defined in the last subsection (4.1.4) and the compliance assessment of the pilot. This value is important to consider for future production applications of the pilots and to provide recommendations. The application of the assessment methodology and the result of the compliance value will be reported in D9.16 [14] (second version of Business Modelling and Innovation Management) at the end of the project.

4.1.1.1 Regulation and directive compliance

Regulation and directive compliance must be evaluated in order to assess if the pilot complies or not (it is a binomial condition) considering EU legal framework. This should be evaluated using the results described in D2.7[20], D2.8 [4] and D2.21 [21] of the Infinitech project and on the section 2 of the current document.

In the case the pilot is not compliant, the analysis will be stopped at this point, until a solution application that solves it is found.

4.1.1.2 Internal Compliance framework

Each pilot should follow specific procedures and, have a quality plan and a risk management procedure in place.

Following the guidelines provided by the INFINITECH project that can be found in D1.2 [19]and future versions, the first step is to prepare a risk management procedure for every pilot and a quality plan, which will help to assess the compliance. The selected model is the Compliance Maturity Model [15] summarized in the following sections.

4.1.1.2.1 Risk management

D1.2 [19] defines how risks associated with the INFINITECH project will be identified, analysed, and managed. the risks associated to every pilot will be managed using the same procedure as all the risks identified in the project:

1. Risk identification: the partners involved in the pilot will communicate the risks associated to the concrete pilot
2. Every Risk will be analysed and assessed by valuing the likelihood and impact and using these values in the Risk Matrix following the levels given in section 2.2 of D1.2 [19]
3. The pilot leader will include the risks in the risk register which is included in the INFINITECH APP under “RISKS” [16] filling in the following values that are described in section 2.5 of D1.2[19]. In order to organize the risks per pilot, the table field “Description” will start with the number of the pilot in this way: “Pilot #n: “and after that write the suitable risk description

4.1.1.2.2 Quality plan

The general quality plan for INFINITECH project is defined in section 3 of D1.2 [19] and for every pilot there will be a quality plan following the same guidelines.

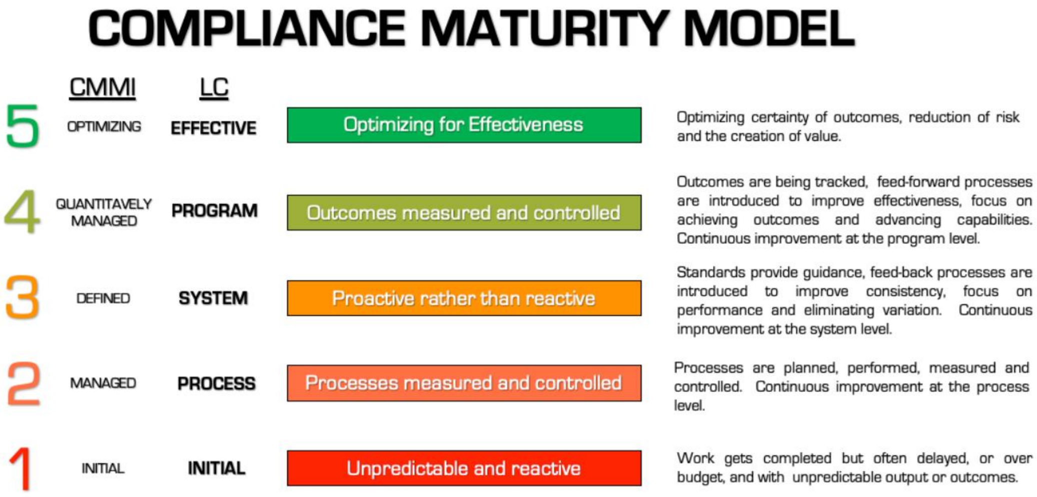
Basically, the quality Management will be based in a set of Actions that will be defined in the INFINITECH App [17] and this will be the procedure:

1. Prepare actions set for the pilot
2. Include them in the INNFINITECH App under the “Actions” folder filling in the items described in section 3.3 of D1.2 [19]. The field “ID” in this case, should start with “Pilot #n” followed by a number

4.1.1.2.3 Internal compliance assessment

It was mentioned before that in INFINITECH the selected model is the Compliance Maturity Model (Figure 1) represented and explained below and was stated in D1.2[19]. It is an adaptation of the widely used CMMI (Capabilities Maturity Model Integration).

Considering the Risk management and Quality plan of the previous steps, it is possible to assess the compliance following the CMMI.



COMPLIANCE MATURITY MODEL

	CMMI	LC		
5	OPTIMIZING	EFFECTIVE	Optimizing for Effectiveness	Optimizing certainty of outcomes, reduction of risk and the creation of value.
4	QUANTITAVELY MANAGED	PROGRAM	Outcomes measured and controlled	Outcomes are being tracked, feed-forward processes are introduced to improve effectiveness, focus on achieving outcomes and advancing capabilities. Continuous improvement at the program level.
3	DEFINED	SYSTEM	Proactive rather than reactive	Standards provide guidance, feed-back processes are introduced to improve consistency, focus on performance and eliminating variation. Continuous improvement at the system level.
2	MANAGED	PROCESS	Processes measured and controlled	Processes are planned, performed, measured and controlled. Continuous improvement at the process level.
1	INITIAL	INITIAL	Unpredictable and reactive	Work gets completed but often delayed, or over budget, and with unpredictable output or outcomes.

Figure 1: Compliance Maturity Model Source: https://www.leancompliance.ca/post/capabilities-maturity-model-for-compliance

For every pilot it is needed to assess the value according with CMMI values of the previous figure.

4.1.1.3 Intellectual property rights

The third value to consider in compliance value is related to the Intellectual Property Rights (Table 6) and the level is self-assessed through the table below.

Table 6: Intellectual Property Rights. Source: https://www.researchgate.net/figure/Intellectual-Property-Readiness-Level-Part-of-TRL-Hasenauer-et-al-Managing_fig4_313063121

Level	Description
1	Hypothesizing on possible IPR (patentable inventions)
2	Identified specific patentable inventions or other IPR
3	Detailed description of possible patentable inventions. Initial search of the technical field and prior art.
4	Confirmed novelty and patentability; decided on alternative IP protection if not patenting
5	FIrst complete patent application filed, Draft of IPR strategy done.
6	Positive response on patent application; initial assessment of freedom to operate, patent strategy supporting business
7	Patent entry into national phase; other formal IPR registered
8	First patent granted, IPR strategy fully implemented, more complete assessment of freedom to operate

9	Patent granted in relevant countries, strong IPR support for business

4.1.1.4 Compliance value

The final compliance value for this methodology will be obtained by calculating the mean value of internal compliance assessment and intellectual property rights. Table 7 shows the mapping of IPR Readiness Level and CM Model to the value (Parameter), that will be used to calculate the mean value of compliance value.

Table 7: Final Compliance value

Parameter	1	2	3	4	5
IPR Readiness Level	1	2-3	4-5	6-7	8-9
CM Model	1	2	3	4	5

The Compliance value is not provided in the present deliverable, but will be calculated in D9.16 [14] being available at the end of the project.

4.1.2 Stakeholders' evaluation

INFINITECH proposes an evaluation and stakeholders' validation in D7.21 [15] from both business and technical/operational perspective. The first one has been achieved by leveraging the stakeholders' feedback through the hosting of workshops; the technical/operational performed considering some metrics such as KPI. The validation for every pilot considers: the INFINITECH Innovation Pentagon that takes into account five aspects, the compliance value is the one considered in the present deliverable and is calculated as the average of the stakeholders' feedback whose scores range from 1 to 5; monitored KPIs for every pilot, risks and critical problems reported.

In the present deliverable, the stakeholders' evaluation coming from [15] is considered. Every pilot has been evaluated by stakeholders assessing the stakeholder's satisfaction with respect to the INFINITECH Pentagon`s.

The following sections in the present deliverable collect, for every pilot, this compliance value inside the section "assessment solution" and it has been considered for providing recommendations per pilot.

4.2 Pilot #1

The pilot has not any update because it is not part of the project anymore

4.3 Pilot #2

As stated in section 4.3.2 of D3.15 [1], the high-level aim of Pilot #2 is to provide "bank trader's 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". Pilot #2 addresses this issue by providing a solution that can provide aggregate market data, trends and predicted risk/yield that updates in real-time.

4.3.1 Pilot #2 solution

The main objective of the Pilot is to provide risk assessments of financial portfolios in (near) real time, allowing faster reactions and enabling adjustments to the portfolio composition if necessary. To this end, the solution provided should be able to make use of the most recent market data and, based on entry trading positions, provide accurate risk measurements. In addition, the pilot should increase user satisfaction (e.g. trader, risk manager) through a friendly user interface that provides features such as sentiment analysis on financial news and pre-trade analysis.

Therefore, Pilot #2 will not use or process personal data. The data used are ''market data'', which is proprietary but not confidential, and transaction date, which is generated synthetically, and ''open text data'' provided by Twitter's API for sentiment analysis in financial news feeds. The functions of the pilot solution do not consider real customers and do not deal with transactions. As a result, no regulatory compliance tools are required for the Pilot 2 pilot solution.

Security and privacy issues and requirements

As stated in D3.15 [1] 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.

High level Architecture

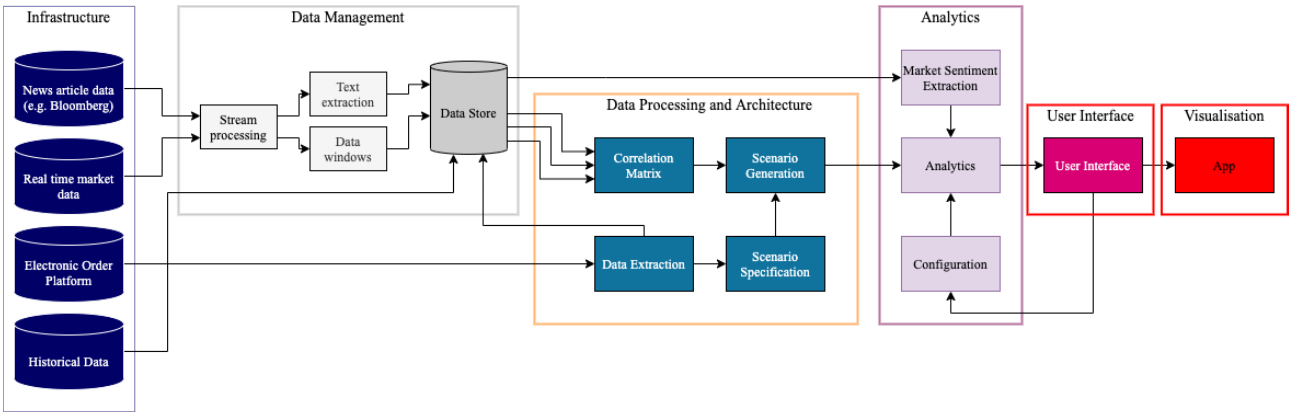


Figure 2: Real-Time Risk Assessment pilot pipeline in-line with the IRA

4.3.2 Pilot #2 real case solution

The Pilot incorporates user authentication and authorization processes in order to regulate access to its user interface. Besides that, the back-end services of the Pilot, deployed in Nova Infrastructure, follow the INFINITECH Way that provides enhanced security.

It should be noted that in a real case scenario, the pilot solution will operate on proprietary infrastructure (cloud or bare metal). Thus, privacy concerns with respect to the trades data will not be an issue

4.3.3 Pilot #2 assessment solution

According to D7.21 [15] the compliance value from Stakeholder’s evaluation is 4.0

Pilot #2 will be offered to enterprises (e.g., banks and asset management firms) and not to the public while the utilized data are open source. Moreover, the Pilot acts as a virtual assistant to the user without taking any automated actions. Thus, this solution does not require special data protection mechanisms.

4.3.4 Pilot #2 recommendations

At production level, the Pilot should examine whether the provided user authentication and authorization processes comply with the customer use case and requirements.

4.4 Pilot #3

Pilot #3 Uses AI to search big data and produce Red Flag typologies that emphasise potential KYC (Know Your Customer) risks regarding human trafficking for Financial Institutions (FI).

TrakffikAnalysis Hub ("TAH"), with the assistance of IBM, has established a data hub (the "TA Hub") for the exchange of information regarding human trafficking, for anti-human-trafficking purposes.

The pilot would rely on a wide number of personal (customer) data.

Data is stored and processed within two distinct environments:

1. Within a financial institution's own secure ringfenced data system
2. Within the TraffikAnalysis Hub.

For addressing the first one, each FI that will use Pilot #3 technology will have all necessary regulatory compliance systems in place as demanded by the appropriate Data regulator

With regard to the second one, Pilot #3 will give access to FIs to TAH's Private Cloud Repository: this consists of data obtained from public / online news media sources plus data contributed by NGO participants in TAH, and includes data relating to identifiable individuals. TA Hub data which are generated from the original source content will allow isolation / highlighting of personal identifiers.
The Private Cloud Repository will be hosted by IBM on behalf of TAH in a single data centre, which may be in Germany, the UK or the US. It will be accessible through a secure set of TAH APIs which are only accessible to regulated FIs (requiring a username, password and unique API key). Access will be limited to law enforcement agencies and KYC risk analysts working within FIs .

4.4.1 Pilot 3 solution

As the FI users of Pilot #3 will be National and International FIs they will also have carried out the necessary DP risk and policy work

The UK Data Protection Act (DPA) permits processing of criminal offence data in a large number of specific circumstances, most of which are not relevant to the TA Hub. Most pertinently, processing (including disclosure) is permitted if it is preventing / detecting unlawful acts and
"...the processing—
(a) is necessary for the purposes of the prevention or detection of an unlawful act [or failure to act;] [and]
(b) must be carried out without the consent of the data subject so as not to prejudice those purposes",

TAH is registered as a controller with the UK Information Commissioners Office and as such is required to meet UK DP laws and, for example, make returns of breaches.

For relevant purposes, data protection and data privacy in the UK are regulated by EU General Data Protection Regulation 2016/679, as implemented into UK law by the European Union (Withdrawal) Act 2018 (the "UK GDPR") [29], and the UK Data Protection Act 2018 (the "UK DPA"), each as amended, to work in UK law in a post-Brexit environment, by the Data Protection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019 (the "UK DP SI") [28]. In this note we refer to the UK GDPR and the UK DPA, including as amended by the UK DP SI, as "UK Data Protection Law".

A copy of the EU version of the GDPR is available at https://UK GDPR-info.eu/. Note, however, that this copy does not take account of the changes made by the UK DP SI when the UK GDPR was re-implemented into UK. The differences are minimal for the purposes of this document.

Architecture including the solution

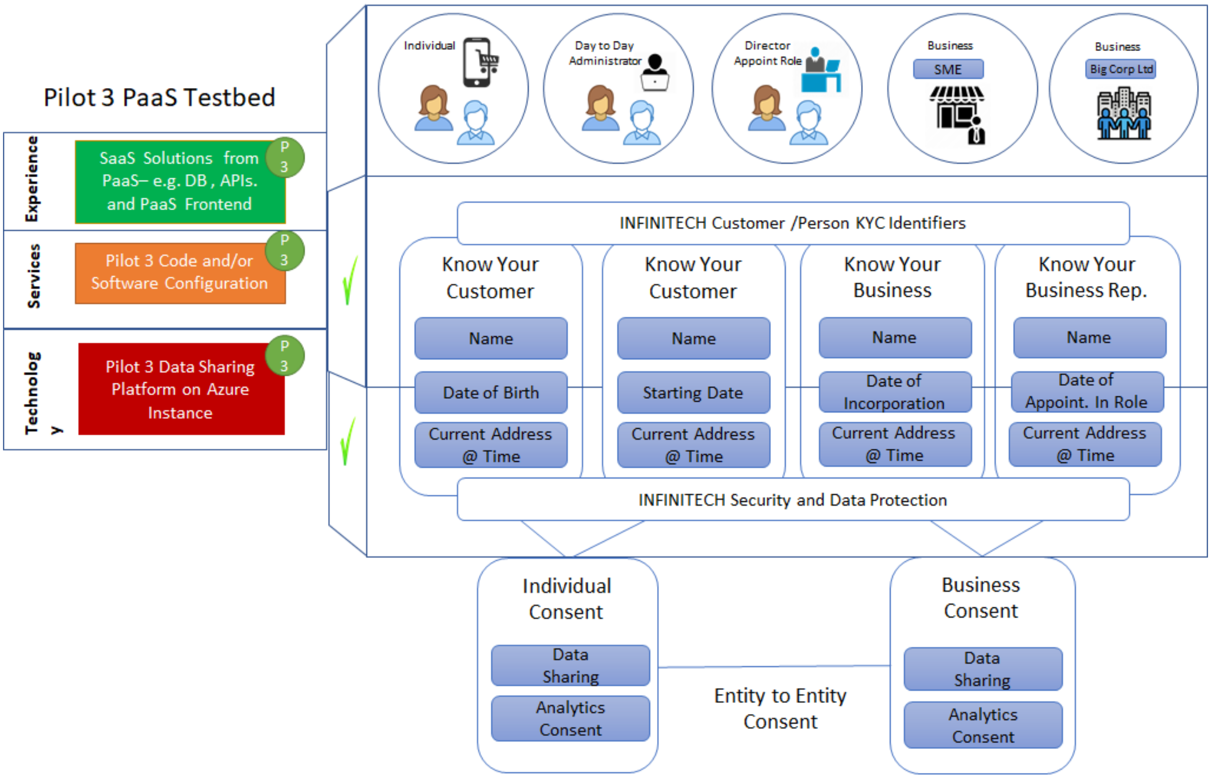


Figure 3: High level architecture diagram for Pilot 3

4.4.2 Pilot 3 real case solution

TAH approached Clifford Chance (CC), an international law firm with expertise in Data regulatory compliance, in order to ensure that TAH's data systems complied with the necessary regulations. (UK data regulations mirror EU GDPR rules). All of CC's advices have been applied by TAH. The result is that TAH takes the view that it needs to process and share data in order to pursue its legitimate interest in combatting human trafficking, and that this interest is not overridden by prejudice to the privacy of the individuals concerned. It allows for the processing of sensitive personal data for "safeguarding" purposes – broadly, because the purpose of the sharing is to combat human trafficking and thereby protect vulnerable individuals.

TAH has already published on its website a privacy notice describing its processing of personal data derived from news stories and contributed data (https://www.traffikanalysis.org/_files/ugd/dfe11e_d6a7fbfd33774bdda96f9df350bb63e4.pdf)

BoI (Bank of Ireland) has also published on its website a privacy notice describing its processing of personal data (https://www.bankofireland.com/privacy/data-protection-notice/)

4.4.3 Pilot 3 assessment solution

According to in D7.21 [15] the compliance value from Stakeholder's evaluation is 3.0

Regarding compliance, the proposed solution in this pilot is addressing current compliance issues, but the solution is evolving over time and then, new issues regarding privacy may arise that could be more challenging

Pilot #3 will ensure that each member of the consortium maintains data privacy compliance by being registered with the appropriate data privacy regulator, keeping records that the regulator demands and taking all action that is necessary to follow the regulator's regulations

4.4.4 Pilot 3 recommendations

TAH has conducted and recorded a data protection impact assessment, taking the approach recommended on the UK Information Commissioner Office's website and recording the potential risks and the steps taken to address them. Regulated FIs are obliged to do the same.
TAH has imposed appropriate contractual privacy undertakings on IBM details can be found at [18]

The data within TAH are protected by strong security arrangements

It is expected that regulated FI users of the Pilot #3 INFINITECH technology have the necessary regulatory systems in place. Pilot #3 carries out substantial due diligence of all users particularly those that have access to the secure APIs within TAH. As secure APIs will only be available to users who are appropriately regulated and registered with a data regulator, Pilot #3 will keep a register of users who have access to the secure APIs together with their regulatory body. Pilot #3 will carry out quarterly due diligence checks with that regulator and with a negative news internet search. In the case of a negative result, the user's access will be immediately suspended and will only be released once a full integration by Pilot #3 has taken place and a satisfactory outcome has been achieved

Pilot #3 takes the view that all data regulatory compliance needs have and will continue to be met because access is only intended to operate within regulated FIs, as well as TAH's Private Cloud Repository is only available to FIs (whether users of Pilot #3 or not)

4.5 Pilot #4

As stated in D3.16 [2], the goal of this pilot is to explore the possibilities of AI-based Portfolio construction for Wealth Management in general regardless of which amount is to be invested. This allows Portfolio construction and optimization for all the customers, not only for the ones with more wealth.

This pilot can be complimentary to potential B2B Customers request with a potential enhancement providing also a Digital onBoarding Authentication Step using the application DUOS (Digital User Onboarding System) provided by Atos. Development of this offering (=DUOS) lies out of the initial Pilot 4 setup within INFINITECH and might potentially be implemented on explicit customer request at later stages. If a potential B2B customer needs such authentication, it will always depend on the customer's existing authentication setup.

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 Portfolio Holdings Data, Financial Market Price Data or Financial Market Asset Master Data. "All datasets will be stored within the Privé SaaS (Software as a Service) solution in a cloud setup. Asset data is mainly fetched from 3rd party databases and from selected market-data providers." [5] Client Data will be provided in all cases directly from the customer's custodian bank.

The output data consists of the single portfolio holdings, their weights and amounts for the proposed portfolios where the advisor or asset manager can finally decide. Fitness Factors Scores and Total Fitness Score will be a further output for both the current and proposed (optimised) portfolio.

4.5.1 Pilot 4 solution

As stated in D3.16 [2], from the point of view of privacy and security, these are different parts of the pilot, which in general could be also operated without any protected personal data included. If personal data shall be provided in a potential B2B Customer setup, then the consequent security and privacy issue can be addressed as described below:

- Customer authentication: in the case that a B2B customer requires (as a precondition) the customer authentication, then it would be possible to provide authentication for the customer in a secure way to get the results of the pilot
- Additional option: If protected data shall not be provided, alternatively an implementation of a Tokenizer (potentially from a third party) would be possible. In this case only anonymized customer data is then provided.
- AI Based Portfolio construction and optimization for Wealth Management: the data source is mostly different price data and newsfeeds which are available depending on a data license. No personal data is needed and collected.

For customer authentication, a possible enhancement might be to adopt the DUOS solution (Digital User Onboarding System - a solution for dealing with virtual identities in a mobile device) on explicit B2B customer request, which lies out of Pilot 4 setup within INFINITECH Project. This solution would come from Atos and it is described in section 2.3 and 3.3 of INFINITECH D3.14 “Data Governance Framework and Tools – III” [3].

Description of a potential setup including DUOS (stated before in D3.16 [2]):

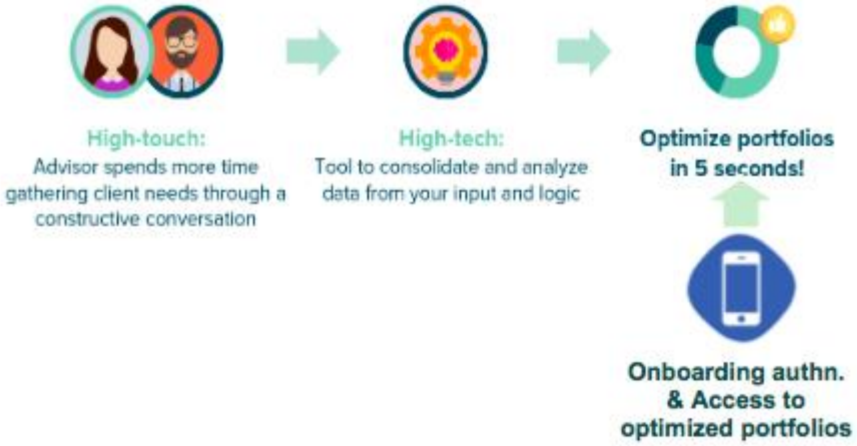
If a B2B customer (Asset Manager, Bank or Advisor company) might require – as a precondition – the identification service for new retail customers, DUOS might be potentially integrated in a real customer application depending on the B2B customer's existing authentication setup.

Based on that potential integration the asset manager/advisor can then provide their new retail-clients the risk-profiling and personalisation service for their investments and will provide a portfolio optimization.

The retail customer would be registered in a portal (defined by the bank/asset manager/Advisor) so that they can access this portfolio construction service.

In this case, the customer-specific “Customer onboarding solution” shall then call the DUOS application (if licensed by the B2B customer).

Figure 4 shows the potential integration of DUOS within the main workflow (only the initial step of authentication) in pilot 4:


High-touch:
Advisor spends more time gathering client needs through a constructive conversation
High-tech:
Tool to consolidate and analyze data from your input and logic
Optimize portfolios in 5 seconds!
Onboarding authn. & Access to optimized portfolios

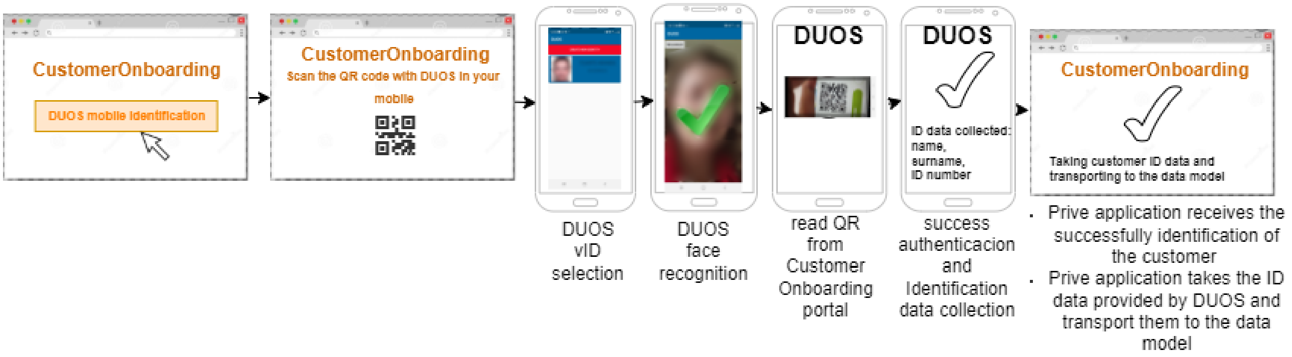

CustomerOnboarding
DUOS mobile identification
CustomerOnboarding
Scan the QR code with DUOS in your mobile
DUOS
DUOS
ID data collected:
name,
surname,
ID number
CustomerOnboarding
Taking customer ID data and transporting to the data model
DUOS vID selection
DUOS face recognition
read QR from Customer Onboarding portal
success authenticacion and Identification data collection
Prive application receives the successfully identification of the customer
Prive application takes the ID data provided by DUOS and transport them to the data model

Figure 4: potential integration of DUOS within the main workflow (only the initial step of authentication) in pilot 4

These are the steps within the DUOS workflow:

1. Customer Onboarding-DUOS mobile identification: the first step is that it is required that the B2B Customer portal offers through DUOS mobile identification by clicking the button.
2. Customer Onboarding- QR: the B2B portal provides a QR code which contains a URL that will be used later to send the identification data obtained in DUOS.
3. DUOS-vID selection: In the mobile phone, DUOS allows users to select between different virtual identities (note that there must be a previous virtual identities registration in the DUOS app that is described in [3]). The user must select one of them; after that selection, all the data are taken (data from the chip, data from the MRZ zone and face image stored in the chip)
4. DUOS-face recognition: The app asks to capture a face image in order to detect that the person that is using the app is the one that is using the selected identity.
5. DUOS-read QR: the B2B customer must read the QR code shown in the Customer onboarding portal
6. Success authentication: in this case the authentication is correct and DUOS sends to the URL (the customer Onboarding portal would sent it before via QR) the identification data agreed between both applications such as the name, surname, ID number and birth date.
7. Successfully identified: the B2B customer onboarding portal receives the data from DUOS and transports them to the data model.

There would be an interface between the B2B Customer Onboarding Portal and DUOS application:

1. Customer Onboarding -> DUOS: DUOS app reads the QR code provided by B2B Customer Onboarding portal, this QR code contains the URL of the web service and a token, and the B2B Customer Onboarding portal then is waiting for the response
2. DUOS performs the authentication and then when it is successful, the DUOS app uses the URL and the token that is captured in the QR code and calls this web service (from the Customer Onboarding portal) and sends the necessary data for identification. A proposal of this data is to send the name and the identity card number in order that the B2B Customer Onboarding portal could perform a “login” that ensures that the data the customer previously filled in the form (in this case name and identity card number) are the same data that DUOS is sending

4.5.2 Pilot 4 real case solution

The concept within Pilot 4 for portfolio construction is based on different personal preferences of mostly anonymous end-clients (multiple portfolio "health/fitness" factors which compete partially with each other - historically only humans can factor this - and now this is available via an automated calculation process), and the full advisory journey for creating these kind of personalised portfolio proposals has no need for any PII(Personal Identifiable Information) data stored on the SaaS infrastructure of the pilot. Generally, the Pilot’s solution would allow financial intermediaries and advisors to offer bespoke wealth management solutions to customers at scale through use of AI based portfolio construction tools while adhering to various regulatory compliances. Both quantitative and qualitative historical data will be leveraged for personalizing the portfolio construction and investment proposals, rendering the journey clear from a regulatory perspective.

4.5.3 Pilot 4 assessment solution

According to in D7.21 [15] the compliance value from Stakeholder's evaluation is 4.2

The Pilot's solution provides clear compliance and would allow financial intermediaries and advisors to offer bespoke wealth management solutions to customers at the same time that provides regulatory compliance. "Both quantitative and qualitative historical data will be leveraged for personalizing the portfolio construction and investment proposals, rendering the journey clear from a regulatory perspective" [15]

4.5.4 Pilot 4 recommendations

In cases, where potentially some parts of PII data may be forwarded to the Privé AIGO SaaS platform, Privé and the respective cloud providers (AWS) follow all required rules for GDPR compliance. For this purpose Privé also has a licensed external Data Protection Officer (DPO), who would be consulted in these setups from the rather beginning.

In addition, Privé is capable of delivering a service that fully obfuscates any or all sensitive data by tokenizing them on the customer premises before they are being sent to the SaaS application of Privé. The tokenization service holds a translation table that creates reliable, static and consistent tokens tied to user specifics required by the SaaS application to verify or authenticate a specific user. On the SaaS side only the token is known and needed.

4.6 Pilot 5a

This pilot is not part of the project anymore

4.7 Pilot 5b

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." [5]

It should be mentioned that the pilot is still under development and so far the already developed services include the transaction categorization, the cashflow prediction, the benchmarking tool and transaction monitoring). The data that is used relates to SME transactions and no specific individuals making compliance with the GDPR easier.

Also the microservices were created following the INFINITECH way and are hosted/deployed in AWS infrastructure and the security measures are focused on data privacy. Furthermore, measures were taken to ensure integrity and robustness of the AI models using XAI techniques to re-evaluate the data models as part of the ML process. [25]

4.7.1 Pilot 5b solution

As stated in D3.16 [2], 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, especially data related to invoices. Furthermore the description of each transaction may include names of individuals or sensitive information.

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 by the tokenized data the identity of the involved persons, which in practice leads to those data being anonymized for UPRC.

The data is under discussion if need to be encrypted at rest and in transit, lowering the risk of data poisoning. That implies that the Infinstore may have to support encrypted queries [13].

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

Reliability and fallback are handled by the INFINITECH way of rest containerized microservices. Also extensive tests for defining the parameters of the model making their results reproduceable were taken into account [13].

Finally, to guard against data poisoning and possible model adversaries the AI models are trained and validated under a golden standard subset of the data that is used to evaluate the outcomes of the AI models as proof of being valid.

The outcomes of each model are stored in the Infinistore so they can be monitored and checked.

Architecture

Figure 5 shows how the pseudonymisation module fits in the system. Note how all data travelling from one module to the other passes through the pseudonymization module. Moreover, open-source data will also be leveraged in the BFM toolkit, as well as open banking data.

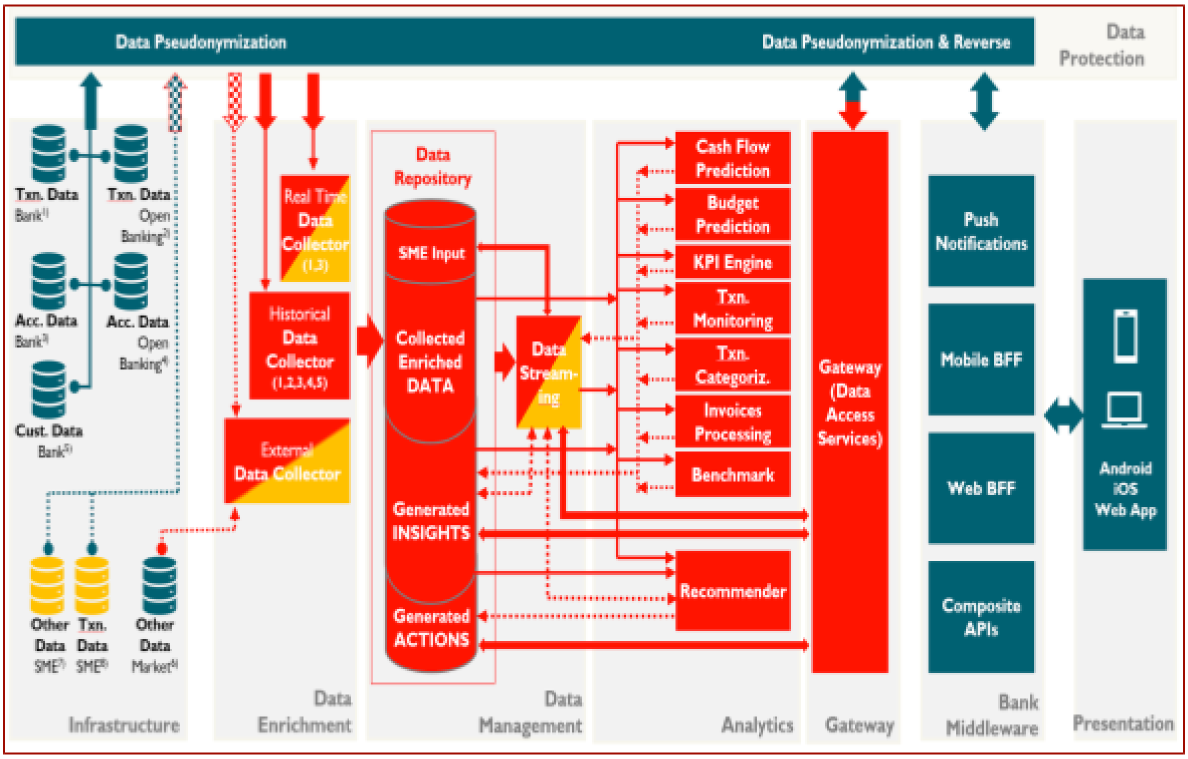


Figure 5: INFINITECH Pilot #5b RA

Expected results

Useful and operational-grade financial advice to BoC customers.

4.7.2 Pilot 5b real case solution

Data tokenization (or encryption when possible) should be continued and performed on all data exiting in the bank's premises, eliminating the chance of externally tracing the end user. Furthermore. security auditing and penetration tests should be applied.

Additionally, the legal department of the bank should thoroughly review the SLAs and the terms of cloud providers before using the offered services.

4.7.3 Pilot 5b assessment solution

According to in D7.21 [15] the compliance value from Stakeholder's evaluation is 4.0.

In this pilot the compliance constraints have been considered at a project level as well as internally at the Bank's level [15].

4.7.4 Pilot 5b recommendations

At a future operational level, the EBA 2017 guidelines [24] for outsourcing to cloud service providers should be given careful consideration. These recommendations are intended to provide guidance on outsourcing by institutions to cloud service providers.

4.8 Pilot 6

As stated in D3.16 [2], 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.

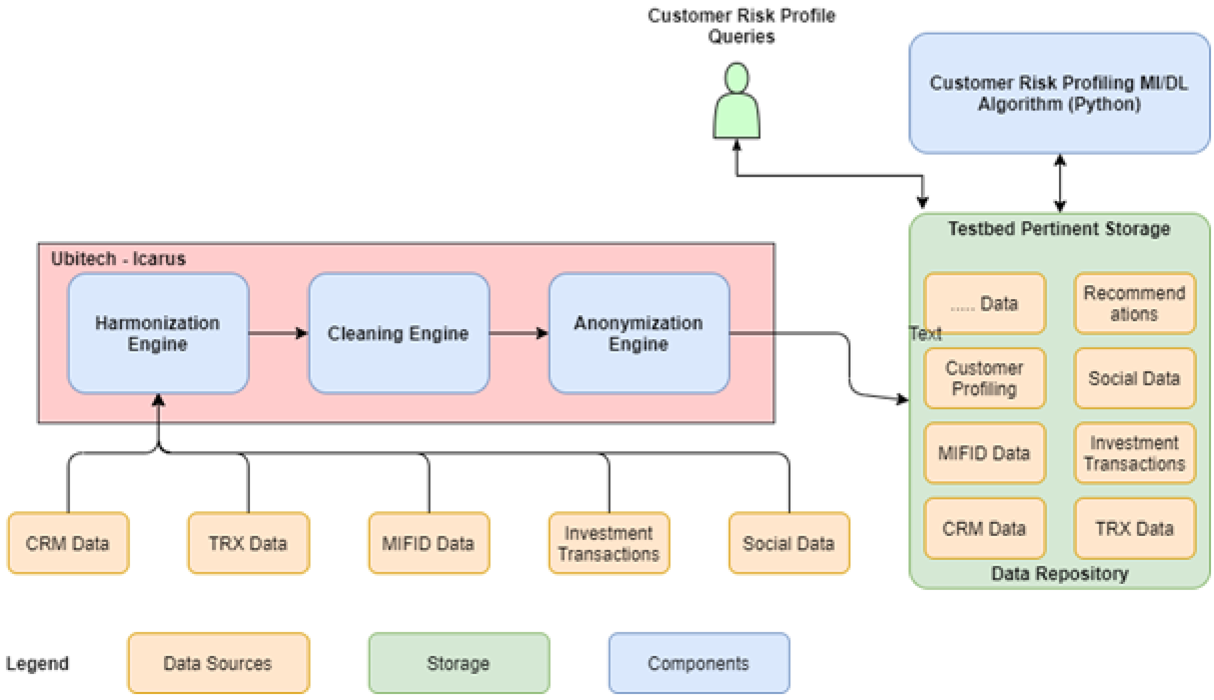
Creation of personalized investment recommendations available for all Retail Customers and not only to highly affluent. Development of algorithms that aim to perform Customer profiling and categorization according to their intention to invest, based not only in questionnaire input but also in transactional activity. The aim is to create a service, available to financial advisors, which not only examines each customer's transactional activity but also takes into account similarities and patterns among customers.

4.8.1 Pilot 6 solution

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

Each customer's personal data is anonymized in order to avoid the identification of individuals.

Figure 6 shows how the anonymization engine fits in the system. Note how all data traveling from one module to other passes through the anonymization.



The architecture in a more detailed version may be also found on the diagram below (Figure 6):

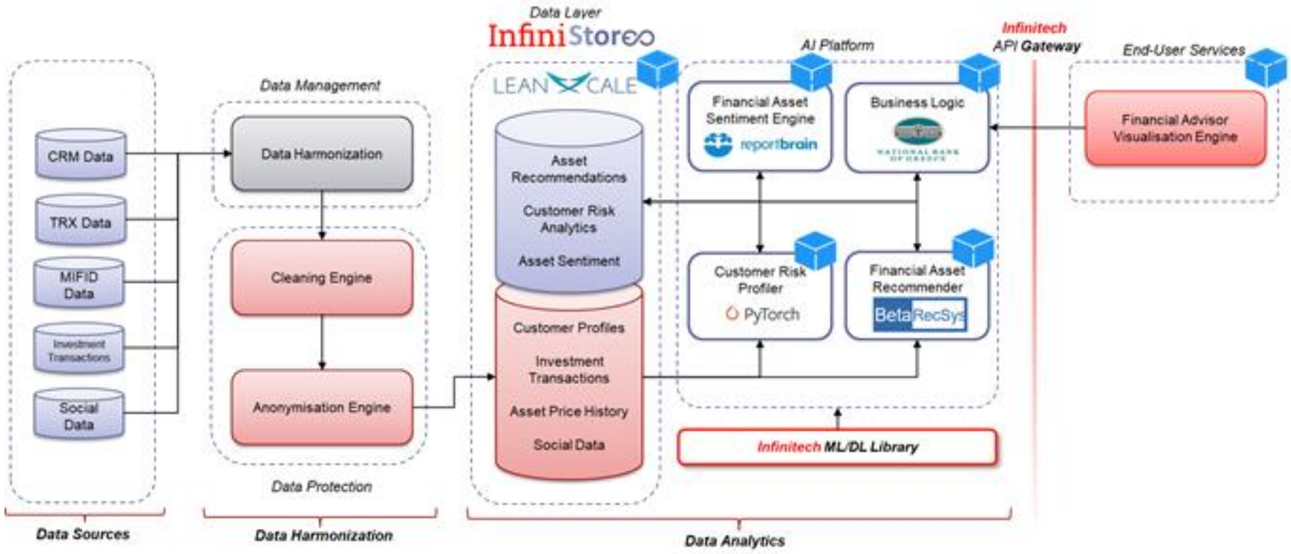


Figure 6: Pilot #6 Architecture

4.8.2 Pilot 6 real case solution

As mentioned in sector 4.8.1, Pilot 6 will not use personal data. Each customer's personal data is anonymized in order to avoid any possible identification of individuals, since in the development are involved third parties, outside from NBG.

In real life, the use of any personal data comes with the respective consent from the individual, following the respective regulations.

4.8.3 Pilot 6 assessment solution

According to in D7.21 [15] the compliance value from Stakeholder's evaluation is 4.0.

Pilot #6 will use anonymized data, which makes GDPR not applicable in that case. The anonymization component will be implemented as described in the architecture approach (figure 6). Moreover, on top of anonymized data, Pilot #6 will use social network information, which in any case does not contain any personal information, but only public information related to assets.

4.8.4 Pilot 6 recommendations

Pilot #6 provides regulatory compliance solutions in both cases, for the INFINITECH pilot case and for the production or real case perspective. Applying the proposed measures in every case, the solutions are regulatory compliant.

4.9 Pilot 7

As stated in D3.16 [2], the main goal of Pilot #7 is to explore how next generation technical solutions like Machine Learning, together with advanced modelling could help to create a more accurate, comprehensive and near real-time picture of suspicious behaviour in the Financial Crime and Fraud with the final objective of stealing the bank customers' identity and money.

In the Financial Crime and Fraud Intelligence scene, Machine Learning has the ground-breaking potential to reveal much more realistic Financial Crime typologies, compared with traditional rule-based systems. Traditional screening systems don't evolve with criminal behaviour and result in high false positive rates, while potentially overlooking the real suspicious behaviour.

The pilot is focused on a particular kind of transaction which ultimately has been targeted by fraudsters. Typically, these transactions are immediate loans. These transactions are important in terms of security because fraudsters have realized that most of the harvested stolen accounts have no money and it is a way to get money by immediate loans to be transferred to other fraudulent accounts outside the bank. Therefore, we can use this type of transaction as a way to determine if a final transaction is fraudulent or not.

Machine learning is crucial to accomplish this objective as to assign them a risk score it is necessary to evaluate a lot of variables and assess different kind of environments. For this reason we are piloting this transaction of requesting an obtaining immediate loans, because if this works, we will be able to adopt the same strategy followed by this INFINITECH's pilot to other kind of transactions important to detect fraud.

4.9.1 Pilot 7 solution

The major regulatory requirements of Pilot #7 beyond the common and regular IT Security Management standards for banking are related to the GDPR. As Figure 7 illustrates, two pathways are followed:

- Synthetic data are generated by utilizing the commercial solution. Mostly.ai (https://mostly.ai) which generates intelligent synthetic data using AI enabling to generate realistic data without compromising sensitive or personal information. Unlike other tools for generating synthetic data, monstly.ai is able to generate realistic data which can be used to do data analytics because it conserves and maintains the relations of the real data. Those data will be used for training the AI model. This way, no personal data are involved while keeping the statistical features of the data as much as possible.

- The GRADIANT anonymization tool will be assessed if feasible to anonymise real data, which shall facilitate a validation of the pilot's solution based on the actual bank's data. In case of legal issues synthetized data shall be a fall-back option for technical validation only.

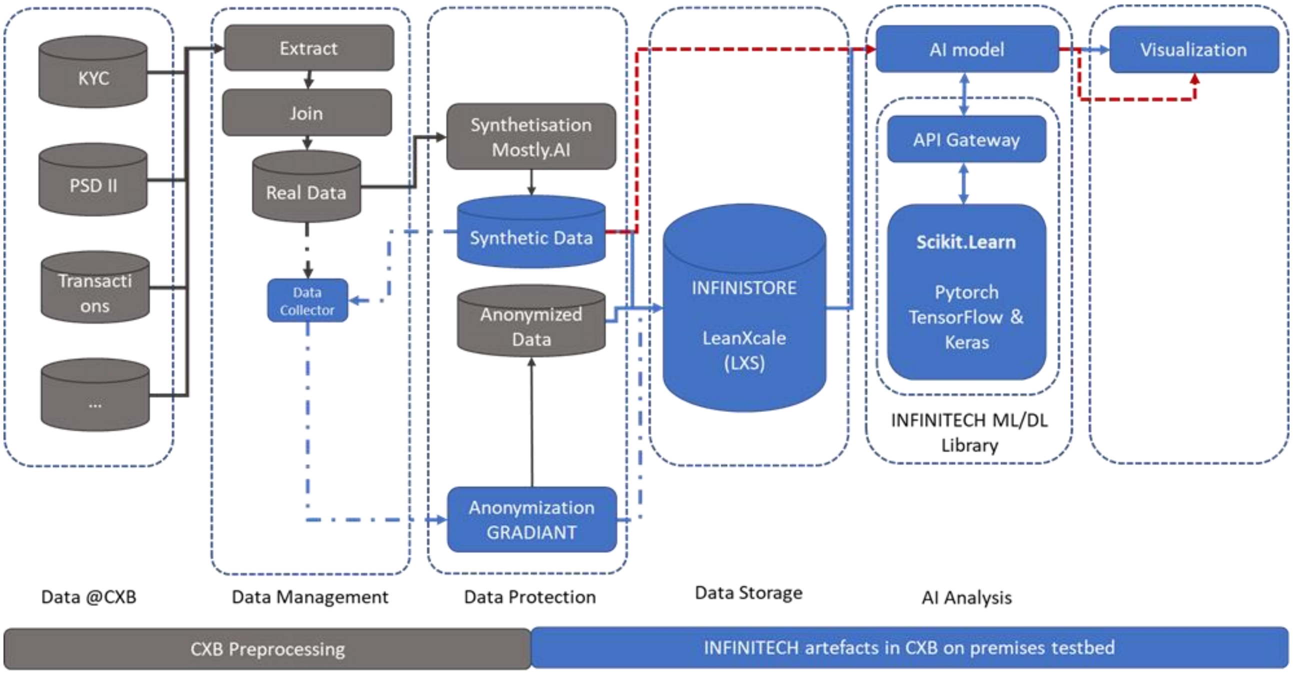


Figure 7: Pilot #7 planned Solution

In either way anonymized or synthetized data shall be used for scoring the risk of a fraudulent request

4.9.2 Pilot 7 real case solution

The assessment of the anonymization tool will provide insights within the context of a real world data application.

This pilot is treating a special case in terms of compliance because there is an exception in the GDPR that would allow the usage of personal data if this is necessary to protect the customers, so for fraud prevention it is possible to use personal data without the final users' consent in the production environment. This exception is recognized in two of its recitals[23]:

- Recital 47: "The processing of personal data strictly necessary for the purposes of preventing fraud also constitutes a legitimate interest of the data controller concerned…"[23]
- Recital 71: "decision-making based on … profiling should be allowed where expressly authorised by … law … including for fraud or tax evasion monitoring and prevention purposes"[23]

Although in the production environment and to prevent fraud it is possible to use the real data without the necessity of getting the consent of the customers, for the pilot we have use synthetic and anonymized data, because the data has been treated outside our premises by third parties (the members of the INFINITECH's consortium). Most importantly, the objective of this Pilot is a demonstration that the technology is working for this objective, but it is not actually a production project.

4.9.3 Pilot 7 assessment solution

According to in D7.21 [15] the compliance value from stakeholders evaluation is 3.0.

"Confidentiality of data is crucial in this case because AI solutions needs almost real data and although GDPR is making an exception for the usage of personal data when protecting the client from fraud, this exception is not applicable if the usage of personal data is in the innovation or development phase, for this reason in this project we have dedicated a lot of efforts to determine which way is the best to extract data that can be used to build AI algorithms without personal and sensitive information. In our case we have seen that the usage of synthetic data is a feasible way to accomplish it. The problem is that normally synthetic data has not valid information to work with AI or Data Analytics tools as synthetic data follows only patterns and losses important relationships to apply AI. In our case we have tried the tools Mostly.ai which provides intelligent synthetics data using AI too, to build a recipe of synthetic data that can also be used to apply AI to the outcome. The results have been very promising in Pilot #7 and seems to be a valid way to work with restricted data outside the organization respecting the GDPR limitations" [15]

4.9.4 Pilot 7 recommendations

For the purpose of piloting and validating in the INFINITECH-way the AI system based on compliant data the usage of synthetic data shall simplify a compliant application. As a 2nd pathway the anonymization tool by GRAD is assessed on its feasibility at least for validation purpose. In the case of feasibility, it is proposed to check the pipeline using this tool even with the usage of synthetized data for simplicity of implementation.

4.10 Pilot 8

The objective of the pilot is to develop a Platform for anti-money laundering Supervision (PAMLS), which will improve the effectiveness of the existing supervisory activities in the area of anti-ML/FT by processing large quantities of data owned by BOS and other competent authorities.

Specific security and privacy issues and requirements for this pilot are already stated in *D3.15 – Regulatory Compliance Tools I - SECRET annex for pilots 8*.

4.10.1 Pilot 8 solution

As stated in D3.16 – Regulatory Compliance Tools – II [2], in order to be compliant with applicable data protection rules at the EU and national level personal data about the individuals and confidential information on legal entities within the transactions will be pseudo-anonymized prior to data delivery to PAMLS. The End user of PAMLS will not be able to identify (directly or indirectly) individuals behind the transactions.

The Pseudo-anonymization tool will be provided by the JSI. Detailed description of the pseudo-anonymization tool is described in section 2.1 of INFINITECH D3.14 "Data Governance Framework and Tools – III" [14].

4.10.2 Pilot 8 real case solution

The main goal of the Pseudo-anonymization tool is to ensure that sensitive data is hidden both during the development and production use, while still ensuring that the underlying structure of the data stays the same and supports analysis and reasoning. Namely, the connections between different subjects should still be visible after pseudo-anonymization, but sensitive data (iban number, names) should either be anonymized, masked or removed to ensure privacy. The Pseudo-anonymization tool will be the first component of the data preparation pipeline to remove and mask identifying information from provided datasets (legal compliance) and to ensure that (as specified by configuration) same subjects will be anonymized in the same (or similar) way to enable the pattern recognition algorithms at a later point in the pipeline. To ensure sufficient and robust anonymization, multiple battle-hardened anonymization algorithms are available and can be individually configured for specific use cases.

Combining already anonymized data with publicly available datasets can introduce security concerns, as properties of a publicly available dataset can be leveraged to (partially) deanonymize existing datasets which can be an important regulatory concern. To combat the possibility of such deanonymization, the Pseudo-anonymization tool can be configured to introduce information specific noise during the anonymization procedure in both a deterministic and non-deterministic way. The pipeline can therefore ingest information about publicly available transactions, mask identifying information (transaction reason) and introduce noise to partially identifying information that is critical for further data analysis. In particular, data of transactions are randomly moved by a few days, transaction amounts are rounded and random noise is introduced.... This causes only minimal data loss, while fully ensuring compliance with regards to the handling of sensitive data.

In a real cases scenario, additional measures to ensure proper handling of data will automatically be enforced by the Pseudo-anonymization tool. The tool keeps an auditable access log, disables anonymization of small datasets and randomly shuffles anonymized data to prevent salt cracking based abuse.

The Pseudoanonymizer also enables the usage and sharing of sensitive data between different stakeholders. Using a predefined seed allows to obtain fully anonymized datasets (for the end user) from different providers and then combine them. The specific use case of Bank of Slovenia data and public data of Office for Money Laundering Prevention was implemented and tested to provide information on regulatory compliance and usability and the results showed that a combination of datasets will greatly enhance the information gained during analysis. All the above-mentioned compliance and monitoring procedures were in place during the test and the analysis showed that the presented steps form a sufficiently stable and robust process that enables data sharing with full compliance of privacy-related rules.

4.10.3 Pilot 8 assessment solution

According to in D7.21 [15] the compliance value from Stakeholders evaluation is 4.0. A compliance strength in this pilot comes from the inclusion in the project of a team of experts from Legal and Compliance departments.

4.10.4 Pilot 8 recommendations

The most suitable way to fulfil the compliance standards is to ensure that the Pseudo-anonymization tool is the first part of the data processing pipeline. Great care must be taken to ensure that sharing of initial configurations and salts used for pseudo anonymization are communicated in a secure manner and preferably never available to human operators. To ensure sufficient separation, the pseudo-anonymization process should be run separately, and the rest of the data pipeline never observes the original data.

4.11 Pilot 9

“Blockchain crypto currencies and tokenized assets that are obtained fraudulently can go through various transfers” [5] on the blockchain and end up as stable coins (e.g. USD, EUR and TRY tokens) in different jurisdictions. Pilot #9 performs (i) parallel transaction graph construction and graph traversal-based analysis on an HPC (High Performance Computing) cluster and (ii) its user interface to provide blockchain transaction graph visualization.

4.11.1 Pilot 9 solution

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

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 (Figure 8):

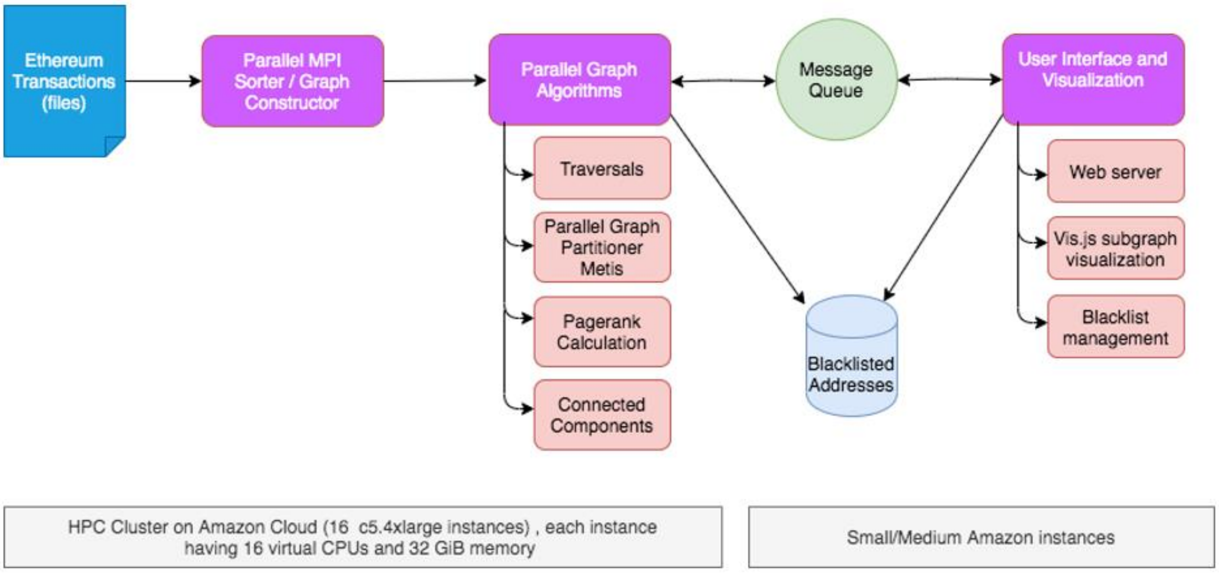

Ethereum Transactions (files)
Parallel MPI Sorter / Graph Constructor
Parallel Graph Algorithms
Message Queue
User Interface and Visualization
Traversals
Parallel Graph Partitioner Metis
Pagerank Calculation
Connected Components
Blacklisted Addresses
Web server
Vis.js subgraph visualization
Blacklist 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

4.11.2 Pilot 9 real case solution

Within the context of a real use case, when the blockchain data, that is used in the pilot, is used within organizations, only the identity of the customer will be linked to the blockchain address that is declared to be owned by the user. Only the specific customer's data will be shown when the transaction graphs are displayed. Other customers' data will not be shown. Hence, existing GDPR/AMLD4 practices, that are routinely carried out on each customer, will be applied as usual. Therefore, no additional practices will be necessary beyond those that are currently in practice.

Examples for existing practices include applying KYC (Know Your Customer), which checks for customer identification and providing secure authentication mechanisms for access control. Since blockchain data that is shown in the transaction graphs are public and blockchain addresses are pseudo-anonymous, there is no additional privacy issue. In Turkey, crypto regulations are still being discussed and not passed. Hence, depending on development of new regulations related to blockchain and crypto assets, further obligations may arise. For example, organizations may be obliged to screen their customers' banking transactions, apply EDD (Enhanced Due Diligence) to their customers and report suspicious transactions to financial crime authorities.

4.11.3 Pilot 9 assessment solution

According to in D7.21 [15] the compliance value from Stakeholder's evaluation is 4.0.

4.11.4 Pilot 9 recommendations

Blockchain regulations are still evolving. Global money laundering and terrorist financing watchdog FATF has published guidelines [22]. They can be adopted in the future, provided these guidelines are approved by the governmental authorities.

4.12 Pilot 10

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.

.

4.12.1 Pilot 10 solution

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.

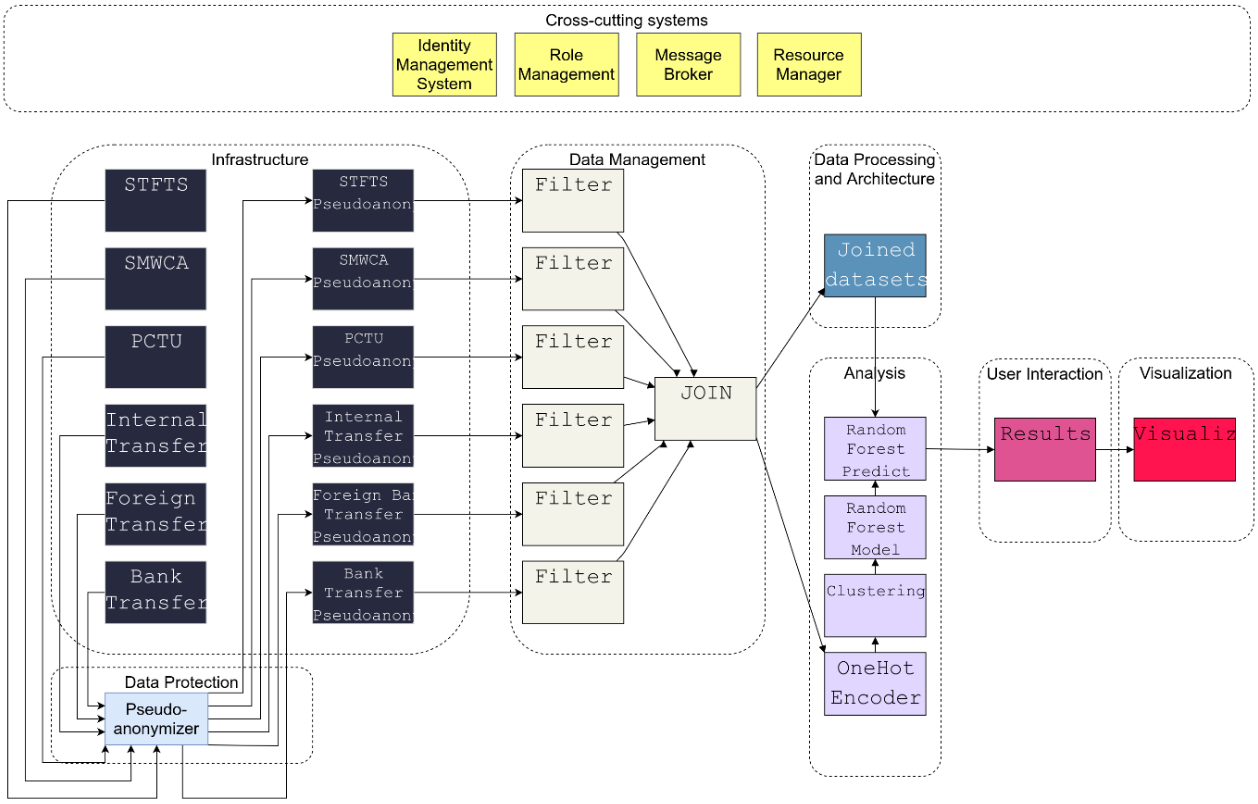


Figure 9: Pilot #10 INFINITECH RA

Expected results

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

4.12.2 Pilot 10 assessment solution

According to in D7.21 [15] the compliance value from Stakeholder's evaluation is 4.5.

A compliance strength in this pilot comes from that the fraud detection system under development, is designed by considering the banking data sovereignty and such prototype will be released to be integrated on premise so to keep all the existing internal procedures and avoid the external data processing.

4.13 Pilot 11

As previously exposed in D3.16 [2] Pilot #11 aims to improve the analysis, definition and assignment of risk profiles in car insurance scenarios. To do so, the pilot combines the information collected from connected real vehicles in real time (including speed, fuel consumption, acceleration, or geo-spatial information, among others) with related context information such as weather parameters, traffic alerts and roads' information, to enhance Usage Based Insurance products by developing two main services: a) "Pay as you drive" service which adapts insurance premiums according to the drivers' profile; and b) the "Fraud detection" service, which exploits driving profiles to identify possible undeclared drivers and driving risks. These services are defined and implemented by the pilot's insurance company, DYN.

The pilot develops an AI powered driver profiling model, supported by Atos, fed by a data collection and homogenisation (based on ETSI NGSI standard and the FIWARE Smart Data Models) framework (Atos' SmartFleet).

The information from connected cars is captured by a device developed and installed by CTAG, which directly connects with the vehicle's sensor network (CAN Bus). The collected data is grouped into "routes", as technical vehicle's data during a given time window. These are used as the basic data information set to train and test the AI models. Context information is later added to the model definition and training process to enhance and extend its capabilities.

Since geo-located data is considered sensitive by the General Data Protection Regulation (GDPR), it must be properly protected. The Pilot foresees two main security and privacy issues: unauthorized access to the different modules of the platform, and the use of sensitive data to train AI models.

4.13.1 Pilot 11 solution

Pilot #11 implements different solutions to address the identified security and privacy issues.

First, prevention of unauthorized access to the platform will be provided by ATOS with OAuth 2.0 based authorization mechanisms to access the platform [12].
Regarding the collection and processing of sensitive data, the driver of the connected car will answer an "ask for consent" for the data treatment. Certain identifiers, such as the user identifier, will be directly pseudo-anonymized by the driver profile collection tool from CTAG, while other sensitive information such as GPS location will be anonymized to protect user privacy by using the Regulatory Compliance Tool by GRAD described in section 3.1.9 of deliverable D2.6 [7]. In addition to apply privacy enhancing technologies like anonymization, the data will be stored and classified in the AI INFINITECH Pilot #11 platform.

The anonymization component and security framework will be integrated in the INFINITECH Pilot #11 architecture as shown in Figure 10.

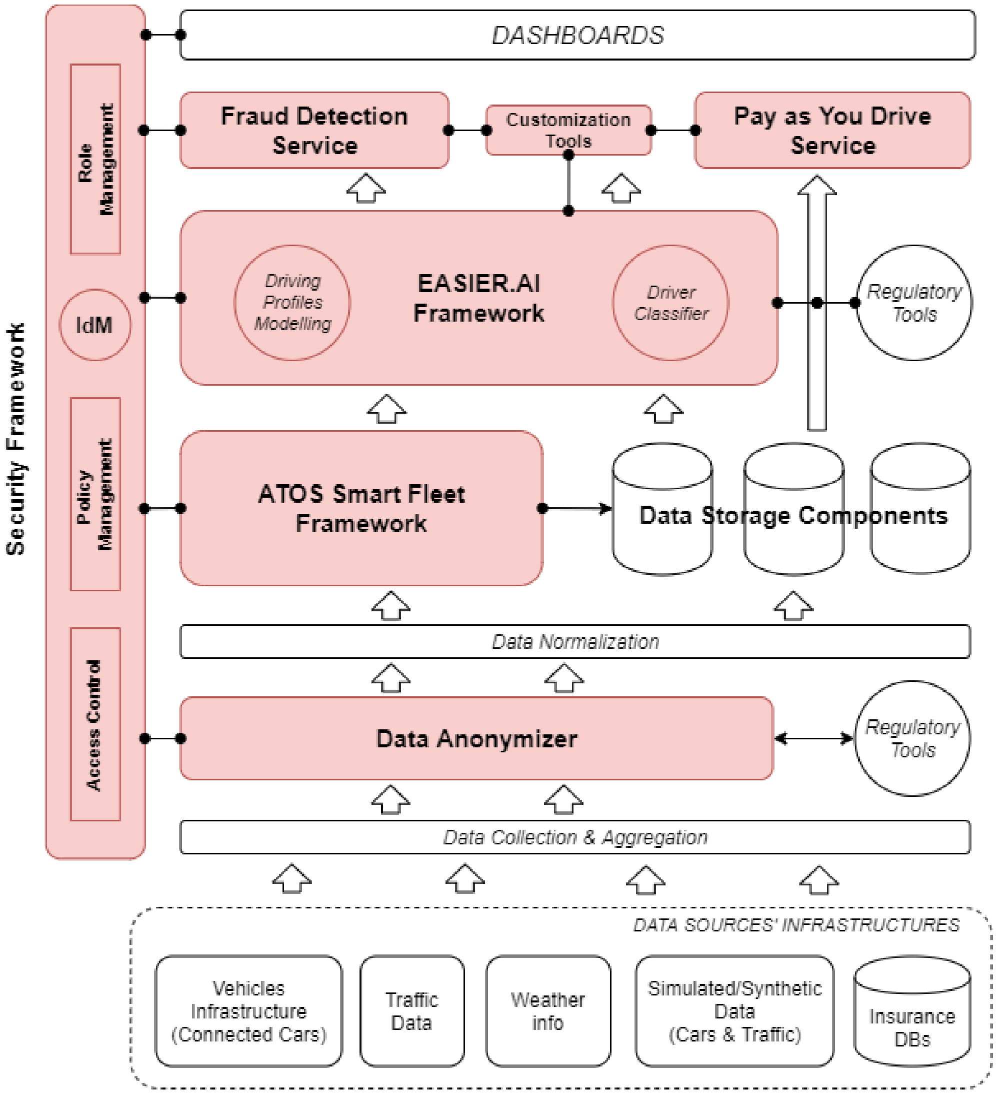


Figure 10: Pilot #11 – High level architecture

The anonymization component (within Pilot #11) is intended to anonymize GPS data from connected cars in real time. First, we will use already collected data from ATOS Smart Fleet historical storage in order to develop the location data anonymization algorithm and decide the best anonymization configuration that meets the privacy and utility requirements of the pilot. The set of specific techniques to be applied is to be defined, but it includes the anonymization of the GPS position of the connected vehicles by using either Differential Privacy (geo-indistinguishability) and/or Spatial Cloaking. Both techniques aim at modifying the location of the vehicles by introducing controlled statistical noise (in the first case) or by grouping the locations of nearby vehicles together in order to achieve privacy. The ultimate goal is to either obfuscate the precise location of the user (to protect individual GPS points) and/or to hide the location of individual vehicles by reporting the same location for a set of vehicles. The pilot will experiment with different levels of anonymization in order to determine the best approach to fulfil the privacy and data quality requirements, this is, to provide a certain level of privacy to end users, but without altering the data in a way that makes it unusable for artificial intelligence algorithms.

Once the anonymization algorithm is developed, the anonymization component will be integrated on the real-time data collection flow (by deploying an endpoint to receive data from connected cars). The anonymization component will apply the selected anonymization configuration to the data in real time, and store the anonymized data using the ATOS Smart Fleet framework.

]

4.13.2 Pilot 11 real case solution

From Pilot's #11 production perspective, we consider two different scenarios that requires different sensitive data from involved individuals:

1. The development of the AI Driver Profiling Model, which may include the design and development of a new model or the enhancement of the provided one. Both solutions will cover the training and testing stages of the AI model. For these purposes, technical information from real connected vehicles will be required. The approach exploited within Pilot #11 considers only vehicles' routes that discards any information directly related to the identification of the driver or the car itself. Only GPS information is considered as sensitive in this scenario. Depending on the desired accuracy of the Driver Profiling model, this information can be anonymised (e.g. by the GRAD Anonymizer tool). In the case that these GPS datasets are completely anonymised before entering the AI model development, training or testing process, there is no initial GDPR issue affecting the scenario, since the data is not considered personal anymore. In the case that raw GPS data is required (because the impact of the anonymization procedure is too high), all participant drivers involved in any of the AI development stages MUST be informed and their corresponding explicit consent (from each individual) is required.
2. In the usage of the AI model for drivers' profiling, by, e.g. the Pay as You Drive or the Fraud Detection services scenario, the services are completely hosted by the insurance company (or the entity providing the service). The insurance company owns the data, the infrastructure and the outcomes of the services and NOTHING is shared with any other third party. In this sense, the AI model is distributed as a standalone module, deployed by the insurance company within the insurance company premises, and it doesn't consume nor share any other data different from those managed by the insurance company. Considering all this, the results of the services may refer, produce or consume sensitive data from insured clients. In this sense, and in this scenario, all GDPR constraints MUST be considered by the insurance (or the entity providing the services) company

4.13.3 Pilot 11 assessment solution

According to in D7.21 [15] the compliance value from Stakeholders evaluation is 4.0.

4.13.4 Pilot 11 recommendations

Pilot #11 is provides regulatory compliance solutions in both cases, for the INFINITECH pilot case and for the production or real case perspective. Applying the proposed measures in both cases, the solutions are regulatory compliance.

4.14 Pilot 12

As stated in D3.16 [2], Pilot #12 aims to improve the analysis, definition and assignment of risk profiles in health insurance, by using the information collected from IoT devices and questionnaires. The pilot applies Machine Learning to develop two different services: a risk assessment service and a fraudulent behaviour detection service.

To this end, the Healthentia app is leveraged to collect data from real users, by means of different activity trackers (such as Fitbit devices, Android sensors and Apple Health Kit data) and questionnaires (from psychological to social and environmental aspects). Synthetic data (simulated lifestyle) will be collected in the context of the pilot. Therefore, the collected data from the devices are sensitive since it is 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 risk score for each user. The health insurance companies will use this score to adapt the price of the customers' premium.

Given the nature of the collected data, the Pilot foresees two main security and privacy issues: user authentication and the use of personal data to train AI models.

4.14.1 Pilot 12 solution

As a solution for regulatory compliance, the Healthentia application's users will each sign a consent form in order to allow the collection and processing of their data. Regarding user authentication, the pilot solves this internally by means of the access control framework.

As this pilot uses personal data for training the ML models, we are developing a regulatory compliance tool to call an anonymization tool developed by GRAD; this regulatory tool will protect the user's privacy. The regulatory tool uses the General INFINITECH Regulatory Compliance tool which is based on the DPO developed by Atos that orchestrates the calls to different security or privacy tools.

The following diagram shows the Architecture of this pilot:

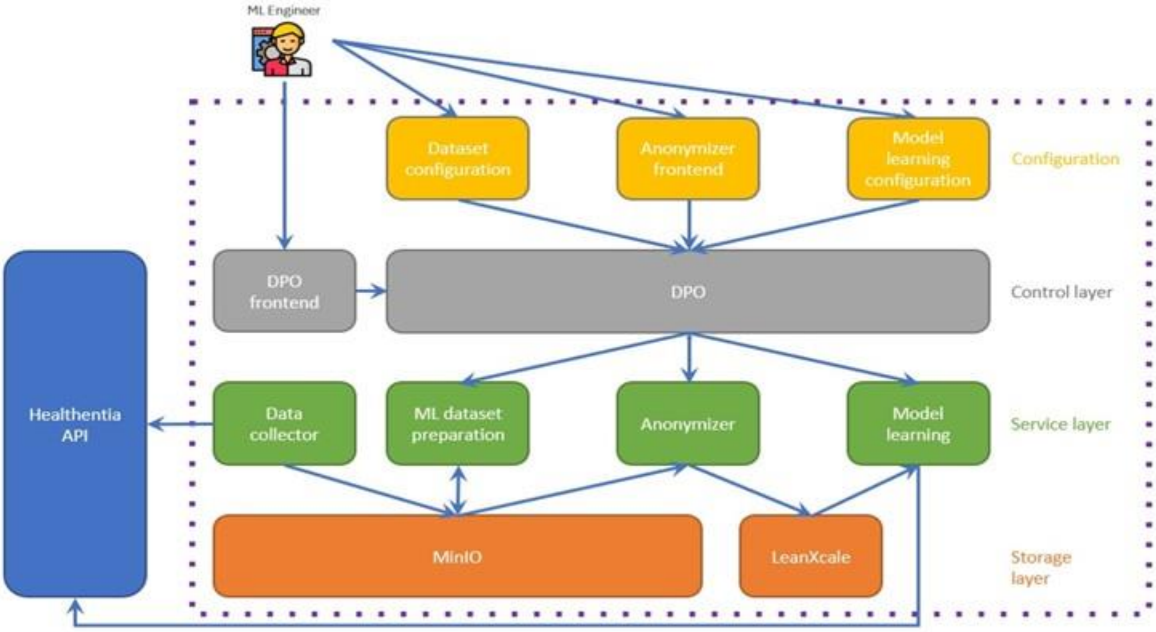


Figure 11: INFINITECH Pilot #12 Architecture

These are the interactions between the components in more detail:

1. Data collector runs autonomously, asking Healthentia API for data and storing daily CSVs in MinIO
2. A Machine Learning Engineer decides the datasets needed to learn the model, decides on the set of anonymization operations, uses the DPO frontend to start the anonymization process and decides on the model learning parameters.
3. The Machine Learning dataset preparation service prepares data for ML experiment, reads daily CSVs from MinIO and writes single CSV
4. The DPO controls the call to Anonymizer in order to anonymize data, it is activated by ML engineer and notifies ML engineer once the data is ready.
5. Anonymizer service anonymizes data to be used in offline model learning, reads single CSV file from minIO and writes data in different anonymization level in LeanXcale.

6. Model learning service learns and validates model, reads vectors at certain anonymization level from Lean Xcale and writew model to Healthentia via the API.

The DPO will have a key role in regulatory compliance as it collects the configuration and the parameters and feeds the Anonymizer with these data.

4.14.2 Pilot 12 real case solution

When Pilot #12 will be implemented in real life there won't be significant changes in the whole process and the solutions for regulatory compliance described in previous paragraph should adequate. Although for pilot purposes synthetic datasets are used to train the models, real data are also acquired in the process and they are handled taking into consideration regulatory compliance. Since this pilot collects data about physical activity, habits and mood these are considered as sensitive personal data and is mandatory to have full consent for the process of these datasets. During the pilot this problem is solved with the consent form signed electronically by participants. In real life this will be handled by the end user of the services which is the insurance company, asking their clients to give their consent for the use and process of their personal data.

In the production environment, the final service for the insurance company will adjust health insurance premiums according to the outcomes of the collected data from IoT devices and questionnaires is omitted. In real life this will be in place to satisfy the need of the end users, which is to improve risk analysis and continuously monitoring insureds' lifestyle ta adjust their premiums according to the updated risk. In this sense the insurance companies have to verify that the consent they ask from their insureds cover this part of the process as well.

4.14.3 Pilot 12 assessment solution

According to in D7.21 [15] the compliance value from Stakeholders evaluation is 4.0.

4.14.4 Pilot 12 recommendations

In Pilot #12 there aren't any ignored gaps in terms of regulations. As it is already mentioned the only regulation applied is GDPR and the solution is the full consent of the participant in the study if we are referring to pilot's life circle, or of insurance companies' clients if we are referring to real life implementation. The best way to assure the compliance to regulation in this pilot is to create all the necessary procedures to terminate instantly the services provided and erase all collected data for those who will cease their consent at any time, creating a gap in regulatory compliance.

4.15 Pilot 13

As stated in D3.16 [2], "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 Pilot #13 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 Pilot #13 Platform, this will receive the name of the 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 the services provided by the platform 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 *datastore Targets table* (Infrastructure).” [5]

4.15.1 Pilot 13 solution

As stated in D3.16 [2] Pilot 13 does not use personal data, there is no need to find a solution for regulatory compliance in this regard.

For user access to the platform, the solution is to apply IAM authorization and access control with Role management through using standard access security measures delivered by Amazon Web Services.

The high-level architecture for this pilot, plus where the security access and data interchange can be found, are shown in Figure 10.

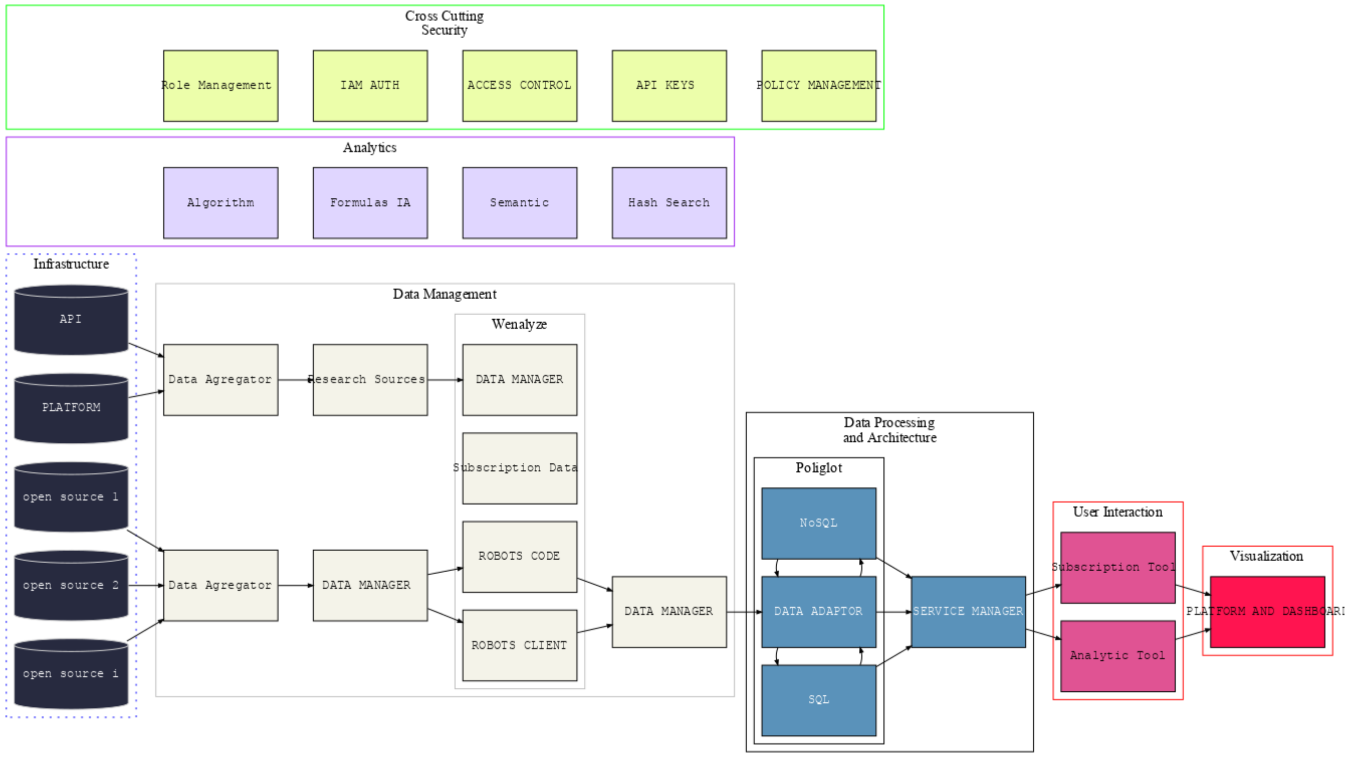


Figure 12: INFINITECH Pilot #13 RA

Pilot #13's connectivity with clients and the data to be analysed will be through an API that will connect client systems to Pilot #13's analytics platform.

4.15.2 Pilot 13 real case solution

Pilot #13 develops an architecture to capture data from public and open sources in large quantities and in real time. These sources range from official records, company websites, social media, company geo-positioning and other socio-economic and financial variables. Artificial intelligence techniques will be applied to this data to obtain a profile of the risks of SMEs and a system for updating and recommending insurance products for optimal coverage.

As mentioned above no changes have been made regarding developing of the Pilot#13 or the system or the data that is used to get the analysis therefore no changes regarding compliance or using new tools to reach compliance. We want to highlight that the pilot does not apply personal data

The data in the pilot flows as follows. First, a data set of companies including SMEs from different European countries is chosen, as one of the aims of the pilot #13 is to be an international pilot. The second step is to analyse the level of digital footprint among the companies in the different countries in order to know if the level of presence is homogeneous. Finally the sources are selected. The data sources are divided into two types, global and local, the global ones are cross-border and therefore affect equally the collection of information from all countries, and the local ones are where adjustments must be made to the search robots. The aim is to minimize the use of local sources. The process of data collection and storage takes place in two different places. The data collection is done through the Wenalyze robot module and once collected it is stored in the Nova testbed on the dedicated servers for the INFINITECH consortium. For the storage structure, the NoSQL architecture provided by our technology partner LeanXcale will be used.

The system does not use any personal data, only data of legal entities are analysed and the results do not affect individuals in any case it will be the insurer who will be delegated the control of privacy. The AI system has integrated an explainable artificial intelligence (XAI) module in order to explain the classification. The module provides the most significant features to perform the prediction.

Bias is avoided in the model in several ways. First, by using balanced datasets for training, secondly the models are always based on supervised machine learning, so that the results obtained can be explained and analysed.

4.15.3 Pilot 13 assessment solution

According to in D7.21 [15] the compliance value from Stakeholders evaluation is 5.0.

"The services contained and developed in the Pilot comply with applicable regulations in the banking and insurance sectors. In fact, demos and projects have already been carried out and have had to be approved beforehand by the compliance departments of the insurance companies and other stakeholders" [15].

4.15.4 Pilot 13 recommendations

In conclusion, we must say that the pilot respects the legislation in force in terms of personal data protection, ethics and bias in the algorithms and in the models on which these algorithms are based.

The data collected are only from legal entities, companies, and are therefore not regulated by GDPR Regulation (EU) 2016/679. Nevertheless, in the design, definition and development of the Pilot#13, all the provisions of this regulation have been taken into account with regard to the traceability of the processes, data flows and processes used.

4.16 Pilot 14

As stated in D3.15 [1], the objective of Pilot #14 "Big Data and IoT for the Agricultural Insurance Industry" is to deliver a commercial service module INFINITECH Agri-Insurance toolbox 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.

- Earth Observation data products will act as a complementary source to the information used by insurance companies to design their products and assess the risk of natural disasters.
- The Weather Intelligence Engine is used to verify the occurrence of catastrophic weather events and to predict future perils that could threaten the portfolio of an agricultural insurance company.
- These services are combined with a state-of-the-art user interface which also provides simplified portfolio management and business intelligence tools.

In more details, the pilot will provide Insurance companies with a robust and cost-effective toolbox of functions and services allowing them to alleviate the effect of weather uncertainty when estimating risk of AgI products, reduce the number of on-site visits for claim verification, reduce operational and administrative costs for monitoring of insured indexes and contract handling, and design more accurate and personalized contracts. [6].

Security and privacy issues and requirements

Insurance companies have to handle personal data from their potential, clients, i.e. the parcels' geolocation (polygon), in case this pilot is used to estimate the risk of AgI products or from their insureds in case this pilot is used to estimate damage after a claim. The regulation applied in this case is GDPR. However, in the pilot to be deployed by the handler of these services no personal data are required, instead each field can be accompanied by a specific id (e.g. 1, 2, 3, etc.) connected to the location where the claim was made or the wider region where risk estimation is required.

4.16.1 Pilot 14 solution

As stated in the 5.16, the insurance company which is the end-user of the pilot must handle with GDPR regulation and personal data of their clients, i.e. the parcels' geolocation (polygon), so the procedure to have a signed consent from the potential clients is entirely insurance companies' responsibility and is mandatory regardless of the pilot. In case of a claim and since no personal data are revealed in third parties no other specific consent is required.

The high-level architecture of this pilot and all its components are shown in the Figure 13:

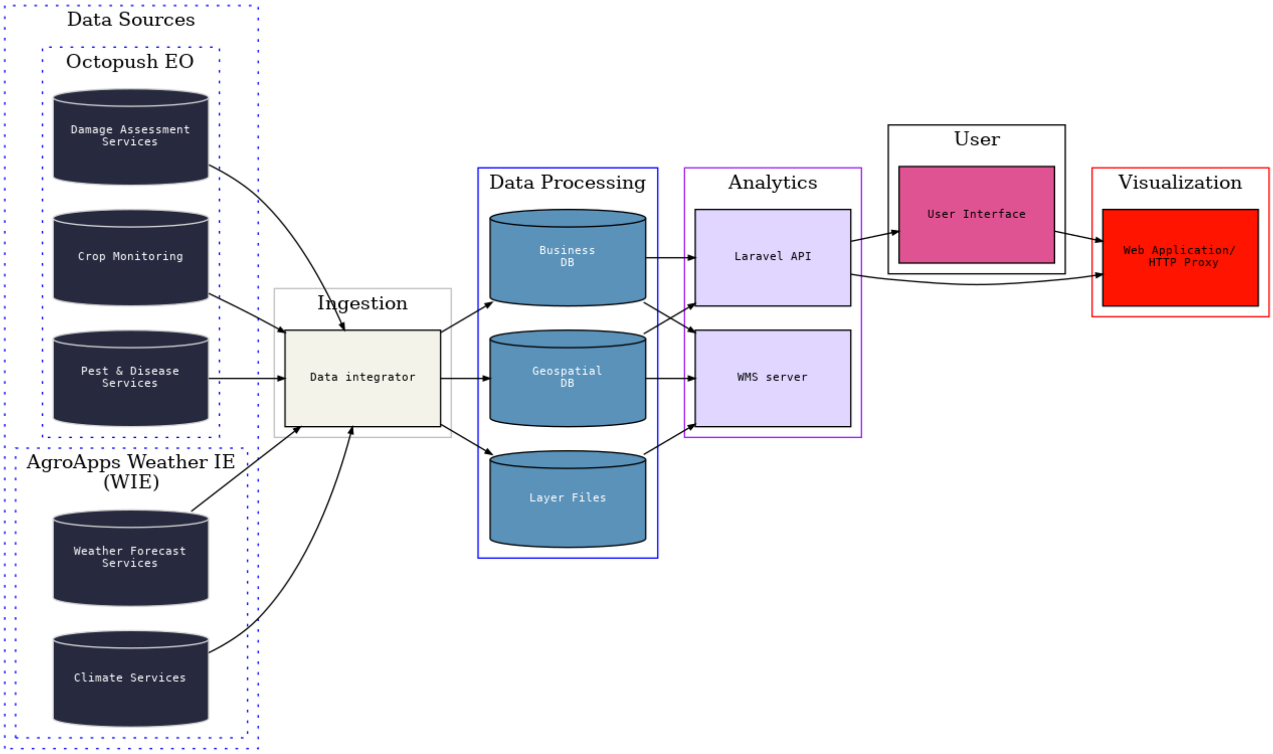


Figure 13: INFINITECH Pilot #14 RA

The pilot's components are [2]:

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

User interface (Interface in RA): The front-end user interface is the gateway responsible to present all the system data through user-friendly controls and web mapping interfaces

With regards to the personal data, all the information derived from the contracts is stored on the insurance companies' servers, the needed information in order for the services to be delivered is the geolocation. Therefore, the platform stores only the geolocation of the parcels and anonymise the rest information by giving an id in each parcel. This information is stored in the data processing and specifically, in the geospatial DB

4.16.2 Pilot 14 real case solution

The services will be served through a Web API via HTTPS, which supports Transport Layer Security (TLS) encryption. This ensures that the data-transfer is end-to-end encrypted. On top of that it authorizes the HTTPS requests by validating the API token, a mandatory parameter on all the HTTPS calls, against the Authorization Server registry. This process verifies that the identity making the request is authorized to receive the particular set of data.

In terms of data retention and destruction, data will be deleted or fully anonymised as soon as the relevant purpose as stated in the DoA is fulfilled. Regarding data processing, the collected data will be immediately pseudonymised and aggregated, and the original data will not be stored whatsoever.

Furthermore, a "Personal Data Protection Policy" and "Terms and Conditions" documents will be prepared, in order to inform the users of the purposes of data collection.

4.16.3 Pilot 14 assessment solution

According to in D7.21 [15] the compliance value from Stakeholders evaluation is 3.0.

"Strict Compliance: current legislation schemes oblige, in some countries, in field visits of personnel to perform the damage evaluation (confirm the extent and severity of damage). Nevertheless, future legislation updates may allow the completion of the process cycle 100% remotely. Even in those cases the proposed solution can bring added value, in assisting companies to prioritize in-field inspections and perform partial payments/compensations. Its use can be seen as a disruption of existing internal underwriting procedures. Hence, the service needs to be very flexible and easily adaptable to the operative IT-system of the insurance company and should be accompanied by staff training efforts."[15]

4.16.4 Pilot 14 recommendations

The specific pilot case will be implemented under the fundamental rights of the European Union (posed by the Lisbon Treaty) focusing on the right to the integrity of a person, protection of personal data and family issues, as well as rights in the freedom of scientific research. Moreover, the data will be handled by the following legislative documents:

1. Regulation (EU) 2016/679 - General Data Protection Regulation (GDPR)
2. Directive (EU) 2016/680 on the protection of natural persons regarding processing of personal data connected with criminal offences or the execution of criminal penalties, and on the free movement of such data
3. Regulation (EU) 2018/1725 on the protection of natural persons regarding the processing of personal data by the EU institutions, bodies, offices and agencies
4. Guidelines on Transparency under Regulation 2016/679 (wp260rev.01)
5. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01)
6. Guidelines on Personal data breach notification under Regulation 2016/679 (wp250rev.01)
7. Guidelines on Consent under Regulation 2016/679 (wp259rev.01)

8. Guidelines on the application and setting of administrative fines (wp253).
9. Guidelines on the Lead Supervisory Authority (wp244rev.01)
10. Guidelines on Data Protection Officers ('DPOs') (wp243rev.01)
11. Guidelines on the right to "data portability" (wp242rev.01)
12. Guidelines on Data Protection Impact Assessment (DPIA) (wp248rev.01)
13. The data will be collected for internal use in the project, and not intended for long-term preservation. No personal information will be kept after the end of the project. Furthermore, the respective partners handling personal data pay special attention to security and respect the privacy and confidentiality of the users' personal data by fully complying with the applicable national, European, and international framework, including the European Union's GDPR 2016/679.

14. Even though the participation is voluntary, informed consent will be sought from each individual user before his or her data is even stored. All data subjects are adults and therefore no issues on handling data of children are raised.

15. The best way to access the compliance in the pilot is to provide the insurance companies on what type of data can be considered personal (i.e. geolocation data in the case of open LPIS) and assure them that these data will be handled during the processing and development of the services anonymized. Nevertheless, insurance companies are obliged to follow GDPR principles and store/ handle/ share their clients' data under a secure way. Owners of personal data will be able to withdraw their consent for processing their personal data in accordance to the guidance of the GDPR principles of as depicted in their consent document under contract.

4.17 Pilot 15

ABI Lab, together with five (5) banks part of the AI HUB, a centre of excellence on AI chaired by ABI Lab, studied an application to design and to process management tasks in banks and tested the method on documents from the Italian banking community.

The solution proposed, DEcoDE - Documents Enhancement & COncept Detector), is capable of reading, analysing, filtering and organising banks' internal documents to support the development of an innovative taxonomy. The resulting module represents a solution that will allow the screening of extensive documentation in real time.

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

The measured impact of the proposed approach to process-related metadata creation has confirmed its applicability.

4.17.1 Pilot 15 solution

As stated in D3.16 [2], since the service assessment application analyses a subset of process operating documents for classification purposes, without ever accessing the data about the customers, GDPR is not applicable and there are no privacy issues.

4.17.2 Pilot 15 real case solution

In developing the experimentation, it was paramount to undertake an accurate activity of data selection, identifying only data that didn't involve personal or sensitive information of the banks participating in the Pilot.

The data selection process was defined and subscribed by all the participants to the initiative.

The common policy that regulates the data governance includes also:

a) restriction of the scope (usage of information only within the context of the experimentation);

b) limit the dissemination and access to the content of Confidential Information exclusively to the staff directly involved in the development activities

4.17.3 Pilot 15 assessment solution

According to in D7.21 [15] the compliance value from Stakeholders evaluation is 3.0.

In designing the solution architecture, Data Protection Requirements represents a core aspect.

In particular, a model was designed that includes:

- The infrastructure protection of the back-ends limiting risks of potential data leakage
- Access to machines protected by a firewall
- The documents and related metadata are stored on premises on protected servers
- Documents and data uploaded through Secure File Transfer Protocol (SFTP) with dedicated accounts

4.17.4 Pilot 15 recommendations

Due to the nature of the Pilot #15, where there is no sensible data the data governance policy signed by all the participants takes into consideration and fulfills the compliance regulation.

4.18 Pilot 16

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

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

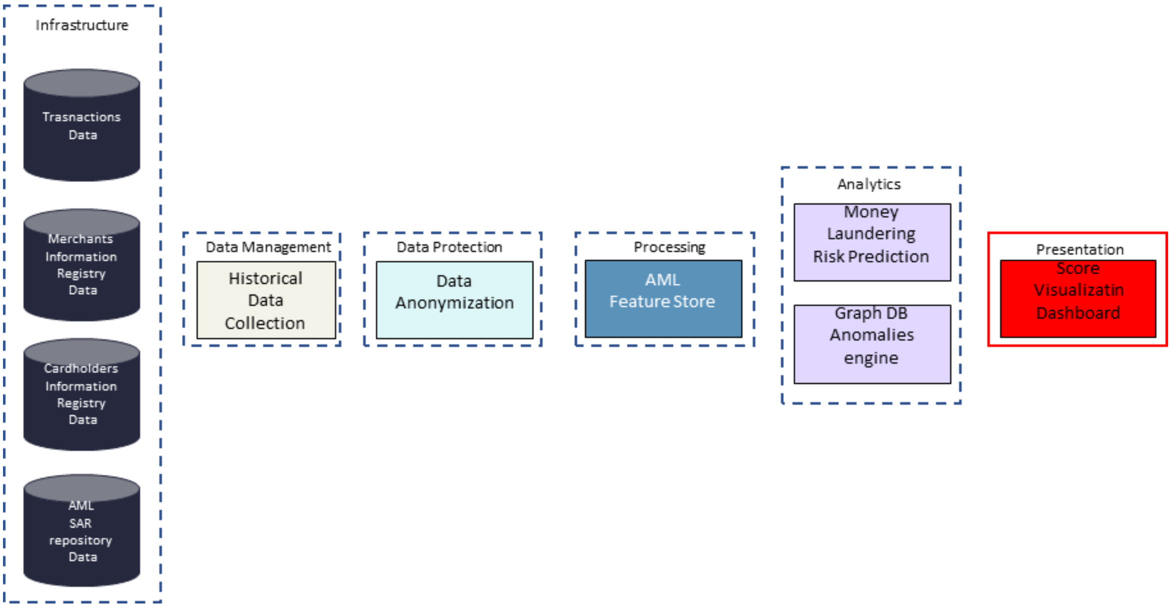
The adoption of pilot platform will improve quality and efficiency of AML users work and, at the same time, will concur in reducing risk of unmatched scenarios related to money laundering events.

4.18.1 Pilot 16 solution

All data related to subjects (both legal and individual persons) that will be used during the pilot are anonymized; it is not possible to identify the individuals involved in the transaction and that are analyzed by anomaly detection algorithm. This allows us to be compliant with GDPR and other regulatory requirements, like PCI and other security protocols.

The figure below, represents Pilot Reference Architecture.

In this particular Pilot, all Nexi customers and transactions data sources are processed through Data Anonymization stack. Due to this, all Machine learning and graph-based algorithms applied in the Analytics step are processed without the possibllity to know identity of the involved subjects.


Infrastructure
Trasnactions Data
Merchants Information Registry Data
Cardholders Information Registry Data
AML SAR repository Data
Data Management
Historical Data Collection
Data Protection
Data Anonymization
Processing
AML Feature Store
Analytics
Money Laundering Risk Prediction
Graph DB Anomalies engine
Presentation
Score Visualizatin Dashboard

The solution requirement is that the user, in pilot case Nexi Payments, would apply algorithm processing only to anonymized data. After that algorithm returns anomalous cases, only the authorized users (f.i. Anti Money Laundering analysts) can detokenize the anonymized data to know the identity of subject involved in the anomaly, using appropriate and compliant tools (out of scope from the Pilot solution).

For instance: once the pilot 16 returns an anonymized ID cardholder has performed anomalous self-financing transactions, the AML analyst / operator can detokenize the ID, and notify to Regulatory Financial Units the anomalous case (transaction operations) + all subject required personal information.

4.18.2 Pilot 16 real case solution

In a real case scenario, the regulatory requirements are guaranteed as in the case of synthetic data use. This is because even during the pilot development only anonymized data are being used.

To learn from data, our AI algorithm doesn't need cleartext data, but only transactional information: for instance, a list of numerical features used by machine learning algorithms is: charge amount for cardholder, number of transactions received by a merchant, percentage of transactions in gambling merchants, average ticket processed by a merchant POS system.

Only Anti Money Laundering analysts needs to know who has performed an anomalous transaction.

4.18.3 Pilot 16 assessment solution

According to in D7.21 [15] the compliance value from Stakeholders evaluation is 4.0.

The stakeholders consider the data management constraints to be respected in almost every case. Regarding the internal procedures the solution is coherent with the actual regulations.

4.18.4 Pilot 16 recommendations

The only needed requirement is that data that are going to be used by Pilot 16 anomaly detection tools, must be in an anonymized format to be compliant with GDPR regulation. Once the pilot solution will be released into the INFINITECH Marketplace, it won't provide an anonymization processing tool. It will ownership of the user to anonymize data before using anomaly detection tools deployed by the solution.

5 General INFINITECH Regulatory Compliance Tool

As stated in D3.16 [2], regulatory compliance issues may arise with any application, service or component that aims to provide technological solutions specially when it combines data subjects' expectation requirements and needs with the objectives of data controllers and service providers. It is critical to combine correctly the technology and the data in order to maintain compliance with the regulations which would constitute the regulatory requirements of the solution ensuring the protection rights of data subjects and of data controllers obligations. Every technological solution is responsible for implementing controls that ensure the regulatory compliance provides technical solutions that match different and adequate levels of privacy, as well as that consider data subjects' preferences and business objectives.

In this chapter, a General Regulatory Compliance Tool is proposed and described. This tool helps to solve privacy and/or security issues by using the DPO (Data Protection Orchestrator) tool provided by Atos. This tool is able to interact with different Protection Enhancing Technologies or Services that provide security or privacy by following a business process that calls to pre-defined tools. By using the DPO, the regulatory compliance tool is capable of preparing and executing privacy, security and data protection processes which ensures these aspects by design and by default.

The general philosophy followed in INFINITECH project to comply with the regulations is to use preferably solutions that are already being used by the pilots and provide new regulatory compliance tools only for those functionalities that are not already covered by their existing tools. In addition, the INFINITECH General Regulatory Compliance tool is able to provide solutions to fill the gap between the INFINITECH pilot cases and the real cases.

This chapter will firstly describe the Data Protection Orchestrator, including interfaces and technical designs, the solution provided of a General INFINITECH Regulatory Tool and finally the integration with Anonymization tool from Gradiant, a real privacy enhancing technology that participates in INFINITECH

5.1 Data Protection Orchestrator (DPO)

This section introduces the Data Protection Orchestrator (DPO). This component is responsible for coordinating the invocation of components that implement privacy, security or data protection techniques as well as other external services in order to provide a suitable privacy, Security and data protection level specified by a secure service provider compliance to regulations. DPO has been created in Witdom European ICT Project, and the general information on Witdom comes from [8].

5.1.1 DPO Description

The Data Protection Orchestrator coordinates several privacy, security and data protection components and services to ensure that the successive use of the data that have been protected can be processed or stored preserving their privacy and Security. It also allows the removal of the protection of the results (if required) before delivering them to the end user.

The Data Protection Orchestrator uses processes in the Business Process Model and Notation 2 (BPMN2) format. BPMN is a XML-based standard for business process modelling that provides graphical representation for specifying business processes, similar to UML diagrams.

The business process guides and establishes all the steps in the adequate order that must happen in order to ensure the security privacy and data protection.

The use of Data Protection Orchestrator provides the following benefits:

- Helps secure service developers and protection component's providers to ease the provision of the process for protection configurations

- Allow the combination of individual privacy, security or data protection components creation complex protection processes.
- Provides the needed business logic that allows to ensure that the regulations are fulfilled.
- BPMN diagrams can be visually showed, providing a clear view of the protection process.

The business processes will include workflows similar to the ones presented in Figure 14.

In general, it is common to have a first step that would trigger the invocation of a transformation service, which would transform data in domain-specific standard formats (e.g., CSV files) to a table format suitable to store it in a regular SQL database. The protection configuration will choose the preferred available algorithm and will invoke the components with parameters regarding the location of the data, where it must be output and metadata characterizing the data input.

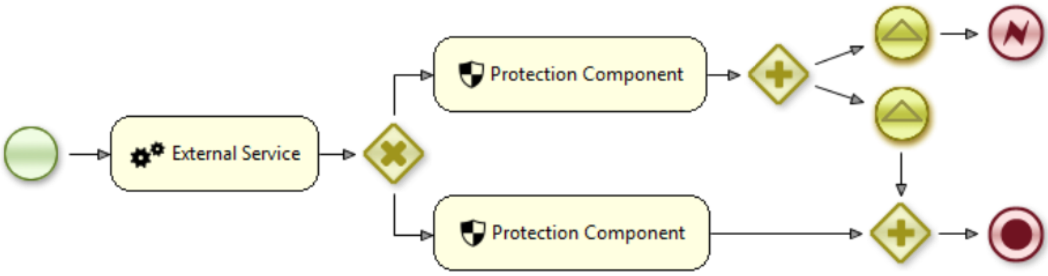


Figure 14: Example of business flow

The protection orchestrator will require interfaces with every component that will participate in the business process and will be provided by developing domain-specific

5.1.2 DPO Architecture Technical design

The Data Protection Orchestrator can receive requests from secured and reliable components

DPO can receive three types of requests:

- **Protection configuration management requests**: aimed to deploy protection configurations and manage them.
- **Protection configuration execution requests**: they are triggered by the secure components that calls the DPO. In this case, the DPO create a process to address the required protection configurations.
- **Protection configuration events**: the engine receives these events that can come from other components that are participating in the business process (e.g. wait for a review from a privacy expert, wait for data subject's consent or for an asynchronous response from a protection component)

The engine interacts with the protection components and other services through domain-specific tasks, which are basically Java classes following jBPMN specific interfaces that solve the communication particularities of the components or services that the DPO needs to invoke. Figure 15 depicts the DPO subcomponents and their relations.

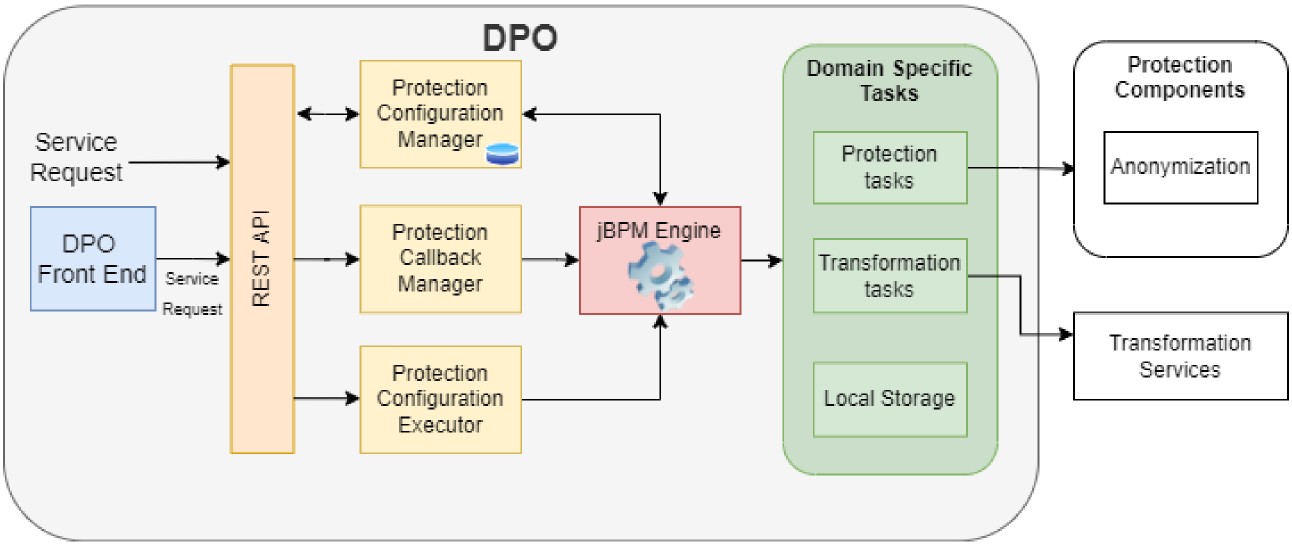


Figure 15: DPO architecture

The protection orchestrator has the following main subcomponents:

- The **Front End** to select and execute a business process.
- The **REST API** to receive each request regarding managing and executing business processes in order to protect the data.
- The **Protection Configuration Manager** receives all the requests coming from the REST API regarding the management of the protection configurations. It uses storage capabilities (in files) to store the configurations allowing the jBPM engine to take them. It will use database or filename to keep track of the deployed configurations allowing their invocations.
- The **Protection Callback Manager** accepts external events that will be inputs for the engine such as inputs from a privacy expert that would interact approving some tasks.
- The Protection Configuration Executor processes requests to execute the business processes. It will be in charge of choosing the suitable configuration and execute the new protection process in the engine.
- The **local storage** can be used within the business processes to store information regarding the calling application or the user which launched the business process and it can be utilized in other calls.
- The **Domain Specific Tasks** constitutes the interface between the engine and other Security or privacy components such as anonymization. They follow the interfaces of jBPM and implements the calls to these components

5.1.3 DPO Interfaces

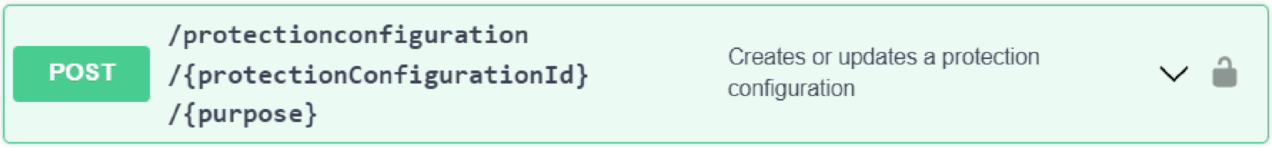
DPO interfaces can be divided into the input interfaces to call the DPO and the communication with other components required to apply a security or privacy measure.

5.1.3.1 DPO Input interfaces

The DPO will only accept HTTPS requests from a component which is configured in it. Moreover, a Front End DPO has been prepared to access directly the DPO.

5.1.3.1.1 DPO API

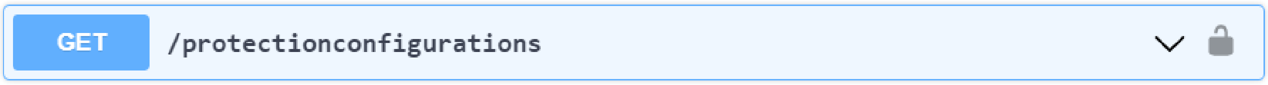
The DPO provides a REST API to be accessed by an external component configured in it. The API supports the following operations:



POST /protectionconfiguration/{protectionConfigurationId}/{purpose}

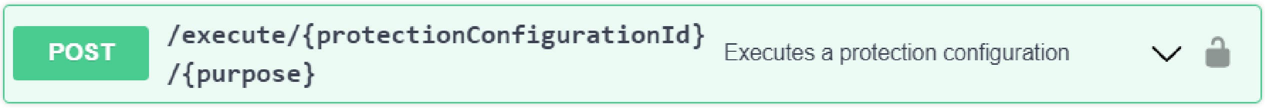
Creates or updates a protection configuration

This operation enables the uploading of a business process The business process will be defined using BPMN2 in XML notation. The business process can be generated with any BPMN2 editor such as such as BPMN.IO [https://demo.bpmn.io/] or Eclipse editor (jBPM Eclipse Plugin should be previously installed). This operation is used by the DPO Front End to upload a new BPMN business process definition.



GET /protectionconfigurations

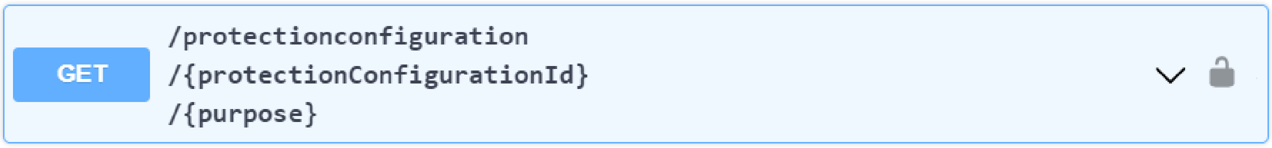
This operation provides the complete set of business processes uploaded in the DPO



POST /execute/{protectionConfigurationId}/{purpose}

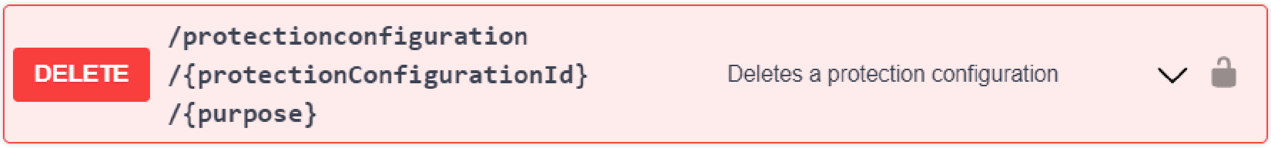
Executes a protection configuration

This operation executes a protection configuration that has been previously uploaded in the DPO. This method is used by the DPO Front End to execute the business processes.



GET /protectionconfiguration/{protectionConfigurationId}/{purpose}

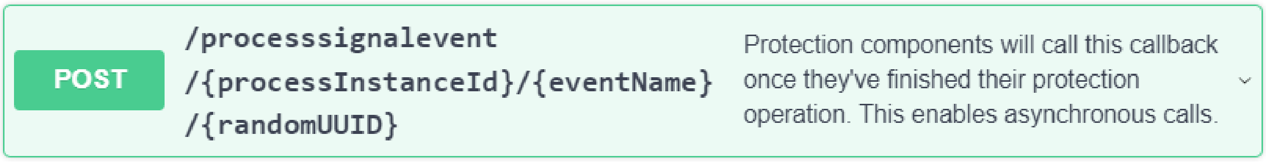
This operation provides the stored business process definition



DELETE /protectionconfiguration/{protectionConfigurationId}/{purpose}

Deletes a protection configuration

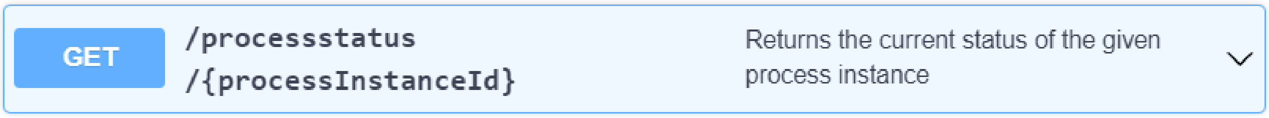
This operation deletes the business process that was previously uploaded



POST /processsignalevent/{processInstanceId}/{eventName}/{randomUUID}

Protection components will call this callback once they've finished their protection operation. This enables asynchronous calls.

This operation can be used in asynchronous calls. An asynchronous call has a first call to a component, and after that the DPO needs to know when the results are available once the call has finished. This can be performed by a call started by the component to the DPO, and this operation performs a callback mechanism.

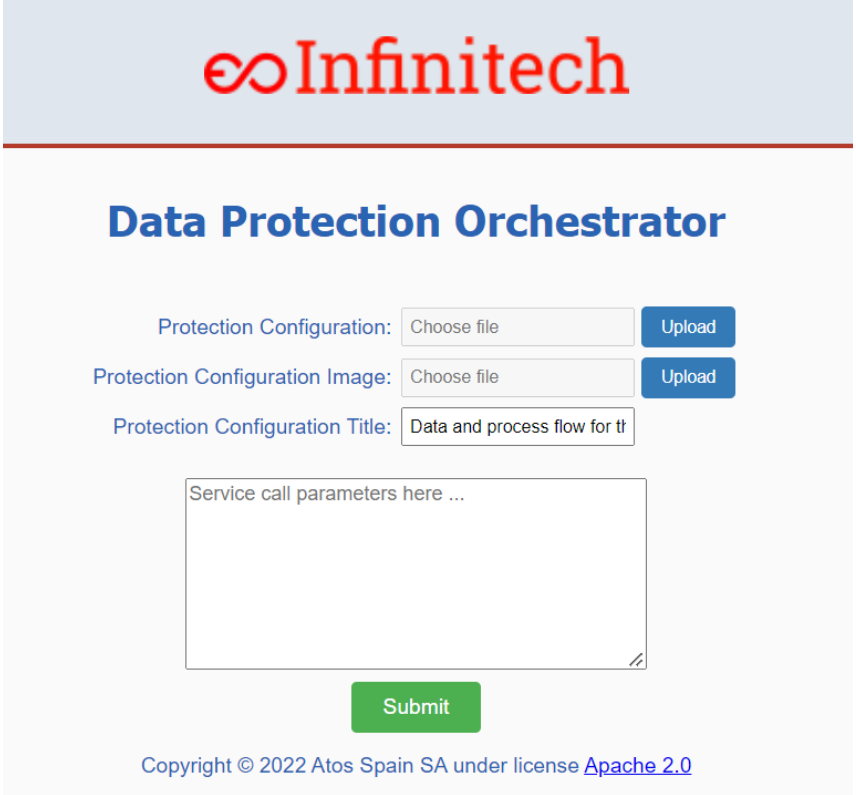


GET /processstatus /{processInstanceId} Returns the current status of the given process instance

This operation provides the current status of the process in order to know if it has finished

5.1.3.1.2 DPO Front End

The Front End DPO provides a web interface showed in figure 14 that can be accessed through https:// https://rancherproxy.vps.uninova.pt:31087/:8085/po-server/demo/index:



Protection Configuration: Choose file Upload

Protection Configuration Image: Choose file Upload

Protection Configuration Title: Data and process flow for tl

Service call parameters here ...

Copyright © 2022 Atos Spain SA under license Apache 2.0

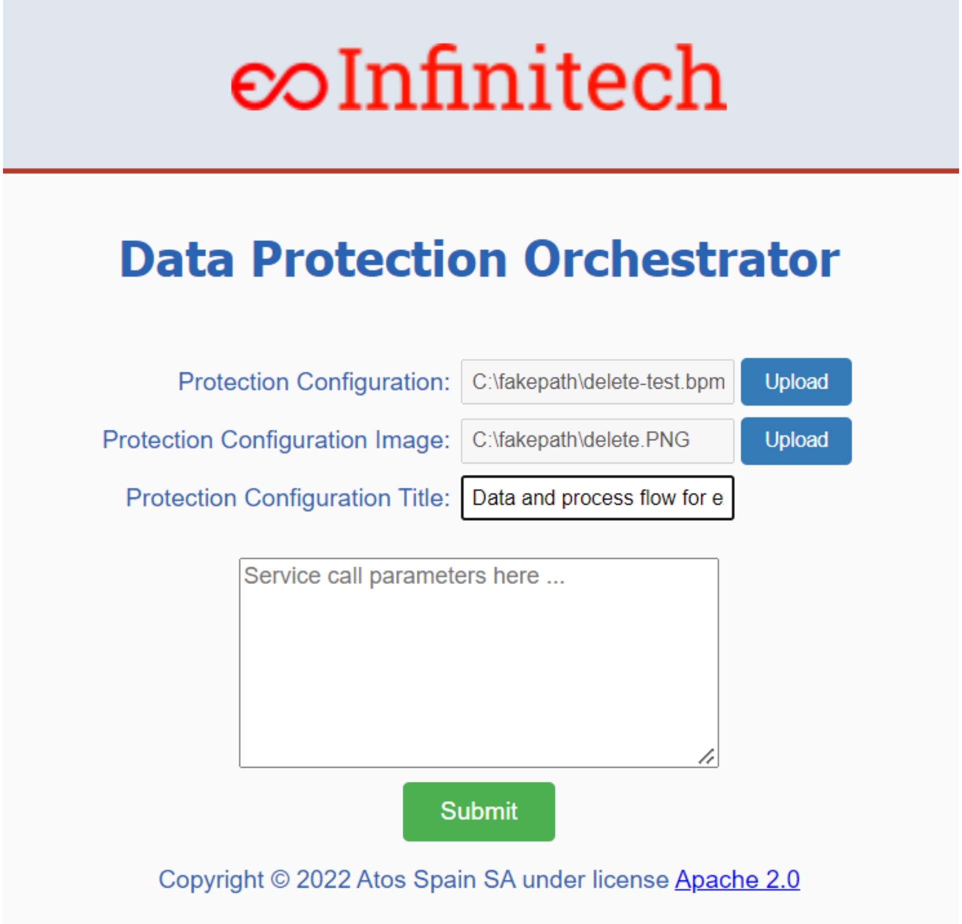
Figure 16: DPO Front End interface

This interface allows to charge and execute a concrete business process through uploading a prepared BPMN file. The following parameters should be filled in:

- **Protection Configuration**: this field must be filled in with a BPMN file prepared previously. This BPMN file defines the business flow to be executed by the DPO. The BPMN files are prepared by a privacy expert using a BPMN flow editor such as https://demo.bpmn.io/ or Eclipse editor (jBPM Eclipse Plugin should be previously installed)
- **Protection Configuration image**: it's the image file of the business process that was charged in the previous field. This field is for presentation purpose.

- **Protection configuration title**: it's the title of the business flow.
- **Service call parameters**: Input parameters needed to execute concrete business flows that have to be filled at execution time.

Once the parameters are filled in (Figure 17), and after clicking the submit button, the DPO starts processing the operation by firstly uploading the business flow (using the POST operation /protectionconfiguration/{protectionConfigurationId}/{purpose} and finally executing the business flow (using the POST operation /execute)



Data Protection Orchestrator

Protection Configuration: C:\fakepath\delete-test.bpm

Protection Configuration Image: C:\fakepath\delete.PNG

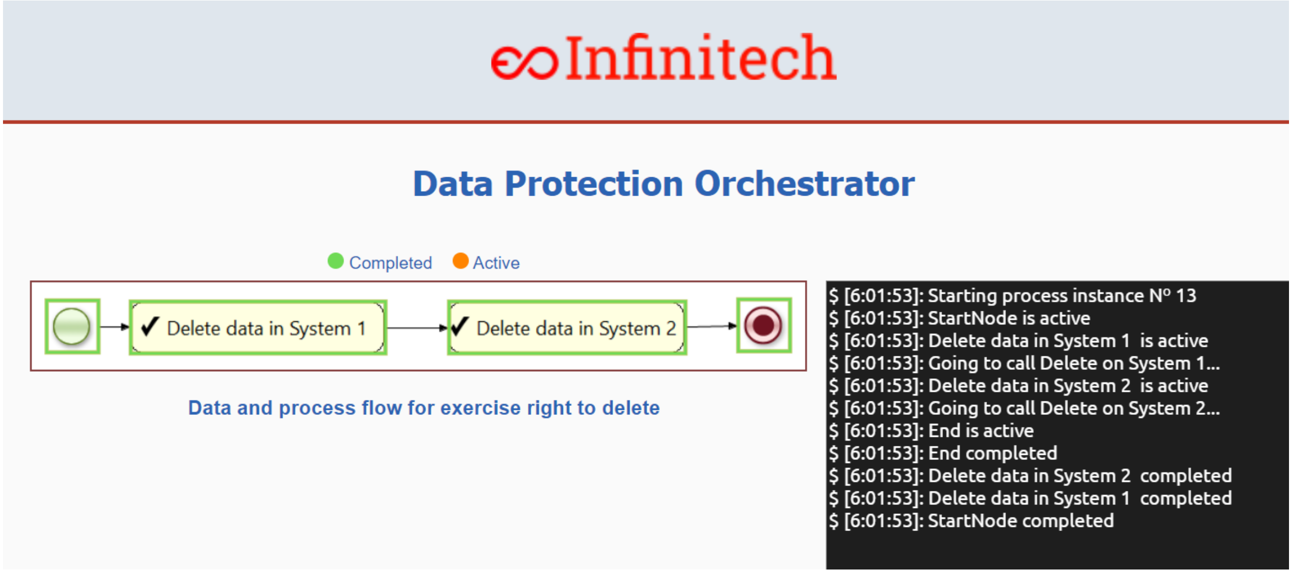
Protection Configuration Title: Data and process flow for e

Service call parameters here ...

Copyright © 2022 Atos Spain SA under license Apache 2.0

Figure 17: DPO FE – parameters filled in

In this example, it has been uploaded a business flow that aims to perform the right to delete of a customer, in this case it is required to delete data in two systems. The flow execution can be showed in the interface, the steps are highlighted first in red and finally in green when they are finished. The web interface shows then the execution completed in the diagram picture and also in a short text box with the main flow steps(Figure 18):


Infinitech
Data Protection Orchestrator
Completed
Active
✓ Delete data in System 1
✓ Delete data in System 2
Data and process flow for exercise right to delete

```
$ [6:01:53]: Starting process instance Nº 13
$ [6:01:53]: StartNode is active
$ [6:01:53]: Delete data in System 1  is active
$ [6:01:53]: Going to call Delete on System 1...
$ [6:01:53]: Delete data in System 2  is active
$ [6:01:53]: Going to call Delete on System 2...
$ [6:01:53]: End is active
$ [6:01:53]: End completed
$ [6:01:53]: Delete data in System 2  completed
$ [6:01:53]: Delete data in System 1  completed
$ [6:01:53]: StartNode completed
```

Figure 18: DPO interface – execution result

5.1.3.2 DPO Interfaces with security and privacy components

The DPO is able to orchestrate calls among several components. The DPO will require interfaces with each of the components that orchestrates. The way the DPO communicates with the components is through the domain-specific tasks that uses the mentioned interfaces. The domain-specific tasks provide integration with the component API.

The communication between the DPO and the components, requires that the components provide information regarding the location of the data to be protected. The DPO would organize data logic and create tables for leaving the data. In the case that a business process requires the creation of a table, there is a domain-specific task to perform this for which it needs to inform about the location of the table and structure (using a JSON object including the table name and the database location, detailing the columns name, default value, nullable and primary key).

The DPO will prepare the protection parameters which will be sent as a JSON object in the request to the Protection components and will inform them about the protection algorithms, the data origin and destination and the data structures.

The domain specific tasks that provide integration with the components are called work item handlers and they are implemented through a class that extends the interface WorkItemHandler from KIE API 6.3.0 Drools & jBPM provide a knowledge-centric API [27]. In this class is implemented the calls to the REST web services that provides the components. Then when a call to a new web service is needed, a new handler has to be created. All the handlers must be registered in a configuration file of the DPO, also, the handlers are linked with the boxes of the BPMN file in another configuration file, then they are ready to use by any business flow that could be needed.

In the work item handlers we can include code or logs or execution of some instructions. The handlers are executed when the BPMN process is being executed and reaches the box that links with the handler.

In the figure 18 business flow execution showed before, two boxes have been implemented to link two different work item handlers that enable the communication with the REST API of the security component that implements the services.

5.2 Definition of general INFINITECH Regulatory Compliance tool

As previously commented in the introduction of this chapter, the INFINITECH project provides an approach of a general Regulatory Compliance Tools that solves privacy and/or security issues by using the DPO. DPO is able to interact with different Protection Enhancing Technologies or Services that provide security or privacy by following a business process that calls tools.

The aim of the General INFINITECH Regulatory Compliance tool is to provide a mechanism that easily implements new regulatory compliance tools that calls privacy or security technologies.

Although the DPO is able to implement complex business flows, the approach that has been considered for this general definition is to prepare simple business flows that eases the integration with the technologies.

The technologies that can be called can be divided in two types, mainly depending on the time that takes to execute the process of the privacy or security technology. The first type would be a synchronous process in which there would be only one box that calls to the REST API of the technology. The second type would be an asynchronous implementation that would be for the most complex processes that require high computational load and therefore longer time.

The general business flow provides the following elements:

- A general synchronous business flow
- A general asynchronous business flow
- General Work item handlers that manage the calls to the security and privacy tools REST APIs

Figure 19 shows two business flows: the general synchronous and asynchronous business flow.

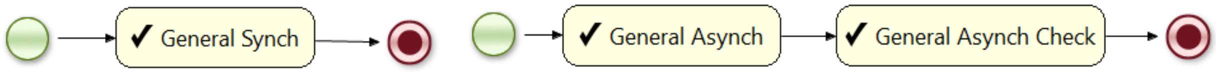


Figure 19: general synchronous and asynchronous regulatory tool

5.3 Integration with Anonymization

The anonymization component is developed as a service and provides different anonymization techniques and algorithms that can be applied to a dataset to protect privacy. The tool computes the different possible anonymization configurations over a dataset, and automatically determines which one better fits the user's privacy and utility goals.

The anonymization tool is intended to be used in two modes: **analysis** and **anonymization**.

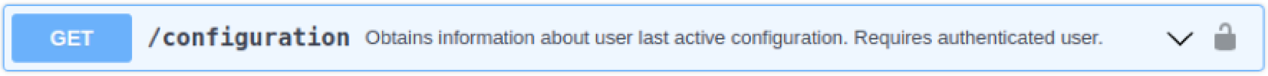
First, the **analysis mode** takes the set of operations and privacy/utility metrics that the user desires to apply to the different columns of the data and computes all the possible anonymization configurations. The tool executes the different anonymization operations and computes the privacy and utility metrics for each case. This allows the user to discover the set of anonymization operations that better fit its privacy and utility needs. This process is time consuming, and usually takes place over a subset of the final data. Secondly, the **anonymization mode** takes the selected anonymization configuration, and applies it to the final dataset, storing the results in a destination database or file, or in a data streaming queue.

The **analysis process requires human interpretation of the results** and will be executed by a human operator (for instance, a Machine Learning Engineer that wants to anonymize a dataset) in the Anonymization Tool interface. The human operator will define the different anonymization operations that the operator wants to apply to their dataset; the metrics to be computed, and the results, will be presented to him/her for interpretation and selection. Once the user selects the operation(s) that fit their privacy/utility needs, the Anonymization Tool stores **the selected configuration, so the DPO** can retrieve it using and API call for further use in the anonymization operation. The DPO and Anonymization component must agree on a common schema to i) exchange the configuration files and ii) map end users or components and configuration files.

The Data Protection Orchestrator will communicate with the Anonymization Tool through a REST API interface to orchestrate the anonymization operations from different tools and users**.** The API works in an asynchronous way: since the anonymization and analysis operations are very time consuming, when the API receives a valid petition, returns a **task identifier** that can be used to track the progress (see */progress* petition) of a particular anonymization task. The DPO can subscribe to the responses as an Event Stream or check the status of the anonymization task periodically. The asynchronous operation of the anonymization component allows the execution of multiple tasks in parallel, without blocking the execution of the DPO.

In this way, the DPO orchestrates the anonymization petitions from other components by making API calls to the anonymization component using pre-generated configuration files or firstly calling /configuration to obtain the last available anonymization configuration generated by the human operator. As explained above, this configuration defines the anonymization operations to be applied to the dataset, and the destination database to store the results.

Henceforth we define the API calls and format of the required input (large JSON schemas are provided as appendix links), together with the different responses and their format:


GET /configuration Obtains information about user last active configuration. Requires authenticated user.

Description	Obtains information about user’s last active configuration			
Parameters	Type	Name	Description	Schema
	application/json	Anonymization configuration	Configuration file containing the anonymization parameters, pre-loaded in the DPO by the Anonymization Tool	
	Authentication Token	Auth Token	Token for authentication against the REST service	
Responses	HTTP Code	Description	Schema	
	200	OK, Returns last active configuration	Appendix B: Anonymization configuration	

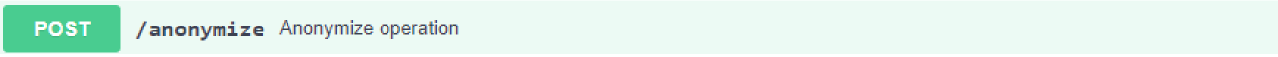
	400	Bad Request	{
			“status”: “Bad Request”
			“message”: “error message”
			}	

	401	Unauthorized	{ “status”: “Unauthorized” “message”: “error message”	

			}

	403	Forbidden	{
			“status”: “Forbidden”
			“message”: “error message”
			}

	404	Not Found	{
			“status”: “Not Found”
			“message”: “error message”
			}



POST /anonymize Anonymize operation

Description	Performs anonymization according to certain given parameters			
Parameters	Type	Name	Description	Schema
	application/json	Anonymization configuration	Configuration file containing the anonymization parameters, pre-loaded in the DPO by the Anonymization Tool	APPENDIX B: ANONYMIZATION CONFIGURATION
	Authentication Token	Auth Token	Token for authentication against the REST service	
Responses	HTTP Code	Description	Schema	

	200	OK	{
			“task_id”: “task id”
			}	

	400	Bad Request	{
			“status”: “Bad Request”
			“message”: “error message”
			}	

	401	Unauthorized	{
			“status”: “Unauthorized”
			“message”: “error message”
			}	

	403	Forbidden	{
			“status”: “Forbidden”
			“message”: “error message”
			}

	404	Not Found	{
			“status”: “Not Found”
			“message”: “error message”
			}

GET /progress/{task_id} Get the status of a task.

As explained above, the asynchronous API allows the execution of multiple parallel tasks. The */progress* endpoint receives a task identifier as parameter, and returns the current status of the task, namely:

- **Received**: Task was correctly received by the anonymization tool, but it did not start.
- **Started**: The task started its execution. The response includes information about the current progress of the task (for instance, N steps completed out of M).
- **Success**: The task was correctly finished. The response includes the result of the task (if it is an anonymization operation, returns the values of the computed privacy and utility metrics).
- **Failure**: The task execution failed. The response includes the reason for the failure

Description	Obtains current status of an anonymization or analysis task			
Parameters	**Type**	**Name**	**Description**	**Schema**
	String	Task ID		UUID
	Authentication Token	Auth Token	Token for authentication against the REST service	
Responses	**HTTP Code**	**Description**	**Schema**	

	200	OK	{
			"state": "RECEIVED",
			"info": "None"
			}
			
			{
			"state": "STARTED",
			"info": {
			"completed": "1",
			"total": "5"
			}
			
			{	

			"state": "SUCCESS",
			"info": {
			"result": {
			"working_points_info": {}
			}
			}
			
			{
			"state": "FAILURE",
			"info": "Exception"
			}
			
			{
			"state": "PENDING",
			"info": "None"
			}
			]

	400	Bad Request	{
			"status" : "Bad Request”,
			“message”: “error message”
			}

	401	Unauthorized	{
			"status" : "Unauthorized”,
			“message”: “error message”
			}

	403	Forbidden	{
			"status" : "Forbidden”,
			“message”: “error message”
			}

	404	Not Found	{
			"status" : "Not Found”,
			“message”: “error message”
			}

In order to implement the communication between DPO and Anonymization, it has been used the general asynchronous INFINITECH regulatory tool described in Section5.2 and customized for the Anonymization tool. It is depicted in Figure 20. This implementation would define the Anonymization Regulatory Tool.

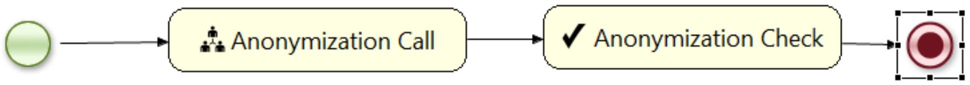
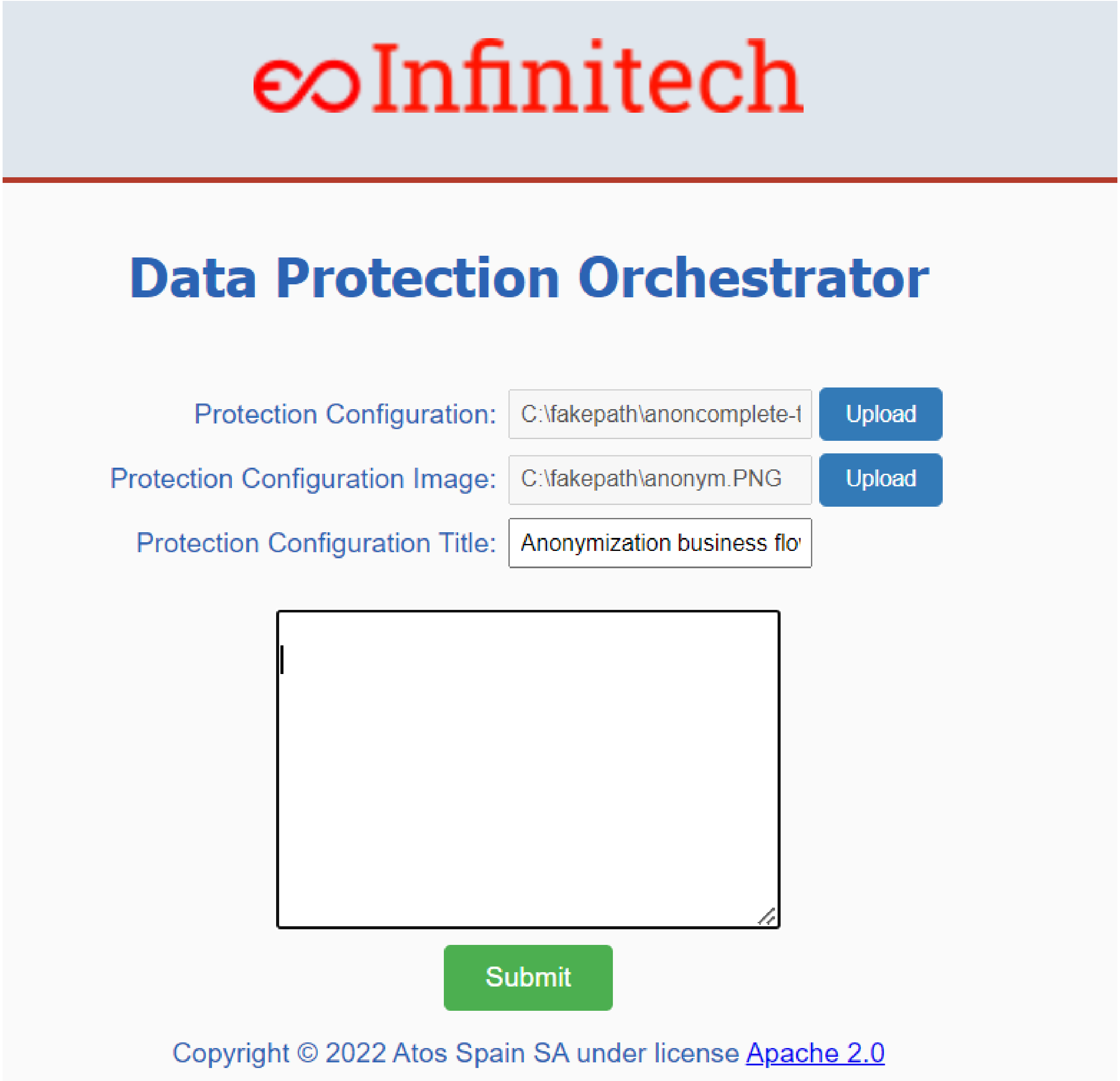


Figure 20: Anonymization Regulatory Tool

The first box “Anonymization Call” would use a new prepared work item handler that implements the call to the POST operation /anonymize. This call would start the anonymization process. In the call it will be sent the configuration file containing the anonymization parameters, retrieved by the DPO using the /configuration endpoint.

The second box “Anonymization Check” would use another work item handler implementing the call to the GET operation /progress that would track the status of the Anonymization process, allowing to monitor the process and get notified when it is finished and the anonymized data are ready to be used.

The execution of this process can be launched from the DPO Front End. First it is needed to fill in with the business flow that has been prepared (figure 21):


Data Protection Orchestrator

Protection Configuration: C:\fakepath\anoncomplete-t

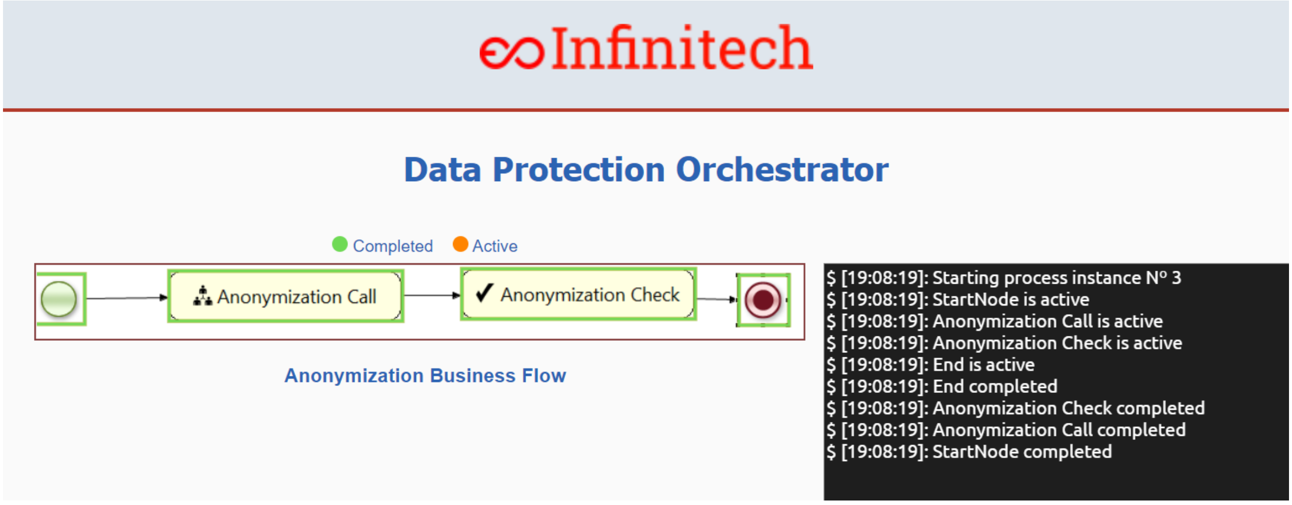
Protection Configuration Image: C:\fakepath\anonym.PNG

Protection Configuration Title: Anonymization business flo

Copyright © 2022 Atos Spain SA under license Apache 2.0

Figure 21: Front End DPO with the fields filled in

And after executing it, the interface shows that all the steps in the anonymization process have been successful (figure 22)



```
$ [19:08:19]: Starting process instance Nº 3
$ [19:08:19]: StartNode is active
$ [19:08:19]: Anonymization Call is active
$ [19:08:19]: Anonymization Check is active
$ [19:08:19]: End is active
$ [19:08:19]: End completed
$ [19:08:19]: Anonymization Check completed
$ [19:08:19]: Anonymization Call completed
$ [19:08:19]: StartNode completed
```

Figure 22: Business Flow Execution

This definition of General INFINITECH regulatory compliance tool is being used in Pilot #12.

6 Conclusions

The present deliverable D3.17 is devoted to assessing regulatory compliance in INFINITECH and the tools needed to ensure it. This deliverable documents an analysis that is fundamental to ensuring that all the pilots comply with relevant regulations, updating the results found in D3.16 (which was the previous edition of this work). It also extends D3.16 in its analysis of the regulations for every pilot and provides a review of all the technologies that the partners are bringing to INFINITECH, aiming to find possible technologies that could help to give solutions for regulatory compliance.

The analysis identified possible privacy and security issues for each pilot and the deliverable offers possible solutions. In some pilots, a key value is that they are providing solutions that directly ensure regulatory compliance and this deliverable describes the solutions adopted by them. The analysis of the pilots is completed by encompassing the solutions that should be applied in case they work in a production environment. It also provides the compliance assessment results coming from D7.21 [15]. Considering this analysis for every pilot, this deliverable suggests recommendations for compliance with the regulations in the most appropriate way.

To facilitate the provision of solutions for regulatory compliance, the deliverable describes the prototype implementation of a General INFINITECH Regulatory Compliance Tool, which is a general solution based on the DPO (Data Protection Orchestrator) from Atos that is capable of orchestrating technologies for preserving privacy, data protection and security. This tool facilitates the provision of possible solutions for new, modified pilots, or more complex solutions for real cases, helping future compliance with new or changed or freshly-identified regulations. The definition of this INFINITECH General Regulatory Compliance Tool as a prototype based on the DPO is provided in the deliverable, including its architecture, the interfaces and the integration with the Anonymization tool from Gradiant.

Table 8 – Conclusions (TASK Objectives with Deliverable achievements)

Objectives	Comment
Applications in the finance and insurance sectors have to comply with many and quite complex regulations. This holds for most Big Data and IoT applications, which tend to be data-intensive and to involve complex data processing across multiple systems and stakeholders.	The deliverable shows the assessment (coming from WP2 deliverables) and complete it to ensure that all the pilots comply with relevant regulations. It provides a review of all the technologies related to regulatory compliance that the partners are bringing to INFINITECH, aiming to find possible technologies that could help to give solutions for regulatory compliance. The analysis identified possible privacy and security issues for each pilot, offers possible solutions for the pilot case and production case (in some pilots, the proposed solutions directly ensure regulatory compliance) and maps the regulations with the technologies that provide solutions in line with INFINITECH-RA.

Table 9 – (map TASK KPI with Deliverable achievements)

KPI	Comment
Number of supported regulations >=4	The regulations GDPR, MIFID II, PSD 2, and AMLD4 were assessed and for the INFINITECH Pilots.

	Technologies for Security, Privacy and Data protection provided in INFINITECH project were mapped with the regulations and also matched to the pilots.
Policy Rules Orchestration Enforcement Framework >= 1	The most common Technological solutions provided in INFINITECH for regulatory compliance such as Anonymization, Pseudonymization, IAM or Consent Management have been mapped to solve issues related with regulations (considering GDPR, MIFID II, PSD 2, and AMLD4) that applies to all the INFINITECH pilots. Also, general rules and particular rules have been developed to orchestrate technologies such as Anonymization inside the General INFINITECH Regulatory Tool

Appendix A: Literature

[1] INFINITECH consortium, “INFINITECH D3.15 – Regulatory Compliance Tools – I”, 2020.

[2] INFINITECH consortium, “INFINITECH D3.16 – Regulatory Compliance Tools – II”, 2021.

[3] INFINITECH consortium, “INFINITECH D3.14 – Data Governance Framework and Tools – III”, 2022.

[4] INFINITECH consortium, “INFINITECH D2.8 – Security and Regulatory Compliance Specifications – II”, 2020.

[5] INFINITECH consortium, “INFINITECH D2.14 – Reference Architecture – II”, 2021.

[6] INFINITECH consortium, “INFINITECH D7.15 – Configurable and personalized insurance products – I, 2021.

[7] INFINITECH consortium, “INFINITECH D2.6 – Specifications of INFINITECH Technologies – II, 2020.

[8] Marcus Brandenburger, Eduarda Freire, “Witdom Project, D4.2 – Final specification of an end-to-end secure architecture”, August 2016. [Online]. Available: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ac7baf1a&appId=PPGMS [Accessed 10-July-2021].

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

[10] INFINITECH consortium, “INFINITECH D4.9 – Permissioned Blockchain for Finance and Insurance – III”, 2022.

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

[12] “Keycloak – Authorization Services Guide”, https://www.keycloak.org/docs/latest/authorization_services/#authorization-services, [Accessed: March 2022].

[13] INFINITECH consortium, “INFINITECH D9.15 – Business Models and Innovation Management - I, 2022.

[14] INFINITECH consortium, “INFINITECH D9.16 – Business Models and Innovation Management - II, will be delivered in December 2022.

[15] INFINITECH consortium, “INFINITECH D7.21 –Pilots’ Evaluation and Stakeholders’ Feedback - II , 2022.

[16] INFINITECH Risks, https://app.infinitech-h2020.eu/risks

[17] INFINITECH Actions, https://app.infinitech-h2020.eu/actions

[18] IBM – Security, Policy and Compliance, https://www.ibm.com/cloud/garage/architectures/securityArchitecture/security-policy-governance-risk-compliance [accessed March 2022]

[19] INFINITECH consortium, “INFINITECH 1.2 – Risk Management and Quality Plan, 2021.

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

[21] INFINITECH consortium, “INFINITECH D2.21 – Security and Regulatory Compliance Specifications - II”, 2020.

[22] FAFT, Updated Guidance for a Risk-Based Approach to Virtual Assets and Virtual Asset Service Providers, https://www.fatf-gafi.org/publications/fatfrecommendations/documents/guidance-rba-virtual-assets-2021.html [accessed March 2022]

[23] EU General Data Protection Regulation (EU-GDPR), https://www.privacy-regulation.eu/en/index.htm [accessed March 2022]

[24] Recommendations on outsourcing to cloud service providers, https://www.eba.europa.eu/sites/default/documents/files/documents/10180/2170121/5fa5cdde-3219-4e95-946d-0c0d05494362/Final%20draft%20Recommendations%20on%20Cloud%20Outsourcing%20%28EBA-Rec-2017-03%29.pdf?retry=1) [accessed March 2022]

[25] INFINITECH consortium, “INFINITECH D9.13 – Exploitation and Sustainability Plan - II, 2022.

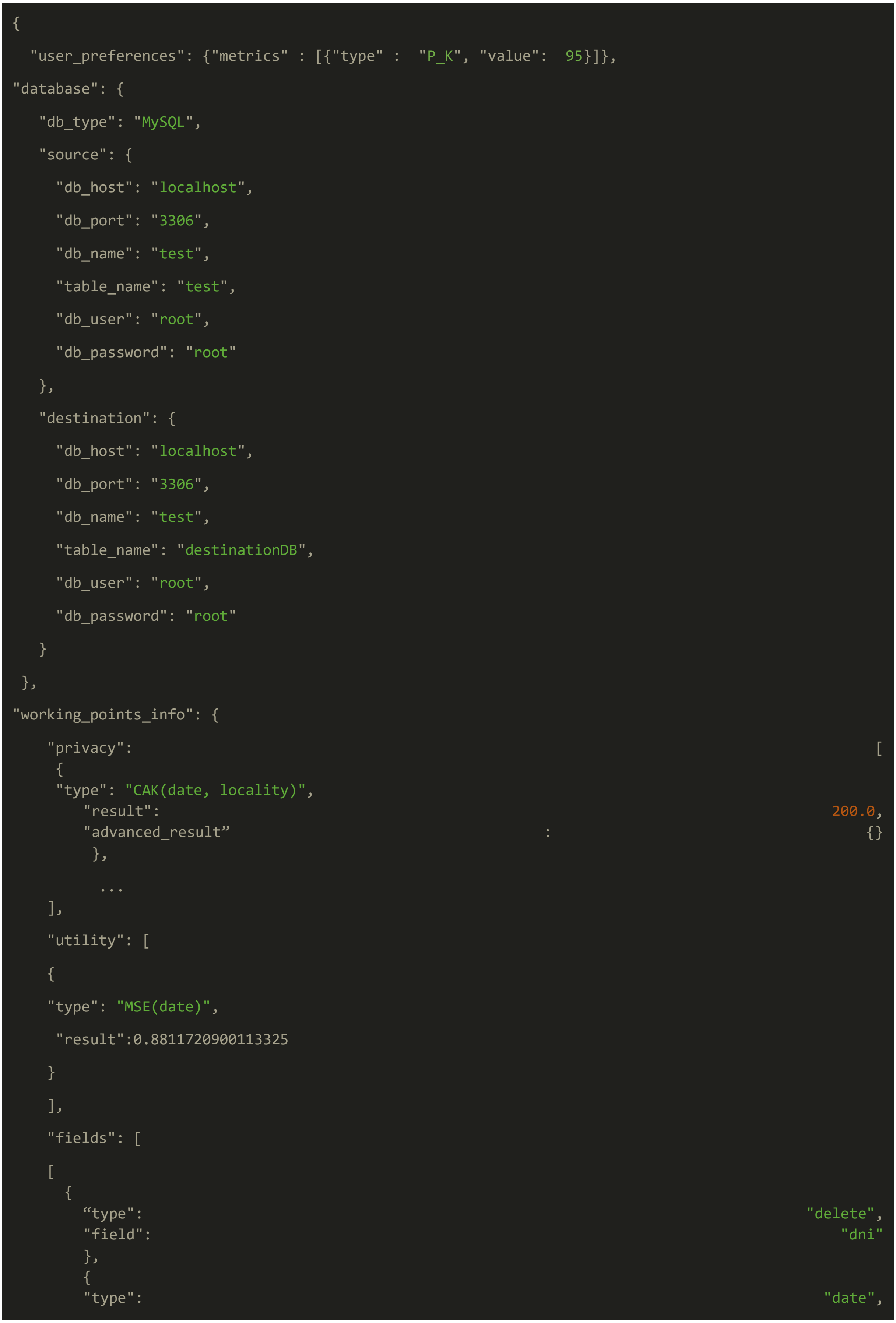
[26] INFINITECH consortium, “INFINITECH D2.19 “Reference Scenarios and Use Cases – Version II”, 2021

[27] KIE API, https://docs.jboss.org/drools/release/6.3.0.Final/kie-api-javadoc/overview-summary.html, [accessed March 2022]

[28] Data Protection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019, https://www.scl.org/news/10384-data-protection-privacy-and-electronic-communications-amendments-etc-eu-exit-regulations-2019 [accessed March 2022]

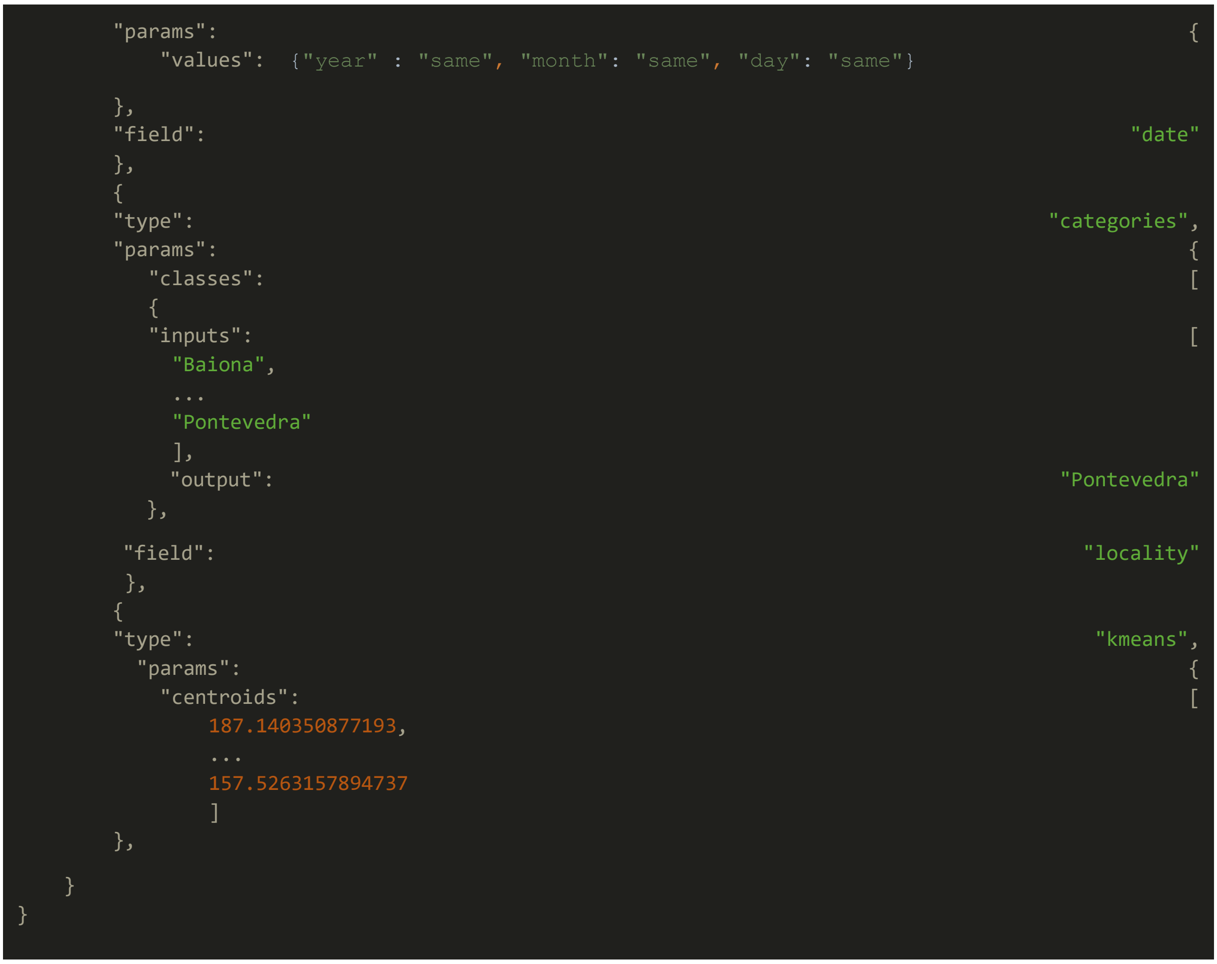
[29] Recital 19 EU GDPR, https://www.privacy-regulation.eu/en/recital-19-GDPR.htm, [accessed March 2022]

Appendix B: Anonymization Configuration



```
{
  "user_preferences": {"metrics" : [{"type" :  "P_K", "value":  95}]},
"database": {
   "db_type": "MySQL",
   "source": {
     "db_host": "localhost",
     "db_port": "3306",
     "db_name": "test",
     "table_name": "test",
     "db_user": "root",
     "db_password": "root"
   },
   "destination": {
     "db_host": "localhost",
     "db_port": "3306",
     "db_name": "test",
     "table_name": "destinationDB",
     "db_user": "root",
     "db_password": "root"
   }
 },
"working_points_info": {
    "privacy":                                                                                       [
     {
     "type": "CAK(date, locality)",
        "result":                                                                              200.0,
        "advanced_result”                                     :                                     {}
         },

          ...
    ],
    "utility": [
    {
    "type": "MSE(date)",
     "result":0.8811720900113325
    }
    ],
    "fields": [
    [
      {
        “type":                                                                             "delete",
        "field":                                                                                "dni"
        },
        {
        "type":                                                                               "date",
```



```
        "params":                                                                                   {
            "values":  {"year" : "same", "month": "same", "day": "same"}

        },
        "field":                                                                               "date"
        },
        {
        "type":                                                                         "categories",
        "params":                                                                                   {
           "classes":                                                                               [
           {
           "inputs":                                                                                [
             "Baiona",
             ...
             "Pontevedra"
             ],
             "output":                                                                   "Pontevedra"
           },

         "field":                                                                          "locality"
         },
        {
        "type":                                                                             "kmeans",
          "params":                                                                                 {
            "centroids":                                                                            [
                187.140350877193,
                ...
                157.5263157894737
                ]
        },

    }
}
```