

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



D2.6 - Specifications of INFINITECH
Technologies - II

Title	D2.6 - Specifications of INFINITECH Technologies - II
Revision Number	3.0 [build#37]
Task reference	T2.3
Lead Beneficiary	GFT
Responsible	Maurizio Megliola
Partners	ATOS CP FBK FTS GFT GLA GRAD HPE LXS NOVA ORT RB UNIC UPRC WEA
Deliverable Type	Report
Dissemination Level	PU
Due Date	2020-12-31 [M15]
Delivered Date	2021-03-31
Internal Reviewers	CTAG RB
Quality Assurance	INNOV
Acceptance	Accepted by Coordinator
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	Role ¹	Author(s) ²	Section(s)
GFT	Lead Beneficiary	Maurizio Megliola	
GFT	Contributor		0. 1. 2. 3. 3.1. 4. 5.
ATOS	Contributor		2. 3. 3.1.
CP	Contributor		2. 3. 3.1.
FBK	Contributor		2. 3. 3.1.
FTS	Contributor		2. 3. 3.1.
GLA	Contributor		2. 3. 3.1.
GRAD	Contributor		2. 3. 3.1.
HPE	Contributor		2. 3. 3.1.
LXS	Contributor		2. 3. 3.1.
NOVA	Contributor		2. 3. 3.1.
ORT	Contributor		2. 3. 3.1.
RB	Contributor		2. 3. 3.1.
UNIC	Contributor		2. 3. 3.1.
UPRC	Contributor		2. 3. 3.1.
WEA	Contributor		2. 3. 3.1.
CTAG	Reviewer		all sections
RB	Reviewer		all sections
INNOV	Quality Assurance	Dimitris Drakoulis	

1. Lead Beneficiary, Contributor, Internal Reviewer, Quality Assurance

2. Can be left void

Revision History

Version	Date	Partners	Description
0.1	2020-12-02	GFT	First version with TOC / from D2.5
0.2	2020-12-19	GFT	2nd Version for Work in Progress
0.3	2021-01-19	GFT	3rd Version Work in Progress
0.4	2021-03-26	GFT	Version in APP
0.5	2021-03-26	GFT	Version for Work in Progress
1.0	2021-03-31	GFT	Version for Internal Review
2.0	2021-03-31	GFT	Version for Quality Assurance
3.0	2021-03-31	GFT	Version for Submission

Executive Summary

The Deliverable D2.6 “Specifications of INFINITECH Technologies - II” is the second and final edition of Deliverable D2.5 about the technological components used in the project and indexed in the Reference Architecture. Deliverable D2.6 supersedes completely the previous version D2.5.

As part of the Horizon 2020 Programme Innovation Action, the INFINITECH project has the ambition to advance the state of the art of the technology integrating and developing the components needed to demonstrate the feasibility of the different pilots of the project. Moreover, the developed solutions will allow to approach a larger class of use cases related to data processing and smart analytics within the Financial and Insurance Sectors to solve similar problems.

The Deliverable has the objective to specify the technological components (mature products or artefacts or poc) in their actual status of development, providing the specifications of the technologies that constitute the building blocks utilized to solve the different use cases of the pilots, but more specifically to provide a reference description of the components that along with the Reference Architecture in Deliverables D2.13 and D2.14 form the INFINITECH components' library.

The specifications take into account the requirements of the previous two tasks of the project, namely T2.1 and T2.2 that deals with User Stories and Reference Scenarios, while at the same time aligning to the INFINITECH-RA developed as part of T2.7.

In particular, the deliverable lists the tools and technologies currently available and in development by the technology partners. The deliverable also contains detailed specifications of the component characteristics , including Input and Output , functionalities and specifications about the implementation technologies (e.g., BigData/IoT platforms, AI/ML toolkits, HPC infrastructures).

As a methodology, the project maintains updated a database of assets information and in particular those about the components used in pilots' sandboxes. This deliverable exposes this information in a explainable tables. However much more information is kept in the source of information referred by each single components (e.g. documentation).

With respect to its previous version (D2.5 “Specifications of INFINITECH Technologies - I”), the deliverable enriches the content so far provided by updating it with new techtools and new pilot components as well as by providing additional information for each pilot component, in particular related to their availability in the Marketplace, their availability as containerized components (e.g. dockerized microservices), along with the specification, if any, of their endpoints or REST API URLs.

The deliverable is also part of a larger picture as it also provides the basis for future work and material for the Marketplace and the Virtualized Digital Innovation Hub of WP8. In particular most of the technological components will be included as assets in the Marketplace.

Rather than reinventing the wheel, the INFINITECH project will build on the shoulders of the giants of the many and different assets brought from the technological partners and fintech companies as well as the assets provided by end-users and stakeholders involved in the project.

Table of Contents

1 Introduction	9
1.1 Objective of the Deliverable	9
1.2 Insights from other Tasks and Deliverables	9
1.3 Structure	10
1.4 Differences and Updates between previous version	10
2 Available Tools and Technologies	11
2.1 Car data ingestion	11
2.2 Data Protection Orchestrator (DPO)	11
2.3 Digital User Onboarding System (DUOS)	12
2.4 EASIER-AI	12
2.5 Driver profile classifier	12
2.6 Distributed near-real-time HPC processing and exchange of IoT streaming data	13
2.7 Botakis Chatbot Development Network	13
2.8 Crowdpolicy Open (innovation) banking solution	14
2.9 AI-Engine-for- Psychometric-Profilng and Personalization	14
2.10 OpenSource AI/ML frameworks	14
2.11 Data Layer - REST API	15
2.12 Terrier Information Retrieval Platform	15
2.13 Anonymization tool	15
2.14 Polyglot Database Management System	16
2.15 HTAP Database for the financial and insurance sector	16
2.16 Natural Language Processing for real-time, high-accuracy Credit Risk assessment	17
2.17 Machine learning algorithms for health related data	17
2.18 Wenalyze Big data analytics platform	18
2.19 Octopush geospatial enabling framework	18
2.20 AgroApps Weather Intelligence Engine	19
2.21 Sentiment Analysis Tool	19
2.22 Partitioned and Distributed Transaction Graphs	20
2.23 ALIDA: A Micro-service based platform for composition, deployment and execution of BDA applications	20
2.24 Text Analysis Tool	20
2.25 Blockchain tokenization	21
2.26 Healthentia LifeSciences BigData Platform	21
2.27 Event-registry	21
2.28 Stream Story	22
2.29 Qminer	22
2.30 SSC - Super Stream Collider, a multiformat Data Management and Query System	22
2.31 The Global Engine with Neural Network Intelligence (GENNI)	23
2.32 Data Check-in mechanism	23
2.33 IoT-Catalogue	24
2.34 Analytics Library	24
2.35 Catalogue of Objects	24
2.36 Portfolio Optimization Tool	25
2.37 Sentiment Analysis for Financial news	25
2.38 AI model for VaR prediction	26
3 INFINITECH components specifications	27
3.1 Data Component Description	28
3.1.1 Component Template	28
3.1.2 Relational Database	28
3.1.3 Polyglot Query Processing	29
3.1.4 Incremental Analytics	29

3.1.5 OneHotEncoder	30
3.1.6 Stream Processor	31
3.1.7 Online Aggregates	31
3.1.8 Data Collection	32
3.1.9 Anonymization tool	33
3.1.10 DUOS (Digital User On-boarding Services)	33
3.1.11 DPO (Data Protection Orchestrator)	34
3.1.12 Blockchain Reader	34
3.1.13 Blockchain Writer	35
3.1.14 Smart Contract Executor	36
3.1.15 Blockchain Data Visualizer	37
3.1.16 Blockchain Authenticator	37
3.1.17 Blockchain Encryptor	38
3.1.18 Blockchain Decryptor	39
3.1.19 Blockchain Transaction Dataset Preparation Component	39
3.1.20 Scalable Transaction Graph Analysis Component	40
3.1.21 Semantic Streams Analyzer	40
3.1.22 Semantic Reasoner	41
3.1.23 Ontology Mapping	42
3.1.24 Semantic Annotator-Preprocessing	42
3.1.25 Smart Fleet (IoT Context Management and Historical data component)	43
3.1.26 Fraud Detection Service Training	44
3.1.27 Fraud Detection Service Execution	44
3.1.28 Pay As You Drive Service Training	45
3.1.29 Pay As You Drive Service Execution	45
3.1.30 Investment Recommendation Engine Training	46
3.1.31 Investment Recommendation Engine Execution	46
3.1.32 Recommender	47
3.1.33 Cash Flow Prediction	48
3.1.34 Budget Prediction	48
3.1.35 KPI Engine	49
3.1.36 Transaction (Txn) Monitoring	50
3.1.37 Transaction (Txn) Categorization	50
3.1.38 Invoice Processing	51
3.1.39 KMeans	52
3.1.40 Random Forest (Model)	52
3.1.41 Random Forest (Predict)	53
3.1.42 Client Contextual Information	54
3.1.43 Financial Fraud/Crime Risk Score	54
3.1.44 Anomaly Analysis	55
3.1.45 Pattern Analysis	55
3.1.46 Stream Story	56
3.1.47 Open API Gateway	56
3.1.48 User Interface for Blockchain Transaction Reports and Visualization Component	57
3.1.49 Visualization Preparation	57
3.1.50 Real Time Visualization	58
3.1.51 INFINISTORE	59
3.1.52 UI Risk Assessment based on VaR	59
3.1.53 Pseudo-anonymization tool	60
3.1.54 Health insurance risk assessment service	60
3.1.55 Health insurance fraud detection service	61
3.1.56 Well-being outlook classifiers	61
3.1.57 Synthetic RWD for well-being analytics	62

3.1.58 Open Banking Agreggator Solution	62
3.1.59 Big data analytics platform	63
4 Conclusions	64
5 References	65

List of Figures

Figure 1 - Tasks interactions in WP2	9
Figure 2 - Logical Schema of Data Processing Components in INFINITECH RA	27

Abbreviations

Abbreviation	Definition
ACID	Atomicity, Consistency, Isolation, Durability
AI	Artificial Intelligence
API	Application Programming Interface
BDA	Big Data Association
BDVA	Big Data Value Association
BOC	Bank of Cyprus
BOI	Bank of Ireland
BOS	Bank of Slovenia
BPMN	Business Process Model and Notation
CEP	Center (for) Energy Policy
CPU	Central Processing Unit
CSV	Comma Separated Value files
DB	Data Base
DL	Deep Learning
DPO	Data Protection Officer
ERP	Enterprise Resource Planning
ES	End System Expert System
ETSI	European Telecommunications Standards Institute
EU	European Union
FIBO	Financial Industry Business Ontology
FIGI	Financial Instrument Global Identifier
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HDFS	Hadoop Distributed File System
HPC	High Performance Computing
HTAP	Hybrid transaction/analytical processing
I/O	Input Output
IBM	International Business Machines
ICT	Information Communication Technologies
IN	Intelligent Network
IOT	Internet of Things (also IoT)
IP	Internet Protocol
IRA	INFINITECH REFERENCE ARCHITECTURE

Abbreviation	Definition
IT	Information Technology
IoT	Internet of Things
JDBC	Java Database Connectivity
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
KYC	Know Your Customer
ML	Machine Language
MPI	Message Passing Interface
N/A	Not Available / Not Applicable
NLP	Natural language processing
ODBC	Open Database Connectivity
OLAP	On Line Analytical Processing
OLTP	On Line Transaction Processing
OWL	Web Ontology Language (W3C)
PSD2	Second Payment Service Directive
RA	Reference Architecture
RDF	Resource Description Framework
REST	Representational State Transfer
RWD	Rear Wheel Drive
SAR	Synthetic Aperture Radar
SME	Small and Medium-Sized Enterprises
SPARQL	Simple Protocol and RDF (Resource Description Framework) Query Language
SQL	Structured Query Language
SSC	Super Stream Collider
TBD	To Be Determined
UI	User Interface
URL	Uniform Resource Locator
USD	United States Dollar
VAT	Value Added Tax
WP2	Work Package 2 dealing with Requirements and Specifications
WP3	Work Package 3 dealing with BigData/IoT Data Management
WP5	Work Package 5 dealing with AI Algorithms (ML/DL) and Blockchain
XML	Extensible Markup Language

1 Introduction

Deliverable D2.6 is about technology specifications of the building blocks that will be used within the pilots and in particular in the pilots' sandboxes of the INFINITECH Project. It is conceived as a reference resource of information for for the entire project about the components used and/or developed within the project.

In this respect, the deliverable is a snapshot of the current database kept to contain the technical information of the technologies used within the project. The database will be maintained and updated throughout the entire project life time and beyond in the Marketplace.

1.1 Objective of the Deliverable

The objectives and scope of this deliverable are to detail the BigData/IoT technological building blocks that will be developed in the scope of the project, in the areas of data management, semantic interoperability, cost-effective real-time BigData analytics, elastic cloud storage, integrated (declarative) data querying, AI/ML algorithms and more.

Compared to its previous version (D2.5), this deliverable is an update with new techtools and pilot components and provides additional information for each pilot component, in particular related to their availability in the Marketplace build in WP8, their availability as a containerized application (e.g. dockerized microservice), and finally the specification, when available, of its Endpoints/REST API URLs.

The deliverable is also conceived as a specification library that can be used in diverse use cases and scenarios to find out solutions out-of-the-shelf.

1.2 Insights from other Tasks and Deliverables

The specifications included in this deliverable take into account the requirements and specifications of the previous two tasks [1] and [2], while at the same time aligning to the INFINITECH-RA that will be developed as part of T2.7.

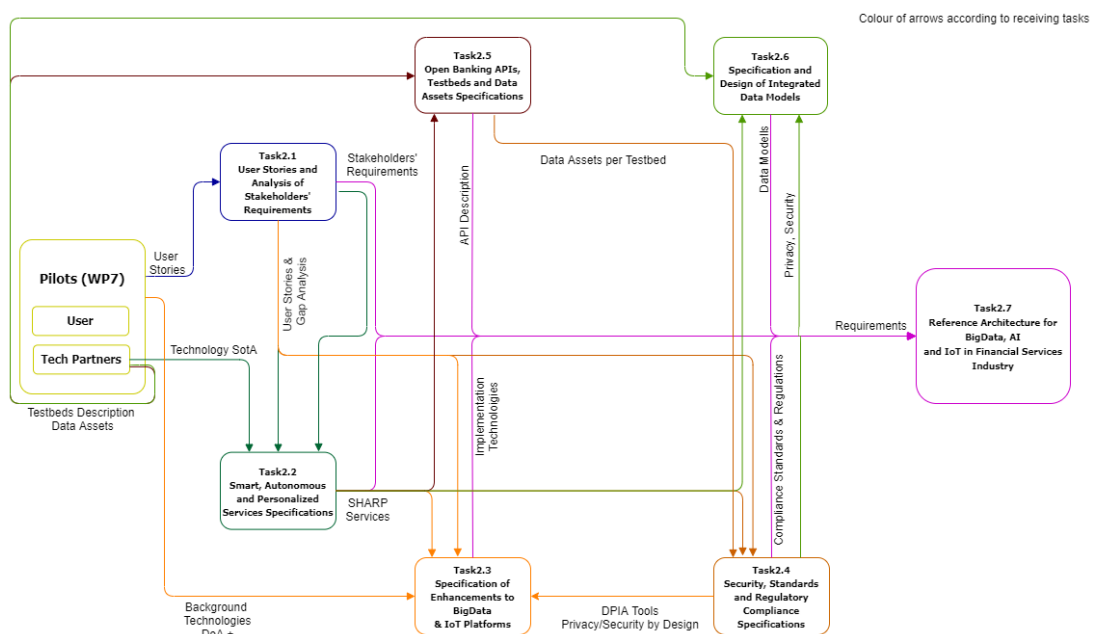


Figure 1 - Tasks interactions in WP2

1.3 Structure

This deliverable is composed of four main sections.

- Chapter 1 is the introduction to the deliverable and includes the description of the objective, insights from other tasks and deliverables and the structure.
- Chapter 2 describes in detail the available technologies within the context of the project.
- Chapter 3 details the specifications of the INFINITECH components.
- Finally, Chapter 4 reports some conclusions.

1.4 Differences and Updates between previous version

The deliverable D2.6 contains revised text from D2.5 as this is an updated version of the latter and supersedes entirely the previous version.

With respect to its previous version (D2.5 “Specifications of INFINITECH Technologies - I”), the deliverable enriches the contents provided by updating it with new techtools and pilot components as well as by providing additional information for each pilot component, in particular related to their availability in the Marketplace, their availability as containerized components (e.g. dockerized microservices), along with the specification, if any, of their endpoints or REST API URLs.

2 Available Tools and Technologies

The following section presents the available tools and applications owned by the Consortium's Partners that constitute the background of the technologies exploited in the Project. These BACKGROUND Technologies can be the basis to build up other components for INFINITECH, by improving them and increasing the related TRL.

The technologies are identified after a process/reasoning based on the input provided by the Consortium Partners.

For each tool or application, the following information are provided:

- Title/Name as a short description of tool/platform;
- Description in particular the characteristics and/or technology the component is based on;
- Documentation or detailed references to links/demo environment;
- ACRONYM of the Company or Partner in the project who has developed tool/platform and is owner;
- TRL Technology Readiness Level current and expected level at the end of the INFINITECH project;
- Ideas for enhancements to be done in the course of the project;
- License schema (e.g. Proprietary, GPL, Apache, MIT, ...)
- Pilots (usually referred as P01, P02, etc)

2.1 Car data ingestion

Attribute	Value
Title	Car data ingestion
Description	This technology is tasked with the capture, homogenization, distribution and storage of the datasets that support the connected car pilot. It involves the design and implementation of the IoT agents that adapt the data available from the Vehicle, as well as deploying the necessary modules for data storage and distribution (FIWARE Orion, FIWARE Cygnus, etc)
Documentation	https://fiware-datamodels.readthedocs.io/en/latest/Transportation/Vehicle/Vehicle/doc/spec/index.html
Company	ATOS
Trl	5->7
Ideas	As the data sources size increases, the technology will have to become more robust by polishing the performance of the deployed IoT agents.
License	Proprietary
Remarks	Based on FIWARE technologies and FIWARE Data models
Pilots	P11

2.2 Data Protection Orchestrator (DPO)

Attribute	Value
Title	Data Protection Orchestrator (DPO)
Description	It is an enabler for embedding and automating the assurance of security and privacy by design and by default in heterogeneous and complex business flow. It orchestrates various privacy and security management functions (such as access control, encryption and anonymization). It will be used as part of the data governance Framework of the Project, and towards establishing the regulatory compliance tools in the project's sandboxes.
Documentation	Witdom Project: http://witdom.eu/sites/default/files/witdom/public/content-files/deliverables (see REF [7])

Company	ATOS
Trl	5-> 6
Ideas	It requires Swagger specification of the components (PETs) that will be called via REST by DPO There will be developed the business flows to address the specific communication with the components
License	Proprietary
Remarks	
Pilots	P12

2.3 Digital User Onboarding System (DUOS)

Attribute	Value
Title	Digital User Onboarding System (DUOS)
Description	It is a solution for dealing with virtual identities in a mobile device. It provides remote user registration using eID or passport
Documentation	ARIES project: https://www.aries-project.eu/content/aries-prototype-instantiation ,
Company	ATOS
Trl	5-> 6
Ideas	It is needed to use eIDs issued by European National authorities according to the EU eID schemas: eID cards and Passports There will be implemented the needed improvements regarding integration with end users application (Bank application) that needs user authentication
License	Proprietary
Remarks	
Pilots	P04

2.4 EASIER-AI

Attribute	Value
Title	EASIER-AI
Description	EASIER-AI is an Hybrid (Cloud/Edge) platform that facilitates to develop, measure, monitor and deploy your AI models. The platform facilitates the data science tasks and it is focused on working on Hybrid Infrastructure and exploiting data generated by IoT. The platform synchronizes Cloud and Edge, keeping the Edge always up to date to run always the most accurate model.
Documentation	
Company	ATOS
Trl	4->6
Ideas	By including this tool in the INFINITECH project, we aim to feed it with new dataset sources, resulting in the development of new ML models for the platform.
License	Proprietary
Remarks	
Pilots	P11

2.5 Driver profile classifier

Attribute	Value
Title	Driver profile classifier

Description	This technology is aimed at the use of high-quality vehicle data allows insurance companies to offer customized products. The application of supervised machine learning techniques is proposed to classify drivers' profiles which generates a customized insurance premium. The resulting model is then deployed with TensorFlow serving and integrated as part of the cloud platform with a wrapper, achieving an accuracy of 85.7%.
Documentation	
Company	ATOS
Trl	5->7
Ideas	The inclusion of this model in the project could bring an improvement on the data analysis of the datasets, bringing an improvement to the model accuracy.
License	Proprietary
Remarks	
Pilots	P11

2.6 Distributed near-real-time HPC processing and exchange of IoT streaming data

Attribute	Value
Title	Distributed near-real-time HPC processing and exchange of IoT streaming data
Description	AI algorithms' optimization exploiting GPUs capabilities
Documentation	
Company	ATOS
Trl	5->6
Ideas	Usage of GPU processors to enhance AI algorithms performance and reinforce CPU capabilities.
License	Proprietary
Remarks	
Pilots	P11

2.7 Botakis Chatbot Development Network

Attribute	Value
Title	Botakis Chatbot Development Network
Description	A tool for rapid development of chatbots applications, which will be used for the development of chatbots features in the INFINITECH pilots (i.e. notably the LIB, BOC and NBG led pilots).
Documentation	https://www.crowdpolicy.com/service/botakis/
Company	CP
Trl	7->8
Ideas	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 be 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.
License	Proprietary
Remarks	
Pilots	P05b

2.8 Crowdpolicy Open (innovation) banking solution

Attribute	Value
Title	Crowdpolicy Open (innovation) banking solution
Description	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. Our solution puts the bank (or any monetary financial institution) in control of the third-party partner relation. The solution is full PSD2 & GDPR Compliance
Documentation	https://www.crowdpolicy.com/fintech-innovation-enabler-en/
Company	CP
Trl	6/7->8/9
Ideas	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. The application worked by enabling users to
License	Proprietary
Remarks	
Pilots	P03

2.9 AI-Engine-for- Psychometric-Profiling and Personalization

Attribute	Value
Title	AI-Engine-for- Psychometric-Profiling and Personalization
Description	AI-driven engine to extract four categories of behavioural features, grouped according to the type of spending behaviour they capture: (i) overall spending behaviour, (ii) temporal spending behaviour, (iii) category-related spending behaviour, and (iv) customer category profile. The engine uses these features to predict the personality of customers.
Documentation	
Company	FBK
Trl	4->5
Ideas	
License	Proprietary
Remarks	
Pilots	no pilots at the current stage

2.10 OpenSource AI/ML frameworks

Attribute	Value
Title	OpenSource AI/ML frameworks
Description	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
Documentation	
Company	FTS
Trl	7->7

Ideas	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.
License	Proprietary
Remarks	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.
Pilots	P01. The frameworks will be mainly applied in the solutions of Pilot#7.

2.11 Data Layer - REST API

Attribute	Value
Title	Data Layer - REST API
Description	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.
Documentation	https://gitlab.infinitech-h2020.eu/datamanagement/infinistore
Company	GFT
Trl	4
Ideas	To complete the wrap with standard I/O
License	Proprietary
Remarks	
Pilots	P15

2.12 Terrier Information Retrieval Platform

Attribute	Value
Title	Terrier Information Retrieval Platform
Description	Search Engine for BigData sets that offers integration with Spark for distributed processing
Documentation	http://terrier.org
Company	GLA
Trl	5->7
Ideas	We plan to extend Terrier with a new open source module that combines real-time data stream ingestion (via Apache Flink) with distributed database access (via LeanXcale) for real-time data indexing and updating from multiple sources, within WP3 and WP5. We will also expand Terrier with enhanced Python integration, allowing easier use from common data science pipelines, such as those involving Pandas. This technology can be used for tasks such as searching/sampling financial product portfolios, user profiles or for providing recommendations.
License	Proprietary
Remarks	
Pilots	P02

2.13 Anonymization tool

Attribute	Value
Title	Anonymization tool

Description	A tool that anonymizes data in order to preserve privacy. It also provides metrics that allow to measure the risk of the anonymized data and the impact of the anonymization process on the utility of the data. The tool will be used in pilot #11 and pilot #12. The component needs a specific configuration/development for each pilot in which it is used
Documentation	There is no public documentation available
Company	GRAD
Trl	5->6-7
Ideas	In T2.3 the REST API of the component will be defined, so it can be instantiated within INFINITECH's platform. The rest of the enhancements of the tool will be related to developing new anonymization algorithms and metrics, which fall under the scope of T3.5.
License	Proprietary
Remarks	
Pilots	P01, P11, P12

2.14 Polyglot Database Management System

Attribute	Value
Title	Polyglot Database Management System
Description	The LXS DBMS is a polystore database that provides access to different and heterogeneous datastores via a common interface. It allows for the data user to submit a query, whose scan operators can request data that are stored in external datastores, and combine their intermediate results with data coming from other sources, either LXS internal datastore or others. For instance, a JOIN operator might require to JOIN table A (resigned in LXS) and table B (resigned in a MongoDB or a Hadoop DataLake). At this phase, there is a support for a limited target datastores, for Proof-of-Concept of the prototype. Moreover, the user has to write queries for the target datastores in the specific dialect. What is more, JOIN operations are not efficient, as they require all data resigned in an external datastore to be retrieved in the query engine level.
Documentation	There is no public documentation available
Company	LXS
Trl	4->7
Ideas	The enhancements that are planned to be implemented are the following: 1) Provide support for a variety of different datastores, according to the needs of the pilots. 2) introduce a novel SQL-like query language, so that the data user can write a query in a seamless way, and let the polystore interpret it to the target datastore. 3) improve the query engine in order to take into account scan operations (mostly part of the one of the JOIN arguments) that need to retrieve data from an external datastore. This will require the query optimizer to explore equivalent operation graphs based on the nature of the target polystore, the operator to be able to push down operations to the target store, and the query processor to set up the corresponding data pipelines during the execution of the query plan. It is not clear at this moment, if the corresponding wrappers/connectors for each target datastore will be able to retrieve statistics to feed the query optimizer. This depends on the target datastores that will be used by the pilots, and their capabilities to expose this type of information
License	Proprietary
Remarks	
Pilots	P13

2.15 HTAP Database for the financial and insurance sector

Attribute	Value
Title	HTAP Database for the financial and insurance sector

Description	An ultra-scalable SQL Database and real-time big data platform that revolutionize the business database management systems by introducing the next generation business database that can scale in any of the three Vs of Big Data (Volume, Velocity and Variety). In more technical details, it provides an ultra-scalable transactional management system that can scale out to 100s of nodes, which is typically a bottleneck in traditional database systems that provide transactional semantics, while on the same time, is full SQL compatible and ensures all ACID transactional properties. It additionally exposes an interface for direct access of its key-value storage engine, thus providing a dual access without downgrade transactional semantics. It offers OLTP and OLAP integration, thus providing support for HTAP that allows for analytical queries over operational data, which realizes the concept of real time business intelligence. Finally, it enables for the execution of polyglot query processing across different and heterogeneous data sources. Modules of the database will be used for implemented all building blocks of INFINITECH that are related to data management.
Documentation	There is no public documentation available
Company	LXS
Trl	4-6->7-8
Ideas	LXS background technology will be enhanced in order to support the data management building blocks: Mainly, it will be enhanced in order to comply with the requirements for HTAP support, and to be compliant with the target data sources that need to be accessed via the polyglot mechanism. Moreover, it will provide support for real time query processing, enabling queries that combine both streaming data with data at rest. Finally, it will provide the support for incremental and parallel query analytics. However, at this phase of the project, it is not yet defined the exact technical details regarding these enhancements.
License	Proprietary
Remarks	
Pilots	P05b, P13

2.16 Natural Language Processing for real-time, high-accuracy Credit Risk assessment

Attribute	Value
Title	Natural Language Processing for real-time, high-accuracy Credit Risk assessment
Description	ReportBrain's NLP functionalities: -Provide real-time structured feeds on risk-assessment worthy information sourced from the news -Interlink entities, updates and maintains knowledge graphs in real-time with all the interlinking of entities -Use "visual" algorithms to collect and analyse the news in real-time in 65 languages 24/7 -Classify articles in real-time by their content (politics, business etc.) -Use its own models, identifies entities (organizations, persons and locations) in real-time
Documentation	Public Documentation under development
Company	RB
Trl	3->8
Ideas	By using the AI enhancements that will be developed by Reportbrain in financial services & insurance sectors, users will be able to add an extra, yet orthocanonical feed to their existing credit rating models that will provide a real-time understanding of the world and more specifically on what's if happening with specific entities (organizations & persons) of interest.
License	Proprietary
Remarks	
Pilots	P13

2.17 Machine learning algorithms for health related data

Attribute	Value
-----------	-------

Title	Machine learning algorithms for health related data
Description	SILO has implemented such approaches in different health-related projects such as CrowdHEALT
Documentation	There is no public documentation available
Company	SILO
Trl	6->7
Ideas	SILO to make suggestions for the enhancement of the Platform
License	Proprietary
Remarks	
Pilots	P12

2.18 Wenalyze Big data analytics platform

Attribute	Value
Title	Wenalyze Big data analytics platform
Description	Platform that collect and process information from multiple open data sources regarding SMEs and apply cognitive algorithm to detect risk and changes in financial needs. The tool will be use in pilot 13
Documentation	There is no public documentation available
Company	WEA
Trl	
Ideas	As far as I understand, there won't be any relevant enhancements of the tool in T2.3.
License	Proprietary
Remarks	
Pilots	P13

2.19 Octopush geospatial enabling framework

Attribute	Value
Title	Octopush geospatial enabling framework
Description	Octopush is a geospatial enabling framework, developed by AgroApps, allowing the collection, pre-processing, post-processing and distribution of geospatial data products and services, either referred to remote sensing (satellite, drones) acquisitions or multidimensional data outputs from numerical simulations. Octopush allows users to have access through a centralized access point to decentralized services, while Octopush SDK enables IT developers, to easily adapt or expand the provided geospatial services. Octopush was created by AgroApps aiming to adders the company/operations and services need for a modular system, independent from any third-party service provider (excluding those offering raw data like Copernicus, NASA etc.), a framework that will be easily adapted to the market needs and follow the service-oriented business model of the company. Octopush is the baseline framework that addresses the AgI companies need in geospatial information, either through the development of new services and data models or the adaptation of the existing ones. Some of the services that are currently offered by Octopush are: a significant number of Vegetation Indices derived from Optical and SAR imagery and crop specific biophysical parameters including leaf area index, chlorophyll content and above ground biomass; crop specific yield estimation; farm management information services such as irrigation scheduling and variable rate fertilization; weather-driven models of possible pests and diseases outbreaks; high resolution weather forecasts and specific agrometeorological parameters; crop damage assessment services.
Documentation	https://api.agroapps.gr/
Company	AGRO

Trl	9
Ideas	n/a
License	Proprietary
Remarks	
Pilots	P14

2.20 AgroApps Weather Intelligence Engine

Attribute	Value
Title	AgroApps Weather Intelligence Engine
Description	Weather Intelligence Engine, developed by AgroApps, is a numerical weather prediction and atmospheric data assimilation processing chain, based on the WRF numerical weather prediction model. Weather Intelligence Engine is producing operationally all the needed weather data products (Near-real-time, medium-range weather forecasting, subseasonal to seasonal forecasts) by AgroApps offered services
Documentation	https://api.agroapps.gr/
Company	AGRO
Trl	9
Ideas	Weather Intelligence Service could take advantage of the available INFINITECH HPC resources, and pilot test a hybrid ensemble data assimilation scheme in convective scales.
License	Proprietary
Remarks	
Pilots	P14

2.21 Sentiment Analysis Tool

Attribute	Value
Title	Sentiment Analysis Tool
Description	Reportbrain Sentiment analysis tool uses application programming interface (API) calls to search existing news article index - Elastic Search Index. The results of the search are processed in Reportbrain's Sentiment Analyzer and the outcome is returned to the caller as a REST API response. This means that articles that are requested by an authorized caller are evaluated in real-time for sentiment and returned to the API caller. Sentiment evaluation describes sentiment as 0 for neutral, -1 for negative or +1 for positive.
Documentation	Public documentation under development
Company	RB
Trl	5 -> 8
Ideas	The purpose of the RB News Article Sentiment API is to provide the sentiment of articles selected by the user via a Query. In particular, the RB News Article Sentiment API provides several fields, that could be used as filters to query on RB platform. These are: content, language and date. The user can retrieve the sentiment analysis to personalize portfolios of their clients, using valuable insights of positive/negative/neutral evaluation of news articles about the entity of interest.
License	Proprietary
Remarks	
Pilots	P04 P06

2.22 Partitioned and Distributed Transaction Graphs

Attribute	Value
Title	Partitioned and Distributed Transaction Graphs
Description	Ethereum and bitcoin public blockchain transaction datasets.
Documentation	BOUN to provide links / docs that can be shared
Company	BOUN
Trl	4->7
Ideas	BOUN to make suggestions for the enhancement of the Platform
License	Public
Remarks	
Pilots	P09

2.23 ALIDA: A Micro-service based platform for composition, deployment and execution of BDA applications

Attribute	Value
Title	ALIDA: A Micro-service based platform for composition, deployment and execution of BDA applications
Description	A micro-service based platform for composition, deployment, optimisation, execution and monitoring of big data analytics workflows (covering ingestion, preparation, analysis and visualization). It is designed and developed on top of the most cutting-edge open source Big Data technologies and framework.
Documentation	https://home.alidalab.it/
Company	ENG
Trl	6->7
Ideas	ENG to make suggestions for the enhancement of the Platform
License	Proprietary
Remarks	
Pilots	P10

2.24 Text Analysis Tool

Attribute	Value
Title	Text Analysis Tool
Description	Reportbrain text analysis tool generates insights from both structured, semi-structured and unstructured text data using natural language processing (NLP). Such insights include sentiment analysis, key phrases, language, and entities, among others. The Reportbrain Analytics Engine uses advanced parallel processing and combines complex NLP tasks in real-time to produce desired results.
Documentation	Public documentation under development
Company	RB
Trl	5 -> 8
Ideas	Text analysis tool allows companies to better understand all types of data they are interested in. After deep analysis of data lake performed by Text Analysis tool, user gain valuable insights about entity that was used in a query. Performing the same process manually would require tremendous amount of effort and time. The final objective for using the tool is that knowledge is provided and not-relevant information are ignored.
License	Proprietary

Remarks	
Pilots	P01

2.25 Blockchain tokenization

Attribute	Value
Title	Blockchain tokenization
Description	Hyperledger Fabric blockchain support for tokens
Documentation	https://www.ibm.com/blockchain/hyperledger ; D4.10 - Blockchain Tokenization and Smart Contracts - I
Company	IBM
Trl	9->9
Ideas	Enhance Hyperledger Fabric with tokenization capabilities for digital trading of assets
License	open source
Remarks	
Pilots	

2.26 Healthentia LifeSciences BigData Platform

Attribute	Value
Title	Healthentia LifeSciences BigData Platform
Description	BigData platform providing data sources aggregation and management, as well as tools for analytics and visualization. It will be used in the IoT-based Life Insurance pilot (ARI/RRD Lead pilot).
Documentation	https://healthentia.com/
Company	ISPRINT
Trl	6->7
Ideas	Re-purpose the platform to support the insurance-related monitoring of people for pilot 12
License	Proprietary
Remarks	
Pilots	P12

2.27 Event-registry

Attribute	Value
Title	Event-registry
Description	Event Registry (ER) is real-time cross-lingual global media monitoring service for modelling global social dynamics (eventregistry.org) developed by the JSI. ER aggregates and analyses news content for over 120,000 news sources published globally in 100+ languages. Events mentioned in the news are identified and relevant information about them is automatically extracted and stored in a searchable form. The data can be accessed directly on the platform or via the API. ER supports various analytics including deep analytics of the events and correlations between events, extracted entities and financial data extracted from the main financial indexes.
Documentation	http://eventregistry.org/
Company	JSI
Trl	9

Ideas	Eventregistry could be enriched with new insights of the potential use scenarios and would benefit from them to expand the current offering to the fintech industry. In addition, better insight will be given to the team in order to develop further analytics.
License	Proprietary
Remarks	
Pilots	P08

2.28 Stream Story

Attribute	Value
Title	Stream Story
Description	Stream-Story multi-resolution modelling and explanation of (possibly real-time) streaming data: (1) Exploratory data mining - A system for the analysis of multivariate time series. It computes and visualizes a hierarchical Markov chain model which captures the qualitative behaviour of the systems' dynamics.; (2) Multi-scale representation -The hierarchical model allows users to interactively find suitable scales for interpreting the data; (3) Real-time monitoring Visualizes streaming data by mapping it to the hierarchical model. It can provide predictions and alarms for different behaviour.
Documentation	http://streamstory.ijs.si/
Company	JSI
Trl	4
Ideas	
License	Proprietary
Remarks	
Pilots	P08

2.29 Qminer

Attribute	Value
Title	Qminer
Description	QMiner is an analytics platform for large-scale real-time streams containing structured and unstructured data. It is designed for scaling to millions of data points on high-end commodity hardware, providing efficient storage, retrieval and analytics mechanisms with real-time response.
Documentation	https://github.com/qminer/qminer
Company	JSI
Trl	7->9
Ideas	
License	BSD licence
Remarks	
Pilots	P08

2.30 SSC - Super Stream Collider, a multiformat Data Management and Query System

Attribute	Value
Title	SSC - Super Stream Collider, a multiformat Data Management and Query System

Description	The SSC enables distributed cloud- based high-performance processing of semantically linked streams i.e. it is an enabler for semantic analytics. It will be used for analytics over semantically unified/interoperable streams (in WP5), as well as in the KYC and customer-centric services pilot (BOI-led pilot).
Documentation	NUIG to provide links / docs that can be shared
Company	NUIG
Trl	6->7
Ideas	NUIG to make suggestions for the enhancement of the Platform
License	Proprietary
Remarks	
Pilots	P03

2.31 The Global Engine with Neural Network Intelligence (GENNI)

Attribute	Value
Title	The Global Engine with Neural Network Intelligence (GENNI)
Description	AI Engine executing DL (deep neural networks) algorithms over semantically annotated streams. It will be used in the pilot for customer centric analytics (BOI led pilot).
Documentation	NUIG to provide links / docs that can be shared
Company	NUIG
Trl	4->7
Ideas	NUIG to make suggestions for the enhancement of the Platform
License	Proprietary
Remarks	
Pilots	P03

2.32 Data Check-in mechanism

Attribute	Value
Title	Data Check-in mechanism
Description	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 the: a) data mapping of data source entities to the designed common data schema, b) data cleaning operations on the data source entities, c) the anonymisation operations on the data source entities and d) the 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.
Documentation	There is no public documentation available at the moment.
Company	UBI
Trl	7
Ideas	

License	Proprietary
Remarks	The specific technology served as the basis for the design and implementation of the INFINITECH Data Collection component
Pilots	

2.33 IoT-Catalogue

Attribute	Value
Title	IoT-Catalogue
Description	The 'IoT Catalogue' is a web-based catalogue from where to pick & choose IoT solutions; it is an explorer of innovations in IoT applications and technologies. It aims to be single entry point of support to IoT developers/integrators/advisors/end-users in the process of identifying and selecting IoT technologies (ranging from complete end-to-end solutions to tools and components/parts) but also inspecting a wide set of IoT use-cases, their validations, associated contact persons/organisations, detailed characterisation (value propositions, ICT problems, functions, target, domain), supporting technological solutions, and much more.
Documentation	www.iot-catalogue.com
Company	UNP
Trl	9
Ideas	
License	Proprietary
Remarks	
Pilots	

2.34 Analytics Library

Attribute	Value
Title	Analytics Library
Description	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)
Documentation	https://www.atmosphere-eubrazil.eu/
Company	UPRC
Trl	4 -> 5
Ideas	Update the library to include metadata relevant to security and privacy constraints of the INFINITECH algorithms to be made available through the library
License	Open source
Remarks	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
Pilots	All through the INFINITECH marketplace. P01

2.35 Catalogue of Objects

Attribute	Value
Title	Catalogue of Objects

Description	During the 5GTANGO and MATILDA EU Projects, both aiming to enable the flexible programmability of 5G networks and to devise and realize a radical shift in the development of software for 5G-ready applications, UPRC team contributed to the market platform and to the catalogue of services & functions, respectively. Therefore, the relative outcomes and the solutions developed by UPRC for these projects will be utilized as a baseline for the INFINITECH's marketplace, integrating with UNP's IoT-catalogue. (WP8)
Documentation	https://www.5gtango.eu/ https://www.matilda-5g.eu/
Company	UPRC
Trl	5 -> 6
Ideas	Extend to support a variety of assets (e.g. datasets, models, etc) as well as to support composite assets (i.e. analytics pipelines)
License	Open source
Remarks	The catalogue has formed the baseline for the INFINITECH marketplace. The back-end of the marketplace utilizes and extends the catalogue of objects of 5GTANGO to support a variety of assets being offered through the marketplace.
Pilots	All through the INFINITECH marketplace

2.36 Portfolio Optimization Tool

Attribute	Value
Title	Portfolio Optimization Tool
Description	Privé goal is to automatically construct an optimised portfolio by modelizing the investment advisory and decision process using a certain level of AI procedures. The result is a tailored portfolio for each individual investor. All functionalities shall be also available via API access, which should be a kind of "Fintech-as-a-service" (FaaS).
Documentation	In Development
Company	Privé Services Europe
Trl	7
Ideas	Our target market are all financial services intermediaries who provide advisory and wealth management services. Hence Banks, Insurers, Insurance Brokers, EAMs, Securities and Brokerage firms are the target customers for this solution.
License	Proprietary
Remarks	None
Pilots	P04

2.37 Sentiment Analysis for Financial news

Attribute	Value
Title	Sentiment Analysis for Financial news
Description	This is a REST API taking as input financial news in text form and performs Sentiment Analysis on the given text. Additionally, the utilised model is be retrained periodically retrieving historical news from LXS database. It enables classification of financial news according to their impact (i.e. positive, neutral, negative) on given portfolio.
Documentation	None
Company	JRC, INNOV
Trl	0
Ideas	None
License	JRC & INNOV Proprietary
Remarks	This tool can be used in parallel with other quantitative metrics in order to provide a comprehensive risk assessment in the stakeholders (traders, risk managers etc,)

Pilots	P02
--------	-----

2.38 AI model for VaR prediction

Attribute	Value
Title	AI model for VaR prediction
Description	This component is a REST API which predicts Value-at-Risk and Expected Shortfall of several financial Portfolios, utilising both well-established and innovative techniques. AI model for VaR prediction takes as input both the asset prices and the current trading position, and derives the VaR and ES estimates at 95% and 99% confidence level utilizing three different models. The estimation procedures are repeated every minute to take into account the most recently available data providing risk assessment in (near) real time.
Documentation	None
Company	JRC, INNOV
Trl	6->7
Ideas	None
License	JRC & INNOV Proprietary
Remarks	Our target market are the institutions which are exposed to market risk such as commercial and investment banks, insurance companies and institutional investors.
Pilots	P02

3 INFINITECH components specifications

The Infinittech Reference Architecture (IRA) [Deliverables D2.13, D2.14], defines a set of layers and allows flexible **workflows** out of data processing modules, called **components** that performs specific data transformations. A workflow usually consumes one set of data (sources) to produce another set of data (destinations). A component can be identified by the Input/Output interfaces and the functionality (transformation) provided at the edge. In that respect, a data component can be considered as a black box that can be replaced by any technological implementation that performs at the edge the same transformation.

The figure below depicts the different reference layers of the IRA along with examples of data components. Scope of the section is to identify existing and desired data components that will support the design, the development and deployment of the Pilots' sandboxes.

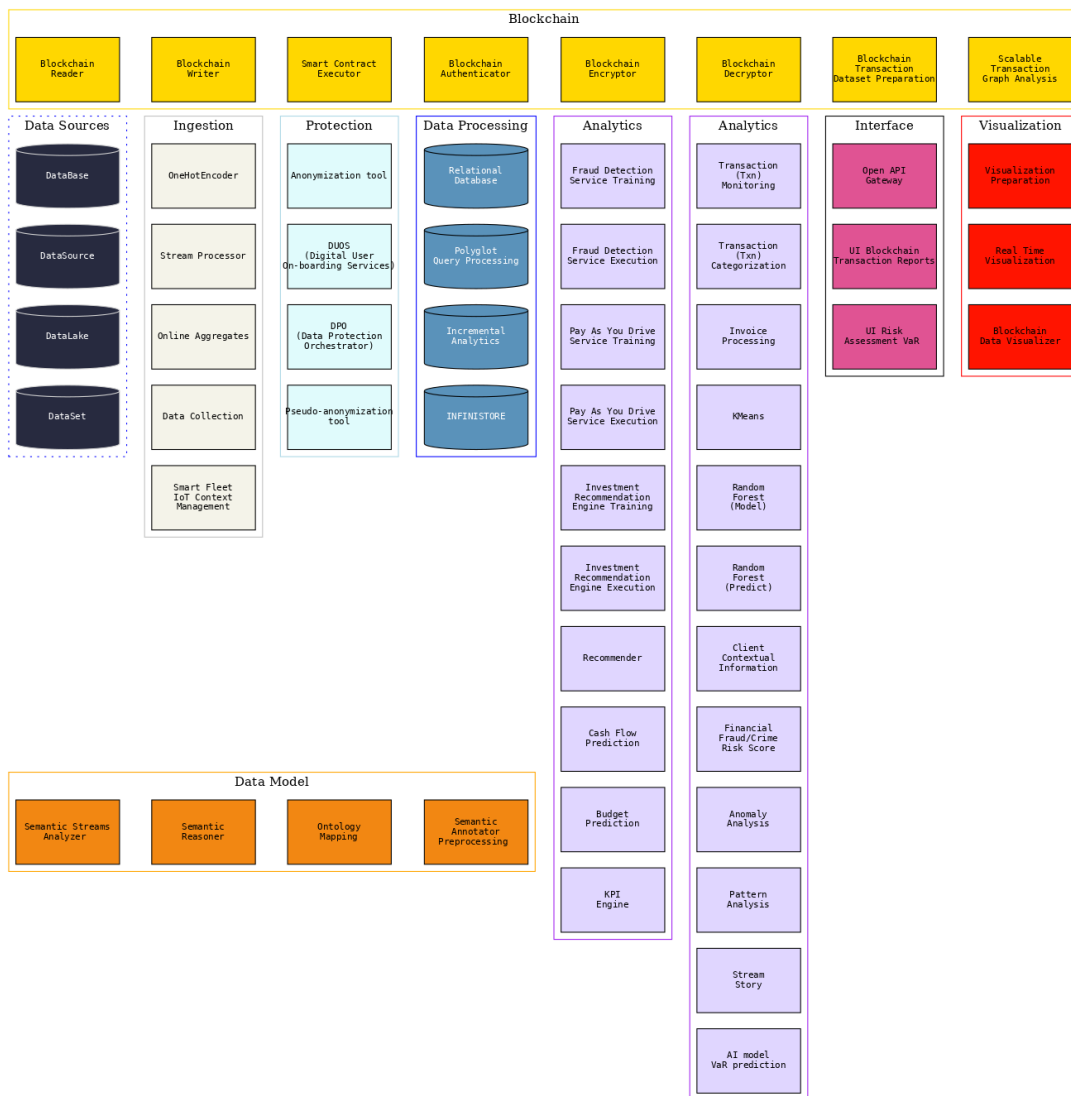


Figure 2 - Logical Schema of Data Processing Components in INFINITECH RA

The components follow predefined colors in order to map components in the different layers.

An INFINITECH solution (sandbox) can be built organizing components in a workflow (sometimes referred to data pipeline) to accomplish a complex transformation from one set of data sources to another set of data solving the business case. Components should therefore be interoperable, with clear interfaces and perform a clear function over the data. The rest of this section finally describes the existing and to be developed components of the project. Further progress in the project will provide the needed details for implementation and deployment.

3.1 Data Component Description

In the following a first list of the identified components, are described using a standard template described in the table below.

3.1.1 Component Template

Attribute	Value
Title	Component Template
Componentid	INF-DCO-001
Description	Short description
Category	Category used
Layer	Infrastructure/Data Management/Data Processing
Input	Description of Input format (File, JSON, ...)
Output	Description of Output format (File, JSON, ...)
Technology	Describe technology used
Core	Yes / No
Marketplace	Yes if it will be part of the Marketplace/ No
Microservice	Yes if it is a dockerized microservice component/ No
Endpoints	If Yes, please provide the Endpoints URLs <code>https://api.example.com/v1/users?role=admin&status=active</code> \ <code>_____</code> / <code>_____</code> / server URL Endpoints query parameters REST API
Owner	Acronym of the Partner that owns the IP
License	Opensource License or Closed
Remarks	E.g. Project's official website
Documentation	Link to reference documentation if any
Pilots	E.g. P01, P02, ...

3.1.2 Relational Database

Attribute	Value
Title	Relational Database
Componentid	INF-DMA-108
Description	This component consists in the central data repository of the platform. It enables transactional semantics and provides capabilities for query processing based on standard SQL statements. It can scale out on the runtime while continuing serving operational workloads. It can support analytical processing in combination with operational data modifications with the level of isolation to be snapshot isolation. That is, it enables real-time business analytics.
Category	Data Management
Layer	Data Processing
Input	A JDBC connection needs to be established in order to execute statements for storing the data.
Output	A JDBC ResultSet or a key-value JSON, based on the OData specification.
Technology	The data repository is written in Java and C, and provides support for JDBC, ODBC and python drivers. It runs on K8S cluster.
Core	Yes
Marketplace	Yes
Microservice	Yes

Endpoints	No
Owner	LXS
License	This component is under proprietary rights of LXS.
Remarks	
Documentation	N/A
Pilots	P13

3.1.3 Polyglot Query Processing

Attribute	Value
Title	Polyglot Query Processing
Componentid	INF- DMA -109
Description	This component enables the query execution over more than one datastore in seamless manner. The data user can submit a single statement and let this component to execute the query by pushing it down to the target database. By doing this, it can process the data on premise, and retrieve only the results, thus it is convenient for cases where data cannot be loaded the platform and they need to be accessed from an external datastore.
Category	Data Management
Layer	Data Processing
Input	This component needs to open a connection with the target database. It will require different input according to the number of target stores that will be supported.
Output	A JDBC ResultSet or a key-value JSON, based on the OData specification.
Technology	This component is implemented in Java. It is an extension of the central repository itself. In fact, it consists of a jar binary which is loaded in the classpath of the query engine of the data repository. As of that, it will be deployed as part of the query engine via K8S cluster.
Core	Yes
Marketplace	Yes
Microservice	Yes
Endpoints	No
Owner	LXS
License	This component is under proprietary rights of LXS.
Remarks	
Documentation	N/A
Pilots	P13

3.1.4 Incremental Analytics

Attribute	Value
Title	Incremental Analytics
Componentid	INF- DMA -110
Description	This component enables the query execution in an incremental fashion. The data user will be able to submit a continuous query to the datastore, which will be continuously and incrementally validated. This means that the initial results will be retrieved first, and as data arrives to the data repository, they will be validated against the submitted query, and if it validates the statement, it will be returned to the user.
Category	Data Management
Layer	Data Processing
Input	A JDBC connection needs to be established in order to execute statements for storing the data.

Output	To be defined
Technology	This component is implemented in Java. It is an extension of the central repository itself. In fact, it consists of a jar binary which is loaded in the classpath of the query engine of the data repository. As of that, it will be deployed as part of the query engine via K8S cluster.
Core	Yes
Marketplace	Yes
Microservice	Yes
Endpoints	No
Owner	LXS
License	This component is under proprietary rights of LXS.
Remarks	
Documentation	N/A
Pilots	P02

3.1.5 OneHotEncoder

Attribute	Value
Title	OneHotEncoder
Componentid	INF- DMA -113
Description	Service to represent categorical variables as binary vectors
Category	Data Management
Layer	Processing
Input	Any CSV or XLS or ORC. Dataset stored into HDFS/PRESTO/HIVE
Output	Any CSV or XLS or ORC. Dataset stored into HDFS/PRESTO/HIVE
Technology	It is a BDA service registered in the ALIDA catalogue. It is a PySpark-based micro-service running on K8S Spark cluster mode, working as part of the ALIDA framework.
Core	This component is general purpose and reusable as a BDA service within ALIDA
Marketplace	No
Microservice	Yes (it is a dockerized microservice component)
Endpoints	No (it is not a standalone service that waits for incoming requests but it is closely related to the lifecycle of an ALIDA workflow.)
Owner	ENG
License	Component developed with open source technologies. ENG Proprietary.
Remarks	In a nutshell, ALIDA is a Micro-service based platform for composition, deployment, optimisation, execution and monitoring of pipelines of Big Data Analytics (BDA) services. ALIDA is a result of previous research activities developed by ENG. Currently, it is a work in progress. ALIDA offers a catalogue of BDA services (ingestion, preparation, analysis, visualization): user designs his own (stream/batch) pipeline by choosing the BDA services from it, indicates which Big Data set he wants to process, launches and monitors the execution of the pipeline and personalizes the results visualization by choosing from a set of available graphs, all this without worrying about having software developer skills or particular knowledge on big data technologies. This service is registered in ALIDA catalogue as Spring Boot Application containing the python code and its dependencies. After implementing the algorithm using Pyspark, creating the Dockerfile and pushing the new image inside a repository, this microservice is registered into the ALIDA catalogue through the GUI. Source: https://home.alidalab.it/
Documentation	N/A
Pilots	P10

3.1.6 Stream Processor

Attribute	Value
Title	Stream Processor
Componentid	INF- DMA -115
Description	This component will provide streaming processing capabilities. The data user can declare continuous queries that will be executed over the data stream. It will also allow to combine streaming data with data at-rest and also enable the storage of data streams even when injected in very high rates.
Category	Data Management
Layer	Data Processing
Input	This component receives data from a stream and can consume data stored at rest in the data repository
Output	Depends on the pilot
Technology	This component will be based on Apache Flink and will be containerized in order to be deploy with a K8S cluster.
Core	Yes
Marketplace	Yes
Microservice	Yes
Endpoints	No
Owner	LXS
License	This component is under proprietary rights of LXS.
Remarks	This component can exploit the capabilities of the declarative real time analytics. This will be very useful in cases the data user wants to calculate a value over a stream that will need to perform an expensive scan operation over a data table (i.e. compare the input with the overall average of a field in a table). As scan operations (and operations that will require a scan operation, as the average etc) have a complexity bigger then $O(1)$, they are timely costly and cannot be executed in a stream. For that, the developer often caches that value and periodically updates the value. With the declarative real-time analytics, the data user is given the opportunity to declare such an analytical operation (i.e. the overall average) with an SQL fashion, and the query will be executed with a complexity of a get operation, which will allow these types of analytical operations to be included in a stream operation, providing to it the current average, with respect to data consistency and isolation in terms of ACID properties and transactional semantics.
Documentation	N/A
Pilots	

3.1.7 Online Aggregates

Attribute	Value
Title	Online Aggregates
Componentid	INF-DMA-108
Description	This component allows for the execution of aggregate processing operators in an online manner. This way, the definition of the aggregate operations can be defined, and the result of the execution can be pre-calculated in an online manner, preserving data consistency and transactional semantics. When requested, the result of this execution can be retrieved with a GET operation, removing the need to scan the whole dataset/
Category	Data Management
Layer	Data Processing
Input	A JDBC connection needs to be established in order to execute statements for storing the data.
Output	A JDBC ResultSet or a key-value JSON, based on the OData specification.

Technology	This component is based on the relation database component, and extends its core storage and query engine. It runs on K8S cluster.
Core	Yes
Marketplace	Yes
Microservice	Yes
Endpoints	No
Owner	LXS
License	This component is under proprietary rights of LXS.
Remarks	
Documentation	N/A
Pilots	P02

3.1.8 Data Collection

Attribute	Value
Title	Data Collection
Componentid	INF- DMA -116
Description	This component provides the data ingestion mechanism that:a) enables the acquisition and retrieval of heterogeneous data from a variety of diverse data sources and data providers,b) facilitates the data annotation of the retrieved data by enabling the mapping between the data entities included in the retrieved data and the provided by the data provider data model,c) enables the design and execution of data cleaning operations towards the increase of the data quality of the retrieved data.
Category	Data Management
Layer	Data Management
Input	Depending on the provided configuration, this component retrieves new data from: a) a data source's API, b) a FTP or HTTP server, c) a Relational Database, d) an HDFS deployment or MinIO storage server. Additionally, the components depending on the provided configuration is able to receive new data from a data source through its well-defined RESTful APIs.
Output	Depending on the provided configuration, the component provides the locally stored retrieved data for ingestion in the Relational Database provided by LXS. If the optional step of data annotation is configured, the mapping is also produced as a JSON file. Additionally, if the optional step of data cleaning is configured, the "cleaned" data are also made available for ingestion into the Relational Database provided by LXS.
Technology	This component is based on Java and Python programming languages. Additionally, Java Spring Boot, Flask, Pandas and NumPy frameworks and libraries are leveraged.
Core	Yes
Marketplace	Yes
Microservice	Yes
Endpoints	https://{{SERVER_URL}}:{{SERVER_PORT}}/handler/api/v1/pullhttps://{{SERVER_URL}}:{{SERVER_PORT}}/handler/api/v1/pushhttps://{{SERVER_URL}}:{{SERVER_PORT}}/mapper/api/v1/maphttps://{{SERVER_URL}}:{{SERVER_PORT}}/cleaner/api/v1/clean/workflow
Owner	UBI
License	This component is under proprietary rights of UBI.
Remarks	This component is providing a highly configurable mechanism capable of addressing the various connectivity and communication challenges raised during data ingestion. Hence, the data provider is able to configure this mechanism in order to execute data collection pipelines that include data retrieval, data annotation and data cleaning operations which are tailored to their needs.
Documentation	To be included in D5.13
Pilots	P06, P12

3.1.9 Anonymization tool

Attribute	Value
Title	Anonymization tool
Componentid	INF- DPR -117
Description	The anonymization tool modifies data in order to preserve privacy. It is especially indicated in those cases where a dataset contains personal data and it has to be outsourced or shared with a third party. The tool includes different anonymization algorithms that aim at avoiding the appearances of data combinations that could lead to a possible re-identification of the data subjects. 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.
Category	Data Protection
Layer	Data
Input	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.
Output	It stores the anonymized data in a database. The location of the database has to be known beforehand (through the configuration file that is taken as an input).
Technology	The tool is based on two modules that can be deployed as a Docker container. The tool requires to retrieve and store the raw and anonymized version of the data from a relational database.
Core	Yes
Marketplace	Yes
Microservice	Yes
Endpoints	TBD
Owner	GRAD
License	Closed
Remarks	Even though the tool will be part of the INFINITECH platform, it will not be available for all the pilots, as it needs a specific configuration (and probably development) for each specific use case.
Documentation	N/A
Pilots	P11 P12

3.1.10 DUOS (Digital User On-boarding Services)

Attribute	Value
Title	DUOS (Digital User On-boarding Services)
Componentid	INF- DPR -118
Description	Provides remote user registration using eID or passport, dealing with virtual identities in a mobile device. It uses various identity proofing and verification services that link new user eID creation (virtual or derived eID) with government issued e-ID. It verifies electronic data stored on chip and machine readable zone. Provides Flexible Multi-factor authentication for different users or identities
Category	Data Protection
Layer	Infrastructure
Input	<ul style="list-style-type: none"> • For Virtual Identity Creation operation: there will be a virtual identity creation request, the data will be the eID or passport (including MRZ zone and chip) and also the user will complete a form to confirm his/her identity. • For authentication operation: there will be an authentication request and the user will select the identity which will be used to perform the authentication.
Output	<ul style="list-style-type: none"> • For Virtual Identity Creation operation: Virtual Identity creation result. • For authentication operation: Authentication result and return to the application that called DUOS.

Technology	Android
Core	Yes
Marketplace	Yes
Microservice	No
Endpoints	Not applicable
Owner	Atos
License	Different underlying licenses (Apache 2.0 license for MRZ reader and license to be decided for biometric checking)
Remarks	
Documentation	
Pilots	P04

3.1.11 DPO (Data Protection Orchestrator)

Attribute	Value
Title	DPO (Data Protection Orchestrator)
Componentid	INF- DPR -119
Description	The DPO embeds and automats the assurance of security and privacy by design and by default in complex business flows. It provides orchestration of Privacy Enhancing Technologies and related services using BPM tools in order to integrate privacy or data protection perspectives into business processes. It can orchestrate any kind of REST service.
Category	Data Protection
Layer	Infrastructure
Input	REST call to the DPO and input data in JSON
Output	Result of the execution in JSON
Technology	The DPO interacts with the privacy expert that prepares the business flow in a BPMN file. The flow interacts with PETs such as anonymization
Core	Yes
Marketplace	Yes
Microservice	Yes
Endpoints	It will be provided during the deployment stage.
Owner	Atos
License	Apache 2.0
Remarks	
Documentation	
Pilots	P12

3.1.12 Blockchain Reader

Attribute	Value
Title	Blockchain Reader
Componentid	INF-BKC-120
Description	Fetches requested data from blockchain ledger. The specific component is part of a Blockchain chaincode. As the chaincode is tightly connected to the business operation that is performed on top of different business objects, different flavours of the chaincode exist depending on the business use cases (Consent Management, Know Your Customer / Know Your Business, Asset Tokenization).
Category	Blockchain
Layer	Infrastructure/Data

Input	A JSON file the requested data details to fetch (query).
Output	A JSON file Including any requested block information (stored in the ledger) that can be retrieved from the blockchain network and the transactions hashes, to verify that they were successfully validated by the blockchain network and stored on the corresponding ledger
Technology	Hyperledger
Core	Yes
Marketplace	No
Microservice	Yes
Endpoints	No
Owner	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
License	Apache 2.0
Remarks	Project's official website: https://www.hyperledger.org/ License: https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)
Documentation	Included in deliverable D4.7. Additional resources: https://hyperledger.github.io/ https://github.com/hyperledger/fabric https://wiki.hyperledger.org/display/fabric https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html https://www.hyperledger.org/projects/fabric https://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html https://github.com/hyperledger/blockchain-explorer https://wiki.hyperledger.org/display/explorer Documentation of Input and Output Parameters: https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html
Pilots	

3.1.13 Blockchain Writer

Attribute	Value
Title	Blockchain Writer
Componentid	INF- BKC -121
Description	Submits transactions on the blockchain ledger. The specific component is part of a Blockchain chaincode. As the chaincode is tightly connected to the business operation that is performed on top of different business objects, different flavours of the chaincode exist depending on the business use cases (Consent Management, Know Your Customer / Know Your Business, Asset Tokenization).
Category	Blockchain
Layer	Infrastructure/Data
Input	A JSON fileIncluding the necessary data for the transaction submission.
Output	A JSON file including the transactions hashes, to verify that they were successfully validated by the blockchain network and stored on the corresponding ledger,(optional) the submitted data just for validation of correct input (the just now stored data in the ledger after it was approved by the consensus mechanism)
Technology	Hyperledger
Core	Yes
Marketplace	No
Microservice	Yes
Endpoints	No
Owner	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
License	Apache 2.0
Remarks	Project's official website: https://www.hyperledger.org/ License: https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)

Documentation	Included in deliverable D4.7. Additional resources: https://hyperledger.github.io/https://github.com/hyperledger/fabrichttps://wiki.hyperledger.org/display/fabrichttps://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.htmlhttps://www.hyperledger.org/projects/fabrichttps://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.htmlhttps://github.com/hyperledger/blockchain-explorerhttps://wiki.hyperledger.org/display/explorer Documentation of Input and Output Parameters: https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html
Pilots	

3.1.14 Smart Contract Executor

Attribute	Value
Title	Smart Contract Executor
Componentid	INF- BKC -122
Description	Executes smart contracts on the blockchain ledger. The specific component is part of a Blockchain chaincode. As the chaincode is tightly connected to the business operation that is performed on top of different business objects, different flavours of the chaincode exist depending on the business use cases (Consent Management, Know Your Customer / Know Your Business, Asset Tokenization).
Category	Blockchain
Layer	Infrastructure/Data
Input	A JSON filethe chaincode or smart contract specific input required to submit transactions and deploy smart contracts.
Output	A JSON file including the smart contracts' transactions hashes, to verify that they were successfully deployed on the respective blockchain network,(optional) the submitted data just for validation of correct input (the just now stored data in the ledger after it was approved by the consensus mechanism)
Technology	Hyperledger
Core	Yes
Marketplace	No
Microservice	Yes
Endpoints	Consent Management: https://{{SERVER_URL}}:{{SERVER_PORT}}/api/createconsenthttps://{{SERVER_URL}}:{{SERVER_PORT}}/api/readconsenthttps://{{SERVER_URL}}:{{SERVER_PORT}}/api/queryconsentbycustomerhttps://{{SERVER_URL}}:{{SERVER_PORT}}/api/QueryConsentByCustomerAndDataControllerhttps://{{SERVER_URL}}:{{SERVER_PORT}}/api/GetConsentHistoryhttps://{{SERVER_URL}}:{{SERVER_PORT}}/api/QueryConsentByControllerID Know Your Customer / Know Your Business: https://{{SERVER_URL}}:{{SERVER_PORT}}/createPartyhttps://{{SERVER_URL}}:{{SERVER_PORT}}/readPartyhttps://{{SERVER_URL}}:{{SERVER_PORT}}/requestAccesshttps://{{SERVER_URL}}:{{SERVER_PORT}}/enabledAccesshttps://{{SERVER_URL}}:{{SERVER_PORT}}/fetchPartiesAssetTokenization:https://{{SERVER_URL}}:{{SERVER_PORT}}/api/account/minthhttps://{{SERVER_URL}}:{{SERVER_PORT}}/api/account/GetBalancehttps://{{SERVER_URL}}:{{SERVER_PORT}}/api/account/GetAllowancehttps://{{SERVER_URL}}:{{SERVER_PORT}}/api/account/transferhttps://{{SERVER_URL}}:{{SERVER_PORT}}/api/account/approvehttps://{{SERVER_URL}}:{{SERVER_PORT}}/api/account/transferFrom
Owner	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
License	Apache 2.0
Remarks	Project's official website: https://www.hyperledger.org/ License: https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)

Documentation	Included in deliverable D4.7. Additional resources: https://hyperledger.github.io/ https://github.com/hyperledger/fabrichttps://wiki.hyperledger.org/display/fabric https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html https://www.hyperledger.org/projects/fabrichttps://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.htmlhttps://wiki.hyperledger.org/display/fabrichttps://github.com/hyperledger/fabric-chaincode-nodehttps://github.com/hyperledger/blockchain-explorerhttps://wiki.hyperledger.org/display/explorer Documentation of Input and Output Parameters: https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html
Pilots	

3.1.15 Blockchain Data Visualizer

Attribute	Value
Title	Blockchain Data Visualizer
Componentid	INF- BKC -123
Description	Queries and displays information about blocks, transactions, chaincodes and transaction families, network name, status and nodes list, organizations list and peers list.The specific component is part of a Blockchain chaincode. As the chaincode is tightly connected to the business operation that is performed on top of different business objects, different flavours of the chaincode exist depending on the business use cases (Consent Management, Know Your Customer / Know Your Business, Asset Tokenization).
Category	Blockchain
Layer	Infrastructure/Data
Input	A JSON file including the credentials pair (username & password) that authenticate the user to have access to this component's "Output",the blockchain network name andthe component's URL.
Output	A web pageIncluding statistics of the blockchain network activity in charts and pies andability to view and query blocks, transactions, chaincodes and transaction families, network name, status and nodes list, organizations list and peers list.
Technology	Hyperledger
Core	Yes
Marketplace	No
Microservice	Yes
Endpoints	No
Owner	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
License	Apache 2.0
Remarks	Project's official website: https://www.hyperledger.org/ License: https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)
Documentation	https://github.com/hyperledger/blockchain-explorer https://wiki.hyperledger.org/display/explorer https://medium.com/@thanawitsupinnapong/setting-up-hyperledger-explorer-on-fabric-5f1f7cda73b3 https://hyperledger.github.io/ https://github.com/hyperledger/fabrichttps://wiki.hyperledger.org/display/fabric https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html https://www.hyperledger.org/projects/fabric https://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html https://wiki.hyperledger.org/display/fabric Documentation of Input and Output Parameters: https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html
Pilots	

3.1.16 Blockchain Authenticator

Attribute	Value
Title	Blockchain Authenticator

Componentid	INF- BKC -124
Description	Grants access to specific channel(s) of the blockchain network.The specific component is part of a Blockchain chaincode. As the chaincode is tightly connected to the business operation that is performed on top of different business objects, different flavours of the chaincode exist depending on the business use cases (Consent Management, Know Your Customer / Know Your Business, Asset Tokenization).
Category	Blockchain
Layer	Infrastructure/Data
Input	A JSON file including necessary Hyperledger Fabric channel authentication keys and certificates,organization identifier and its CA certificates andparticipant metadata and other relevant metadata.
Output	A JSON file including a unique token confirming that authentication is permitted by the subject entity at the subject time.
Technology	Hyperledger
Core	Yes
Marketplace	No
Microservice	Yes
Endpoints	No
Owner	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
License	Apache 2.0
Remarks	Project's official website: https://www.hyperledger.org/ License: https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)
Documentation	Included in deliverable D4.7. Additional resources: https://hyperledger.github.io/ https://github.com/hyperledger/fabric https://wiki.hyperledger.org/display/fabric https://hyperledger-fabric.readthedocs.io/en/release-2.0/deployment_guide_overview.html https://www.hyperledger.org/projects/fabrichttps://hyperledger-fabric.readthedocs.io/en/release-2.0/prereqs.html https://wiki.hyperledger.org/display/fabrichttps://github.com/hyperledger/fabric-chaincode-nodehttps://github.com/hyperledger/blockchain-explorer https://wiki.hyperledger.org/display/explorer Documentation of Input and Output Parameters: https://hyperledger-fabric.readthedocs.io/en/release-2.0/txflow.html
Pilots	

3.1.17 Blockchain Encryptor

Attribute	Value
Title	Blockchain Encryptor
Componentid	INF- BKC -125
Description	Encrypts the clients' sensitive data within the smart contract using AES256. The specific component is part of a Blockchain chaincode. As the chaincode is tightly connected to the business operation that is performed on top of different business objects, different flavours of the chaincode exist depending on the business use cases (Consent Management, Know Your Customer / Know Your Business, Asset Tokenization).
Category	Blockchain
Layer	Infrastructure/Data
Input	Type "string" variables: A 32-byte key is required to encrypt the sensitive data. The actual plaintext data that is going to be encrypted.
Output	None
Technology	Hyperledger
Core	Yes
Marketplace	No
Microservice	Yes

Endpoints	No
Owner	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
License	Apache 2.0
Remarks	Project's official website: https://www.hyperledger.org/ License: https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)
Documentation	Included in deliverable D4.7. Additional resources: https://www.solarwindsmisp.com/blog/aes-256-encryption-algorithm https://www.atpinc.com/blog/what-is-aes-256-encryption https://searchsecurity.techtarget.com/definition/Advanced-Encryption-Standard
Pilots	

3.1.18 Blockchain Decryptor

Attribute	Value
Title	Blockchain Decryptor
Componentid	INF- BKC -126
Description	Decrypts data retrieved from the blockchain ledger.
Category	Blockchain
Layer	Infrastructure/Data
Input	Type "string" variables: Each user's specific 32-byte key, the same key that is used during encryption. The ciphertext or encrypted user's data.
Output	A JSON struct: The plaintext message which consists of the user's actual data after the response from the AES256 decryption process.
Technology	Hyperledger
Core	Yes
Marketplace	No
Microservice	Yes
Endpoints	No
Owner	Indicatively: IBM (n/a), UBITECH (n/a), INNOV (n/a)
License	Apache 2.0
Remarks	Project's official website: https://www.hyperledger.org/ License: https://tldrlegal.com/license/apache-license-2.0-(apache-2.0)
Documentation	Included in deliverable D4.7.
Pilots	

3.1.19 Blockchain Transaction Dataset Preparation Component

Attribute	Value
Title	Blockchain Transaction Dataset Preparation Component
Componentid	INF- BKC -127
Description	This component will be responsible for retrieving raw transaction blocks from the Bitcoin and Ethereum blockchains and parsing the blocks in order to extract Bitcoin, Ethereum and major token transactions. After retrieving all the blocks up until now, this component will be run periodically (e.g. once a week or as needed) to retrieve newly generated blockchain blocks during the period.
Category	
Layer	Data Processing Architectures
Input	Bitcoin and Ethereum blockchain raw blocks, Major Ethereum ERC20 Token contract addresses (such as USD, EURO stable coin contract addresses).
Output	Scalable Transaction Graph Analysis

Technology	Blockchain blocks must be retrieved from either nodes attached to the blockchain (e.g. running Parity) or from blockchain data supplier gateways (Google, Infura, Cloudflare). Web3.js Ethereum Javascript API and Consensys abi-decoder for smart contract call data parsing
Core	Will be part of the INFINITECH platform.
Marketplace	Yes
Microservice	TBD
Endpoints	TBD
Owner	BOUN
License	Open source, MIT license
Remarks	N/A
Documentation	N/A
Pilots	P09

3.1.20 Scalable Transaction Graph Analysis Component

Attribute	Value
Title	Scalable Transaction Graph Analysis Component
Componentid	INF- BKC -128
Description	This component will be responsible for taking massive bitcoin and ethereum public transaction data. Since transaction graph size massive and growing, it will use parallel algorithms to achieve scalability. It will utilize graph and machine learning algorithms to analyse fraudulent transactions
Category	
Layer	Data Processing Architectures
Input	Bitcoin and Ethereum blockchain raw blocks, Major Ethereum ERC20 Token contract addresses (such as USD, EURO stable coin contract addresses).
Output	Transaction Graph Analysis Results that will be presented to the User Interface for Blockchain Transaction Reports and Visualization Component
Technology	Parallel processing (high performance computing) technologies. Message Passing Interface (MPI). Graph Partitioning Software such as Scotch and/or Metis. Distributed graph algorithms. Machine Learning.
Core	Will be part of the INFINITECH platform
Marketplace	Yes
Microservice	TBD
Endpoints	TBD
Owner	BOUN
License	Open source MIT License
Remarks	N/A
Documentation	N/A
Pilots	P09

3.1.21 Semantic Streams Analyzer

Attribute	Value
Title	Semantic Streams Analyzer
Componentid	INF-SEM-130

Description	Semantic Streams Analyzer Middleware-Engine - SeSAME, The SeSAME component is a data mashup builder for the financial sector that can be used as a data processing component for your data management application, it enhances the capacity to process financial and insurance data in the form of batches and provides a single output, it is ideal when multiple sources have different data formats, it is built to be compatible with the most common data formats in the financial and Insurance sector i.e. FIBO, FIGI and LKIF and additionally it uses INFINITECH Core Graph Data Model to enhance performance. The SeSAME component is designed as a dataflow/workflow execution framework connecting various data input/ outputs through the concept of pipelines for creating the data mashup. Conceptually, each financial operator has input data or streams and SeSAME provides one output data or stream. The multiple inputs can be used simultaneously while a single output in RDF is provided. Only the final operator of a workflow can return a format other than RDF, if necessary by defining and transforming the data into the desired format. The data Operators can be of three modes via APIs: API 1) a data acquisition operator is used to collect or receive data from data sources or gateways and can be pull-based or push-based. API 2) a stream processing operator defines stream processing functionalities in a declarative language, e.g., CQELS. API 3) a streaming operator streams the outputs of the final operator of a workflow to the consuming applications. In these three API modes operators of the data transformations and alignments can be done to produce a normalised RDF output format.
Category	Semantics & Graph Data Model Tools
Layer	Data Processing Frameworks/Middleware Architectures
Input	Java Database Connectivity (JDBC) providing Historical Data sets but also Streams based on J-SON and RDF
Output	A Java-Based application programming interface (API) with RESTful capabilities based on RDF Results
Technology	JDBC, JAVA
Core	RDF-SPARQL
Marketplace	Yes - As Engine
Microservice	Yes - As Middleware
Endpoints	Pending
Owner	NUIG
License	Under Analysis for an Open Source Licence (e.g. Apache 2.0, EUPL, Eclipse, LGPL V3.0)
Remarks	N/A
Documentation	Documentation is provided in Gitlab
Pilots	P03

3.1.22 Semantic Reasoner

Attribute	Value
Title	Semantic Reasoner
Componentid	INF-SEM-131
Description	Enhanced Distributed Reasoner over FinTech Ontologies - EnDoRFIN Semantic Reasoner, Enhanced Distributed Reasoner over FinTech Ontologies - EnDoRFIN Semantic Reasoner, The EnDoRFIN component is a tool for inferring knowledge from data streams, it uses some rules as conditions for defining logics conditions and as a result logical consequences are provided as outcomes. The inference rules are defined based on the most commonly used financial and insurance vocabularies i.e. FIBO, FIGI and LKIF and the way to process the rules is using APIs for defining the logical descriptions for the data applications it is introduced. This component allow the use of other languages but need to be upgraded to the target vocabulary and additionally the EnDoRFIN uses INFINITECH Core Graph Data Model to ensure the inference is applicable to all the involved domains in FINTECHs.
Category	Semantics & Graph Data Model Tools
Layer	Data Processing Frameworks

Input	Java Database Connectivity (JDBC) Data sets but also Streams based on J-SON and RDF
Output	A Java-Based application programming interface (API) based on SPARQL-RDF Results
Technology	JDBC, JAVA, JENA
Core	JENA API
Marketplace	Yes
Microservice	Yes
Endpoints	Pending
Owner	NUIG
License	Under Analysis for an Open Source Licence (e.g. Apache 2.0, EUPL, Eclipse, LGPL V3.0)
Remarks	N/A
Documentation	Documentation is provided in Gitlab
Pilots	P03

3.1.23 Ontology Mapping

Attribute	Value
Title	Ontology Mapping
Componentid	INF-SEM-132
Description	INFINITECH GRAPH Data Model - Online Ontology Mapping Framework and Toolkit, The INFINITECH GRAPH Data Model is the set of online tools referring to the graphs, formats, vocabularies and ontologies used in the INFINITECH project. The INFINITECH GRAPH Data Model is provided in the form of a set of online accessible files, schemas and metadata model diagrams that represent the way the INFINITECH data can be organised and structured, it also contains the metadata in two different formats, json-ld and owl. The Ontology Mapping Ontologies section contains online machine-readable files both in OWL and JSON-LD format for online accessibility, both files are maintained and updated regularly to keep the latest version of the ontology files up to date using a versioning method..
Category	Semantics & Graph Data Model Tools
Layer	Data Modelling
Input	OWL Data
Output	JSON-LD, OWL, Graphs Data and Ontologies
Technology	OWL, Onlien Web Services
Core	Apache TomCat Web Server
Marketplace	Yes
Microservice	No
Endpoints	Pending
Owner	NUIG
License	Under Analysis for an Open Source Licence (e.g. Apache 2.0, EUPL, Eclipse, LGPL V3.0)
Remarks	N/A
Documentation	Documentation is provided online at INFINITECH GRAPH Data Model Online Framework and Toolkit
Pilots	P03 and All Others

3.1.24 Semantic Annotator-Preprocessing

Attribute	Value
Title	Semantic Annotator-Preprocessing
Componentid	INF-SEM-133

Description	Semantic Annotator-Middleware Preprocessing Layer for FinTechs - SAMPLE-FIN, The INFINITECH SAMPLE-Fin is the support online tool for transforming datasets into RDF-compatible format, beside the online available tool, a set of documentation is provided providing the necessary steps to transform data sets from any data-exchange format i.e. CSV, XLS, etc into RDF. This tool is provided as enabler for a semantic layer where enriched data can be processed more efficiently. INFINITECH SAMPLE-Fin is the mechanism INFINITECH uses for addressing cross domain and cross pilot Data Interoperability and Data Exchange and it also provides the pre-processing layer for the interoperability requirements in INFINITECH project.
Category	Semantics & Graph Data Model Tools
Layer	Data Modelling
Input	Java Database Connectivity (JDBC) Data sets but also Streams based on J-SON and RDF
Output	A Java-Based application programming interface (API) based on SPARQL-RDF Results
Technology	RDF
Core	CSV, XLS
Marketplace	Yes
Microservice	No
Endpoints	Pending
Owner	NUIG
License	Under Analysis for an Open Source Licence (e.g. Apache 2.0, EUPL, Eclipse, LGPL V3.0)
Remarks	N/A
Documentation	Documentation is provided in Gitlab
Pilots	P03 and All Others

3.1.25 Smart Fleet (IoT Context Management and Historical data component)

Attribute	Value
Title	Smart Fleet (IoT Context Management and Historical data component)
Componentid	INF-IOT-133
Description	A FIWARE-Based framework designed to capture, homogenise, process and distribute real time traffic and smart vehicle's information (it will also allow other related context information). It will implement Pub/Sub mechanisms and support Geolocation and Time series tools. Additional tools to build custom dashboards will be included.
Category	
Layer	Data Processing Architecture
Input	Inputs from Data Management and Data protection layers (Nodes providing Traffic and Vehicles datasets):Standard/DataModel: based on NGSI-v2 FIWARE data modelsFormat: JSONProcotols: HTTP/HTTPS/MQTT
Output	Provides information to Data Analytics components, as the AI framework (EASIER.AI), Real Time analytics nodes and End-user dashboards.Standard/DataModel: based on NGSI-v2 FIWARE data modelsFormat: JSONProcotols: HTTP/HTTPS/MQTT
Technology	Customised technological building blocks based on FIWARE architecture:Context Broker (Orion) [NGSI & ETSI NGSI-LD]: distribute context data between any possible combination of data-producers and data-consumersHistorical Data (Quantumleap) [ETSI NGSI-LD]: persistence of the data managed by the Context Broker.
Core	Offered as part of the INFINITECH platform
Marketplace	Yes
Microservice	Yes
Endpoints	End point to be provided during deployment stage. NGSIv2/NGSI-LD API supported [https://fiware.github.io/specifications/ngsiv2/stable/]
Owner	ATOS

License	Context Broker – Orion: AGPL-v3MongoDB Database Server and Tools - Server Side Public License (SSPL)Historical Data - Quantumleap: MIT LicenseCrateDB – Apache License V2.0IoT-Agents: It will depend on the agent, but these are encouraged to use AGPL. In any case, OpenSource IoT Agent
Remarks	It will be composed by three main technological building blocks:Context Broker (Orion) [NGSI & ETSI NGSI-LD]: distribute context data between any possible combination of data-producers and data-consumersHistorical Data (Quantumleap) [ETSI NGSI-LD]: persistence of the data managed by the Context Broker.IoT-Agents [NGSI, MQTT, HTTP, UL2.0, JSON]: act as an intermediate between a data-producer (IoT, data-sources, streams) and the Context Broker. Translate data and protocols to the common data model and protocol of the Context Broker. Different IoT-Agents will be used depending on the protocols and data to manage as inputs.
Documentation	N/A
Pilots	Pilot 11

3.1.26 Fraud Detection Service Training

Attribute	Value
Title	Fraud Detection Service Training
Componentid	INF- MLE -135
Description	This component trains the corresponding Machine Learning model that will assign a driver's profile and helps to identify the driver's behaviour using the provided driving routes and vehicle's technical data.
Category	Analytics & ML Algorithms.
Layer	Analytics
Input	A historic dataset of drivers' data captured on each route they drive. Related context information (weather, traffic incidents, etc.)
Output	A trained model that will classify the drivers' way of driving on each route.
Technology	TensorFlow, Keras, Docker, SKLearn
Core	Pilot specific (for Pilot 11 "Personalized insurance products based on IoT connected vehicles").
Marketplace	Yes
Microservice	Yes
Endpoints	TBD
Owner	ATOS
License	ATOS Proprietary.
Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	Pilot 11

3.1.27 Fraud Detection Service Execution

Attribute	Value
Title	Fraud Detection Service Execution
Componentid	INF- MLE -136
Description	This component provides the drivers' profile and classification (behaviour) for the last given driving route. It refers to the last uses the model trained in component 3.6.1.
Category	Analytics & ML Algorithms.
Layer	Analytics

Input	Drivers' data captured on one route made by the car. Related context information (weather conditions, roads data, traffic incidents)
Output	A profile of the driver for the particular route provided.
Technology	TensorFlow, Keras, Docker, SKLearn
Core	Pilot specific (for Pilot 11 "Personalized insurance products based on IoT connected vehicles").
Marketplace	Yes
Microservice	Yes
Endpoints	TBD
Owner	ATOS
License	ATOS Proprietary.
Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	Pilot 11

3.1.28 Pay As You Drive Service Training

Attribute	Value
Title	Pay As You Drive Service Training
Componentid	INF- MLE -137
Description	This component trains the Machine Learning model that will classify the drivers' behaviour while driving, according to the data collected from the driven cars (driving routes)
Category	Analytics & ML Algorithms.
Layer	Analytics
Input	A historic of drivers' data captured from their car, labeled as good or bad drivers.
Output	A trained model that will classify the driver's behaviour as good or bad.
Technology	TensorFlow, Keras, Docker, SKLearn
Core	Pilot specific (for Pilot 11 "Personalized insurance products based on IoT connected vehicles").
Marketplace	Yes
Microservice	Yes
Endpoints	TBD
Owner	ATOS
License	ATOS Proprietary.
Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	Pilot 11

3.1.29 Pay As You Drive Service Execution

Attribute	Value
Title	Pay As You Drive Service Execution
Componentid	INF- MLE -138
Description	This component will classify the drivers' behaviour while driving according to the data collected from their car and exploiting the Driver Classifier/Driving Profiling ML Models
Category	Analytics & ML Algorithms.
Layer	Analytics

Input	Driver's data captured from their car from different trips. Related relevant Context Information (roads, weather, traffic incidents)
Output	Driver's classified behaviour while driving.
Technology	TensorFlow, Keras, Docker, SKLearn
Core	Pilot specific (for Pilot 11 "Personalized insurance products based on IoT connected vehicles").
Marketplace	Yes
Microservice	Yes
Endpoints	TBD
Owner	ATOS
License	ATOS Proprietary.
Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	Pilot 11

3.1.30 Investment Recommendation Engine Training

Attribute	Value
Title	Investment Recommendation Engine Training
Componentid	INF- MLE - 139
Description	This component trains the personalized Investment Recommendation engine, to provide a set of recommendations for the financial instruments categories suitable for each customer and his/her investment profile, based on Market Index & Financial Instruments Sentiment data and Customer Risk Profiling ML Models.
Category	Analytics & ML Algorithms.
Layer	Analytics
Input	Customer Risk Profile, Financial Instruments vs Risk Profile and Market Index & Financial Instruments Sentiment Analysis data
Output	A trained model that provides personalized Financial Instrument recommendations for bank's customers, based on the risk profile and financial instruments sentiment analysis
Technology	Machine Learning Models in Python mainly based on scikit-learn
Core	Pilot specific (for Pilot 6 "Personalized Closed-Loop Investment Portfolio Management for Retail Customers ").
Marketplace	No
Microservice	No
Endpoints	TBD
Owner	NBG
License	NBG Proprietary.
Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	P06

3.1.31 Investment Recommendation Engine Execution

Attribute	Value
Title	Investment Recommendation Engine Execution
Componentid	INF- MLE -140

Description	Financial Instruments Personalized Investment Recommendation engine suitable for each customer and his/her investment risk profile, based on Market Index & Financial Instruments Sentiment data.
Category	Analytics & ML Algorithms.
Layer	Analytics
Input	Customer Risk Profile, Financial Instruments vs Risk Profile and Market Index & Financial Instruments Sentiment Analysis data
Output	Personalized Financial Instrument recommendations for bank's customers, based on the risk profile and financial instruments sentiment analysis.
Technology	Machine Learning Models in Python mainly based on scikit-learn
Core	Pilot specific (for Pilot 6 "Personalized Closed-Loop Investment Portfolio Management for Retail Customers").
Marketplace	No
Microservice	No
Endpoints	TBD
Owner	NBG
License	NBG Proprietary.
Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	P06

3.1.32 Recommender

Attribute	Value
Title	Recommender
Componentid	INF- MLE -147
Description	Generates an actionable insight depending on the output from the remaining components within the Data Analytics layer, i. e. components P5b_Analytics_02 to P5b_Analytics_08.
Category	Analytics & ML Algorithms.
Layer	Analytics
Input	P5b_Data_04 (Collected Enriched Data) / P5b_Data_02 (Generated Insights)A JSON/CSV / XML file from components P5b_Data_04 (Collected Enriched Data) / P5b_Data_02 (Generated Insights) directly or indirectly via component P5b_Data_01 (Data Streaming).
Output	P5b_Data_03 (Generated Actions).A JSON/CSV/XML file containing the obtained insights and recommendation to be provided to the respective SME via component P5b_Data_03 (Generated Actions).
Technology	Python libraries (Pandas, PySpark, Scikit-learn, Tensorflow, Keras)
Core	Pilot specific (for Pilot 5b "Business Financial Management (BFM) tools delivering a Smart Business Advise").
Marketplace	No
Microservice	Yes
Endpoints	tbd
Owner	BOC, UPRC (Business and ML Model).
License	BOC & UPRC Proprietary.
Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	P05b

3.1.33 Cash Flow Prediction

Attribute	Value
Title	Cash Flow Prediction
Componentid	INF- MLE -148
Description	ML/AI model used to indicate and predict the available working capital (operating cash flow) of the SME (AS-IS & near term future).Alerts/notifications (via BOC Middleware PushNotifications) to be pushed to the respective SME in case of potential lack of liquidity and/or balance moving below a threshold.Cashflow Data/Insights to be provided to BOC Middleware Mobile/Web BPF.In order to provide valuable insights to the SME the data should be collected and streamed in real (or near real) time, since whenever a new object/entry appears the model should retrain and adapt.
Category	Analytics & ML Algorithms.
Layer	Analytics.
Input	P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data).A JSON/CSV / XML file directly or via indirectly via component P5b_Data_01 (Data Streaming).Data required will originate from the following data sources:· Transaction (Txn.) Data from the Bank.· Transaction (Txn.) Data from Open Banking.· Transaction (Txn.) Data from SME.
Output	P5b_Data_02 (Generated Insights).Operating Cash Flow statement (flows, values, time) matrix represent history, actual and near-term future prediction.Matrix will be updated in real (or near real) time to provide accurate insights.Matrix will allow parameterization based on time period and interval (e.g. weeks, months).A JSON/CSV / XML file containing the obtained cash flow related data to be provided to SME directly from component P5b_Data_04 (Collected Enriched Data) or indirectly via component P5b_Data_01 (Data Streaming).
Technology	Python libraries (Pandas, PySpark, Scikit-learn, Tensorflow, Keras)
Core	Pilot specific (for Pilot 5b "Business Financial Management (BFM) tools delivering a Smart Business Advise").
Marketplace	No
Microservice	Yes
Endpoints	tbd
Owner	BOC, UPRC (Business and ML Model).
License	BOC & UPRC Proprietary.
Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	P05b

3.1.34 Budget Prediction

Attribute	Value
Title	Budget Prediction
Componentid	INF- MLE -149
Description	AI (ML) model is used to support the budget target setting for the various categories used by the respective SME. Doing so by providing budget predictions for each utilized category. The underlying model will take into consideration the cash flow analysis output, benchmark, macroeconomic and other available SME data (Business Plan).
Category	Analytics & ML Algorithms.
Layer	Analytics

Input	P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data) / P5b_Data_05 (SME Input).A JSON/CSV / XML file directly or via indirectly via component P5b_Data_01 (Data Streaming). Data required will originate from the following data sources: Transaction (Txn.) Data from the Bank. Transaction (Txn.) Data from Open Banking. Transaction (Txn.) Data from SME. Customer (SME) Data from the Bank. Other Data from SME. SME Input from Bank (Mobile app).Other Data from market (e.g. Benchmarking & Macroeconomic).and enginesCash Flow PredictionBenchmark(s)
Output	P5b_Data_02 (Generated Insights).A JSON/CSV / XML file containing the derived budget target for each category used by the respective SME. Output to be provided to SME via component P5b_Data_02 (Insights) / P5b_Data_04 (Collected Enriched Data) directly or indirectly via component P5b_Data_01 (Data Streaming).
Technology	Python libraries (Pandas, PySpark, Scikit-learn, Tensorflow, Keras)
Core	Pilot specific (for Pilot 5b “Business Financial Management (BFM) tools delivering a Smart Business Advise”)
Marketplace	No
Microservice	Yes
Endpoints	tbd
Owner	BOC, UPRC (Business and ML Model).
License	BOC & UPRC Proprietary.
Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	P05b

3.1.35 KPI Engine

Attribute	Value
Title	KPI Engine
Componentid	INF- MLE -150
Description	The KPI engine calculates KPIs with regards to the Financial Health and Performance of the respective SME.Doing so by taking into consideration the respective SME profile (e.g. maturity stage), accounting-wise optimal KPI values and how other similar SMEs perform.
Category	Analytics & ML Algorithms
Layer	Analytics
Input	P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data) / P5b_Data_05 (SME Input).A JSON/CSV / XML file directly or via indirectly via component P5b_Data_01 (Data Streaming).Data required will originate from the following data sources: Account (Acc.) Data from the Bank. Account (Acc.) Data from Open Banking. Customer (SME) Data from the Bank. Other Data from SME. SME Input from Bank (Mobile app). Other Data from market (e.g. Benchmarking & Macroeconomic).and engines Invoices Processing Benchmark(s)
Output	P5b_Data_02 (Generated Insights).A JSON/CSV / XML file containing results on Financial Health and Performance matrix. Output to be provided to SME via component P5b_Data_02 (Insights) / P5b_Data_04 (Collected Enriched Data) directly or indirectly via component P5b_Data_01 (Data Streaming).
Technology	Python libraries (Pandas, PySpark, Scikit-learn, Tensorflow, Keras)
Core	Pilot specific (for Pilot 5b “Business Financial Management (BFM) tools delivering a Smart Business Advise”).
Marketplace	No
Microservice	Yes
Endpoints	TBD
Owner	BOC, UPRC (Business and ML Model).
License	BOC & UPRC Proprietary.

Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	P05b

3.1.36 Transaction (Txn) Monitoring

Attribute	Value
Title	Transaction (Txn) Monitoring
Componentid	INF- MLE -151
Description	A dynamic complex event processing (CEP) mechanism that monitors the transactions of the user. In case transaction amount or type deviates from normal behaviour the user will be informed of abnormal transactions in order to be safeguarded from double payment mistakes and potential fraud attempts. In addition, expense pattern are also analysed to identify potential savings for instance multiple subscription spending or high ATM fees.
Category	Analytics & ML Algorithms
Layer	Analytics
Input	P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data) / P5b_Data_05 (SME Input). A JSON/CSV / XML file directly or via indirectly via component P5b_Data_01 (Data Streaming).
Output	P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data). A JSON/CSV / XML file containing results on abnormal transactions and suspicious expenses. Output to be provided to SME via component P5b_Data_02 (Insights) / P5b_Data_04 (Collected Enriched Data) directly or indirectly via component P5b_Data_01 (Data Streaming).
Technology	Python libraries (Pandas, PySpark, Scikit-learn, Tensorflow, Keras)
Core	Pilot specific (for Pilot 5b "Business Financial Management (BFM) tools delivering a Smart Business Advise").
Marketplace	No
Microservice	Yes
Endpoints	TBD
Owner	BOC, UPRC (Business and ML Model).
License	BOC & UPRC Proprietary.
Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	P05b

3.1.37 Transaction (Txn) Categorization

Attribute	Value
Title	Transaction (Txn) Categorization
Componentid	INF- MLE -152
Description	Smart transaction auto-classification which would also allow the user to manually override the given transaction category and define a new one (re-classify). The categorization performed will be based on the needs of the individual SME.
Category	Analytics & ML Algorithms
Layer	Analytics
Input	P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data) / P5b_Data_05 (SME Input). A JSON/CSV / XML file directly or via indirectly via component P5b_Data_01 (Data Streaming). Data required will originate from the following data sources: Transaction (Txn.) Data from the Bank. Transaction (Txn.) Data from Open Banking. Customer (SME) Data from the Bank. Other Data from SME. § SME Input from Bank (Mobile app).

Output	P5b_Data_02 (Generated Insights) / P5b_Data_04 (Collected Enriched Data).Output to be provided to SME via component P5b_Data_02 (Insights) / P5b_Data_04 (Collected Enriched Data) directly or indirectly via component P5b_Data_01 (Data Streaming).
Technology	Python libraries (Pandas, PySpark, Scikit-learn, Tensorflow, Keras)
Core	Pilot specific (for Pilot 5b “Business Financial Management (BFM) tools delivering a Smart Business Advise”).
Marketplace	No
Microservice	Yes
Endpoints	TBD
Owner	BOC, UPRC (Business and ML Model).
License	BOC & UPRC Proprietary.
Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	P05b

3.1.38 Invoice Processing

Attribute	Value
Title	Invoice Processing
Componentid	INF- MLE -153
Description	Processes transaction data and ERP data in order to keep record of the invoices which have been partially or fully paid by the SME. Provides insights regarding the respective VAT amount payable and the optimization of cash flow by providing background info on invoices such as paying a invoice at the “right” time.In cases where available invoice data is limited the engine will utilize a simplified approach in order to derive the expected VAT amount to be paid by the SME at the next VAT due date. Doing so by utilizing the banks transaction info or/and past VAT payment amounts.
Category	Analytics & ML Algorithms.
Layer	Analytics
Input	P5b_Data_04 (Collected Enriched Data).A JSON/CSV / XML file directly from component P5b_Data_04 (Collected Enriched Data) or indirectly via component P5b_Data_01 (Data Streaming).Data required will originate from the following data sources:§ Transaction (Txn.) Data from the Bank.§ Transaction (Txn.) Data from Open Banking.§ Customer (SME) Data from the Bank.§ Other Data from SME.§ SME Input from Bank (Mobile app).
Output	P5b_Data_02 (Insights) and P5b_Data_04 (Data Repository). A JSON/CSV / XML file containing Matrix with invoice information and payment prioritization. Output to be provided to other component P5b_Analytics_02 (Cash Flow Prediction) and SME via component P5b_Data_02 (Insights) and P5b_Data_04 (Collected Enriched Data) directly or indirectly via component P5b_Data_01 (Data Streaming).
Technology	TBD
Core	Pilot specific (for Pilot 5b “Business Financial Management (BFM) tools delivering a Smart Business Advise”).
Marketplace	No
Microservice	Yes
Endpoints	TBD
Owner	BOC, UPRC (Business and ML Model).
License	BOC & UPRC Proprietary.
Remarks	TBD
Documentation	Link to reference documentation if any
Pilots	P05b

3.1.39 KMeans

Attribute	Value
Title	KMeans
Componentid	INF- MLE -155
Description	Batch BDA Service: Given a set of observations (x_1, x_2, \dots, x_n), where each observation is a d -dimensional real vector, k -means clustering aims to partition the n observations into k ($k \leq n$) sets $S = \{S_1, S_2, \dots, S_k\}$ so as to minimize the within-cluster variance.
Category	ML Algorithm
Layer	Analytics/User
Input	Any CSV or XLS or ORC. Dataset stored into HDFS/PRESTO/HIVE Input type: Quantitative
Output	Any CSV or XLS or ORC. Dataset stored into HDFS/PRESTO/HIVE
Technology	It is a BDA service registered in the ALIDA catalogue. It is a PySpark-based micro-service running on K8S Spark cluster mode, working as part of the ALIDA framework.
Core	This component is general purpose and reusable as a BDA service within ALIDA
Marketplace	No
Microservice	Yes. It is a dockerized microservice component
Endpoints	No (it is not a standalone service that waits for incoming requests but it is closely related to the lifecycle of an ALIDA workflow.)
Owner	ENG
License	Component developed with open source technologies. ENG Proprietary
Remarks	In a nutshell, ALIDA is a Micro-service based platform for composition, deployment, optimisation, execution and monitoring of pipelines of Big Data Analytics (BDA) services. ALIDA is a result of previous research activities developed by ENG. Currently, it is a work in progress. ALIDA offers a catalogue of BDA services (ingestion, preparation, analysis, visualization): user designs his own (stream/batch) pipeline by choosing the BDA services from it, indicates which Big Data set he wants to process, launches and monitors the execution of the pipeline and personalizes the results visualization by choosing from a set of available graphs, all this without worrying about having software developer skills or particular knowledge on big data technologies. This service is registered in ALIDA catalogue as Spring Boot Application containing the python code and its dependencies. After implementing the algorithm using Pyspark, creating the Dockerfile and pushing the new image inside a repository, this microservice is registered into the ALIDA catalogue through the GUI. Source: https://home.alidalab.it/
Documentation	N/A
Pilots	P10, P13

3.1.40 Random Forest (Model)

Attribute	Value
Title	Random Forest (Model)
Componentid	INF- MLE -156
Description	Batch BDA Service: An ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set
Category	ML Algorithm
Layer	Analytics/User
Input	Any CSV or XLS or ORC. Dataset stored into HDFS/PRESTO/HIVE
Output	model object stored into HDFS.

Technology	It is a BDA service registered in the ALIDA catalogue. It is a PySpark-based micro-service running on K8S Spark cluster mode, working as part of the ALIDA framework.
Core	This component is general purpose and reusable as a BDA service within ALIDA.
Marketplace	No
Microservice	Yes. It is a dockerized microservice component
Endpoints	No (it is not a standalone service that waits for incoming requests but it is closely related to the lifecycle of an ALIDA workflow.)
Owner	ENG
License	Component developed with open source technologies. ENG Proprietary
Remarks	In a nutshell, ALIDA is a Micro-service based platform for composition, deployment, optimisation, execution and monitoring of pipelines of Big Data Analytics (BDA) services. ALIDA is a result of previous research activities developed by ENG. Currently, it is a work in progress. ALIDA offers a catalogue of BDA services (ingestion, preparation, analysis, visualization): user designs his own (stream/batch) pipeline by choosing the BDA services from it, indicates which Big Data set he wants to process, launches and monitors the execution of the pipeline and personalizes the results visualization by choosing from a set of available graphs, all this without worrying about having software developer skills or particular knowledge on big data technologies. This service is registered in ALIDA catalogue as Spring Boot Application containing the python code and its dependencies. After implementing the algorithm using Pyspark, creating the Dockerfile and pushing the new image inside a repository, this microservice is registered into the ALIDA catalogue through the GUI. Source: https://home.alidalab.it/
Documentation	N/A
Pilots	P10, P13

3.1.41 Random Forest (Predict)

Attribute	Value
Title	Random Forest (Predict)
Componentid	INF- MLE -157
Description	Streaming random forests algorithm
Category	ML Algorithm
Layer	Analytics/User
Input	This component is a consumer who subscribes to Kafka's topic, so it reads the messages that will be sent to the topic. Useful information to take stored machine learning model created by Random Forest Model.
Output	This component is also a producer, who write the results of the prediction to another Kafka topic.
Technology	It will be a BDA service registered in the ALIDA catalogue. It is a PySpark-based micro-service running on K8S Spark cluster mode, working as part of the ALIDA framework.
Core	This component is general purpose and reusable as a BDA service within ALIDA
Marketplace	No
Microservice	Yes. It is a dockerized microservice component
Endpoints	No (it is not a standalone service that waits for incoming requests but it is closely related to the lifecycle of an ALIDA workflow.)
Owner	ENG
License	Component developed with open source technologies. ENG Proprietary

Remarks	In a nutshell, ALIDA is a Micro-service based platform for composition, deployment, optimisation, execution and monitoring of pipelines of Big Data Analytics (BDA) services. ALIDA is a result of previous research activities developed by ENG. Currently, it is a work in progress. ALIDA offers a catalogue of BDA services (ingestion, preparation, analysis, visualization): user designs his own (stream/batch) pipeline by choosing the BDA services from it, indicates which Big Data set he wants to process, launches and monitors the execution of the pipeline and personalizes the results visualization by choosing from a set of available graphs, all this without worrying about having software developer skills or particular knowledge on big data technologies. This service is registered in ALIDA catalogue as Spring Boot Application containing the python code and its dependencies. After implementing the algorithm using Pyspark, creating the Dockerfile and pushing the new image inside a repository, this microservice is registered into the ALIDA catalogue through the GUI. Source: https://home.alidalab.it/
Documentation	N/A
Pilots	P10, P13

3.1.42 Client Contextual Information

Attribute	Value
Title	Client Contextual Information
Componentid	INF-MLE-158
Description	The component generates and updates a relevant client contextual information related to the clients' data and behaviour
Category	Analytics/ML
Layer	Analytics/ML - Data Processing
Input	Input format tbd. Examples include json, csv, parquet, xml, npz, etc.
Output	Output format tbd. Probably json or CSV
Technology	open source AI/ML frameworks
Core	No
Marketplace	No
Microservice	No
Endpoints	not yet specified
Owner	CXB / FBK / FTS
License	Closed
Remarks	
Documentation	
Pilots	P07

3.1.43 Financial Fraud/Crime Risk Score

Attribute	Value
Title	Financial Fraud/Crime Risk Score
Componentid	INF-MLE-159
Description	Using the clients' contextual information and transactions a risk score of a fraudulent request of an online instant loan is evaluated
Category	Analytics/ML
Layer	Analytics/ML
Input	Input format TBD. Examples include json, csv, parquet, xml, npz, etc.
Output	Output format TBD. Probably json or CSV
Technology	Streaming analytics

Core	Yes
Marketplace	TBD
Microservice	No
Endpoints	not yet specified
Owner	CXB / FBK / FTS
License	closed
Remarks	
Documentation	
Pilots	P07

3.1.44 Anomaly Analysis

Attribute	Value
Title	Anomaly Analysis
Componentid	INF-DSM-130
Description	Anomaly Analysis provides two main functionalities:- Anomaly detection- Anomaly prediction for time series data.
Category	Analytics & ML algorithms
Layer	Data Analytics
Input	Data Preprocessing
Output	User interaction
Technology	QMiner
Core	Will be part of the INFINITECH platform
Marketplace	Yes
Microservice	Yes
Endpoints	Yes, but needs to be developed
Owner	JSI
License	Open source MIT License
Remarks	N/A
Documentation	N/A
Pilots	P08

3.1.45 Pattern Analysis

Attribute	Value
Title	Pattern Analysis
Componentid	INF- DSM -131
Description	Pattern Analysis provides two main functionalities:- Pattern matching- DiscoveryThe component will provide support for detection of complex patterns on data graphs
Category	Analytics
Layer	Data Analytics
Input	Data Preprocessing
Output	User interaction
Technology	SNAP, QMiner
Core	Yes
Marketplace	Yes
Microservice	Yes

Endpoints	No
Owner	JSI
License	Open source, MIT License
Remarks	N/A
Documentation	N/A
Pilots	P08

3.1.46 Stream Story

Attribute	Value
Title	Stream Story
Componentid	INF- DSM -132
Description	Stream Story is a component for the analysis of multivariate time series. It computes and visualizes a hierarchical Markov chain model which captures the qualitative behaviour of the systems' dynamics, where system is described with a group of timeseries
Category	Analytics & ML algorithms
Layer	Data Analytics
Input	Data Preprocessing
Output	User interaction
Technology	Stream Story, QMiner
Core	Will be part of the INFINITECH platform
Marketplace	Yes
Microservice	No (docker container)
Endpoints	No (available GUI)
Owner	JSI
License	Open source, MIT License
Remarks	N/A
Documentation	N/A
Pilots	P08

3.1.47 Open API Gateway

Attribute	Value
Title	Open API Gateway
Componentid	INF-INT-001
Description	This component provides the single point of entry for the added-value functionalities (such as the Machine Learning (ML) / Deep Learning (DL) analytics functionalities) of INFINITECH which are based on microservices. The specific component enables the discovery and invocation of the dynamically registered microservices, effectively handling the incoming requests towards these microservice instances.
Category	Interface
Layer	Communication and Connectivity
Input	The component receives as input:a) the Endpoints or the list of Endpointss of the desired microservice(s) which should be prefixed with the names that the corresponding microservices used to register with the registry.b) the required data that will be handled by the microservice(s)c) the JSON Web Token required for the client's authentication
Output	Depending on the invoked microservice, the outputs of the microservice are returned to the requestor.

Technology	This component is based on Java programming language and the Java Spring Boot framework. Furthermore, the open-source service discovery solution Consul is leveraged.
Core	Yes
Marketplace	Yes
Microservice	Yes
Endpoints	https://{{Server_URL}}:{{Server_PORT}}/gateway/api/v1/gateway
Owner	UBI
License	This component is under proprietary rights of UBI.
Remarks	This component is effectively handling the discovery of dynamically-deployed microservices and their effective invocation from client applications. To achieve this, the microservices self-registration process offered by this component, should be followed and adopted by the respective microservices.
Documentation	Included in deliverable D5.10
Pilots	To be defined - Probably most of the pilots

3.1.48 User Interface for Blockchain Transaction Reports and Visualization Component

Attribute	Value
Title	User Interface for Blockchain Transaction Reports and Visualization Component
Componentid	INF- BKC -129
Description	This component will be responsible for providing user interaction with the Scalable Transaction Graph Analysis component within the bank and collect/manage user as well as annotated blacklisted blockchain addresses . It will utilize OpenAPIs (REST APIs) to submit queries consisting of customer blockchain addresses and blacklists to transaction graph analysis component and generate web based reports and visualization based received results
Category	
Layer	Data Visualization and User Interaction
Input	Scalable Transaction Graph Analysis component results
Output	Web interfaces and reports showing customer blockchain addresses to fraudulent transactions if any and web based visualization of transaction graph paths and subgraphs
Technology	OpenAPIs / REST APIs, Web servers, Javascript, Vis.js graph drawing library (community version) https://github.com/visjs-community/visjs-network
Core	Will be part of the INFINITECH platform
Marketplace	Yes
Microservice	TBD
Endpoints	TBD
Owner	AKTIF
License	Open source MIT License
Remarks	N/A
Documentation	N/A
Pilots	P09

3.1.49 Visualization Preparation

Attribute	Value
Title	Visualization Preparation
Componentid	INF- DMA -114

Description	Stream BDA Service: Service to prepare data to the visualization depending on the type of incoming data or the data you want to view
Category	Data Management
Layer	Processing
Input	This component is a consumer who subscribes to Kafka's topic, so it reads the messages that will be sent to the topic.
Output	This component is also a producer, who writes the results to other one Kafka topic.
Technology	It will be a microservice belonging to the ALIDA core.
Core	This component is part of ALIDA core
Marketplace	No
Microservice	Yes. It is a dockerized microservice component
Endpoints	No (it is not a standalone service that waits for incoming requests but it is closely related to the lifecycle of an ALIDA workflow.)
Owner	ENG
License	Component developed with open source technologies. ENG Proprietary
Remarks	In a nutshell, ALIDA is a Micro-service based platform for composition, deployment, optimisation, execution and monitoring of pipelines of Big Data Analytics (BDA) services. ALIDA is a result of previous research activities developed by ENG. Currently, it is a work in progress. ALIDA offers a catalogue of BDA services (ingestion, preparation, analysis, visualization): user designs his own (stream/batch) pipeline by choosing the BDA services from it, indicates which Big Data set he wants to process, launches and monitors the execution of the pipeline and personalizes the results visualization by choosing from a set of available graphs, all this without worrying about having software developer skills or particular knowledge on big data technologies.Source: https://home.alidalab.it/
Documentation	N/A
Pilots	P10

3.1.50 Real Time Visualization

Attribute	Value
Title	Real Time Visualization
Componentid	INF-VIS-158
Description	Useful tool for displaying charts through web application
Category	Visualization
Layer	Interaction/Visualization
Input	Metadata to handle resulting datasets as data source and query.
Output	Aggregated information chart plotting
Technology	Microservices deployed through Kubernetes● Docker● Data Source Connectors
Core	Standalone
Marketplace	Yes
Microservice	Yes. It is a collection of dockerized microservices
Endpoints	No
Owner	ENG
License	Component developed with open source technologies. ENG Proprietary
Remarks	
Documentation	N/A
Pilots	P10

3.1.51 INFINISTORE

Attribute	Value
Title	INFINISTORE
Componentid	INF-DMA-161
Description	This is a GENERIC DATA STORE implementation for the INFINITECH Project as a microservice on top of a noSQL DB (MongoDB) instance. It is fed by different data ingestions servers and supports all other services.
Category	Data Management
Layer	Data Management
Input	
Output	
Technology	Microservice wrap on top of MongoDB. The microservice is implemented as a python-flask web server application on top of MongoDB instance. The API are defined with SWAGGER Open Api 3.0
Core	No
Marketplace	Yes
Microservice	Yes
Endpoints	https://gitlab.infinittech-h2020.eu/datamanagement/infinistore/-/blob/master/infinistore.yaml
Owner	GFT
License	Opensource License
Remarks	https://gitlab.infinittech-h2020.eu/datamanagement/infinistore
Documentation	https://gitlab.infinittech-h2020.eu/datamanagement/infinistore
Pilots	P15

3.1.52 UI Risk Assessment based on VaR

Attribute	Value
Title	UI Risk Assessment based on VaR
Componentid	INF-VIS-159
Description	Web application to monitor portfolio risk in real time, perform what-if analysis, providing also several statistics of the underlying financial assets.
Category	Analytics & Visualisation
Layer	Interaction/Visualisation
Input	This component reads VaR predictions (produced by VaR Predictor component), which are stored in the Relational DB component, via JDBC connection.
Output	Dashboards and Charts
Technology	Python libraries (Pandas, Numpy, Scipy, Flask, Dash, Plotly)
Core	No, Part of Pilot 2
Marketplace	No
Microservice	Yes
Endpoints	TBD
Owner	JRC, INNOV
License	JRC & INNOV Proprietary
Remarks	
Documentation	N/A
Pilots	P02

3.1.53 Pseudo-anonymization tool

Attribute	Value
Title	Pseudo-anonymization tool
Componentid	INF-DCO-055
Description	A tool that pseudo-anonymize data in order to preserve privacy. The component needs a specific configuration/development for each pilot in which it is used.
Category	Data protection
Layer	Data Processing
Input	Description of Input format (File, JSON, ...)
Output	Description of Output format (File, JSON, ...)
Technology	Docker container
Core	Yes
Marketplace	Yes
Microservice	Yes
Endpoints	If Yes, please provide the Endpoints URLs https://api.example.com/v1/users?role=admin&status=active _____/_____/ _____/\ server URL Endpoints query parameters REST API
Owner	BOS, JSI
License	Opensource License
Remarks	E.g. Project's official website
Documentation	Link to reference documentation if any
Pilots	P08

3.1.54 Health insurance risk assessment service

Attribute	Value
Title	Health insurance risk assessment service
Componentid	INF-DCO-056
Description	Algorithm yielding risk based on user RWD and a pre-trained classifier. Implemented as a Python script
Category	Analytics & ML Algorithms
Layer	Analytics
Input	The previous risk assessment value for the given person. None if this is the first calculation. A collection of attribute vectors, in the form of a CSV file, each row corresponding to a new observation about the person of interest.
Output	A number corresponding to the updated risk assessment
Technology	Currently the assessment classifiers are: Random Forests or Logistic regression (using scikit-learn for inference), or neural networks (using Tensorflow for inference)
Core	No
Marketplace	Yes
Microservice	No
Endpoints	
Owner	ISPRINT
License	Closed
Remarks	The component is still under preparation
Documentation	A description of the risk assessment service can be found in this paper: https://www.mdpi.com/2227-9091/9/3/46

Pilots	P12
--------	-----

3.1.55 Health insurance fraud detection service

Attribute	Value
Title	Health insurance fraud detection service
Componentid	INF-DCO-057
Description	Algorithm detecting fraudulent behavior of insurance company customers. Implemented as a Python script
Category	Analytics & ML Algorithms
Layer	Analytics
Input	Models of the behavioral attributes. A collection of attribute vectors, in the form of a CSV file, each row corresponding to a new observation about the person of interest.
Output	A number corresponding to the trust to be put on the attribut vectors under examination
Technology	Scikit-learn and/or Tensorflow
Core	No
Marketplace	Yes
Microservice	No
Endpoints	
Owner	ISPRINT
License	Closed
Remarks	The component is still under preparation
Documentation	
Pilots	P12

3.1.56 Well-being outlook classifiers

Attribute	Value
Title	Well-being outlook classifiers
Componentid	INF-DCO-058
Description	Classifiers to be used by the health insurance risk assessment algorithm. The current set includes Random Forest, Logistic Regression and Neural Network classifiers of varying complexity. The format of the classifier depends on its type.
Category	Analytics & ML Algorithms
Layer	Analytics
Input	A collection of attribute vectors, in the form of a CSV file, each row corresponding to a new observation about a person
Output	An array of predictions
Technology	Scikit-learn and/or Tensorflow
Core	No
Marketplace	Yes
Microservice	No
Endpoints	
Owner	ISPRINT
License	Not applicable - a binary file
Remarks	Multiple classifiers are being made available
Documentation	A description of the well-being outlook classifiers can be found in this paper: https://www.mdpi.com/2227-9091/9/3/46

Pilots	P12
--------	-----

3.1.57 Synthetic RWD for well-being analytics

Attribute	Value
Title	Synthetic RWD for well-being analytics
Componentid	INF-DCO-059
Description	Synthetic data of the same format as those collected in Pilot 12. The current version of the data span 1,000 people simulated for 116 weeks each
Category	Analytics & ML Algorithms
Layer	Analytics
Input	The data are provided as different CSV files containing: (a) the behavioral phenotype of each person, (b) the aggregated daily data, (c) the autodetected exercise sessions and (d) a large collection of intra-day data, currently at a 5-minute resolution
Output	Not applicable
Technology	The iSprint RWD simulator
Core	No
Marketplace	Yes
Microservice	No
Endpoints	
Owner	ISPRINT
License	Opensource License
Remarks	
Documentation	A description of the RWD simulator and the data obtained can be found in this paper: https://www.mdpi.com/2227-9091/9/3/46
Pilots	P12

3.1.58 Open Banking Agreggator Solution

Attribute	Value
Title	Open Banking Agreggator Solution
Componentid	INF-DCO-060
Description	Crowdpolicy Open Banking Agreggator Solution is a modular architecture (UIs, connectors & APIs) platform so that it can be integrated into web / mobile banking applications, by the existing provider of the Bank in the form of API integration, but also as a separate application that could be made available to the users of Bank's online services Compatibility with best known market standards based on the European PSD2 Directive (Berlin Group, Open Banking UK, STET) Support for PISP & AISP services based on the PSD2 European Directive: - Payment Initiation Services - Account information Services
Category	Open APIs
Layer	Infrastructure/Data Management/Data Processing
Input	N/A
Output	N/A
Technology	Java Springboot, NodeJS, Firebase,
Core	No
Marketplace	No
Microservice	Yes if it is a dockerized microservice component
Endpoints	N/A
Owner	CP Crowdpolicy

License	Closed License per year
Remarks	
Documentation	https://finteliox.com/fintelio-x-core-products/fintelio-x-crossover/
Pilots	

3.1.59 Big data analytics platform

Attribute	Value
Title	Big data analytics platform
Componentid	INF-DCO-061
Description	Platform that collect and process information from multiple open data sources regarding SMEs and apply cognitive algorithm to detect risk and changes in financial needs. The tool will be use in pilot 13
Category	Category used
Layer	Data Management/Data Processing
Input	API Rest, CSV, JSON
Output	API Rest, CSV, JSON
Technology	Describe technology used
Core	No
Marketplace	Yes
Microservice	No
Endpoints	If Yes, please provide the Endpoints URLs https://api.example.com/v1/users?role=admin&status=active _____/ _____/ server URL Endpoints query parameters REST API
Owner	Wenalyze
License	Closed
Remarks	https://infinitech.wenalyze.com/
Documentation	Link to reference documentation if any
Pilots	P13

4 Conclusions

Task T2.3 has the objective of provide the specification of BigData/IoT technological building blocks to be developed in the scope of the project, especially in the areas of cost-effective real-time BigData analytics, data and semantic interoperability, elastic cloud storage, integrated (declarative) data querying, AI/ML algorithms.

More precisely, the detailed specifications of a large set of innovative components that are illustrated in Section 3, including specifications about the implementation technologies (e.g., BigData/IoT platforms, AI/ML toolkits, HPC infrastructures) will be used to realize and guide the implementation and integration. Overall, the INFINITECH pilot requirements reflect the State of the Art of the application of BigData, IoT and AI in the Financial Services and contribute to the latest trends. Within this deliverable, the final services related to the landscape of technology and data driven innovation of Financial Services are identified based on the broad use cases and reference scenarios provided by the INFINITECH pilots. Based on existing services and processes (e.g., KYC/KYB, fraud detection, customer service, portfolio management, asset management, usage based insurance) it has been outlined in the pilots' scenarios how these could become more autonomous, personalized and context-aware. In addition the pilots workflows design have been analyzed to provide the required functional services for their fulfilment. Also, functional components are mapped to the BDVA reference model [4] and the updated IRA design facilitates the next steps within the task T2.7.

With respect to its previous version (D2.5 "Specifications of INFINITECH Technologies - I"), this deliverable enriches the content so far provided by updating it with new techtools and pilot components as well as by providing additional information for each pilot component, in particular related to their availability in the MARKETPLACE, their availability as a dockerized microservice component, and finally the specification, if any, of their Endpoint/REST API URLs.

Based on the outcomes of the previous deliverable D2.5, the following further activities have been updated and completed:

- Analyze state of the art technologies to complete components specifications
- Complete components surveys and map requirements to components' specification
- Find communalities of the features reducing the number of components needed
- Complete component identification, correct mapping and identification.

5 References

1. INFINITECH Deliverable D2.1 – User Stories and Stakeholders' Requirements - I
2. INFINITECH Deliverable D2.3 – Reference Scenarios and Use Cases – I
3. INFINITECH Deliverable D2.13 – Reference Architecture – I
4. BDVA: “BDV SRIA – European Big Data Value Strategic Research and Innovation Agenda”, Version 4.0, October 2017
5. INFINITECH Deliverable D2.5 - Specifications of INFINITECH Technologies - I
6. INFINITECH Deliverable D2.14 – Reference Architecture – II
7. WITDOM Project Documentation <https://bit.ly/3cEA75e>