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

Infinitech

D2.9 – Initial Specification of Testbeds, Data Assets and APIs - I

Lead Beneficiary	CP
Due Date	2020-06-30
Delivered Date	2020-06-30
Revision Number	4.0
Dissemination Level	Public (PU)
Type	Report
Document Status	Final
Review Status	Internally Reviewed
Document Acceptance	WP Leader Accepted and/or Coordinator Accepted
EC Project Officer	Pierre-Paul Sondag

HORIZON 2020 - ICT-11-2018

This project has received funding from the European Union's horizon 2020 research and innovation programme under grant agreement no 856632

Contributing Partners

Partner Acronym	Role[1]	Name Surname[2]
CP	Lead Beneficiary	Marinos Xynarianos
GFT	Contributor & Internal Reviewer	Marcelo Colomer, Marina Cugurra, Martin Felix de Miguel Lillo, Ernesto Troiano, Maurizio Megliola
BOI	Contributor	Eymard Hooper
NBG	Contributor	Manos Silignakis, Nikos Droukas
AKTIF	Contributor	Orkan Metin
BPFI	Contributor	Gavin Purtill, Richard Walsh
JRC	Contributor	Petra Ristau
PRIVE	Contributor	Pablo Carballo, Roland Meier
WEA	Contributor	Carlos Albo Portero
RB	Contributor	John Kaldis
UNP	Contributor	Tiago Teixeira
NUIG	Contributor	Martin Serrano
ENG, BS, CPH, FBK, UNIC, GLA	Contributor	
Pilots (BANKIA, JRC, BOI, PRIVE, LIB, BOC, NBG, FBK, BOS, AKTIF, PI, ATOS, SILO, WEA, GEN AGRO)	Contributors	Massimiliano Aschi, Pablo Carballo, Nikolaos Droukas, Elena Femenía, Eymard Hooper, Klaudija Jurkosek-Seitl, Bruno Lepri, Lukas Linden, Giorgos Marinos, Orkan Metin, Grigoris Mykdakos, Paul O'Connel, Aristodemos Pnevmatikakis, Carlos Alberto Portero, Petra Ristau, Silvio Walser
DWF	Internal Reviewer	

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

[2] Can be left void

INNOV	Quality Assurance	

Revision History

Version	Date	Partner(s)	Description
0.1	2020-04-30	CP	ToC Version
0.2	2020-05-28	CP	Revised ToC & Initial Draft Version
0.3	2020-06-11	CP, ALL	Second Draft Version
1.0	2020-06-15	DWF, GFT	Internal Reviewer
2.0	2020-06-24	INNOV	Version for Quality Assurance
3.0	2020-06-29	INNOV	Version for Quality Assurance, including comments from GFT, FTS (WP Leader)
4.0	2020-06-30	CP, GFT	Version for Submission

Executive Summary

The goal of task T2.5 "Open Banking APIs, Testbeds and Data Assets Specifications" is the specification of the Open APIs that will be implemented as part of the INFINITECH-RA, the project's technologies and the project's testbeds. These APIs will take into account the functionalities of the technological building blocks of the project, along with the needs for compliance to certain regulations (e.g., GDPR, PSD2, MiFiD, 4MLD). Furthermore, the task will produce the specification of the testbeds of the project in terms of functionalities, data assets, as well as in terms of the regulatory tools that they will offer.

The document aims to provide specifications for the **testbeds** that will be implemented as part of the INFINITECH Reference Architecture (INFINITECH-RA), regarding the relevant **functionalities**, **Open APIs**, **data assets**, that will be used and developed from the relevant pilots that will be hosted in each testbed, as well as the **regulatory tools** that they will offer.

This deliverable is the first version of a total of two deliverables which are meant to provide the outcome of task T2.5. This version of the document will consider the **technology building blocks** that will be developed as part of each pilot, along with the **data assets** that will be used. Also, a first approach of the **Open APIs** that each pilot will make available to be used for giving access to the main or part of the functionality that each pilot will provide, based on the relative user stories (use cases) that will be addressed.

In particular, the deliverable contains initial specifications for:

- The testbeds that will be implemented as part of the project for hosting the relative pilots;
- The data assets that will be used for each pilot and the relative sandboxes that will be hosted from each testbed, based on the relative use cases and stakeholders' requirements;
- The Open APIs that will be available for accessing the data assets or will be developed and be available to be used for each pilot sandboxes and testbeds, that may be reusable from other INFINITECH Pilots;
- Regulatory Compliance Tools that will be used in order to meet the requirements of certain regulations (e.g. GDPR, PSD2, MIFID,4MLD, etc)

The work related to task T2.5 will continue until Month 18, when the 2nd version of this deliverable will be submitted (D2.10), with the updates on the specification, based on the progress of the INFINITECH Project.

Table of Contents

List of Figures

List of Tables

Abbreviations

API	Application Programming Interface
DL	Deep Learning
GDPR	General Data Protection Regulation
HTTP	Hypertext Transfer Protocol
IoT	Internet of Things
KYC	Know Your Customer
KYB	Know Your Business
MiFID	Markets in Financial Instruments Directive
MiFIR	Markets in Financial Instruments and Amending Regulation
ML	Machine Learning
NDA	Non-Disclosure Agreement
NIS	Network and Information Systems
OES	Operators of Essential Services
PAN	Primary Account Number
PaaS	Platform as a Service
PCI DSS	Payment Card Industry Data Security Standard
PIA	Privacy Impact Assessment
PSD2	Payment Service Directive 2
PSP	Payment Service Provider
PSU	Payment Service User
P2PP	Peer-to-Peer Payment
QTSP	Qualified Trust Service Provider
RA	Reference Architecture
REST	Representational state transfer

RTS	Regulatory Technical Standard
SCA	Strong Customer Authentication
SHARP	Smart, Holistic, Autonomy, Personalized and Regulatory Compliance
SME	Small and Medium-Sized Enterprises
SA	Supervisory Authority
SECaaS	Security-as-a- Service
TI	Threat Intelligence
VDIH	Virtualized Digital Innovation Hub
XML	Extensible Markup Language
3DS	Three-Domain Secure
4MLD	Fourth Money Laundering Directive

1. Introduction

Task 2.5 of INFINITECH Project, provides the initial specification of the advanced experimentation infrastructures (testbeds & sandboxes), which shall provide access to resources for application development and experimentation of BigData, IoT and AI-based innovations, as well as the specifications of relative data assets, regulatory tools, libraries of ML/DL algorithms, Open APIs and more, that will be implemented as part of the INFINITECH-RA. Such experimentation infrastructures (testbeds & sandboxes) should be available, based on the deployment of the relative technical building blocks, in various configurations.

The INFINITECH project's results will be validated in the scope of Several Large-Scale Innovative Pilots in Finance and Insurance, which will leverage both the technological developments of the project and the testbeds/sandboxes in order to deploy and validate novel use cases in real-life environments based on realistic datasets. The pilots will span a wide array of areas covering the most prominent processes of the financial and insurance sectors, including KYC (Know Your Customer)/ KYB (Know Your Business) and customer centric analytics, fraud detection and financial crime, credit risk assessment, risk assessment for capital management, personalized portfolio management, risk assessment in investment banking, personalized usage-based insurance, insurance products recommendations and more. The pilots will demonstrate the added-value of the project's technologies and testbeds, while at the same time showcasing the project's disruptive impact on Europe's financial and insurance sectors.

1.1. Objective of the Deliverable

This deliverable describes the initial specifications of the advanced experimentation infrastructures (testbeds & sandboxes) that will be implemented as part of the INFINITECH- RA, taking into account the functionalities of the technological building blocks of the project, along with the relative data assets and needs for compliance to certain regulations (e.g., GDPR, PSD2, MiFiD, 4MLD). Also, the deliverable provides the initial specification of Open APIs and other resources that will be available or being will be developed for validating autonomous and personalized solutions to be implemented through the relative pilots and respective testbeds, including a unique collection of data assets for finance/insurance.

The overall main objectives of this deliverable are:

- Provide the **general guidelines** of the deployment of **advanced experimentation infrastructures (testbeds & sandboxes)** that will be implemented for hosting each pilot related to BigData, IoT and AI-based innovations, based on the required significant testing and validation efforts;
- Provide **initial specifications** of the **data assets** that will be selected **and bundled together** to **flexibly configure and provision sandboxes** over these testbeds;
- Enable financial organizations **through Open APIs, and regulatory tools** needed for testing/ validating the use cases in the target (finance/insurance) sector.

1.2. Insights from other Tasks and Deliverables

The deliverable will utilize pilot descriptions and user stories developed in Task 2.1 and available through deliverable D2.1 with a functional-services view from all pilots as part of Task 2.2, described in deliverable D2.3. Moreover, partners involved in Task 2.3 also contributed with respect to the INFINITECH background technologies that will be included in deliverable D2.5. Finally, Task 2.4, deliverable D.2.7 Regulatory and compliance requirements for each pilot (e.g., GDPR, PSD2, MiFiD, 4MLD, or others) and Deliverable 2.13 as part of Task 2.7, will be used as input (see Figure 1 below):

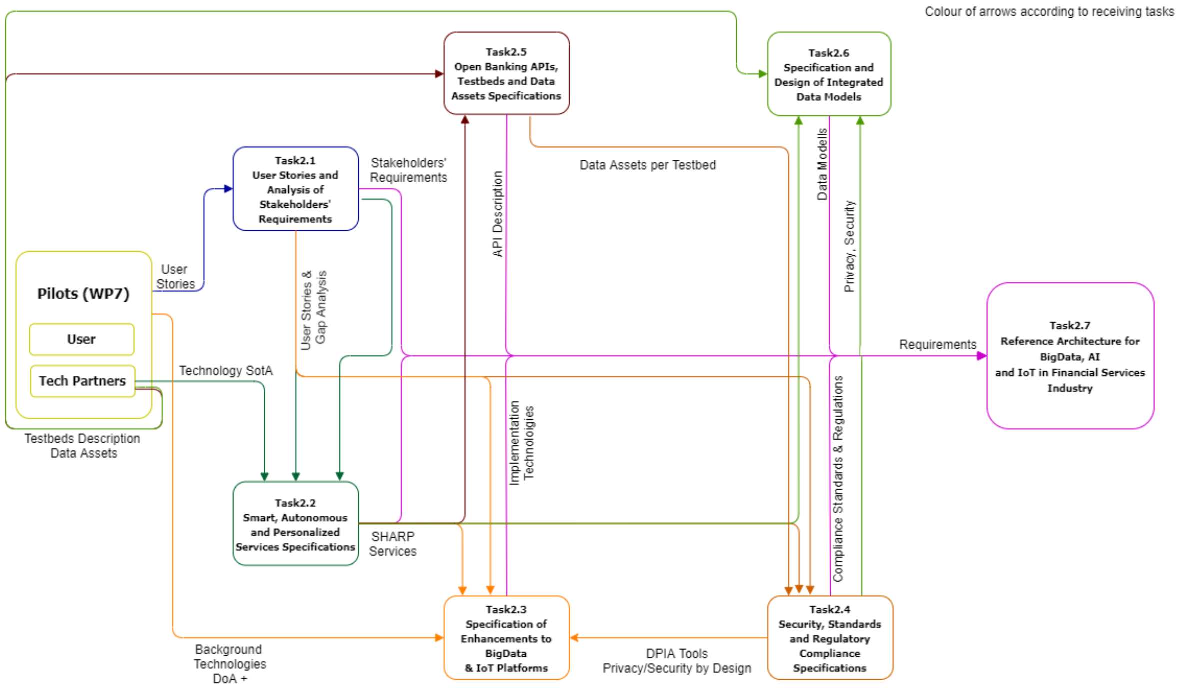

Figure 1– Schema of the links among Tasks

Also, the deliverable will rely on a feedback requested from all pilots to contribute regarding data assets that will be used for each pilot and the relative testbed host.

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 the methodology followed for the collection of relative information from all INFINITECH Partners included in the deliverable. Chapter 3 contains the Initial Specifications regarding Testbeds, Data Assets and Open APIs that will be implemented as part of all INFINITECH Pilots execution. Finally, Chapter 4 reports some conclusions.

2. Methodology

The major source of insights of deliverable D2.9 is the INFINITECH Pilots and their contributions to other deliverables of WP2:

- D2.1 User Stories and Stakeholders' Requirements - I
- D2.3 Reference Scenarios and Use Cases - I
- D2.5 Specifications of INFINITECH Technologies - I
- D2.7 Security and Regulatory Compliance Specifications – I

The Data Assets Specifications are gathered based on a relative spreadsheet that all INFINITECH Pilots updated with the relative information, described in Appendix B.

Testbed Initial Specifications are based on all INFINITECH Pilots contributions for deliverables D2.5 and D2.13 - INFINITECH Reference Architecture – I, that also is the basis for all Pilot implementation. The Initial Specifications for each Testbed are described in Appendix A.

The results of those contributions are explained in the following sections.

3. Testbeds, Data Assets, Open APIs, initial Specifications for the Pilots

3.1 Initial Testbed Specification

3.1.1 Testbed Definition

A testbed is an environment or a platform where the correct blend of components - operating system, servers, database, network configurations, browser installation and so on- are put to be used in order to verify the accuracy of a product being tested (see Figure 2).

Be it automation or manual testing, configuring the required components is an essential step to efficiently test an application under test, by incorporating the correct blend of features.

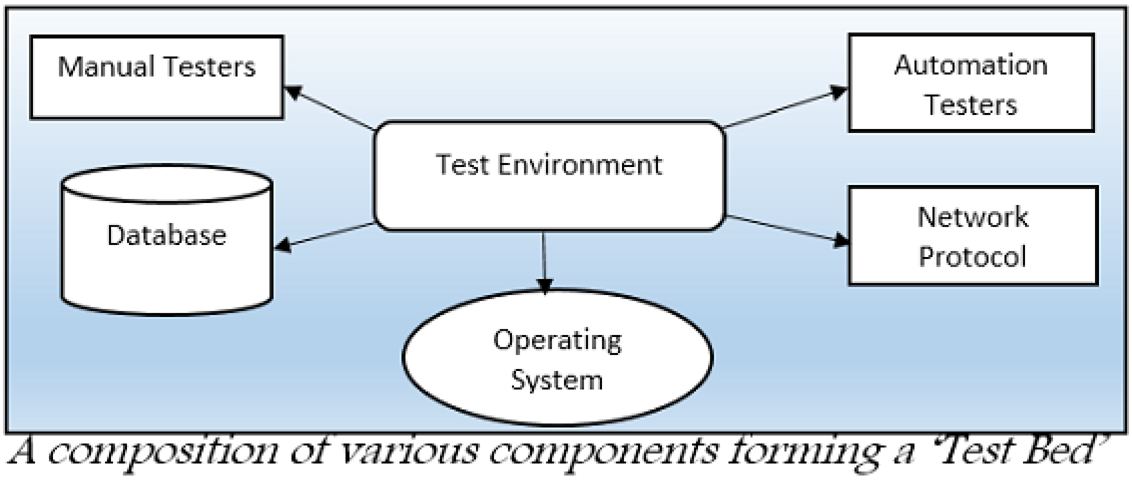

Figure 2 – A composition of various components forming a "Test Bed"

Design considerations for Test Environment:

Testbed - Challenges:

- **Setting up a Test Bed at a Remote Location:** The test environment could be at various locations, either on premise or cloud based, which means that it may not take place at a centralised location. Given the diversity of the modern environment, where businesses are deploying their plans irrespective of local boundaries, the need to pace up and work coherently has become a need of the hour.

  Thus, setting up a centralised environment despite the geographical boundaries is an imperative ladder to prosper.

- **Collaboration Among Teams:** In the context of software testing, developers, testers and business analysts form the core team working on a software development project. Therefore, assimilating resources in a centralised location is a great way to minimise cost, time and the overall effort required to complete a project's objectives.

- **Complex Test Configuration:** 'Test Configuration' takes into account all the components/artefacts that are essential in conducting tests for a system, including human skills in terms of technical knowledge. Employing the right people at the right place is what really matters to configure an apt test configuration.

- **Time Consuming:** Test case implementations may vary as they require various test configurations. The reason could be the availability of a wide range of technologies that are required to be applied to the current testing. The complexity lies in assimilating the different aspects together, as they must follow a consistent approach.

Environmental Configuration:

Examples of a test bed configuration is somewhat like the following:

- **Hardware:** On premise or Cloud Based (AWS, MS Azure, Google Cloud, Oracle Cloud, etc)

- **Operating System:** Unix, Linux, Windows Server, etc.

- **Database:** Oracle, IBM DB2, MS SQL Server, MongoDB, MariaDB, etc.

- **Web Server:** Apache Tomcat, IBM WebSphere, Oracle WebLogic, Microsoft IIS, etc.

- **Browser:** Internet Explorer, Microsoft Edge, Chrome, Firefox, Safari, etc.

- **Programming Languages:** Java, Python, C#, C++, etc.

- **Data Analysis Tools:** Apache Hadoop, Apache Spark, Python, etc.

Note:

The aforementioned list is just a sample depiction of a typical pattern of test bed configuration. Combination of all the components depends upon project specific requirements.

For INFINITECH Project Pilots Testbeds implementation will mainly refer to the following aspects:

a) **Cloud specific** infrastructure, instead of **on-premise** (depending of course on the specific requirements of each Pilot) (e.g. AWS, MS-Azure, etc)
b) **Application containerization (app containerization)** an OS-level virtualization method used to deploy and run distributed applications without launching an entire virtual machine (VM) for each application (e.g. Kubernetes, Docker, etc)

c) INFINITECH available **Datastore Technologies** (e.g. LXS Data Store)

d) INFINITECH **BigData, AI/ML Tools** developed

All the above will be provided to all Pilots as part of the INFINITECH Platform in order to achieve the highest level of similarity among Testbeds relative sandboxes. More details for **INFINITECH Testbeds, Data assets** and **relative technologies** are available in the next **chapter**, as well as in more detail in **Appendix A and B** of this deliverable.

More details regarding the Testbeds Implementation will be available into the upcoming WP6 deliverables (in particular D6.1 and D6.4).

3.1.2 Testbeds concept of INFINITECH Project

The INFINITECH project aims to the **development** of **BigData, IoT and AI-based innovations** requiring significant testing and validation efforts, such as testing for regulatory compliance and optimizing Machine Learning (ML) and Deep Learning (DL) models. Therefore, there is a need for advanced **experimentation infrastructures (testbeds & sandboxes)**, which shall provide access to resources for application development and experimentation, such as datasets, regulatory tools, libraries of ML/DL algorithms, Open APIs and more. Such experimentation infrastructures should be available in appropriate testbeds, based on the deployment of the technical building blocks in various configurations.

INFINITECH will provide **10+2 testbeds** for experimentation, testing and validation of BigData and IoT applications in the financial and insurance sectors, including:

(i) **Ten testbeds (10)** that will be established in incumbent financial organizations of the consortium **and**

(ii) **Two testbeds (2)** that will be established and made available to Financial/FinTech/InsurTech enterprises of the consortium for their pilots

Each one of the INFINTECH experimental infrastructures will comprise the following elements:

a) **Open APIs** (see section 3.2.1 below) for accessing data assets and the INFINITECH data management, analytics, data governance, interoperability and data exchange building blocks**;**

b) **Data assets** for experimentation**,** notably **real anonymized datasets and synthetics/simulated datasets;**

c) **Regulatory tools** for ensuring the compliance of innovative developments with **regulations (e.g., 4MLD, GDPR, PSD2, MiFID2);**

d) Access to a library of **ML/DL algorithms** for finance and insurance applications

A key innovation of INFINITECH is that it will provide the means for provisioning and configuring **tailored sandboxes** over the **project's testbeds**, which will comprise specific data sources, ML/DL algorithms, APIs and regulatory compliance algorithms. **The INFINITECH sandboxes and testbeds** will facilitate innovators in their efforts to produce BigData/ IoT applications that disrupt the sector based on their **SHARP (Smart, Holistic, Autonomy, Personalized and Regulatory Compliance)** properties.

The set of hardware resources of each Data Centre (Storage, Compute and Network) will be considered as a **testbed**, as shown in Figure 3 below.

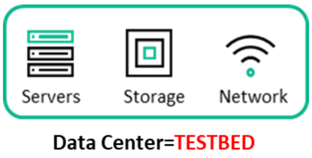

Figure 3- Testbed

INFINITECH project will deliver **15 Pilots**: 10 out of 15 will be carried out on **dedicated Data Centres**, while the remaining 5 out of 15 will be carried out (the selection of the pilots and the location still needs to be defined) on the **NOVA Data Centres** (**shared Data Centres**).

Therefore the 15 pilots will be executed in 10+2=12 testbeds (see figure 4 below).

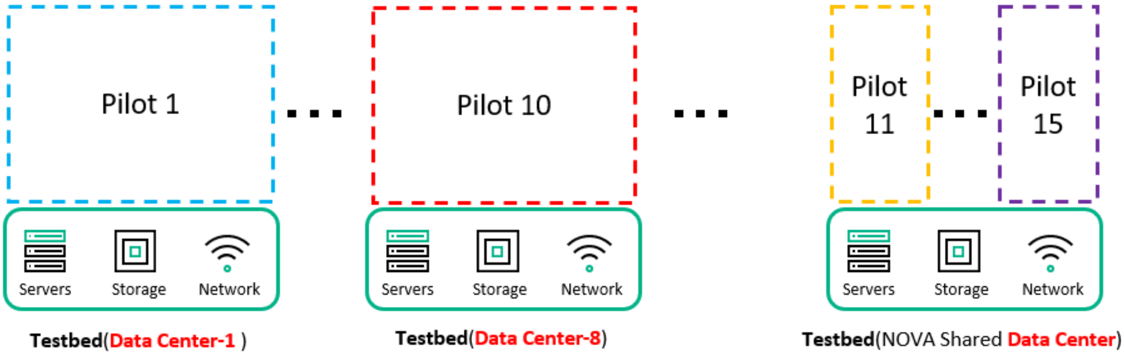

Figure 4- Pilots vs dedicated of shared testbeds

The following table 1 describes the **initial list** of the **testbeds and the pilots** that will be implemented as part of the INFINITECH Project:

Table 1 – INFINITECH Initial list of Testbeds & hosting Pilots

Testbed Host	Pilot No	Title	Leader
Category 1 - T7.2 Smart, Reliable and Accurate Risk and Scoring Assessment			
BANKIA (Spain)	Pilot 1	Invoices Processing Platform for a more Sustainable Banking Industry	BANKIA
NOVA (Portugal)	Pilot 2	Real-time risk assessment in Investment Banking	JRC
Category 2 - T7.3 Personalized Retail and Investment Banking Services			

BOI (Ireland)	Pilot 3	Collaborative Customer-centric Data Analytics for Financial Services	BOI
PRIVE (Austria)	Pilot 4	Personalized Portfolio Management ("Why Private Banking cannot be for everyone?")	PRIVE
LIB (Spain)	Pilot 5a	Smart and Personalized Pocket Assistant for Personal Financial Management	LIB
BOC (Cyprus)	Pilot 5b	Business Financial Management (BFM) tools delivering a Smart Business Advise	BOC
NBG(Greece)	Pilot 6	Personalized Closed-Loop Investment Portfolio Management for Retail Customers:	NBG
Category 3 - T7.4 Predictive Financial Crime and Fraud Detection			
Selected Bank	Pilot 7	Operation Whitetail – Avoiding Financial Crime	FTS
BOS(Slovenia)	Pilot 8	Platform for Anti Money Laundering Supervision (PAMLS)	BOS
AKTIF (Turkey)	Pilot 9	Analyzing Blockchain Transaction Graphs for Fraudulent Activities	AKTIF
ENG (Italy)	Pilot 10	Real-time cybersecurity analytics on Financial Transactions' BigData	PI
Category 4 - T7.5 Personalized Usage-Based Insurance Pilots			
NOVA (Portugal)	Pilot 11	Personalized insurance products based on IoT connected vehicles	ATOS
NOVA (Portugal)	Pilot 12	Real World Data for Novel Health-Insurance products	SILO
Category 5 - T7.6 Configurable and Personalized Insurance Products for SMEs and Agro-Insurance			
NOVA (Portugal)	Pilot 13	Alternative/automated insurance risk selection - product recommendation for SME	WEA
NOVA (Portugal)	Pilot 14	BigData and IoT for the Agricultural Insurance Industry	GEN

For each Testbed, the initial specifications are described in **Appendix A.**

3.1.3 INFINITECH Sandboxes

Each pilot will have one or more Use Cases, based on the user stories already described in Deliverable *D2.1 - User Stories and Stakeholders' Requirements – I* (implemented by one or more **pilot Apps – Technology blocks**), each Use Case will be a **Sandbox**. A **Sandbox** is planned that will be orchestrated by means of underlying technology like Kubernetes (www.kubernetes.io), by the use of Kubernetes Namespace.

In fact, the Kubernetes Namespace feature allows to logically isolate the objects (mainly PODs) from other Namespaces. So, each (10 out of 15) **dedicated testbeds** will only have one Kubernetes cluster with as many Namespaces as the number of Use Cases (**Sandboxes**) to be implemented for a single pilot (see Figure 3). In the other case, each (5 out of 15) **shared testbed** will have one Kubernetes cluster for each pilot it has to host and manage. Each Use Case (**Sandbox**) could have one or more **pilot Apps – Technology blocks** and each App will be a POD.

A **POD** is the **basic execution unit** of a **Kubernetes** application--the **smallest and simplest unit** in the Kubernetes object model that **you create or deploy**. A POD is a higher level of abstraction grouping containerized components. A POD consists of one or more containers that are guaranteed to be co-located on the host machine and can share resources. The basic scheduling unit in Kubernetes is a POD.

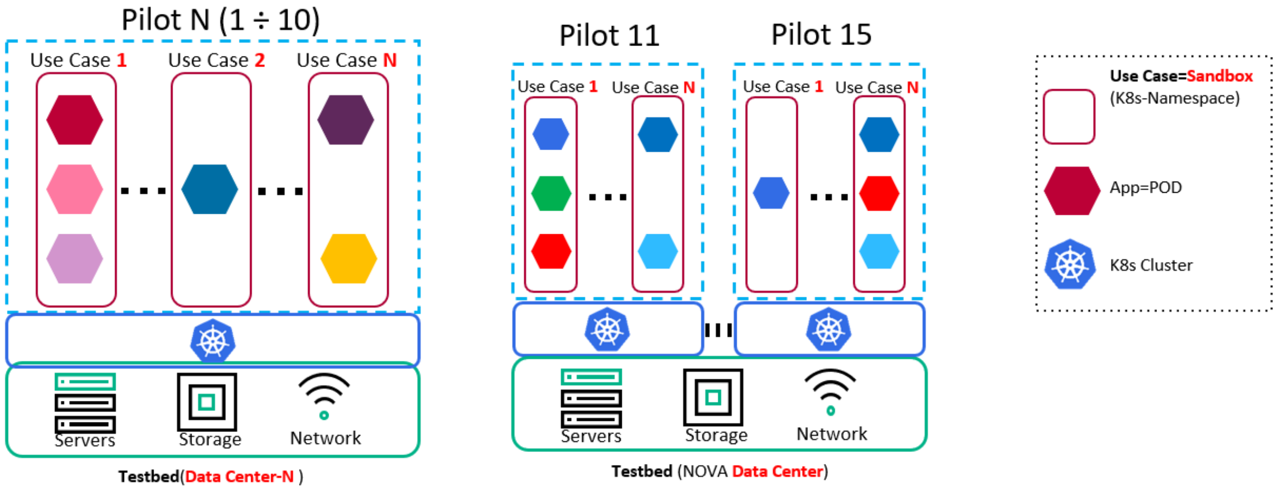

Figure 5 - Pilot Sandboxes in a dedicated or shared testbed

3.1.4 Initial Data Assets Specification

Based on the fact that the pilots will validate the INIFINITECH developments from both a technical/technological and a business/economic perspective, they will leverage readily available **BigData datasets** (i.e. data assets) that are available in the organizations of the consortium, **high velocity data from IoT devices** (i.e. connected cars, medical devices, smart phones), along with **alternative data** from a wide array of **open sources** like **news and social media**.

These data assets include **many millions of customer records**, **<u>billions of customer transactions</u>**, **<u>streams</u>**, **<u>billions of alternative data items</u>** (e.g., news and social media) and more. These data will be used in the pilots, while some of them will be also made accessible as part of the project's VDIH in order to facilitate the rapid development of BigData/IoT applications in finance/ insurance and to boost the innovation capital of SMEs (including FinTech/InsurTech).

The main categories of the Data assets that will be utilized for INFINITECH Pilots are the following:

a) **Invoice Data**

b) **Financial Market Data (Real Time or History)**

c) **Customer Info Data (Anonymized, Pseudo – Anonymized) – Retail & Corporate (SMEs)**

d) **Customer Behavioural Profile Data**

e) **Customer Risk Analysis Data**

f) **Customer investment Profile Data**

g) **Transactions Detail Data (Retail & Corporate)**

h) **Sentiment analysis Data (News & Social Feeds)**

i) **Financial market price data**

j) **Financial Instruments Characteristics Data**

k) **TARGET2/SEPA transactions Data**

l) **Local or International Business Registries Data**

m) **Bitcoin & Ethereum Blockchain Data**

n) **Black List Data (Individuals, Corporates)**

o) **Traffic & Vehicle IoT Data**

p) **Vehicle & Life Insurance Data**

q) **Health Data**

r) **SMEs Geolocation & Characteristics Data**

s) **Gridded Climate Indices Data (History)**

t) **Earth Observation IoT Data**

These data assets will be available using different sources, e.g. Data Lakes, Data Warehouses, Operational Databases, Social & Internet Feeds, Data extraction files in various formats, IoT Devices Data and other forms.

A description of the data assets of each pilot and test-bed, along with in-depth details is described in **Appendix B.**

3.2 Initial API Specification

3.2.1 API Specification Introduction

Nowadays, experimentation and testing of digital finance and FinTech applications takes place within specialized environments that are termed sandboxes. Sandboxes come typically with Open APIs that enable innovators and testers to mimic the characteristics exhibited by the production environment. They can operate on a real-time basis and help simulate responses from all the systems that are involved in delivering the service under test, which can facilitate pilot testing and reduce risks associated with novel technologies such as BigData and AI.

Open API (often referred to as a **public API**) is a publicly available application programming interface that provides developers with **programmatic access** to a **proprietary software application** or **web service**. **APIs** are **sets of requirements** that govern how one **application** can **communicate and interact with another**. APIs can also allow developers to access certain internal functions of a program, although this is not typically the case for web APIs. In the simplest terms, an API allows **one piece of software to interact with another piece of software**, whether within a single computer via a **mechanism** provided by the **operating system** or **over** an **internal** or **external TCP/IP-based** or **non-TCP/IP-based network**. In the late 2010s, many APIs are provided by organisations for access with HTTP. APIs may be used by **both developers inside the organisation** that published the API or by **any developers outside** that organisation who wish to **register for access** to the interface.

INFINTECH will provide, enhance, customize and validate in realistic pilots a range of novel technologies, which will facilitate financial and insurance organizations to innovate based on BigData, IoT and AI. These Pilots will utilize **Open APIs** for experimentation connected to each pilot, but also for supporting innovation beyond the specification pilot requirements (e.g., in the scope of hackathons). The Open APIs will be a customized version of the Open APIs of the INFINITECH technology enablers (i.e. the APIs for Data Management and Analytics).

Also, INFINITECH Project aims the establishment of a **market platform** and of an EU- **Virtualized Digital Innovation Hub (VDIH)**, which will offer financial/insurance organizations and innovators with a unique blend of innovation management services (e.g., training, consulting, development based on **Open APIs**) that will enable them to innovate with BigData and IoT in the finance sector.

3.2.2 API Design Best Practices

API design is the collection of planning and architectural decisions you make when building an API. Your basic API design influences how well developers are able to consume it and even how they use it. Just like website design or product design, API design informs the user experience. Good API design principles meet initial expectations and continue to behave consistently and predictably. Our API design guide assists in supporting this theme throughout your API design process.

There is not a single approach on how to design good APIs “the right way.” Instead, we need to lean on good industry basic API design guidelines, best practices and patterns where relevant, then take cues from those who will use our APIs.

The **Open APIs** specification that will be available through the INFINITECH Pilots will be based on the Open API Initiative (https://www.openapis.org/).

The **OpenAPI Specification (OAS)** defines a standard, programming language-agnostic interface description for REST APIs, which allows both humans and computers to discover and understand the capabilities of a service without requiring access to source code, additional documentation, or inspection of network traffic. When properly defined via OpenAPI, a consumer can understand and interact with the remote service with a minimal amount of implementation logic. Similar to what interface descriptions have done for lower-level programming, the OpenAPI Specification removes guesswork in calling a service.

OpenAPI allows you to define how your REST API works, in a way that can be easily consumed by both humans and machines. It serves as a contract that specifies how a consumer can use the API and what responses you can expect.

OpenAPI v3.0 was released in July 2017, by the **OpenAPI Initiative**, a consortium of member companies who want to standardize how REST APIs are described. There are various other approaches to API description:

- **OpenAPI v2**, (formerly known as **Swagger v2.0**) is still widely used, but increasingly being replaced by OpenAPI v3.0;
- **JSON Schema**, very similar to OpenAPI, but able to describe any JSON-like data, not just APIs;
- **RAML**, the RESTful API Modeling Language, focuses on the planning stage of API design;
- **Web Services Description Language (WSDL),** an XML-based interface description language that is used for describing the functionality offered by a web service.**;**
- **Web Application Description Language (WADL)**, a machine-readable XML description of HTTP-based web services;
- **Open Data Protocol (OData)**, an open protocol which allows the creation and consumption of queryable and interoperable REST APIs in a simple and standard way. Microsoft initiated OData in 2007. Version 4.0 was standardized at OASIS and in April 2015 OASIS submitted OData v4 and OData JSON Format v4 to ISO/IEC JTC 1 for approval as an international standard;
- **RESTful Service Description Language (RSDL),** a machine- and human-readable XML description of HTTP-based web applications (typically REST web services);

While OpenAPI v3.0 is the way forward, each of these alternative formats has tooling associated. Pilots implementors may find themselves converting between them, especially OpenAPI v2.0, until the tools catch up.

Your API design requires a way to define how the API will be used. The future-thinking approach is to select OpenAPI v3.0 to describe your API.

3.2.3 Open APIs Identified from INFINITECH Platform

The details of the Open APIs that will be available from all INFINITECH Platform will be described on a later version of this deliverable, based on the details of the technology building blocks that will be used for each pilot and being described *in D2.5 - Specifications of INFINITECH Technologies – I.* Also, this section will be enhanced based on the feedback of Deliverable D5.10 of Task 5.5 (WP5), that will include specifications of Open APIs for accessing the added-value analytics functionalities of INFINITECH such as the declarative analytics functionalities, as well as the ML/DL algorithms of the project.

4. Conclusions and next steps

The scope of task T2.5 is the initial specification of the advanced experimentation infrastructures (testbeds & sandboxes) which shall provide access to resources for application development and experimentation of BigData, IoT and AI-based innovations, as well as the specifications of relative data assets, regulatory tools, libraries of ML/DL algorithms, Open APIs and more, that will be implemented as part of the INFINITECH-RA.

In particular, **the initial specifications** of testbeds and relative **data assets** used for all INFINITECH Pilots that will be hosted, are described in Section 3, including also the initial specifications about the technologies (e.g., BigData/IoT, AI/ML toolkits, HPC infrastructures) will be used to realize and guide the implementation and integration of INFINITECH Pilots.

Overall, the INFINITECH Testbeds, Data assets and Open APIs initial requirements reflect the State of the Art of the application of BigData, IoT and AI in the Financial Services and contribute to the latest trend, that INFINITECH-RA will envisage as part of all pilots and technologies integration.

Within this deliverable, based on the initial feedback from all Pilots, the initial specifications for advanced **experimentation infrastructures (testbeds & sandboxes)**, which shall provide access to resources for application development and experimentation, such as datasets, regulatory tools, libraries of ML/DL algorithms, Open APIs and more, are described.

As we are running the initial phases of the INFINITECH Project the Open APIs that will be available, will be described in the next version of the deliverable.

Based on the outcomes of this deliverable, the next steps of task T2.5 are:

- Specify in more detail the INFINITECH Testbed hosts, Data Assets, Technologies, Regulatory Tools, based also on the feedback from D2.5, D2.7 and D2.13, for supporting resources for application development and experimentation of all INFINITECH Pilots

- Based on the INFINITECH Technologies that will be implemented as part of project, more detailed Open API specifications for usage of BigData, IoT and AI technologies in the Financial Services will be designed and implemented.

Appendix A: Testbed Initial Specifications

#	Testbed Host	Pilot	Testbed Description	Lead Partner	Cloud/Physical (On premise)	Technologies	Estimated Volume pf Data Assets
1	BANKIA (Spain)	Pilot 1	Invoices Processing Platform for a more Sustainable Banking Industry	BANKIA	Cloud (AWS Bankia Private Cloud)	- Data management: Ignite, elastic search - Data processing: kafka, bert, setokeeper - Data analytics: python - Data visualization: kibana,Floent, prometheus	2TB
2	BOI (Ireland)	Pilot 3	Collaborative Customer-centric Data Analytics for Financial Services	BOI	Physical (On premise)	- Data management: tbd - Data processing: tbd - Data analytics: tbd - Data visualization: tbd	tbc
3	LIB (Spain)	Pilot 5a	Smart and Personalized Pocket Assistant for Personal Financial Management	LIB	Data for Pilot 5a are Confidential		
4	BOC (Cyprus)	Pilot 5b	Business Financial Management (BFM) tools delivering a Smart Business Advise	BOC	Data for Pilot 5b are Confidential		
5	NBG (Greece)	Pilot 6	Personalized Closed-Loop Investment Portfolio Management for Retail Customers:	NBG	Cloud - MS Azure or IBM Cloud	- Data management: Leanxcale Datastore, - Data processing: Icarus (Ubitech), - Data analytics: Python Libraries, - Data visualization : Express, Node JS, Kong, Nginx	- CRM Data: ~500 Mb, - Deposit Account Transactions:10-20 Gb, Cards -Transactions:10-20 Gb, Instruments Historical Prices:~500 Mb, - Investment Related - Transactions:~500 Mb, - Instruments Characteristics:~100 Mb, - News feeds & Blogs (tbd)

#	Testbed Host	Pilot	Testbed Description	Lead Partner	Cloud/Physical (On premise)	Technologies	Estimated Volume pf Data Assets
6	Selected Bank	Pilot 7	Operation Whitetail – Avoiding Financial Crime	FTS	tbd	- Data management: Open Source extraction Tool (tbd), - Data processing: Anonymization Tool (GRAD), - Data analytics: ML/Python Libraries, - Data visualization: Open Souce Visualization Tool (tbd)	Advanced Customer Data Lake
7	BOS (Slovenia)	Pilot 8	Platform for Anti Money Laundering Supervision (PAMLS)	BOS	Data for Pilot 7 are Confidential		
8	AKTIF (Turkey)	Pilot 9	Analyzing Blockchain Transaction Graphs for Fraudulent Activities	AKTIF	Data for Pilot 8 are Confidential		
9	ENG (Italy)	Pilot 10	Real-time cybersecurity analytics on Financial Transactions' BigData	PI	tbd	- Reguratory Tools: Data Protection Orchestrator (DPO) - ATOS, eIDAS Integration (SPEiDI) - ATOS, - Anonymization tool - GRAD, Data Check-in mechanism - UBI, - Data management: ALIDA (ENG), - Data processing: ALIDA (ENG), - Data analytics : ALIDA(ENG), Data visualization : ALIDA(ENG), - Pseudonymization tool: (?)	3.5TB /Year

#	Testbed Host	Pilot	Testbed Description	Lead Partner	Cloud/Physical (On premise)	Technologies	Estimated Volume pf Data Assets
10	NOVA (Portugal)	Pilot 2	Real-time risk assessment in Investment Banking	JRC	Physical (On premise)	- Data management: Portfolio input-stream (tbd), Data windows (tbd), - Data processing: Correlation matrix (Java), Scenario generation (Java) - Data analytics: Analytics-montecarlo simulations (Python Libraries), Market sentiment extraction(tbd), - Data visualization: Numeric values and graph(s) (tbd)	1,2TB + 50GB/Day
	NOVA (Portugal)	Pilot 11	Personalized insurance products based on IoT connected vehicles	ATOS	Physical (On premise)	- Data processing:Context Broker (Orion) [NGSI & ETSI NGSI-LD],Historical Data (Quantumleap),[ETSI NGSI-LD], IoT-Agents [NGSI, MQTT, HTTP, UL2.0, JSON] - Data analytics: EASIER-AI (TensorFlow, Elasticsearch, Keras, Kibana) - Data visualization: ML/AI technologies and models created by EASIER.AI	- Simulated Urban Mobility Dataset: This is on-demand data streaming CAN Data (Historical Data):Estimated 368 GB, - Traffic Events (Historical data): Estimated 900 GB, - NMEA Data for vehicles (Historical): Estimated 120 GB, - CAN Signals (Live): Estimated 150 GB, - Traffic Events (Live) :Estimated 250 GB, - NMEA Data for vehicles (Livel):Estimated 50 GB, - Motor Insurance Data : Estimated 500 MB
	NOVA (Portugal)	Pilot 12	Real World Data for Novel Health-Insurance products	SILO	Physical (On premise)	- Anonymization tool - GRAD, - Data processing: Model subject data and simulate synthetic data from the models,	- Healthentia (Live) Average 720kB per user per week - Helathentia (simulated): Average 720kB per user per week

						- Data analytics: Python libraries such as Tensorflow and Keras, - Data visualization: JS (most probably ECharts library)	
	NOVA (Portugal)	Pilot 13	Alternative/automated insurance risk selection - product recommendation for SME	WEA	Physical (On premise)	- Data management: Leanxcale Datastore in Nova - Data processing: Wenalyze AWS - Data analytics: javascrip, Json - Data visualization: Node JS,	500GB
	NOVA (Portugal)	Pilot 14	Big Data and IoT for the Agricultural Insurance Industry	GEN	Physical (On premise)	- Data management: THREAD Data Server, - Data processing: Weather Intelligence, Octopush (AGROAPPS), - Data analytics: Python Libraries, - Data visualization: Geoserver, Chart JS,	500 GB/ dailly
11	PRIVE (Austria)	Pilot 4	Personalized Portfolio Management ("Why Private Banking cannot be for everyone?")	PRIVE	AWS Cloud	- Data management: tbd - Data processing: tbd - Data analytics: tbd - Data visualization: tbd	tbc

Appendix B: Data Assets Details

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#1 BANKIA	Notary invoices	* 32.300 invoices documents, from 3.000 different Notaries * Dataset TableBank: Table Benchmark for Image based Table Detection and Recognition, 500.000 documents	BANKIA	Confidential data (Notary invoices) & Public Data (TableBank)	No	Not sure	PDF/Image /Text	PDF / PNG / TXT	Apache Hadoop HDFS, Elastic Search
Pilot#2 JRC	real time financial market data	price data for the most liquid Forex, Stocks, Stock Indices and Derivatives	JRC	proprietary data from provider	No	Yes	numeric	CSV	MySQL
Pilot#2 JRC	Derived analysis data	risk measures, correlation matrices	JRC	open	No	Yes	numeric	CSV	
Pilot#2 JRC	Existing historical data	price tick data for the most liquid Forex, Stocks, Stock Indices and Derivatives	JRC	proprietary data from provider	No	Yes	numeric	CSV	MySQL
Pilot#2 JRC	News articles data	A database of 1.5 billion news articles from 95,654 global news sources will be also used in terms of alternative sources.	JRC	open	No	No	Text	TXT	

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#3, BOI	Synthetic Customer - Bank A (<100 Records)	Small dataset of manually produced data to mimic profiles and characteristics of banking customers.	BOI	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#3, BOI	Synthetic Applicant - Bank A (<100 Records)	Small dataset of manually produced data to mimic profiles and characteristics of banking applicants/prospects.	BOI	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#3, BOI	Synthetic Customer - Bank B (<100 Records)	Small dataset of manually produced data to mimic profiles and characteristics of banking customers.	BOI	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#3, BOI	Synthetic Applicant - Bank B (<100 Records)	Small dataset of manually produced data to mimic profiles and characteristics of banking applicants/prospects.	BOI	N/A	N/A	Yes	Text	CSV or JSON	TBD

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#3, BOI	Synthetic Customer - Non-Bank Organisation (<100 Records)	Small dataset of manually produced data to mimic profiles and characteristics of a Non-banking customer or prospect.	BOI	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#3, BOI	Synthetic Applicant - Non-Bank Organisation (<100 Records)	Small dataset of manually produced data to mimic profiles and characteristics of a Non-banking customer or prospect.	BOI	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#3, BOI	Synthetic Account (<200 Records)	Small dataset of manually produced data to mimic accounts of banking customers.	BOI	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#3, BOI	Synthetic Transaction Data (<500 Records)	Small dataset of manually produced data to mimic transcations on accounts of banking customers.	BOI	N/A	N/A	Yes	Text	CSV or JSON	TBD

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#3, BOI	Consent Records (<100 Records)	Small dataset of manually AND/OR POC produced data to mimic history data sharing consent transactions between sharing parties.	BOI & Pilot 3	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#3, BOI	Consent History (<100 Records)	Small dataset of manually AND/OR POC produced data to mimic the history of data sharing consent transactions between sharing parties.	BOI & Pilot 3	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#3, BOI	Data Sharing Log (Parties, Metadata and Data Types) (<500 Records)	Small dataset of manually AND/OR POC produced data to mimic of data sharing transactions between sharing parties. Maybe stored on Blockchain or as GraphDB (TBD).	BOI & Pilot 3	N/A	N/A	Yes	Text	CSV or JSON	TBD

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#3, BOI	Synthetic Customer Input Data. (<1000 Records)	Small dataset of manually AND/OR POC produced data to mimic minimal/basic data sharing application customer/user data (Data TBC).	BOI & Pilot 3	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#3, BOI	Optional Storage of Shared Data Set	TBD - Dependent on design. Duplication of records from other datasets.	BOI & Pilot 3	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#3, BOI	Invite to Share Data Set	TBD - Dependent on design. Proposition dataset based on consnet to analyse data sharing log.	BOI & Pilot 3	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#3, BOI	Ref. Data. (<1000 Records) (for standardisation of data)	Very small dataset of manually produced reference data to support data sharing application(s).	BOI & Pilot 3	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#3, BOI	Customer Linking Data (e.g. iD, Mobile No.) (<1000 Records)	Customer Keys required for data sharing application.	BOI & Pilot 3	N/A	N/A	Yes	Text	CSV or JSON	TBD

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#3, BOI	Synthetic or Open Organisational Data e.g. Banks Credentials (TBC)	Very small dataset of manually producedorganisatoin creditionsl data to support data sharing application(s) including api authenticaton.	BOI	N/A	N/A	Yes	Text	CSV or JSON	TBD
Pilot#4 PRIVE	Customer Transaction dataset	Customer securities and cash transactions through their deposit accounts	PRIVE	Confidential data	Yes	Not sure	Text/ Numeric	CSV	
Pilot#4 PRIVE	Financial market price data	Price data for Stocks, Bonds, Mutual Funds and or other assets like certificates/warrants	PRIVE	Open, partially license agreements with data providers needed	No	No	Text/ Numeric	TXT	
Pilot#4 PRIVE	Financial market asset master data	Asset related characteristics (e.g. expiration date, minimum investment amount, asset class breakdowns)	PRIVE	Open, partially license agreements with data providers needed	No	No	Text/ Numeric	TXT	
Pilot#4 PRIVE	Customer Risk Profile Data	Customer Risk Profile Data through their account data and profiling, based on B2B customers parameters	PRIVE	Confidential data	Yes	Not sure	Text/ Numeric	CSV	

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#4 PRIVE	Mutual Fund, ETF and Structured Products Breakdown	Asset Breakdowns based on bank data or market data providers breakdown	PRIVE	Open/Confidential data, partially license agreements with data providers needed	No	Not sure	Text/ Numeric	CSV	
Pilot#4 PRIVE	Customer EcoNomic Outlook	Customer EcoNomic Outlook data based on questionnaire engine	PRIVE	Confidential data	Yes	Not sure	Text/ Numeric	CSV	
Pilot#4 PRIVE	Account & Investors data	19484 accounts for about 15400 investors (live data) 94.407 different securities available; Investors serviced by 309 different advisor companies; Accounts in 28 different custodian banks (Data from 2019)	PRIVE	Confidential data	Yes	No	Text	TXT	
Pilot#4 RB	News articles data	5000 articles per day (future) plus 4 million articles (existing) GR News	RB	Open	No	Yes-limited	Text	TXT	Apache Hadoop HDFS, Elastic Search

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#4 RB	News articles data	Up to 1.2 Million Articles per day from Global News Sources (ENG Language)	RB	Open	No	Yes-limited	Text	TXT	Apache Hadoop HDFS, Elastic Search
Pilot#5a LIB	Pilot 5a Data Assets Information is Confidential								
Pilot#5b BOC	Pilot 5b Data Assets Information is Confidential								
Pilot#6 NBG	CRM Data	Customer related data like demographics, product ownership and responses to MIFID questionnaires.	NBG	Confidential data	Yes	Not sure	Text/ Num eric	TXT/CS V	Not yet decided
Pilot#6 NBG	Deposit Account Tra nsactions	Customers' transactions through their deposit accounts	NBG	Confidential data	Yes	Not sure	Text/ Num eric	TXT/CS V	Not yet decided
Pilot#6 NBG	Cards Transactions	Customers' transactions through their cards	NBG	Confidential data	Yes	Not sure	Text/ Num eric	TXT/CS V	Not yet decided
Pilot#6 NBG	Instruments Historical Prices	Historical prices of investment instruments	NBG	Open	No	Not sure	Text/ Num eric	TXT/CS V	Not yet decided
Pilot#6 NBG	Investment Related Transactions	Customers' transactions related to investment products	NBG	Confidential data	Yes	Not sure	Text/ Num eric	TXT/CS V	Not yet decided

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#6 NBG	Instruments Characteristics	Detailed characteristics for all instruments that will be considered in the pilot, including asset class, currency, ISIN, maturity etc.	NBG	Open	No	Not sure	Text/ Numeric	TXT/CSV	Not yet decided
Pilot#6 NBG	News feeds & Blogs	Unstructured data collected from public news feeds & Blogs	NBG/RB	Open	No	Not sure	Text	TXT/CSV	Not yet decided
Pilot#7 FTS	Pilot 7 Data Assets Information is Confidential								
Pilot#8, BOS	Pilot 8 Data Assets Information is Confidential								
Pilot#9, BOUN	Bitcoin Blockchain Data	Bitcoin transfers (send);	Open data	Public blockchain data	Yes	Not sure	Text/ Numeric	TXT	blockchain files
Pilot#9, BOUN	Ethereum Blockchain Data	Ether transfers (send); ERC20 Token Smart contract transactions (35 popular tokens including stable coins like EURS, GUSD, USDT, TRYB, PAX,TUSD, QCAD, XAUT)	Open data	Public blockchain data	Yes	Not sure	Text/ Numeric	TXT	blockchain files

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#9, AKTIF	Blacklisted Addresses Data	Small list of blockchain addresses that will be obtained from Internet by manual search for hacked/fraudulent accounts as well as compiled internally from problematic customers. Bitcoin & Ethereum addresses.	Open/AKTIF	Open/Confidential	Yes	Not sure	Text	TXT	csv file
Pilot#10 PI	Financial transactions Data	Financial transactions in operational systems	PI	Confidential	No	Yes	Text/ Numeric	CSV/ORC	HDFS (batch); REDIS/CASSANDRA (streaming)
Pilot#11, ATOS	Simulated Urban Mobility Dataset	Simulated Urban mobility data (mainly vehicles CAN Signals) through different scenarios (cities). Captured from SUMO tool)	ATOS	Open	N/A	Yes	Text	JSON	Context Information (Orion CB - NGSI) and Historical data (FIWARE QL)

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#11, CTAG	CAN Data (Historical Data)	Data collected from vehicle's CAN Bus (80 vehicles driving 4 h/day 1 year). Historical data coming from existing deployments	CTAG	Confidential	Yes	Yes	Text	CSV	Mongo DB (sugges ted)
Pilot#11, CTAG	Traffic Events (Historical data)	Traffic events published by the city of Vigo and DGT (Historical data related to captured CAN Data)	CTAG	Open	N/A	Yes	Text	JSON	Mongo DB (sugges ted)
Pilot#11, CTAG	NMEA Data for vehicles (Historical)	Complementary location (GPS, Timestamp, speed, heading...) for Vehicles' CAN Data (Historical data related to captured CAN Data)	CTAG	Confidential	Yes	Yes	Text	CSV	Mongo DB (sugges ted)
Pilot#11, CTAG	CAN Signals (Live)	CAN data + Driving style info (revolutions, gear, hard breaking...)+ Parking (close doors, windows...) + Maintenance	CTAG	Confidential	Yes	Yes	Text	CSV	Mongo DB (sugges ted)

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#11, CTAG	Traffic Events (Live)	Traffic events published by the city of Vigo and DGT	CTAG	Open	N/A	Yes	Text	JSON	Mongo DB (sugges ted)
PIlot#11, CTAG	NMEA Data for vehicles (Live)	Complementary location (GPS, Timestamp, speed, heading...) for Vehicles' CAN Signal	CTAG	Confidential	N/A	Yes	Text	CSV or JSON	Mongo DB (sugges ted)
Pilot#11, DYN	Motor Insurance Data	Data concerning motor insurance including data from the policies (duration, covers), data from vehicles (licence No, VIN etc.) and data from drivers (age, experience etc.)	DYN	Confidential & Open	Yes	No	Text	CSV	
Pilot#12, iSprint	Healthentia Live	Measured physical activity (steps, floors, sleep and heart rate) and user reported data from users of Healthentia SaaS who have given consent	iSprint	Confidential	Yes	Models trained for simulator	Text	JSON	MySQL
Pilot#12, iSprint	Healthentia Simulated	Simulated physical activity and reported data	iSprint	Open	N/A	Already synthetic	Text	JSON	MySQL

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anony mized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#13, WEA	SMEWIF	SMEs website information and functionalities	WEA	Confidential	N/A	No	Text	S3 / Dynamo DB	AWS
Pilot#13, WEA	ROPS	Review and opinions platforms	WEA	Confidential	N/A	No	Text	S3 / Dynamo DB	AWS
Pilot#13, WEA	EUBD	European SMEs Business Directories	WEA	Confidential	N/A	No	Text	S3 / Dynamo DB	AWS
Pilot#13, WEA	GIO	SMEs geolocation information and characteristics	WEA	Confidential	N/A	No	Text / Image	S3 / Dynamo DB	AWS
Pilot#13, WEA	SMSIP	Social media SMEs information and presence	WEA	Confidential	N/A	No	Text	S3 / Dynamo DB	AWS
Pilot#13, WEA	I&R	Key performance indicators and insurance needs	WEA	Confidential	N/A	No	Text / Image	S3 / Dynamo DB	AWS
Pilot#14, GEN	Gridded Climate Indices (1/1/1979 to 31/12/2019)	Climate Indices based on the ERA-5 Land and ERA-5 Reanalysis Data	AgroApps	Bilateral agreement to issue username/password	N/A	Yes	3D Gridded Data	NETCDF-CF	THREADDS Data Server
Pilot#14, GEN	EO Data	Earth Observation Data (Sentinel-1,2,3/LandSat-8, MODIS, PROBA-V) for remote damage and crop loss assess<ment	AgroApps	Bilateral agreement to issue username/password	N/A	Yes	Geotiff	GeoJSON	GeoServer

INFINITECH Pilot# Dataset Provider	Dataset Name	Dataset (short) description	Owner	License/ Privacy	Anonymized	Capability of Synthetic Data Production	Data Type	Data format	Data store
Pilot#14, GEN	Numerical Weather Predictions	Very High-Resolution Weather Predictions for the Pilot Areas	AgroApps	Bilateral agreement to issue username/password	N/A	Yes	4D Gridded Data	GRIB-2, NETCDF-CF	THREADDS Data Server