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

Infinitech

D3.16 – Regulatory Compliance Tools - II

Revision Number	1.0
Task Reference	T3.6
Lead Beneficiary	ATOS
Responsible	Nuria Ituarte Aranda
Partners	AKTIF, ASSEN, ATOS, BOS, BPFI, DYN, GRAD, JSI, NBG, PI
Deliverable Type	Report
Dissemination Level	Public
Due Date	2021-07-31
Delivered Date	2021-07-23
Internal Reviewers	ISPRINT, CCA
Quality Assurance	CCA
Acceptance	WP Leader Accepted and Coordinator Accepted
EC Project Officer	Pierre-Paul Sondag
Programme	HORIZON 2020 - ICT-11-2018
	This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no 856632

Contributing Partners

Partner Acronym	**Role**[1]	**Author(s)**[2]
ATOS	Lead beneficiary	Nuria Ituarte Aranda
AKTIF	Contributor	Orkan Metin
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
ISPRINT	Internal Review	Aristodemos Pnevmatikakis
CCA	Internal Review & QA	Paul Lefrere

Revision History

Version	Date	Partner(s)	Description
0.1	2021-02-31	Atos	ToC Version
0.9	2021-06-20	Atos, All	Version with partners contributions
1.0	2021-07-14	Atos	First Version for Internal Review
2.0	2021-07-16	Atos	Version for Quality Assurance
3.0	2021-07-23	Atos	Version for Submission

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

[2] Can be left void

Executive Summary

Disclaimer

The starting point for this deliverable is D3.15 – Regulatory Compliance Tools – I [1]. Compared to its predecessor, this deliverable has been extended and improved. Specifically, Section 2 regarding the regulations applicable to INFINITECH pilots has been updated, adding to the table from D3.15 the regulations that apply to Pilot #15, which has been incorporated in INFINITECH after the release of D3.15. Section 3 regarding the technologies, has an updated technologies' table. Section 4 regarding the Regulatory Compliance Tools for every pilot has been extended based on the increased depth of the definition of the pilots at this stage. Extending D3.15 into D3.16 in those 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 continues the analysis of regulatory compliance throughout the INFINITECH project and specifically in every pilot. It starts with the regulations for the financial sector and the available technologies in INFINITECH. It provides an update on the analysis for every pilot, as well as the introduction of the new Pilot #15 starting from the results of D3.15 "Regulatory Compliance Tools – I" [1]. The aspects analyzed 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.

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 analyses in detail all the pilots, and the pilot names will be used recursively, the short names of pilots will be used (in the format "Pilot #1" for example). Table 2 shows the mapping of short names versus long names of the INFINITECH pilots and will be used throughout this deliverable to refer to the long names of the pilots.

Table 2

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

Pilot	Applicable Regulations	Compliance Solution	Technologies for Regulatory Compliance
#1	None: GDPR is not applicable because the system does not ever access the data about the customers and the only natural persons involved are notaries, who are considered as legal persons.	Compliance based on the use of synthetic and aggregated data for both the production case and the pilots.	Not Applicable because no regulation applies
#2	BASEL IV and MIFID II	Compliance based on the use of synthetic and aggregated data for the pilots.	IAM and Audit logs (MIFID II)
#3	GDPR (Need for authentication and authorization)	Compliance based on the use of synthetic and aggregated data for the pilots.	IAM and Cryptography
#4	MIFID II and GDPR	Compliance based on INFINITECH and legacy technologies to provide	IAM

		secure access to the personal portfolio internally by allowing secure onboarding authentication to the investor through DUOS	DUOS: Digital onboarding Authentication
#5b	MIFID II and GDPR	Compliance based on the use of synthetic and aggregated data for the pilots.	IAM and Cryptography (GDPR) Audit logs (MIFID II)
#6	GDPR	Compliance based on INFINITECH and legacy technologies: Need to either secure the data or anonymize them. Use of Icarus platform to anonymize the data	Anonymization
#7	This pilot "Avoiding Financial Crime" concerns how to handle confidential data . Thus, in case of specific need to know, please contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us .		
#8	This pilot concerns security issues in a Platform for Anti Money Laundering Supervision (PAMLS). The issues are confidential. Thus, in case of specific need to know, please contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us .		
#9	Production: GDPR and AMLD4 Pilot: none	Compliance based on the use of synthetic and aggregated data for the pilots.	IAM and Cryptography (GDPR)
#10	Production: GDPR. Pilot: none	Compliance based on the use of synthetic and aggregated data for the pilots.	IAM and Cryptography (GDPR)
#11	GDPR	Compliance based on INFINITECH technologies: security framework (IAM and Consent Management) and anonymization.	Anonymization tool IAM, Consent management
#12	GDPR	Compliance based on INFINITECH technologies: The pilot will incorporate a Security framework that will provide IAM and authentication capabilities. Moreover, a regulatory compliance tool that anonymize personal data will be applied.	Anonymization tool, General regulatory 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	TSL IAM

		Management and anonymization features	
#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

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 in 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 a general definition of the requirements for a regulatory compliance service, 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

The general approach of the project towards regulatory compliance tools is to use the ones already in use by the users 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. The summary of INFINITECH Security, Privacy and Data Protection technologies is in Table 4 collects a review of the technologies for security, privacy and data protection available in INFINITECH project included the ones described in section 3.1 of [1]. Contents of this table have been taken from [1], [4] and [7] which explains all the technologies available in INFINITECH. With respect to the previous version of this deliverable [1], some tools listed in [7] have been added, specifically:

- OpenSource AI/ML frameworks;
- Data Layer - REST API;
- Data Check-In Mechanism;
- Analytics Library;
- Pseudonimization Tool.

These are thoroughly described in the table below.

Table 4 and is also described in more detail in deliverable D2.5 "Specifications of INFINITECH Technologies – I" [4].

However, due to the limited resources of the project to develop new features 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
AML IV	Anti-money Laundering
BFM	Business Financial Management
DPO	Data Protection Orchestrator
DUOS	Digital User Onboarding Services
eID	electronic IDentification
eIDAS	electronic IDentification, Authentication and trust Services
EO	Earth Observatory
FATF	Financial Action Task Force (FATF)
GDPR	General Data Protection Regulation
IAM	Identity and Access Management
MDM	Mobile Device Management
MiFID	Markets in Financial Instruments Directive
MiFIR	Markets in Financial Instruments and Amending Regulation
NDA	Non-Disclosure Agreement
NIS	Network and Information Systems
OES	Operators of Essential Services
PAN	Primary Account Number
PaaS	Platform as a Service
PCI DSS	Payment Card Industry Data Security Standard
PEP	Politically Exposed Person
PET	Privacy Enhancing Technology
PIA	Privacy Impact Assessment
PSD2	Payment Service Directive 2
PSP	Payment Service Provider
PSU	Payment Service User
RA	Reference Architecture
P2PP	Peer-to-Peer Payment
QTSP	Qualified Trust Service Provider
RTS	Regulatory Technical Standard
SA	Supervisory Authority
SCA	Strong Customer Authentication
SECaaS	Security-as-a- Service
SEPA	Single European Payments Area
SME	Small and Medium-Sized Enterprises
SIEM	Security Information Event Management
SSL	Secure Sockets Layer
TI	Threat Intelligence
TRA	Transaction Risk Analysis
3DS	Three-Domain Secure

1. Introduction

The current deliverable is the second 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 are analysed, assessing if they are regulatory-compliant, identifying the need of regulatory-compliance tools and preparing the field for the development of these tools.

This second iteration of Regulatory Compliance Tools is an update of the first one, to consider the new Pilot #15, and also goes more in depth into the solutions applied in every pilot that provides these solutions.

It provides a first approach of the definition of the INFINITECH Regulatory Compliance Tool based on the DPO, and describes this component, its architecture, technical design and interfaces, and finally the integration with the Anonymization tool from Gradiant.

1.1 Objective of the Deliverable

The main objective of this deliverable and of all the deliverables that the task "T3.6 Regulatory Compliance Tools" is producing 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 is providing solutions for several pilots with different business objectives, as specified by the end-users of the project (i.e. financial organizations, banks, insurance companies, or 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 for 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.** 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 updated 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 is one of the most important goals of the task. It concerns the study of security and privacy issues that may arise in every pilot and the subsequent search for applicable regulations. Accordingly, a solution for regulatory compliance in the light of the INFINITECH technologies/pilots is sought. Many INFINITECH 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. In the present deliverable, the analyses are in more depth for the pilots that are providing solutions for regulatory compliance.
- **To produce a general regulatory compliance tool** as needed for the project. This deliverable presents a general definition of regulatory compliance tools for INFINITECH, which will be developed during the task T3.6 "Regulatory Compliance Tools". At the end of this task and in the last deliverable of task T3.6, the project will produce regulatory compliance tools in-line with the needs of the INFINITECH pilots. This deliverable describes the definition of the general regulatory compliance tool that could be used by, and adapted for, all the pilots in the project. This solution uses the Data Protection Orchestrator (DPO) provided by Atos. The DPO embeds and automates the assurance of security and

privacy by design and by default in complex business flows. The DPO tool is described in this deliverable.

All the pilots that incorporate solutions for regulatory compliance are being described in more depth in this deliverable. Also consider that Pilot #7 and Pilot #8 address issues to do with confidentiality, so they are analyzed in separate confidential deliverables.

1.2 Insights from other Tasks and Deliverables

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

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

- **INFINITECH D3.15 "Regulatory Compliance Tools – I**"[1]. The current deliverable starts from the content of deliverable D3.15 [1]; from that base, its coverage has been increased and improved by providing updates on regulations and technologies and also more detailed descriptions of all the regulatory compliance tools of the pilots that are providing them. Also, it describes the general definition of INFINITECH Regulatory Compliance Tool based on DPO.
- **INFINITECH-D2.8 "Security and Regulatory Compliance Specifications** – **II"** [4]. 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 in development by the technology partners of INFINITECH. The deliverable also contains specifications of the components, detailing the APIs, functionalities and specifications of the implementation technologies (e.g., BigData/IoT platforms, AI/ML toolkits, HPC infrastructures) that will be used to realize them. The current deliverable INFINITECH-D3.16 is providing an update on D3.15 [1] by assessing all these technologies to extract the 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-D3.13 "Data governance framework and tools – II"** [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 a description of the tools related with the mentioned technologies. This deliverable is one of the most important and practical inputs for INFINITECH-D3.16, given that the tools developed here will be direct components that will be called in 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-D3.16 to analyze the pilots and also to ensure the integration of regulatory compliance tools or on the other hand, to describe where is the pilot implementing regulatory compliance itself.

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

1.3 Structure

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

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 provides a description of the solutions given in the pilots mapping the regulations with the pilots, analysing the privacy and security issues and trying to give a solution through a regulatory compliance tool.

Section 5 provides a general definition of INFINITECH Regulatory Compliance tool based on DPO. It explains the DPO component describing it and providing its architecture and interfaces. It also provides a description of the integration with the Anonymization tool from Gradiant.

Since this deliverable analyses in detail all the pilots, and the pilot names will be used recursively, the short names of pilots will be used (in the format "Pilot #1" for example). Table 2 shows the mapping of short names versus long names of the INFINITECH pilots and will be used throughout this deliverable to refer to the long names of the pilots.

Table 2: Map of INFINITECH Pilots

Pilot short name	Pilot long name
Pilot #1	Invoices Processing Platform for a more Sustainable Banking Industry
Pilot #2	Real time risk assessment in Investment Banking
Pilot #3	Collaborative Customer-centric Data Analytics for Financial Services
Pilot #4	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

2. Review of main applicable regulations in INFINITECH pilots

The following table summarizes the main regulations applicable to each of the INFINITECH Pilots. It is an update of the previous version, in D3.15 Regulatory Compliance Tools – I [1]. This updates 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 main change in the table below is the introduction of Pilot #15 ""Open Inter-Banking Pilot", which was described in D2.19 "Reference Scenarios and Use Cases – Version II".

Regulatory, Standards and legislative activity will be monitored going forward, including the Proposals for Regulation on crypto-assets and digital operational resilience, as well as the Proposal for a Regulation laying down harmonized rules on artificial intelligence and the Proposal for a Regulation on European data governance. In the case where there are further regulatory developments relevant to INFINITECH Pilots, they will be taken into account in D3.17, to be released at Month 30

Table 3: Main regulations in INFINITECH pilots

Pilot	Regulations
#1	**None:** GDPR is not applicable because the system does not ever access the data about the customers and the only persons involved are notaries, who are considered as legal persons.
#2	**T**he pilot is engaged with financial markets data rather than personal data of individuals, the GDPR will not be applicable to the service. **MIFID II**: "deals with financial analysis and maintains that it has conflict of interest declaration for each employee in place" [1] As the system only provides advice but it does not carry out any operation on its own, there will be no email, phone call or electronic operation to be recorded, nor any need for a recovery system. However, the access to sensitive financial data from the customer still makes 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 analyzed 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 **MIFID II** applies **AMLD4** applies
#9	**GDPR** applies because it will collect data from financial transactions, which identify the persons behind them **MIFID II** does not apply because this use case does not provide financial but security consultancy **AMLD4** applies, because this tool may lead to the discovery of illicit transactions, 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.

3. Review of technologies for security, privacy and data protection

Table 4 collects a review of the technologies for security, privacy and data protection available in INFINITECH project included the ones described in section 3.1 of [1]. Contents of this table have been taken from [1], [4] and [7] which explains all the technologies available in INFINITECH. With respect to the previous version of this deliverable [1], some tools listed in [7] have been added, specifically:

- OpenSource AI/ML frameworks;
- Data Layer - REST API;
- Data Check-In Mechanism;
- Analytics Library;
- Pseudonimization Tool.

These are thoroughly described in the table 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 is needed for the Swagger specification of the components (PETs) that will be called by DPO via REST The business flows must be studied and developed to perform communication with the components
Digital User Onboarding System (DUOS)	Atos	This solution allows dealing with virtual identities in a mobile device. It allows using eID or passport for remote user registration. This solution uses eIDs issued by European National authorities according to the EU eID schemas: eID cards and Passports In order to integrate DUOS, it is necessary to adapt and customize it for a user's context-of-need (e.g., Bank application) that requires user authentication This technology could be used in INFINITECH to implement "anonymous" user on boarding. The user can be securely identified by eID or e-Passport without revealing any detail about his/her identity.
Botakis Chatbot Development Network	CP	"A tool for rapid development of chatbots applications, which will be used for the development of chatbots, features in the INFINITECH pilots. Enhancements expecting to be achieved for Botakis Chatbot Platform, based on INFINITECH pilots (i.e. notably the GFT- and NBG-led pilots): - Built-in dialogs that utilize and are integrated with existing NLP frameworks (open or proprietary) provided by partners or every interested party - Powerful dialog system with dialogs that are isolated and composable. - Built-in prompts for simple things like Yes/No, strings, numbers, enumerations." [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 in the framework of the INFINITECH pilots. Regarding the ability to provide info regarding the 7 points described above, we expect that the relative responses will be included in the questions that the chatbot will cover, so we expect any relative material to be included 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. Enhancement aim through 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 by providing similar user experience as zapier, "yahoo pipes" and "IFTTT". The main objective at the innovation perspective 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 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, deletion, etc.) should be applied in order to preserve the maximum level of privacy for the data. "It also includes a set of privacy and utility metrics that allow to measure the risk that remains after anonymizing the dataset, and the impact of the anonymization process on the quality of the data. The component requires two inputs: the data that has to be anonymized and a configuration file that defines the structure of the data, its location and the privacy requirements. " [9] The anonymization tool is intended to be used in two modes, batch or streaming. In the case of using it in batch mode, the output of the component (anonymized data) is stored in a database. The location of the database has to be known beforehand (through the configuration file that is

		taken as an input). If the streaming mode is used, the output will be the queue of the service.
Blockchain-enabled Consent Management System	UBI, IBM, INNOV	The blockchain-enabled Consent Management System offers a decentralised and robust consent management mechanism that enables the sharing of the customer's consent to exchange and utilise their customer data across different banking institutions. The solution enables the financial institutions to effectively manage and share their customer's consents in a transparent and unambiguous manner. It is capable of storing the consents and their complete update history with complete consents' versioning in a secure and trusted manner. The integrity of customer data processing consents and their immutable versioning control are protected by the blockchain infrastructure [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 D4.7 "Permissioned Blockchain for Finance and Insurance – I" of 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 (comments, transaction descriptions etc.). 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. The AI/ML community is progressing these technologies dynamically. This way it provides the basis for solution development and facilitates the specific solution of a wide range of business problems as in INFINITECH. This way, 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, reference to https://gitlab.infinitech-h2020.eu/datamanagement/infinistore.
Data Check-In Mechanism	UBI	A sophisticated data check-in mechanism that is enabling 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 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) Is based on the idea of Update the library to include metadata relevant to security and privacy constraints of the INFINITECH algorithms to be made 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. More information on the following link: https://www.atmosphere-eubrazil.eu/

4. Solutions of Regulatory Compliance tools in pilots

This section aims to review the solutions on regulatory compliance that INFINITECH is providing. First the solutions given in the pilots are presented in overview, followed by the description of the specific features of each solution that INFINITECH provides to the pilots.

4.1 Pilots: Regulatory issues and solutions

D3.15 [1] already has a table on the regulatory issues and solutions. In this section the table is updated, considering the updates of the pilots and the new Pilot #15.

As stated in D3.15 [1], Table 5 contains a brief summary about all the aspects regarding security, privacy and data protection for every pilot of INFINITECH project. This table has been updated from the one included in D3.15 [1] and it uses the deliverables from WP2, WP3 and WP7 for this update ([2], [3], [4], [5], [6]). Specifically, the table details the security and privacy issues, the regulations to fulfil, the available technologies in the project and the solutions that should be applied in every pilot.

The solutions for solving the issues come from two routes: the pilot brings itself a solution or the pilot needs a solution for regulatory compliance to be given by INFINITECH. The most common case in INFINITECH is that the pilot provides a solution internally; in this case the table shows the solution adopted internally in the project. The other case is that the pilot requires a regulatory tool to solve the issues: in this case, this field explains the desired solution for the pilot. The table can be used by the pilots in order to ensure their regulatory compliance.

Table 5: Summary about security and privacy issues, regulations, technologies and solutions in INFINITECH pilots

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

		Operational and Technological Risk Committee that has assessed data used and has concluded that data is not subject to GDPR. This information has also been communicated to the Bank of Spain.		
#2	“The pilot provides risk-assessment analytics on the fly for bank traders, risk managers and sales negotiators based upon Value-at-Risk and Expected Shortfall procedures. It estimates market risks and pre-trade risks thus facilitating decision making processes for traders.” [1] Since the pilot is engaged with financial markets data rather than personal data of individuals, there are no privacy issues.	Production: GDPR, 4AML, BASEL IV and MIFID II Pilot: none since the pilot does not consider real customers PSDII does not apply because the functionalities of the pilot do not deal with transactions	IAM and Cryptography (GDPR) Audit logs (MIFID II)	GDPR: the pilot does not use sensitive data. (the data used is “market data” which is proprietary, but not confidential and trade date which is generated synthetically). MiFiDII: the pilot does not involve any transparency issues and the pilot does not consider real customers 4AML: the pilot does not consider real customers
#3	Platform for sharing financial data	Production: GDPR: Need for authentication and authorization Pilot: Not Applicable	The use of personal data in the demonstrator would lead to applying all GDPR	Not Applicable: The pilot will work only with synthetic data
#4	The main goal of this pilot is to explore the possibilities of AI Based Portfolio construction for Wealth Management. The investor requires access to his/her personal portfolio in a secure way.	Production; MIFID II and GDPR Pilot: Partial GDPR for strong authentication	DUOS: Digital onboarding Authentication	Pilot #4 solves the secure access to the personal portfolio internally by allowing secure onboarding authentication to the investor providing him/her with access from his/her mobile phone to the bank

				services through DUOS
#5b	Business and financial consulting	Production: MIFID II and GDPR Pilot: None due to using synthetic data	Pilot: None due to using synthetic data	This pilot will use only synthetic data to avoid being subject to MIFID II and GDPR
#6	Personalized and intelligent investment portfolio management for Retail Customer	GDPR	need to either secure the data or anonymize them	Use of Icarus Platform to anonymize the data
#7	This pilot is confidential. Thus, in case of specific interest ("need to know"), please contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us .			
#8	This pilot is confidential. Thus, in case of specific interest ("need to know"), please contact the INFINITECH project at https://www.infinitech-h2020.eu/contact-us .			
#9	Real-Time Cybersecurity analytics on financial transactions' data	GDPR for real-life production, none for pilot.	None	Solutions: the pilot will only use synthetic data that does not originate from individual persons, but is created by a machine. Therefore, there are no privacy issues.
#10	Real time security analytics for financial data	GDPR on production mode.	Financial transactions include the people involved in them, making them become personal data	Use of synthetic data to circumvent GDPR
#11	Driver characteristics data and GPS position of the user will be collected and analysed in the AI INFINITECH platform	GDPR	Anonymization tool, IAM and consent management	The pilot solves this internally by means of the security framework (IAM and Consent management) provided by ATOS however a regulatory compliance tool that anonymizes the location data will be applied.
#12	Physical activity of the user will be collected and analysed in the INFINITECH platform	GDPR	Anonymization tool, Access control	The pilot solves this internally by means of the access control framework, that

				ensures that only the specific user can consult the data. Moreover, a regulatory compliance tool that anonymizes personal data will be applied.
#13	The main goal of Pilot #13 is to develop an insurance product configuration platform for SMEs, which will leverage large amounts of digital data in order to compute the insurance offering. An automation of the subscription process will help the insurance company to reduce costs. No security or privacy issues were found. Privacy Issues: the pilot will only use info from legal persons and it never will use personal data, so there are no privacy issues.	Considered regulations are Not Applicable.	Since the subjects to be analyzed are legal persons there are no privacy issues.	No solution is applicable because no regulation applies due that the pilot doesn't use personal data. For user access to the platform, it is applied IAM authorization and access control with Role management through using standard access security measures delivered by Amazon Web Services.
#14	The main goal of Pilot #14 is configurable and personalized insurance products for SMEs and Agro-Insurance. The pilot will evaluate the risks for insurance companies and offer more personalized products. The data collected for this purpose will be stored in the handler's database and no access will be allowed by third parties. Moreover, Confidentiality Agreements will be signed by the involved parties, clarifying the roles (handler, processor, owner) and how the data will be handled during the project.	Considered regulations are: - GDPR,	The communication with the services will be done through secure and encrypted connection, using the SSL protocol and the access to the data will be done only after authentication through credentials and authorisation of the user. Moreover, the server will be enhanced with firewall and the access will	Data anonymization: The data will be provided to the processor anonymized by the insurance companies since no personal data are required in order to deploy the pilot. Each field (the geospatial information along with the aforementioned details) can be accompanied by a specific id (e.g. 1, 2, 3, etc.) and no personal data can be provided.

			be authorised only to the services and to the SSL protocol and the access will be allowed only to the IP of the service providers company.	The pilot solves this internally by means of the security framework provided by AGA
#15	pilot #15 is the result of an open call to banks to identify a use case of common interest. ABI Lab has created an AI HUB, a community of banks, in which they finally, took the decision to address the need of an intelligent system capable of reading, analysing, filtering and organizing the banks' documents by developing and exploiting an innovative taxonomy. It aims to implement the prototype of a solution based on Machine Learning and Natural Language Understanding paradigms. This prototype will analyze a subset of process operating documents to perform a classification of the information contained in them. It will allow the screening of extensive documentation in real time being the starting point for the banks that can adapt to their context.	Considered regulations are not applicable.	Since the service assessment application analyses a subset of process operating documents for classification purposes, without ever accessing the data about the customers, the GDPR is not applicable and there are no privacy issues.	No solution is applicable because no regulation applies.

Table 5 actually shows that some pilots don't use regulatory tool solutions due to the lack of any applicable regulation; this is the case of Pilot #1, Pilot #2, Pilot #3, Pilot #5b and Pilot #13. In the case of Pilot #12, the solution will be given in section 5 throug the general definition of the INFINITECH regulatory tool that will use the DPO. The following subsections will show the solutions that the rest of the pilots give for regulatory compliance.

4.2 Pilot #4: Preliminary definition for the Solution for regulatory compliance

Pilot overview

As stated in D3.15 [1], the goal of this pilot is to explore the possibilities of AI-based Portfolio construction for Wealth Management in general regardless which amount is to be invested. This allows Portfolio construction and optimization for all the customers, not only for the ones with more wealth.

This pilot can be complimentary on 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. 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 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.

Security and privacy issues and requirements

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 licence. No personal data is needed and collected.

Solution

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. This solution comes from Atos and it is described in section 3.4 of INFINITECH D3.13 "Data Governance Framework and Tools – II" [3].
If a B2B customer (either Asset Manager, Bank, 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).

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

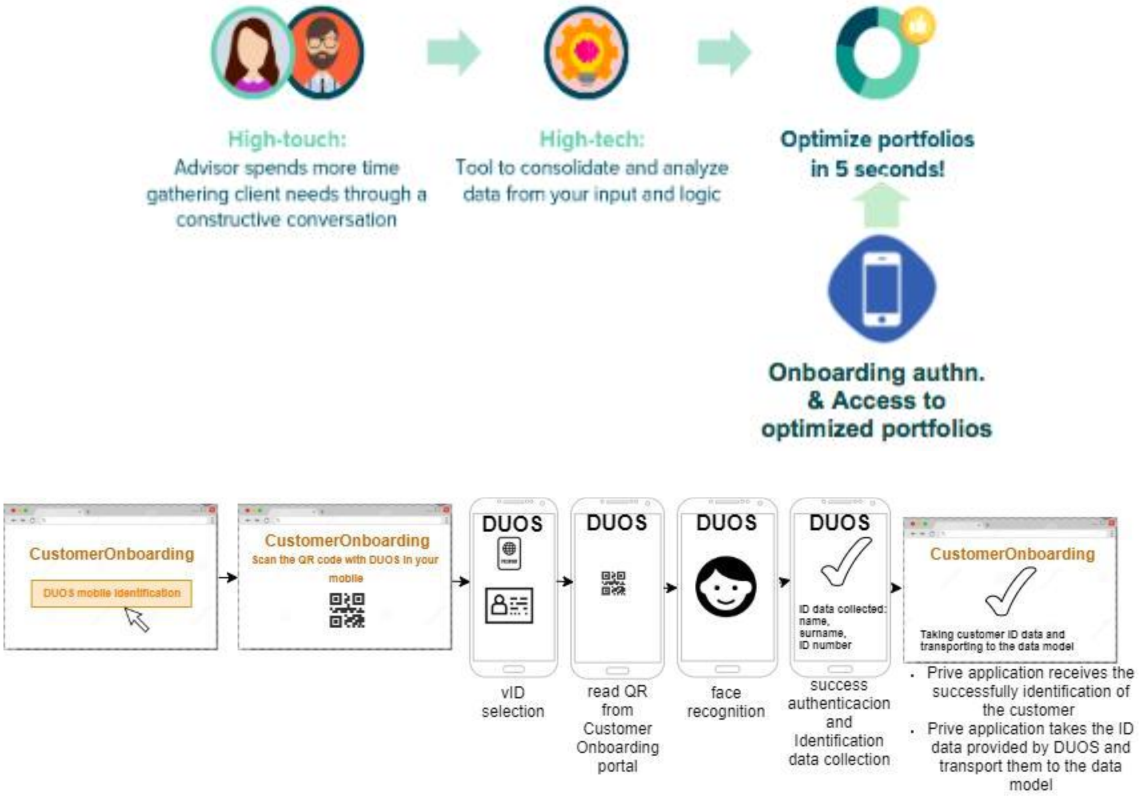

High-touch:
Advisor spends more time gathering client needs through a constructive conversation
High-tech:
Tool to consolidate and analyze data from your input and logic
Optimize portfolios in 5 seconds!
Onboarding authn. & Access to optimized portfolios
CustomerOnboarding
DUOS mobile identification
CustomerOnboarding
Scan the QR code with DUOS in your mobile
DUOS
vID selection
DUOS
read QR from Customer Onboarding portal
DUOS
face recognition
DUOS
ID data collected: name, surname, ID number
success authenticacion and Identification data collection
CustomerOnboarding
Taking customer ID data and transporting to the data model
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 1: 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:

5. 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.
6. CustomerOnboarding- 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.
7. 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 not described in this document yet). 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)
8. DUOS-read QR: the B2B customer must read the QR code shown in the Customer onboarding portal
9. 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.
10. Success authentication: in this case the authentication is correct and DUOS sends to the url (the customer Onboarding portal sent it before via QR) the Identification data agreed between both applications such as the name, surname, ID number, birth date.
11. Successfully identified: the B2B customer onboarding portal receives the data from DUOS and transports them to the data model.

There must 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.

4.3 Pilot #6: Preliminary definition for the Solution for regulatory compliance

Pilot overview

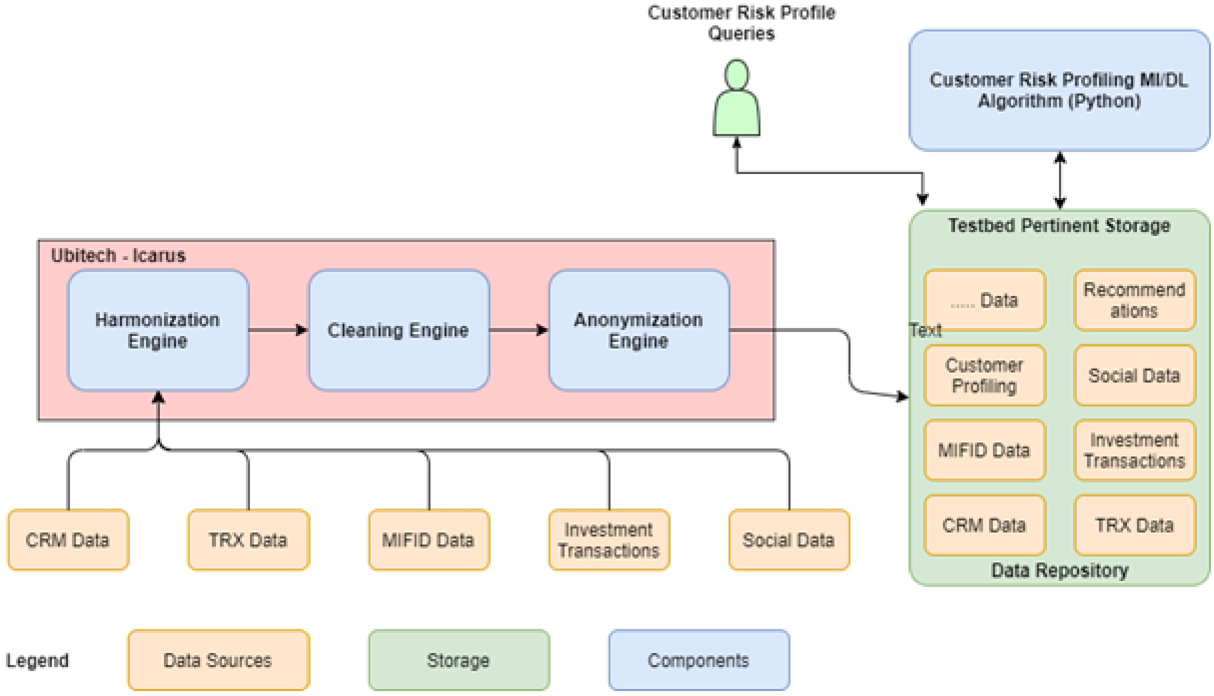
This pilot aims to leverage large customer datasets and large volumes of customer-related alternative data sources (e.g., social media, news feeds, etc) in order to explore the user benefits of the process of providing more targeted, automated, effective, investment recommendations to retail customers.
Creation of personalized investment recommendations available for all Retail Customers and not only to highly affluent. Development of algorithms that aim in 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 takes also into account similarities and patterns among Customers.

Security and privacy issues and requirements

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

Solution

Each customer's personal data is anonymized in order to avoid the identification of individuals.
Figure 2 shows how the anonymization engine fits in the system. Note how all data travelling from one module to other passes through the anonymization.



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

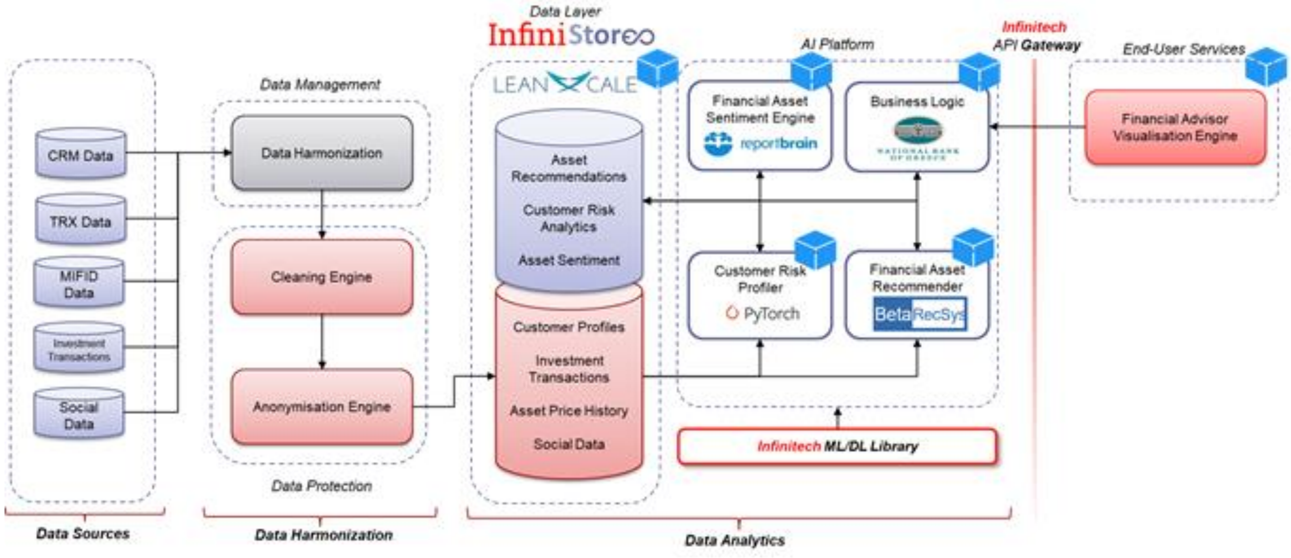


Figure 2: Pilot #6 Architecture

4.4 Pilot #7: Preliminary definition for the Solution for regulatory compliance

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

4.5 Pilot #8: Preliminary definition for the Solution for regulatory compliance

This pilot is confidential. Its detailed description and solutions are delivered as a confidential separate document.
The objective of the Pilot#8 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 AML/CFT by processing Big Data (transaction data, enriched with data from public register). Transaction data will be provided by BOS, financial intelligent unit or third party providers.
To comply 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 data delivery to PAMLS. End user of PAMLS will not be able to identify (directly or indirectly) individuals behind the transactions, therefore the GDPR is not applicable for Pilot#8.
Pseudo-anonymization tool will be provided by the JSI. Detail description of the pseudo-anonymization tool is described on section 2.1 of INFINITECH D3.13 "Data Governance Framework and Tools – II" [3].

4.6 Pilot #11: Preliminary definition for the Solution for regulatory compliance

Pilot overview

As previously exposed in D3.15 [1], this pilot aims to improve the analysis, definition and assignment of risk profiles in car insurance environments. To do so, the pilot will use the information collected from connected vehicles and apply Artificial Intelligence technologies to develop a "Pay as you drive" service (to adapt insurance costs to driver's classification) and a "Fraud detection" service, which exploits driving profiles to identify possible undeclared drivers and driving risks, helping insurance companies detect fraud.

The pilot will use the *driver profile collection* tool from CTAG to collect driving characteristics data from the vehicle, such as speed or acceleration, together with location (GPS position) data. The collected data will create a "route profile" necessary to define, train and test the AI models that the pilot is providing. GPS data is considered sensitive data by the General Data Protection Regulation (GDPR) and must be properly protected.

Security and privacy issues and requirements

This pilot foresees two privacy/security issues: first, unauthorized access to the different modules of the platform, and second the use of sensitive data to train AI models.

Solution

First, prevention of unauthorized access to the platform will be provided by ATOS with OAuth 2.0 based authorization mechanisms to access the platform.

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 will be integrated in the INFINITECH Pilot #11 architecture as follows, where the security framework and the anonymization component can be found:

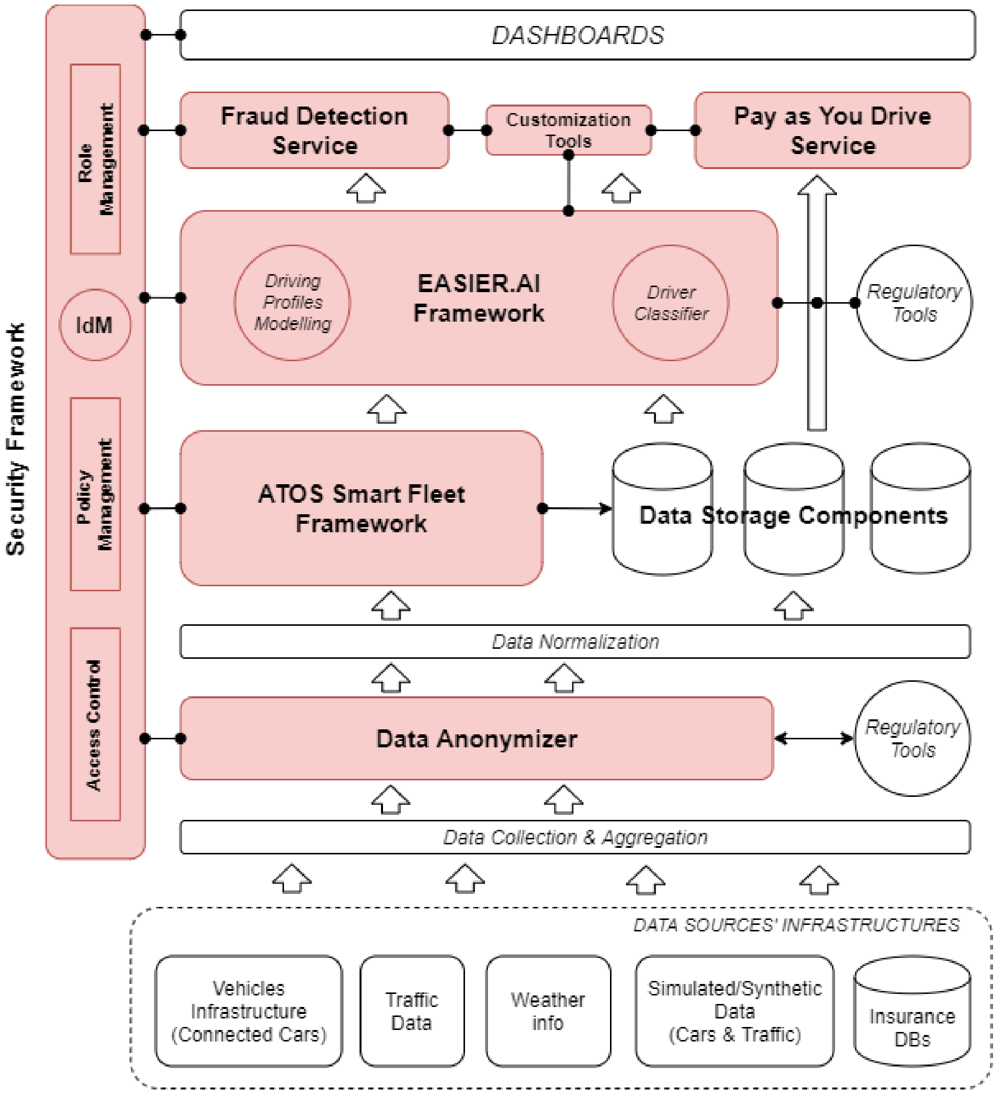


Figure 3: Pilot #11 – High level architecture

The anonymization component will be integrated in Pilot #11 architecture 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.

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

4.7 Pilot #12: Preliminary definition for the Solution for regulatory compliance

Pilot overview

As stated in D3.15 [1] the aim of this pilot is to improve the analysis, definition and assignment of risk profiles in health insurance, by using the information collected from IoT devices and questionnaires, and applying ML technologies. The pilot will develop two distinct services, one performing risk assessment and another one for fraudulent behaviour detection.

To this end, the Healthentia app will collect data from pilot users by means of different activity trackers (Fitbit devices, Android phone sensors and Apple Health Kit) and questionnaires, from psychological to social and environmental aspects. Synthetic data (simulated lifestyle) will be collected in the context of the pilot. Thus, 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 their customers' premium.

Security and privacy issues and requirements

As introduced in D3.15 [1] there exist two different security and privacy issues:

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

Solution

As a solution for regulatory compliance, the Healthentia application's users will each sign a consent form in order to 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 is based on the DPO developed by Atos that orchestrates the calls to different security or privacy tools; in this case it is required to interact with the Data Collector component in order to search the data required to be anonymized and to call the Anonymization tool from Gradiant.
The following diagram shows the role and integration of the regulatory tool to be developed:

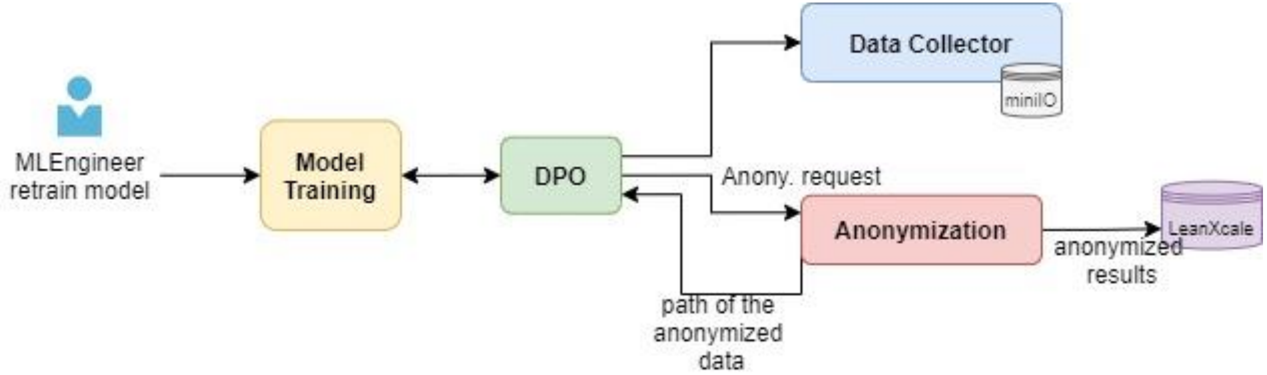


Figure 4: Integration of the Regulatory Tool including Anonymization

The steps are as follows:

1) A Machine Learning Engineer requests to train a ML model with a certain set of data (non-anonymized data)
2) The Model training component calls to DPO requesting to train with a dataset that fulfils concrete features (e.g., files from 1st to 31st of May regarding "physical-measurements")
3) The DPO prepares and configures a business flow to call first the Data Collector to get the desired datasets, performing a filenames search, and obtaining the path and filenames
4) The Data Collector has a miniIO instance which contains the data prepared for using in Machine Learning algorithms. They are organized in files whose names have the data category and dates in which it was collected
5) After that is the DPO business flow call to Anonymization tool (asynchronous API) specifying the paths and filenames obtained in 3)
6) The Anonymization component anonymizes the files requested by the DPO according to a pre-configured setup, and stores the anonymized data in LeanXcale. The dataset returns the path to the anonymized data, and the results of the anonymization in terms of privacy and/or utility metrics.

7) Finally, the Model training component can execute the model training algorithm using the anonymized data stored in LeanXcale.

The DPO will develop interfaces with the Data Collector and with the Anonymization tool.
The high-level architecture for this pilot involves the access control framework, the Regulatory Tools (based on DPO), the Data Collection and the Anonymization tool as shown in Figure 5.

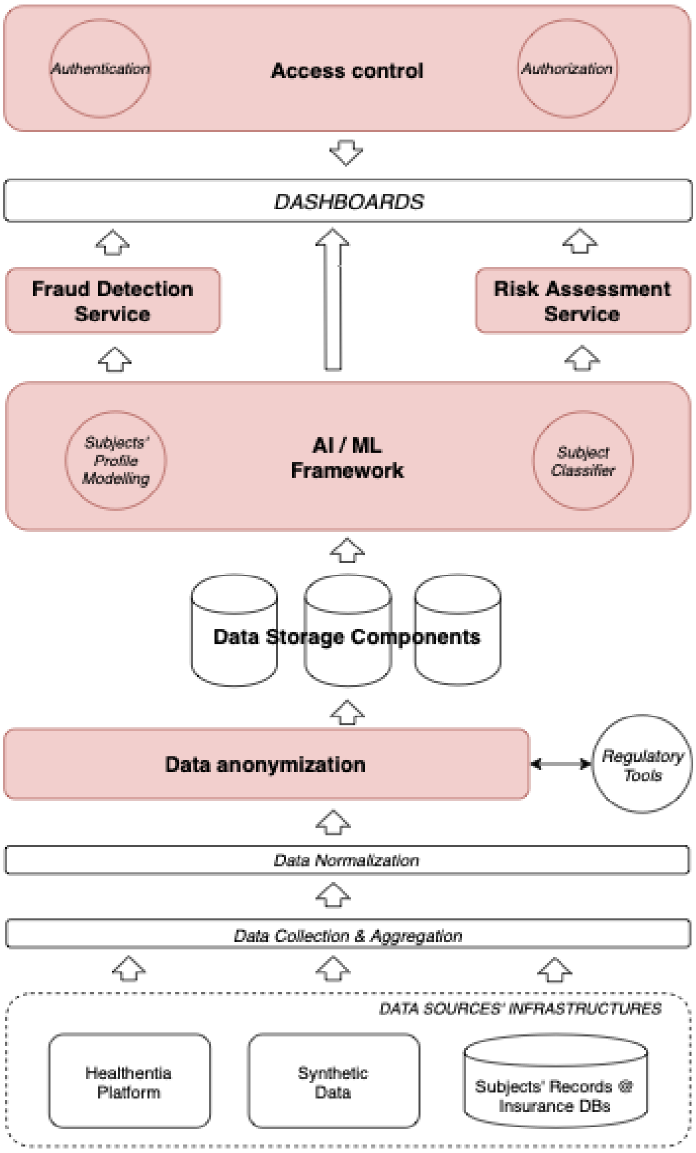


Figure 5: INFINITECH Pilot #12 RA

4.8 Pilot #13: Preliminary definition for the Solution for regulatory compliance

Pilot overview

As stated in D3.15 [1], "The pilot will implement an automation of the subscription process that helps the insurance company reduce costs. In addition, being able to verify that the data entered is correct with a double verification avoids possible errors in the cost of the insurance premium.

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

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

With regard to the use of the information by the companies, the user must load the information they have stored in their systems in our platform, this will receive the name of 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 our service will have a limited number of clients loaded in crude-data, for this, the fields of the Enterprises table, limit, clients_uploaded, total_clients_uploaded will be used in a monitored way.

Each row of this document will identify a client, which can be targeted in different sources of information on the Internet and other open sources in real time, depending on the information available (the quality of information depends on the company), which will be recorded in the *datastore Targets table* (Infrastructure)." [5]

Security and privacy issues and requirements
This pilot does not use personal data, so no regulations apply to it.
As stated in D3.15 [1], there exist two different security issues and a management issue:

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

Solution

As the pilot 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 6.

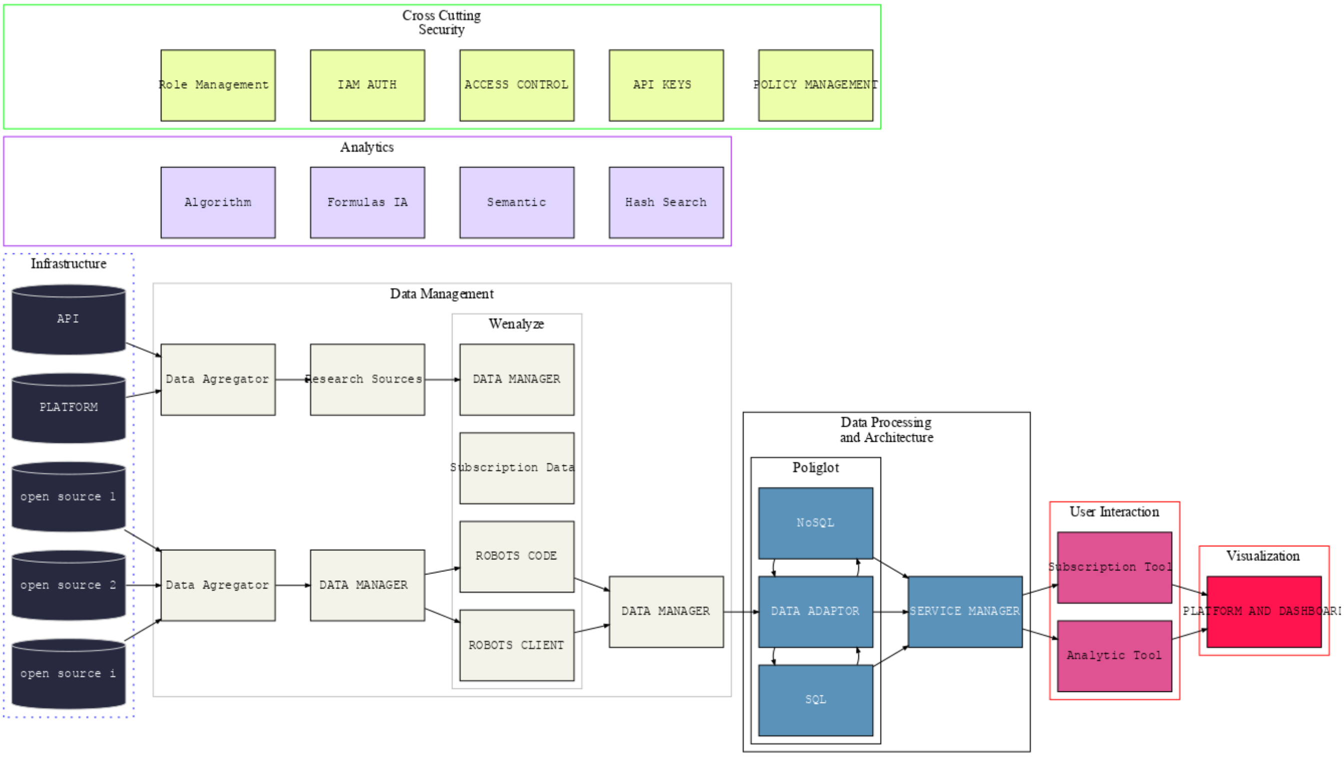


Figure 6: INFINITECH Pilot #13 RA

4.9 Pilot #14: Preliminary definition for the Solution for regulatory compliance

Pilot overview

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 combed 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 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 order 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.

Solution

As stated above the insurance company which is the end-user of the pilot must handle with GDPR regulation and personal data of their clients, 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 following figure:

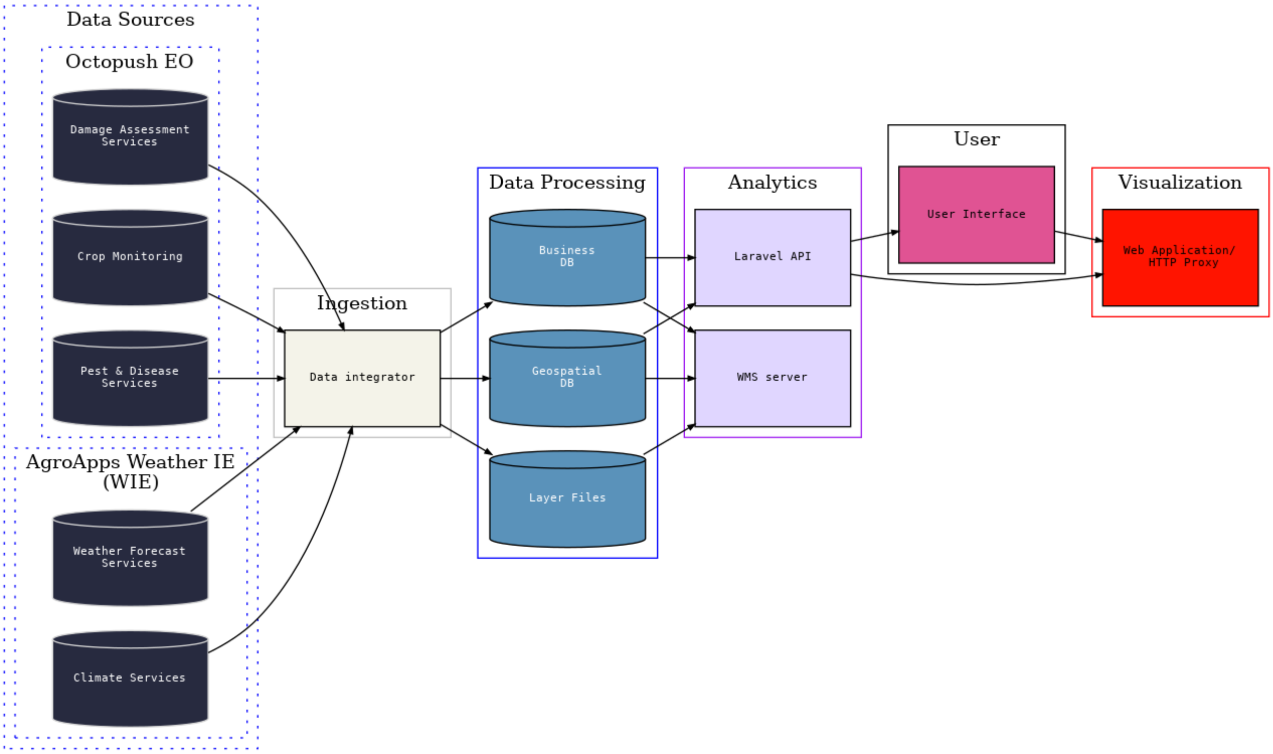


Figure 7: INFINITECH Pilot #14 RA

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.

5 General INFINITECH Regulatory Compliance Tool

5.1 Definition of general INFINITECH Regulatory Compliance Tool

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.

INFINITECH project is providing an approach of a general Regulatory Compliance Tools. 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 dfferent 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.

5.2 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].

5.2.1 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 have to 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 8.

In general, it is usual 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.

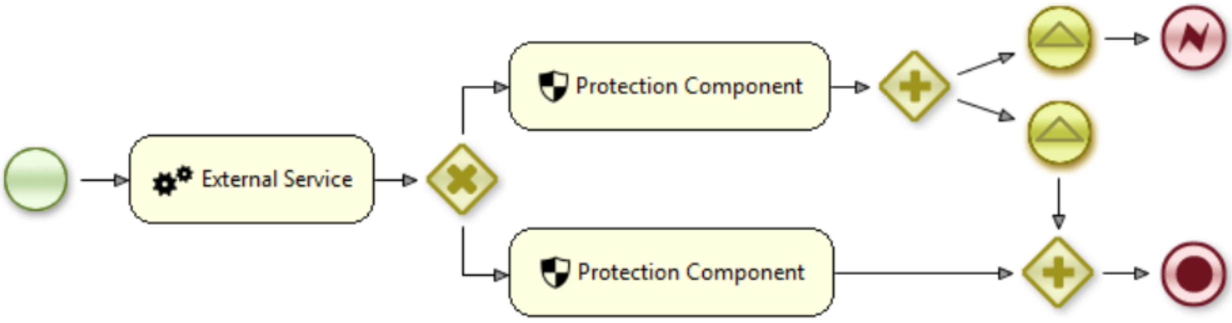


Figure 8: Example of business flow

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

5.2.2 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 9 depicts the DPO subcomponents and their relations.

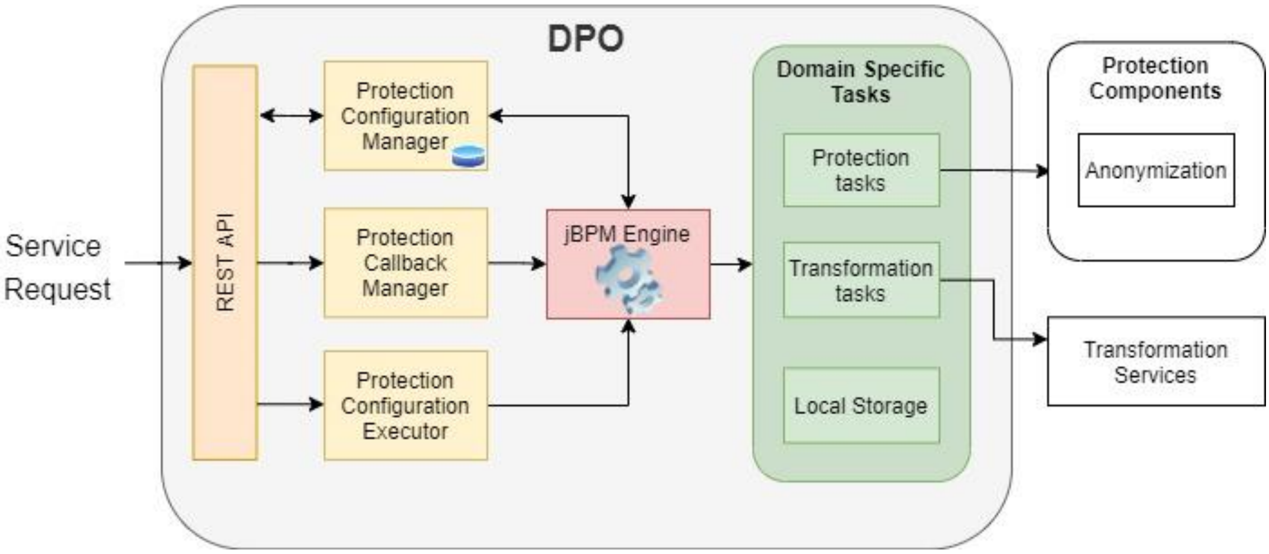


Figure 9: DPO architecture

The protection orchestrator has the following main subcomponents:

- 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.

5.2.3 Security and responsibility considerations for DPO

The Data Protection Orchestrator security model is based on the premise that any request that reaches DPO must be a secure request, that is, it was authenticated and authorized. The DPO is configured in a way that only will accept HTTPS requests from a component which is configured in the DPO

The DPO has a potential security risk which is tampering with the business flows or protection configurations, for example they could be modified removing steps or invoking unauthorized services. That's why it is important that every Protection Configuration management request is only requested by authorized users.

The DPO receives request from an authorized component and the DPO has the responsibility of providing a fault tolerant and real time responsive API. The DPO has a storage service where are stored the protection configurations.

5.2.4 Interfaces

The DPO will only accept HTTPS requests from a component which is configured in it.

The DPO through the domain-specific tasks is able to communicate with other components. 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, it would be called the domain-specific task "Table creation" for which it needs to inform about the location of the table and structure. This would be done using a JSON object which would include information about the table name and the database location. The JSON object would also specify information about the columns, detailing the 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.

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.

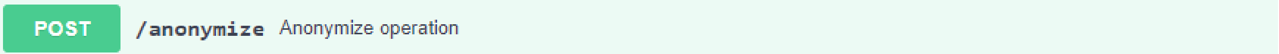
The anonymization tool is intended to be used in two modes: **analysis** and **anonymization**.

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 will allow 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 **sends the selected configuration to the DPO** 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 will orchestrate the anonymization petitions from other components by making API calls to the anonymization component **using pre-generated configuration files**. This section defines 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:



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 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 of 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”
			}

6 Conclusions

The present deliverable D3.16 is devoted to assesing 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 D3.15 (which was the first edition of this work). It also extends the present iteration of D3.15 in its analysis of the regulations for every pilot and its 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 was that they are providing solutions that directly ensure regulatory compliance and this deliverable describes the solutions adopted by them. In other cases, solutions must comply with regulations. To facilitate this, the deliverable introduces 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 would facilitate the provision of possible solutions for new or modified pilots, helping future compliance with new or changed or freshly-identified regulations. The definition of this INFINITECH General Regulatory Compliance Tool as a preliminary prototype based on the DPO is provided in the deliverable, including its architecture, the interfaces and the integration with the Anonymization tool from Gradiant.

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

Appendix A: Literature

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

[2] INFINITECH consortium, "INFINITECH D2.14 – Reference Architecture – II", 2021.

[3] INFINITECH consortium, "INFINITECH D3.13 – Data Governance Framework and Tools – II", 2021.

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

[5] INFINITECH consortium, "INFINITECH D2.13 – Reference Architecture – I", 2020.

[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.

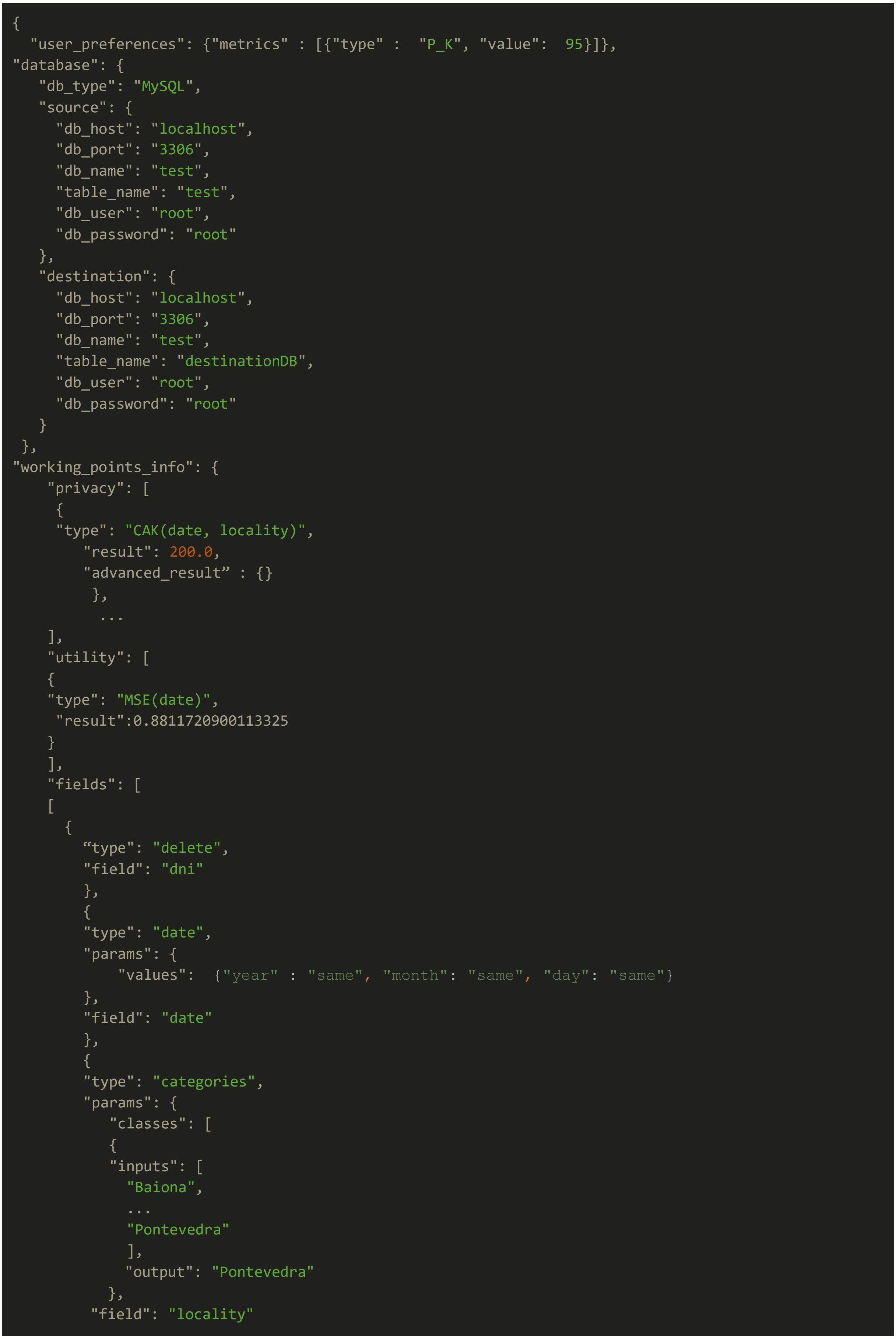
[8] Marcus Brandenburger, Eduarda Freire, "Witdom Project, D4.2 – Final specification of an end-to-end secure architecture", August 2016. [Online]. Available: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ac7baf1a&appId=PPGMS [Accessed 10-July-2021].

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

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

[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].

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
                ]
        },
    }
}
```