IN-HABIT - INclusive Health And wellBeing In small and medium size ciTies

D7.2 IN-HABIT Data Platform

Project Number	869227	**Acronym**	IN-HABIT
Full Title	INclusive Health And wellBeing In small and medium size ciTies		
Project URL	https://www.inhabit-h2020.eu/		
Document Type and Name	Deliverable, D7.2 IN-HABIT Data Platform		
Project Coordinator	University of Cordoba		
Project Call and Funding Scheme	SC5-14-2019 - Visionary and integrated solutions to improve well-being and health in cities H2020-SC5-2019-2 (IA)		
Date of Delivery	M15 – Date 30/11/2021		
WP, WP Leader	WP7, ISIM		
Status	Final version		
Dissemination level (confidentiality)	Confidential, only for members of the consortium (including the Commission Services)		
Authors (names and affiliations)	Jose Antonio Morales (WTG)		

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

EXECUTIVE SUMMARY

This D7.2 IN-HABIT Data Platform consists of the definition of the basic architecture over which the project's platform is based. It is the result of the initial 12 months of work in Task 7.2 IN-HABIT Data Platform, led by Wellness TechGroup.

Accordingly, the aim of the IN-HABIT Data Platform is to provide open and consistent data about the performance and impacts of deployed solutions in the four IN-HABIT cities, aiming to use them for both co-design and monitoring purposes in WPs 1-4, as well as for impact assessment purposes in WP7 and to share bidirectional information with the IN-HABITAPP.

The platform is designed to allow integration, management, and visualisation of data from various sources at city and pilot level (sensor and data bases), aiming to produce interactive scenarios, GIS mapping and data analysis in real time, as well as from other sources such as the IN-HABITAPP, the IN-HABIT website, and primary data collection carried out by project partners at city level.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

TABLE OF CONTENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

1. INTRODUCTION

The main objective of this deliverable is to describe the architecture of the IN-HABIT Platform to have a single big data platform that integrates and turns data from multiple sources, from assets, processes, persons, and external systems of the four pilot cites into actionable intelligence.

IN-HABIT Platform aims to unveil new insights from volumes of multi-structure data, both real-time and historical, enabling the development of Visionary and Integrated Solutions (VIS) to promote inclusive health and wellbeing (IHW) in small and medium-sized cities (SMSC). In each of the four pilot cities, the information collected into the platform will help to investigate how the mobilization of existing undervalued resources, such as culture and heritage, food, human-animal bonds, arts and environment, might contribute to boosting health and wellbeing, with a focus on gender, diversity, equity and inclusion (GDEI).

IN-HABIT Platform provides a single, holistic, robust, and scalable interface to manage the information and enable the analysis of operational KPIs, metrics and related contextual information correlating multiple data sources, delivering user-friendly analytics and dashboard-based monitoring of metrics, whilst supporting decision making processes concerned with the co-design, co-deployment, and co-management of the VIS in the four pilot cities.

2. IN-HABIT PILOT'S FIRST ANALYSIS

This section describes the process of dialogue and consultation with the city partners and WP leaders for the definition of the main data sources, KPIs and types of data that will be integrated by the Platform. This work has been done in several online meetings along these 12 first months of work, resulting in an excel file where these first insights were noted and used to elaborate the information in section 2.2.

Potentially interesting KPIs, and the inputs coming from D9.3 Ethical Requirements, were considered. This information is used to identify the interest of each pilot city and to provide guidance for the development of the services on the IN-HABIT Platform in the following months to achieve the committed goal of MS6 in month 18, where all the sensor network and legacy data must be defined, and of MS19 in month 30 when the platform, services and data collection systems have to be fully operational.

2.1. Common information and first approach to each pilot site

All pilot sites are interested in the continuous monitoring of the deployed solutions in terms of performance and impact on the following aspects:

- Changes in the use of the space and mobility patterns among inhabitants through interviews, satisfaction surveys, data collection, and mobile sampling experience.
- Attendance to events and influx of people and tourists in the public spaces through data collection.
- Perceived inclusiveness, accessibility, and safety of the re-designed areas through interviews, satisfaction surveys, and mobile sampling experience.
- Increased quality of the public environment, through interviews, satisfaction surveys, data collection, street cameras and sensors.
- Economic impacts of the initiatives for local enterprises through satisfaction surveys and primary data collection.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

- Number of new (social, educational, tourist, pet care) services, events and opportunities created by the solutions through satisfaction surveys and primary data collection.
- Increased participation and involvement of local citizens in the decision and management of the co-created urban commons through satisfaction surveys, interviews and focus groups with community representatives.

This monitoring action will also make use of the IN-HABIT APP (for participative Mobile Experience Sampling and gamified assessment) and the IN-HABIT Platform for data integration, scenario development and interactive visualization. Physical devices like street sensors and cameras will be deployed and used as data sources to monitor air quality, temperature, noise, people's movement, and traffic.

This common interest must be adapted to the specific needs of each pilot site, according to the following initial diagnosis:

Córdoba: Cultural and heritage hub

The aim of the pilot is to explore how culture and heritage affects inclusive health and wellbeing by developing green, sustainable, and creative spaces in a deprived neighbourhood and extrapolating previous experiences of "Los Patios" as socio-ecological resilience units to this area. Innovative solutions such as creative lighting, therapy gardens, employment opportunities and business initiatives using new gender and diversity dimensions and inclusive strategies could be further replicated to engage the whole city and beyond.

The pilot is interested in:

- Incrementing health and wellbeing in Córdoba.
- Monitoring weather condition in Córdoba, especially hot temperatures, and actions to mitigate it.
- Measuring spatial and temporal evolution of temperature, and how other parameters such as humidity and solar irradiance can contribute to this evolution
- Relative humidity of air and soil.
- Electrical consumption of HVAC systems.
- Air circulation in patios.
- How patios' architecture, vegetation and irrigation conditions can reduce temperature.
- Exploring other parameters that can affect to this climate change mitigation, such as automated irrigation, people presence, noise, and air quality.

Further analysis must be done to land the pilot concept and know better how to build the required KPIs and what data legacy can be leveraged. Regular meetings will be held in this regard from December 2021 to February 2022.

Riga: Multifunctional food hub

The pilot's goal is to set up a multifunctional food hub creating not only a sustainable food market but also a recreational and educational space integrating a wide range of physical activities for visitors of different ages and social groups. The first conclusions of the dialogue process are that the pilot is interested in:

- Knowing what kind of operations and services are used within the marketplace;
- Detecting the level of noise in the area;

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

- Measuring how the waste management is optimised as well as other market resources such energy use;
- Exploring online shopping services integrated in the IN-HABIT APP.

Deeper analysis will be carried out in further meetings to determine the services to be deployed and the KPIs generation. Regular meetings will be held in this regard from December 2021 to February 2022.

Lucca: Human-animal bonds hub

This pilot seeks to create the first Human-animal smart city in Europe by the deployment of animal path (such cyclable path) to reconnect the historic centre with the ancient walls and the surrounding areas, promoting the use of public spaces to bring human and animals together, involving city infrastructure and activities to improve human wellbeing.

The pilot is interested in:

- Using open data sources to create a community of knowledge;
- Exploring how behavioural games can affect to specific KPIs;
- Including cameras and adaptive lighting to gain security at night;
- Using filling level sensors in certain areas to monitor waste generation.

More details must be landed such as the usage of GPS tracking for pets of phones, and further steps must be done to define all the requirements. Regular meetings will be held in this regard from December 2021 to February 2022.

Nitra: Art and environmental hub

The concept is to create an open-source urban landscape along a cyclable road linking an industrial park with the main city. A series of movable multifunctional elements will be used to provide a wide range of activities along the way.

The pilot is interested in:

- Integrating data from two environmental stations;
- Monitoring people flow counting;
- GIS integration;
- Adaptive lighting;
- Integrating data from behavioural games.

The specific deployment and the associated services will be defined in following conversations, in regular meetings from December 2021 to February 2022.

2.2. Potentially interesting KPIs

As result of these first conversations a list of potentially interesting KPIs was created to be validated by each pilot site, which one could be considered interesting to retrieve data to elaborate on, considering some contextual and impact indicators associated that can condition or affect the KPI.

The following tables lists the potential KPIs, the associated indicators, the method to obtain the data, either from legacy data bases, sensors, surveys, or the IN-HABITAPP, and finally the potential interest of each pilot site to measure it.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

Potential Main KPIs	Associated indicators	Potential data source
Social cohesion	• Satisfaction with personal relationships in the neighbourhood	• Survey/IN-HABITAPP
Perception of security	• Sense of safety at night • Sense of safety in green areas • Perception of crime, violence or vandalism in the living area	• Survey/IN-HABITAPP • Survey/IN-HABITAPP • Survey/IN-HABITAPP
Social inclusion	• Contact with others in public spaces • Sense of inclusion	• Survey/IN-HABITAPP • Survey/IN-HABITAPP
Spatial wellbeing	• Satisfaction with urban green areas • Accessibility of local resources • Inclusiveness of public squares and green areas	• Survey/IN-HABITAPP • Survey/IN-HABITAPP • Survey/IN-HABITAPP
Physical health status	• Self-reported health status	• Survey/IN-HABITAPP
Determinant of health	• Consumption of fruits and vegetables	• Survey/IN-HABITAPP
Sports facilities	• Satisfaction with sports facilities • Practice of sports in public green areas • Benefits from sports	• Survey/IN-HABITAPP • Survey/IN-HABITAPP • Survey/IN-HABITAPP
Cultural consumption and production	• Satisfaction with cultural facilities • Participation in cultural activities within public spaces (outdoor/indoor) • Local cultural engagement	• Survey/IN-HABITAPP • Survey/IN-HABITAPP • Previous data
Leisure/free time	• Time spent playing, relaxing, or doing sports in public green areas • Time spent in social and recreational public spaces • Frequency of use of communal areas (public garden, community centre, etc) • Satisfaction with free time use	• Survey/IN-HABITAPP • Survey/IN-HABITAPP • Previous data (if any) • Survey/IN-HABITAPP
Financial situation	• Satisfactions with one's own financial situation	• Survey/IN-HABITAPP
Employment	• Employment rate • Activity/Participation rate	• Previous data (if any) • Previous data (if any)
Psychological wellbeing	• Life satisfaction perception	• Surveys
Air pollution	• Meteo station to CO2 detection	• Sensors
Temperature	• Meteo station to degree detection	• Sensors
Movements of persons and vehicles	• Vehicles and people flux indicators	• Sensors
Adaptive lighting	• Light intensity • Presence	• Sensors • Sensors

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

Noise	• Decibel detection	• Sensors
Wellness spots	• Citizen wellness spot perception	• IN-HABITAPP
Digital citizens/users	• Number of active users of digital apps	• IN-HABITAPP
Animals GPS location	• GPS tracking of animals	• Sensors
Behavioural games users	• Patterns of active users	• IN-HABITAPP
Online Market shop	• Level of online operations	• To be defined
Waste collection	• Filling level of waste containers and bins	• Sensors

Table 1. List of potential KPIs to measure

Associated indicators	**Córdoba** Culture and heritage hub	**Riga** Multifunctional food hub	**Lucca** Human-animal bonds hub	**Nitra** Art and environment hub
Satisfaction with personal relationships in the neighbourhood	X	X	X	X
Sense of safety at night	X	X	X	X
Sense of safety in green areas	X	X	X	X
Perception of crime, violence, or vandalism in the living area	X	X	X	X
Contact with others in public spaces	X	X	X	X
Sense of inclusion	X	X	X	X
Satisfaction with urban green areas	X	X	X	X
Accessibility of local resources	X	X	X	X
Inclusiveness of public squares and green areas	X	X	X	X
Self-reported health status	X	X	X	X
Consumption of fruits and vegetables	X	X	X	X
Satisfaction with sports facilities			X	X
Practice of sports in public green areas	X	X	X	X
Benefits from sports	X		X	X
Satisfaction with cultural facilities	X	X	X	X
Participation in cultural activities within public spaces (outdoor/indoor)	X	X		X
Local cultural engagement	X			X
Time spent playing, relaxing or doing sports in public green areas	X	X	X	X
Time spent in social and recreational public spaces	X	X	X	X
Frequency of use of communal areas (public garden, community centre, etc)	X		X	X
Satisfaction with free time use	X			
Satisfactions with one's own financial situation	X	X	X	X
Employment rate	X	X	X	X
Activity/Participation rate	X	X	X	X
Life satisfaction perception	X	X	X	X

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

Meteo station to CO2 detection				X
Meteo station to degree detection				X
Vehicles and people flux indicators	X	X	X	X
Light intensity	X			X
Presence for adaptive lighting	X			X
Decibel detection				
Citizen wellness spot perception				X
Number of active users of digital app	X			
GPS tracking of animals			X	X
Patterns of active users	X			
Level of online operations		X		
Filling level of waste containers and bins		X	X	

Table 2. Associated indicators by pilot site

This first approach allows to have some preliminary conclusions of the interest and possibilities of each pilot site.

2.3. Services for the platform.

According to the initial analyses carried out, some services have been already detected, the following table summarizes the initial services to consider for the platform.

SOURCE	POTENTIAL SOLUTION	BENEFITS
Third party	Environmental air quality sensors	Monitoring of the pollution in the pilot areas and the environmental impact of the solutions deployed
Third party	People counting sensors	Sensors deployed at ground level for people counting / traffic monitoring
Third party	Noise sensors	These sensors are addressed to monitor the level of noise produced in an area
WTG's solution	Energy consumption monitoring	Monitoring of the energy consumption of a building. This solution can also interact with other building elements to activate/deactivate lights or HVAC systems accordingly the presence of people (if there are presence sensors deployed)
WTG's solution	Waste collection monitoring	This solution monitors the filling level of the container, providing filling trends and sending alerts regarding temperature, movement, and filling level. There are two versions of this sensor one addressed to containers and another miniaturized version addressed to bins
WTG's solution	Lighting managing system	Enable management of the street lighting, being possible to reduce its consumption and adapt luminosity accordingly the presence of people (if there are presence sensors deployed)
BOT's IN-HABIT APP	Behavioural games	To be implemented in the application through surveys. This will be addressed to ensure the engagement of the users
BOT's IN-HABIT APP	Wellness spots	Monitoring of the impact of the VIS in the pilot through surveys

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

BOT's IN-HABIT APP	Digital citizens/users	Monitoring of the status of defined KPIs through surveys

Table 3. Preliminary services for the platform

The following sections contain the high-level description of the IN-HABIT Platform and the Architecture definition.

3. IN-HABIT DATA PLATFORM HIGH LEVEL DESCRIPTION

3.1. List of main characteristics

- Integrated platform for Holistic 360º View: single interface adapted to visualise information of all the active services
- Web Access (PC, Tablet or Mobile): responsive interface to access from any device from anywhere
- Unified access Interface in a Multitenant Environment: the same interface can be customised to different type of users with different level of access, visualization and editing
- Media Manager: multimedia data integration
- Modular and User Experience (UX) design: different modules to view information, inventory, surveys, generate performance reports…
- User friendly & Easy to use: clear buttons and menus to navigate into:
- Fast and Easy Deployment in Cloud: cloud deployment for a better scale up and security.

3.2. Management modules

- **Datasource Manager**: Work with different type of data sources: databases, Streamings, Sensors, Files…
- **Dashboard Manager:** Customized KPis. Visualizations, Maps and Dashboards.
- **Notification Manager:** Rules for Alert configurations and Reporting of events via emails or to a Ticketing Maintenance System.
- **Support and Administration Tools**: User and Application administration with Access Control List manager for a profiling of access to different datasets based on roles.
- **Vertical Applications:** specific use Case Applications developed natively over the platform framework.

3.3. Technical specifications

- Embedded **Big Data and Machine Learning**, distributed processing historical and real-time engine (time series oriented).
- **Acquisition and Control supports multiprotocol environment:** MQTT, COAP, Modbus, HTTP, SNMP among others.
- SDK and Restful **multilanguage APIs** (swagger) assures wide interoperability for integrations of Smart Services
- **End-to-end Native cybersecurity** mechanisms which analyze data for the detection of vulnerabilities and risk assessment.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

- **GIS engine and repository for asset georeferencing** allowing geographic consultations (OGC compliant) compatible with main map servers in the market as Google Street Map, OpenStreetMap, ArcGIS from ESRI, etc.
- **Support Layer,** with different modules (Logs monitoring, Task Scheduler, ACL with LDAP integration, Semantics Management).
- **Alarm Manager**, including triggering and follow-up of workflows for problem solving (operation/maintenance levels).
- **KPIs manager**, and numerous types of graphs, maps, for a quick and easy creation of dashboard and dynamic reports.

3.4. Benefits expected

- Better response and decision-making to operational demands due to real-time remote monitoring and control of infrastructure and on-field devices.
- Horizontal, Interoperable, scalable, flexible, and modular architecture for a long-term journey in a digital strategy.
- Real-time and historical analysis thanks to a built-in batch, time-series, and streaming process engines.
- Improves quality and control of multi-services scenarios due to dynamic dashboards with customized graphs, KPIs, visualizations, etc.
- Rule-based alerts and notification, based on correlation analysis from the data
- Improved planning, through the creation of simulated environments with what-if analysis getting the optimum set points for operations.
- IN-HABIT specific benefits like monitoring of the effects of the integrated nature based, digital, social, and cultural solutions; improved decision making and co-management of the solutions by the local Public-Private-People Partnerships (IN-HUBs); enhanced integration and access to information and data provided by the project digital innovative solutions like the IN-HABITAPP.
- To deliver datasets and connect users, systems, and data in the areas of interventions, with dashboards, fleet overview of all equipment (IoT, sensors) on a geographical site, with monitoring and notifications and with APIs (programmable interfaces) to interconnect with other systems and with specific user interfaces.
- Operates in the long-term, cybersecure open and consistent data about the impacts of IN-HABIT solutions and will be the core system for communication, exchange of practices and experience sharing.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

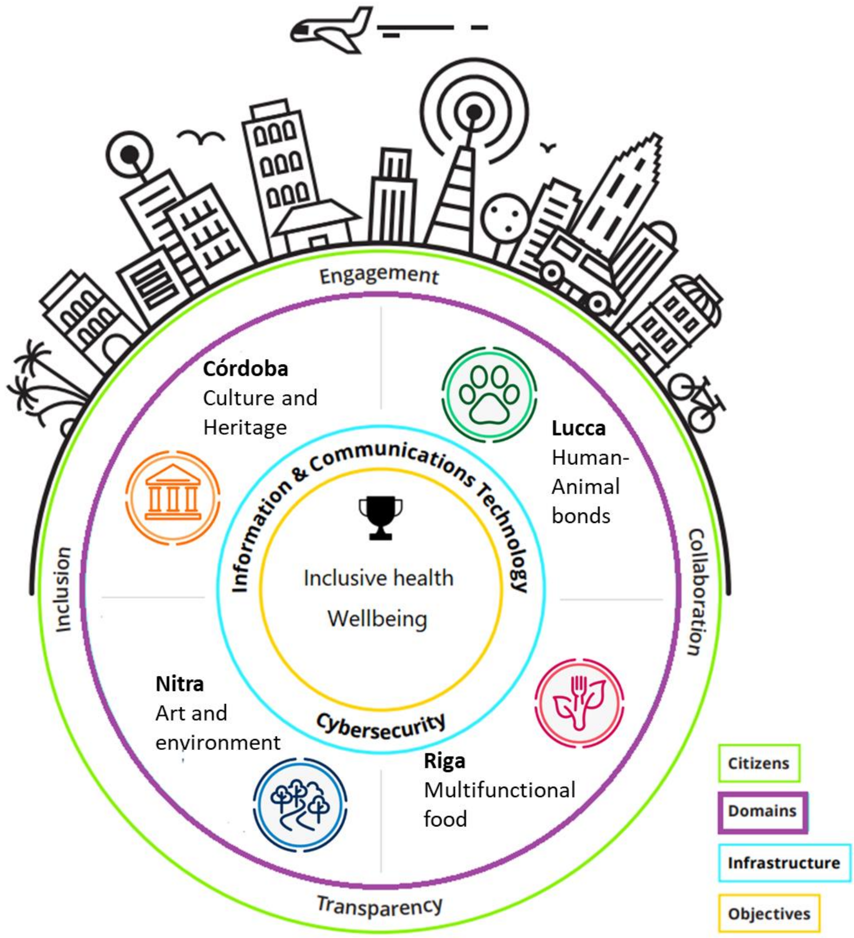

Figure 1. IN-HABIT Platform conceptual benefits

3.5. Holistic approach

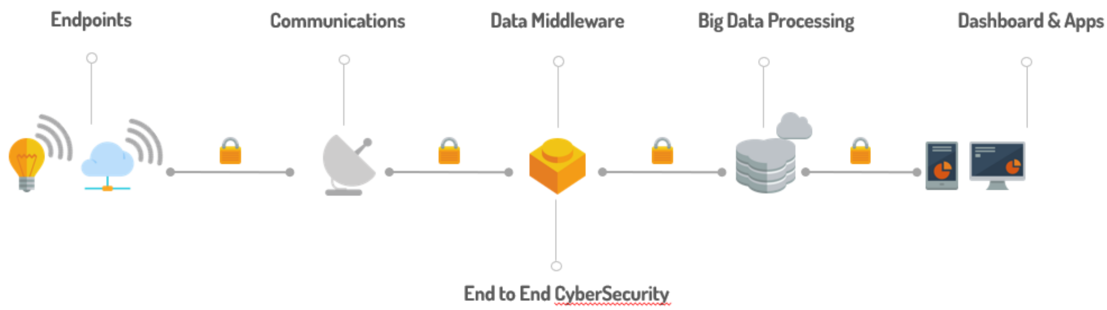

Figure 2. Holistic data platform approach

- **Endpoints**: Our sensors, network nodes, and 3rd party devices capture and transmit a wide variety of data
- **Communications**: Communication/HW agnostic platform, compatible with multiple communication methods (GPRS, Sigfox, LoRa, NB-IoT, 3G/4G, etc)
- **Data Middleware**: Interoperable middleware based on the latest technologies and industry standard protocols, captures, and transmits data to servers (Cloud/on-premises)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

- **Big Data Processing**: Normalizes data, both SQL/NoSQL, for in-batch and data stream analysis, creating metadata for faster access to big volumes of real-time and historical data.
- **Dashboard & Apps:** Visualizes in a simple and intuitive way, the data processed. Customizable Data Base, reports, and Advanced visual Analytics.
- **End to end cybersecurity:** Our advanced services can cover the entire security cycle, from prevention to a response when dealing with today's computing security threats and vulnerabilities.

3.6. Technical details

The platform has been designed following the reference of the UNE 178104: 2017 regulation. Comprehensive systems for a smart city management. Requirements of interoperability for a Smart City Platform.

This section provides a description of the modular architecture of the platform. The architecture of the platform is based on open and non-proprietary standards, which will allow scaling both its functionalities and capabilities. In other words, the platform can be extended to new heterogeneous sources of information or to other management areas that could be integrated in the future.

Specifically, the platform is ideal to serve as the basis of a "city platform" on which to subsequently interconnect other systems and / or services, such as verticals of public lighting management, energy efficiency or waste management.

The structural and functional characteristics of the platform make the solution ready for the Internet of the future. This is:

- **Faster Internet.** Technologies with high-speed data processing and communications capacity. The platform offered has Big Data technologies that allow real-time processing and storage of large volumes of data.
- **Internet of Things.** The platform offered is prepared for the acquisition and treatment of IoT-type sensors that model the behaviour of assets and processes of "things".
- **Open access.** The platform is open, with tools such as REST API to expose data to external agents and third parties to increase the number of application developments.

In addition, it has the capacity to support different areas of application, being possible the simultaneous implementation of several services in the same infrastructure. The following characteristics of the platform make it possible to give a horizontal treatment to the information exploited on the platform:

- **Multi-entity:** The platform allows several entities to connect to the same infrastructure.
- **Multiservice:** The platform allows the management of different services or areas of application simultaneously on the same infrastructure.
- **Transversality:** Both the information collected from different sensors and sources, as well as the devices used by a specific vertical service, can be used by other vertical applications, as well as serve as a basis for developing other advanced applications.

On the other hand, the platform offers support for different technologies, devices, and information capture mechanisms with different communication standards, which allows

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

interaction with different internal / corporate and / or external information systems. To offer this interoperability and heterogeneity, the platform incorporates:

- A REST connector so that any external device or service / platform can easily attack it.
- An IoT Broker for bidirectional connection between IoT devices that implement protocols of this technology (MQTT, CoAP, AMQP, etc.) and that allow the transmission of measurements / events, to the processing layer, to store them, or perform actions, or responses with commands to the sensor / actuator.
- The platform provides in its backend mechanisms for the provision and management of devices,
- The platform also allows the interoperability of applications, since it has mechanisms of extraction, transformation, and loading (ETL) of external data sources that are required.
- The platform provides an open API for Open Data portals at the interoperability layer. This simplifies the integration with other platforms and the development of applications that can be reusable and portable between different platforms.
- Rules creation module based on data or events from devices associated with one or more than one service offered, as well as the integration of other rules.

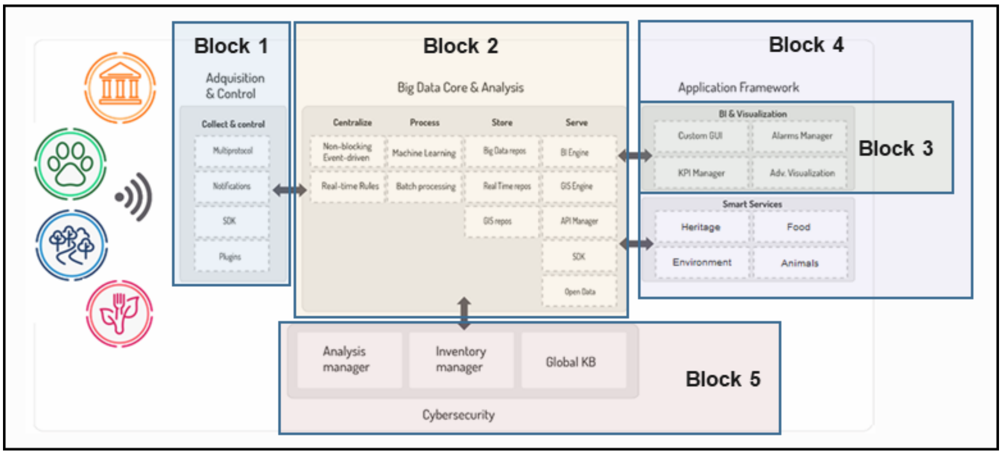

Figure 3. Block diagram of the IN-HABIT platform architecture

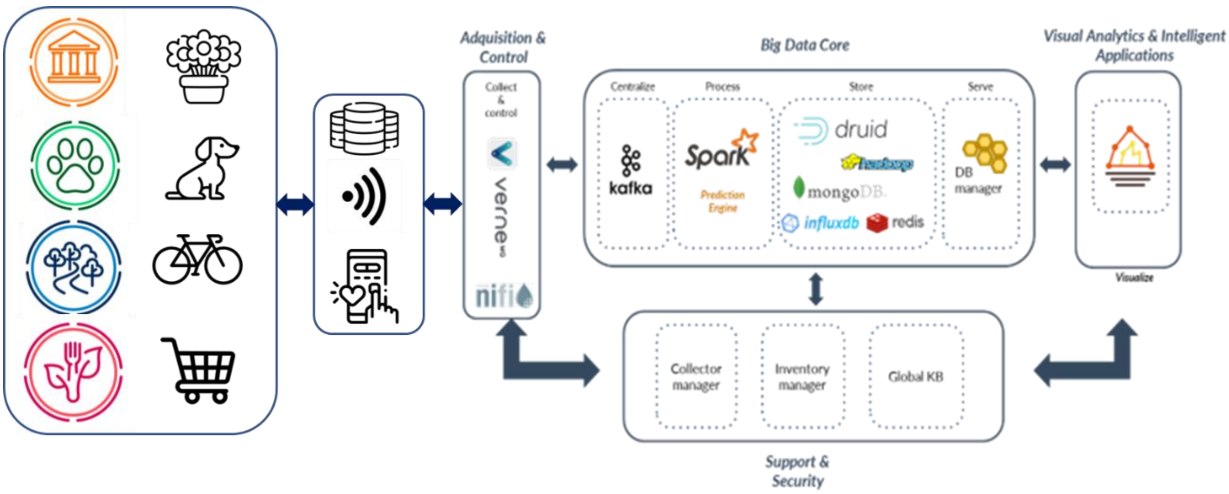

Figure 4. Technologies behind

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

In the first block the platform will collect data from the four pilot sites, from the different alternative we envisage so far, legacy data sources, sensors and IN-HABITAPP, this information will enter the platform via the acquisition and control layer, which collect, unify, pre-process, and control the flux of information.

The second block is the big data core, where the information will be processed to be stored in an enriched way as analytics will be applied to facilitate the generation of valuable information adding specific data of real time and historic events using algorithms for correlation, transformation and modelling of data.

The third block is devoted to the application framework, useful to develop future services easily.

The fourth block is the Business Intelligence and Visualization Layer, this is where the user interface is designed to show the information in an attractive and structured way.

The fifth block is the in-built cybersecurity layer to audit all resources to check for vulnerabilities and establish an assessment of risk level.

Figure 5. Visualization layer example

Each component of the platform is designed to scale automatically based on load. On the other hand, the architecture oriented to events and distributed services provides the platform with high availability of access to its functionalities and data.

Platform's architecture allows adding new elements to support services required in the future, thanks to the extension capacity of the platform through standard strategies. The platform has a modular approach, with clearly differentiated functional units, which allows its modules to be deployed, if necessary, separately.

Controlled access to information is guaranteed, allowing the configuration of different profiles and roles, under a multi-entity environment. Communication between components is carried out in an encrypted and secure manner, when necessary, to guarantee communication security (possibility of using SSL, TLS schemes, with certificate and token authentication).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

The platform includes Big Data capabilities to integrate a large amount of data generated from multiple sources and with different structures. It implements ETL (Extraction-Transformation-Load) integration processes, as well as massive processing of data generated from multiple sources and with different structures. Large volumes of data can be processed and stored with processing, both in real time and in batches, combining different sources. In summary, it is implemented in the proposed platform, mechanisms for:

- Data management
- Data Processing and Exploitation in Real Time
- Data Integration

The system has high performance capabilities to efficiently manage many devices, services, and processes in real time. This number can consist of hundreds of thousands of end-points simultaneously transmitting short messages, in the range of seconds.

For this reason, an architecture based on non-blocking events has been chosen, capable of initially processing millions of concurrent events in real time, which guarantees a minimum latency in their correlation and the obtaining of results or actions.

The platform aims to have a very user-friendly graphical web interface, and completely configurable in terms of control panels, which makes it easy to use with a dashboard with following capabilities:

- Real-time, event-based monitoring of raw and processed indicators
- Key Indicator (KPI) management module, which can be configured from the graphical user interface, not requiring any programming to add a new KPI.
- Customizable dashboards with widgets (drag and drop) and dynamic controls and filters.
- Generation of rules and alarms, sending alerts or emails in response to events.

The KPIs and Visualizations builder module allows complex formulations to be carried out from the graphical interface and without the need for programming, involving data from various sources, being able to add the source data and apply filters (greater than, equal to, is in, etc.) and mathematical operations to the data (sum, max, min, average, log, etc,).

Regarding the visualizations of graphs, there is a wide variety of options (bars, line, area, cakes, etc.), which can be stored in an orderly way once a graph is created, in a repository of objects (KPIs, Visualizations, Maps).

A KPI or visualization can be used as many times as needed in various Dashboards throughout the application, as an add-on widget, being available for user reports, both internal and external according to the privilege levels they have.

The GIS capabilities of the offered platform allow georeferenced information, its geographic query, and its visualization on cartographic viewers. These viewers can display maps, such as the street map, from sources such as Google or Open Street Maps. Open layers are the library most used in projects that have already been implemented on the platform.

4. ARCHITECTURE'S DESCRIPTION

This section provides a description of the software components that make up the architecture of the platform based on the UNE 178104: 2017 standard.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

The architecture brings together all the layers and modules of a Smart City Platform to support and integrate different vertical applications. The various functional blocks are accessible from these verticals in a transversal way, and with common support and administration facilities.

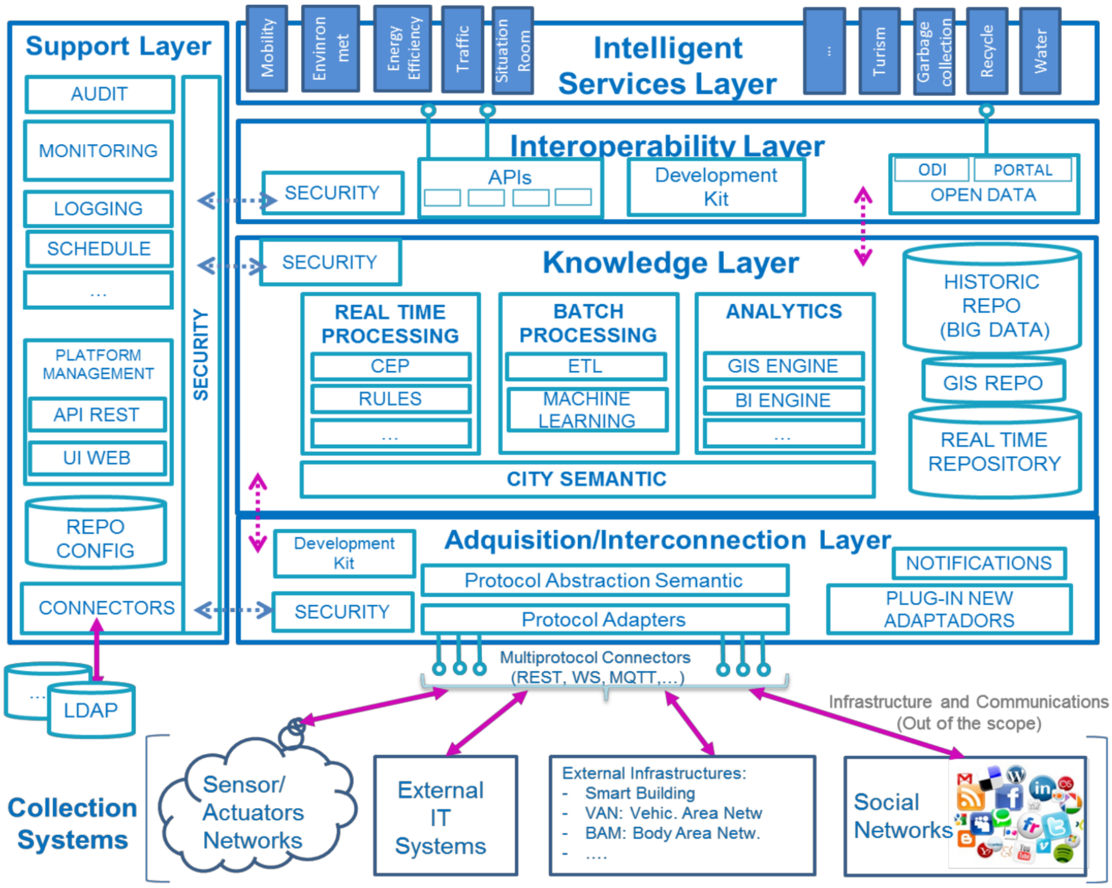

Figure 6. Platform's architecture

A description is made through the different layers of the architecture in order to offer a description of the software elements on which the functionalities of each of the layers are based.

4.1. Collection Systems

The capture systems are composed of any element that can provide data to the platform or on which it is possible to act from the platform, understanding as such elements:

- Sensors, actuators, gateways, and other devices that obtain information or that can be acted upon for the management of services.
- Data sources from external systems.
- Other existing IT systems: applications or legacy solutions.
- Communication with these elements or devices is bidirectional and orchestrated from the acquisition / interconnection layer.

All the endpoints of these architecture capture systems are therefore related to the acquisition and interconnection layer, which is the input of the data captured in the platform. The platform is agnostic from the point of view of communications and technologies and operators and therefore accepts cellular communications (GPRS / 3G / 4G), SigFox, LoRa, LoRaWAN, Wifi, ZigBee, Radio frequency, among others.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

As mentioned before this is the layer that will collect data from the different data sources of each pilot site, either new or existing ones.

4.2. Acquisition/Interconnection Layer

It offers the mechanisms for the correct acquisition and routing of data flows from the capture systems, allowing the integration of heterogeneous data in the platform.

- In charge of allowing interconnection with external systems that only consume data.
- IoT interface: different protocols available for communication, these are:
  - MQTT: based on VerneMQ technology to connect and communicate with devices or gateways that implement the MQTT protocol, with the streaming data bus.
  - IoT Broker / Ingestion Server: based on a proprietary development on Apache NiFi to ingest all types of data sources such as REST, CoAP, WS, HTTP, XMPP, KNX, ModBus, NGSI and ultimately any type of standard protocol. In this layer, the protocol translations are carried out to make the platform independent of the devices, communications technologies and protocols used to transmit information internally to the knowledge layer.

VerneMQ implements a well-proven and robust telecommunications technology stack for continuously operating systems making efficient use of all available resources as the basis for easy vertical scalability. VerneMQ uses a masterless clustering technology. This makes the cluster operation safe and simple. Focused on the collection of information in real time and remote control of endpoints. Thanks to the VerneMQ Webhooks Plugin, you can create powerful plugins using web links (Webhooks). Being able to use Python, Go, C # /. Net and any language with which to develop HTTP request manager.

Apache NiFi is used to define the batch processes of heterogeneous data sources. In this way, interfaces for integration with open data repositories are defined, for example. Formats such as CSV, GeoJSON, SHP, KML, XLSX, etc. can be processed.

The flows resulting from the previous definition go through a data modeler that performs the normalization, filtering, and formatting of the data, following a harmonized semantics throughout the rest of the platform layers, facilitating the analysis in the Knowledge Layer.

The acquisition layer requires a bidirectional orchestrator to control the flow of data. This orchestration is performed by a Streaming Data Bus implemented with Apache Kafka technology.

Kafka is a distributed transmission platform that manages multiple data streams, using a non-blocking, event-oriented bus:

- It allows the construction of real-time transmission data pipelines that reliably obtain data between systems or applications -> Between Acquisition elements (VerneMQ or NiFi) and streaming processing engines (e.g., Streaming Analytics on Spark).
- Applies to real-time processes that transform data or react to streaming data. Knowledge Layer real-time processing.

Additional functionalities are implemented on the previously mentioned technological components to improve their scalability, stability, and security (with encryption). There will be data streams that follow a course for Streaming processing through Kafka, as well as others

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

whose stream is defined in NiFi to feed a batch process. Both flows will be processed, within the knowledge layer, in different databases, due to their different nature and cadence.

Applied to the IN-HABIT project this layer aims to homogenise the way that the information from different data sources is structured to allow following storage and analysis.

4.3. Knowledge Layer

The knowledge layer of the proposed platform includes a scalable back-end for massive and agile storage on which the data is processed for the treatment and extraction of knowledge. It receives data from both the Acquisition Layer and the Interoperability Layer, applying Treatment and analysis processes to these to generate new datasets or modify / complete existing ones.

It incorporates DBmanager technology: a middleware that provides uniform and extensible access to various Big Data databases (Druid, MongoDB, Influx). It allows adding new connectors for the integration of additional databases under the same semantic model, which enables common access to various databases and allows making use of the power and flexibility of each of them in parallel. This layer will oversee housing all the information on which to make OLAP cubes for analytics. Its key features are as follows:

- Advanced and flexible information flow
  - Distributed and flexible management of multiple data streams, via non-blocking, event-oriented bus.
  - Facilitates online integration with other data repositories or geographic information systems.
  - Authentication and encryption of internal communications
- Storage and access to Analytical Databases (Big Data & BI)
  - Efficient persistent storage of large amounts of data with flexible formats, both for real-time information such as Druid (oriented to TimeSeries and OLAP Analytics) as well as that corresponding to historical data. All databases work in clusters and have hot horizontal scalability for the addition of new nodes if necessary.
  - Quick query of aggregated or specific data, historical records, and last events, based on distributed computing cloud, scalable, both in batch (HDFS Hadoop) and in real time (Druid + PostgreSQL), both of type SQL, and for No -SQL includes Mongo DB (non-relational database).
  - Support for storage and specific service of georeferenced data / assets and visualization (Web Map Service). Support of GIS engines on GEOSERVER with relational database POSTGIS (compatible with OGC WMS standards).
- Processing of complex rules in real time (CEP), and application of advanced analytics algorithms with capabilities to carry out Text / Data Mining and Machine / Deep Learning on large volumes of information.
  - Flexible and with algorithms for correlation, transformation, modelling, etc.
  - Data mining techniques to find the correlation between measurements
  - Automatic generation of alarms after detection of abnormal behavior
  - Machine Learning algorithms for pattern detection, clustering, automated clustering, etc., as well as for predictive based on regressions or decision trees.

In addition to the databases Druid, MongoDB, PostgreSQL, HDFS Hadoop, the underlying technologies in this knowledge layer are Apache Spark, a library stack including SQL and

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

DataFrames, MLlib for machine learning. These libraries can be seamlessly combined within the same application for complex analysis. In addition, APIs are provided in Java, Scala, Python, and R to extend the analytics capabilities of the knowledge layer at any time.

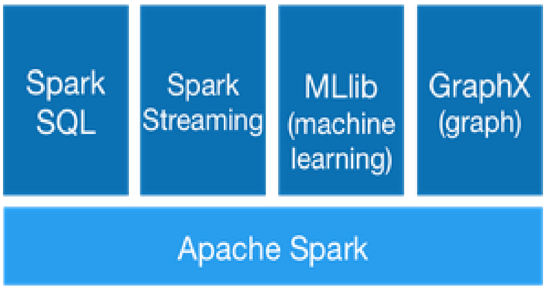

Figure 7. Apache Spark modules

- Spark SQL is Apache Spark's module for working with structured data. DataFrames and SQL provide a common way to access a variety of data sources, including Hive, Avro, Parquet, ORC, JSON, and JDBC.
- Spark Streaming brings Apache Spark's built-in language API for processing streams, allowing you to write streaming jobs in the same way that you write batch jobs.
- Spark MLlib provides both supervised and unsupervised learning algorithms that offer solutions to the three most used techniques in the world of Machine Learning: Classification and Regression, Clustering, and Collaborative Filter for Recommendation Techniques (ALS).
- GraphX is the Apache Spark API for graphing and parallel graphical computing.

The database technologies that support each of the knowledge layer repositories are:

- Druid: is a database for storing, querying, and analysing large streams of events (such as events include user-generated data, such as click sequences, application-generated data, such as performance metrics, and machine-generated data, such as network flows and server metrics). Druid is optimized for sub-second queries to split and give, drill down, search, filter, and aggregate this data.
- PostgreSQL (PostGIS): structured database with support for geographic objects to PostgreSQL and allows analysis through spatial SQL queries or through connection to GIS applications. This database will be consulted by Druid for access to data with geographic coordinates.
- MongoDB: non-relational database of document type, for the storage and treatment of NoSQL data, which allows batch processing of data and aggregation operations
- HDFS (Hadoop): Apache Hadoop is a framework used for distributed information storage and processing of very large data sets. Hadoop provides file system and operating system level abstractions, a MapReduce engine (MapReduce / MR1 or YARN / MR2), and the Hadoop Distributed File System (HDFS). The proposal includes an HDFS cluster, for massive consultation of the BigData heavy data set.

In case of IN-HABIT this is the layer where de data are stored once they have been harmonised, we still need to define how equivalent data from different pilot site will be storage, either by the type of data, by the pilot or by any other criteria that may interest.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

4.4. Interoperability layer

The interoperability layer of the platform facilitates the integration of new intelligent services, granting them access to the data and processes that are implemented in the Knowledge Layer. It offers interfaces on the Knowledge Layer establishing access security policies between the External Verticals to the data sets as well as to the processing engines and databases of the platform. The layer offers these functionalities:

- Open data, with CKAN technology, so that the information from the platform can be generated and published in a portal enabled for this purpose (not included in the platform) as open data according to levels of the 5-star scheme defined for the publication of data open.
- FTP service through which documents of any kind are offered for download. Invocation of FTP services is protected with access credentials that can be defined from the same platform.
- Development kit (APIs and SDKs), Prometheus is an isomorphic web application, developed in Javascript, and with an architecture based on microservices, which allows the implementation of new verticals in the Smart Services layer and advanced applications on the platform. The APIs will give access to the different data through the following methods (an external application consumes the API of the proposed platform):
    - Push mode: In this case, the external system obtains data from the platform by consuming the REST API of the platform in subscription / notification mode.
    - Pull mode: The external system collects data through control orders (response request) that it exchanges with the platform's REST API.

All accesses to data, services, devices / actuators / sensors, and datasets are defined by catalogues of standard primitives widely used throughout Europe. Specifically, use is made of the definitions that the GSMA has published in its repository [1]"HarmonizedEntityDefinitions".

This layer enables the exchange of data among different pilots' services. This way new services can be developed in the platform by the community in future leveraging this interoperability of data.

4.5. Intelligent Services layer

In this layer of the platform, the Services and external Applications are offered connected through the interoperability layer and the APIs provided, to interact with the data and process engines of the platform.

The services layer acts as a business intelligence layer, with metrics, KPIs, and predictive analytics to aid decision making. It is an advanced analytics and reporting Business Intelligence tool that is easy to use and that allows you to work with multiple data sources. The reports it generates are divided into sections or data groups in which the elements of the report can be positioned; The platform incorporates both the possibility of creating transversal dashboards to various verticals and access to the functionality of a specific vertical.

Key features of the layer:

- Real-time, event-driven raw and processed indicator tracking.

[1] Repository of published harmonized entity definitions that have been developed under the IoT Big Data project of the GSMA Connected Living program

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

- Responsive graphic display with any web browser and access through a mobile application, customizable to the desired look and feel.
- Customizable dashboards (by user and role) with widgets and dynamic controls and filters.
- Intuitive exploration and analysis of information
- Key Indicators (KPI) management module, which can be configured from the graphical user interface, and no programming is necessary to add a new KPI once the platform is deployed.
- Georeferencing of the elements that are installed in any of the verticals.
- Generation and management of alerts. Sending alerts within the platform or via email to users as deemed appropriate in response to predefined events and personalized reports.

This layer allows the user to generate rules and alerts according to measures, for example establish some alert if air quality is below a certain threshold.

4.6. Support layer

This transversal layer supports all the previous layers of the platform offering services such as auditing, monitoring, security and access control, management of cybersecurity certificates, API management, management web portal and platform configuration.

- Management and configuration of the platform: Central administration console of services and applications accessible via web and REST API through which any configuration related to the platform can be made. For example: creation of new control panels for new users, creation of rules and automations, editing of semantics, massive loads, new end-points, new connectors to data sources, etc.
- Identity management and access control:
    - Unified access control with role-based authentication and permissions with the possibility of domain server integration (LDAP).
    - Unlimited and hierarchical profiles (super administrator, level administrators, end users, etc.) in a flexible and simple way for the administrator.
    - OAuth / SSO implementation
- Audit: The system records the accesses of the users and components to the stored information, for later analysis.
- Security: The components of the platform store information securely, guaranteeing availability and recovery from system errors, and the confidentiality and integrity of the information stored, transmitting it in encrypted form. In addition to the minimum-security requirements, the platform includes an IT cybersecurity suite for the detection of potential attacks and risk assessment of key assets.
    - IT asset discovery and allocation, including devices and services
    - Network, device, and service analysis with active and passive mechanisms
    - Audit of all resources to verify vulnerabilities
    - Generation of online and offline reports
    - Creation of the global knowledge base: blacklists, signatures, threats, etc.
    - Machine learning to detect anomalous behaviour.
    - IDS optimization for high performance networks.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

- Monitoring: The system monitors the status of its different components by recording the traces left by the different components of the system, to consult its activity and proactively detect errors in its operation. To do this, monitoring tools such as Zabbix are incorporated into the deployments, so that the platform components, by checking their operation, send alarm events to the monitoring centre.

5. GENERAL STRATEGY FOR FUTURE INTEGRATIONS

In this section, a description will be made of the general strategy with which Wellness TechGroup undertakes the integrations carried out with third-party systems. In each case, decisions will be made that seek to preserve the consistency and integrity of the information during loading (Example: perform data extractions at times when the impact is zero or minimal).

Thus, prior to the development of any integration, a detailed functional and technical analysis of the process will be carried out, considering that before starting the integration work, the conclusions must be approved by the partners affected.

In any case, integrations with applications and services can be done through web services. Next, based on the platform architecture, each of the parts in which the integration APIs are used is described and the data flows are defined.

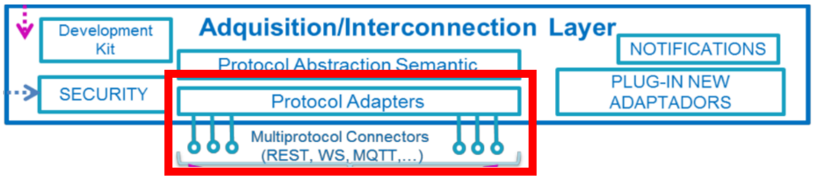

Figure 8. Protocols adapter detail

In the Acquisition / Interconnection Layer, multiprotocol connectors can be divided into two well-differentiated types of protocols: IoT protocol for device data acquisition and APIs for data acquisition from external applications.

5.1. IoT protocol for device data acquisition

The platform together with its IoT Broker, has compatibility with most standard IoT protocols such as MQTT, CoAP, KNX, XMPP, WebSockets, Streaming, NGSI 10 compatible with FiWare, etc.

Following this route, those devices that are contemplated in the future can be integrated, such as filling sensors, electric meters, water meters, etc. The platform also allows interaction with elements and devices in VPN connections with private IP.

5.2. APIs for data acquisition from external applications

In this case, the integration is carried out on the native Data Source Acquisition / Control REST API REST API of the platform. In this configuration, an external application will consume this API to:

- Push mode: In this case, the external system sends data to the platform (push mode) consuming a REST API of the platform.
- Pull mode: Data collection through control orders consuming the REST API.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

Ideally, the third-party system should adapt to the APIs of the platform, thus achieving high interoperability and scalability of the solution. In any case, the platform can integrate third-party systems that do not adapt to the previous data acquisition layer.

For this, the proposed platform has data ingestion methods through generic web services that allow the connection of this type of systems. Under this modality, there are 2 options to be able to interact with the external application:

- Push mode: Sending control orders from the platform to the external system, consuming the API or third-party app interface.
- Pull mode: In this case, the platform consumes the API or interface of the external system.

It is important to note that in all cases, the integrated data sources are normalized through pre-processing, which in this layer perform normalization, filtering, and data formatting, following a harmonized semantics throughout the platform, facilitating this way the analysis that is carried out from the Knowledge Layer.

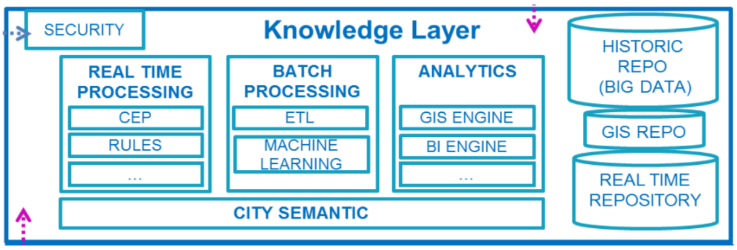

Figure 9. Knowledge layer detail

The data acquisition layer gives way to the Knowledge Layer in which once the data has been acquired, it follows a flow in which it is processed and stored as described below:

- Processing and analysis, some functionalities that are carried out at this stage:
    - Internal data bus between the different blocks of the platform (Kafka).
    - Information enrichment. Once the data have been acquired, they follow a flow in which they are processed and stored as will be seen below.
    - Filtering information fields.
    - Correlation of information from the same or different sources.
    - Machine Learning applied to each component.
    - Predictive analysis based on PMML.
    - Other processing tasks that apply to a specific component.
- Data persistence:
    - Storage of information in real and historical time, and presentation of data through a uniform access interface.

These considerations must be considered once new services over the platform need to be added, for example if we want to extract GPS position of any pilot asset to a specific mobile application.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

6. CONCLUSIONS AND NEXT STEPS

The current platform description shows the basic components of the architecture on which the IN-HABIT platform is built, based on an existing development of Wellness Techgroup's city platform. An open standards-based platform able to be adapted to the project's requirements.

These are the foundations over which the adaptations and vertical services inside IN-HABIT project will be defined and built, according to pilot sites requirements. In the following months (December 2021 to February 2022), the interaction among technical partners and pilot sites leaders will be intensified to achieve a complete definition and understanding of what data will be required to create the needed KPIs that enables IN-HABIT consortium to measure the impact of IN-HABIT methodology and solutions. Data collection will feed the platform from two main sources:

- Deployment of new sensors that will measure new variables not available before IN-HABIT actuation.
- Retrieval of legacy data from existing data bases or sensor already deployed.

Future improvement and evolutions of the platform will be developed as well as the required verticals to satisfy each pilot site requirements and provide the precise set of data to analyse them and create valuable information for decision makers.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

APPENDIX

VERSION LOG

Issue Date	Rev. No.	Author
20/09/2021	V00	Jose Morales (WTG)
30/09/2021	V01	Jose Morales (WTG)
12/11/2021	V02	Jose Morales (WTG)
15/11/2021	V03	Jose Morales (WTG), revised by Roberta Cocchioni (ISIM).
17/11/2021	V04	Jose Morales (WTG), revised by Mihaela Vancea (UCO).
29/11/2021	V05	Jose Morales (WTG), revised by BOT team.
29/11/2021	V06	Jose Morales (WTG), final revision by Mihaela Vancea (UCO)

LIST OF TABLES

LIST OF FIGURES

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227

ABBREVIATIONS

ACL	Access Control List
AMQP	Advanced Message Queuing Protocol
API	Application Programming Interface
CEP	Complex Event Processing
COAP	Constrained Application Protocol
ESRI	Environmental Systems Research Institute
ETL	Extract, Transform and Load
GPRS	General Packet Radio Service
GSMA	Global System for Mobile Communication Association
HDFS	Hadoop Distributed File System
HTTPS	HyperText Transfer Protocol Secure
HVAC	Heating, ventilation, and air conditioning
IoT	Internet of Things
JDBC	Java Database Connectivity
JSON	JavaScript Object Notation
KNX	Open standard for commercial and domestic building automation
KPIs	Key Performance Indicators
LDAP	Lightweight Directory Access Protocol
LORA	Long Range Communication Protocol
MQTT	Message Queue Telemetry Transport
NB-IOT	Narrow Band Internet of Things
NGSI	Next Generation Service Interfaces
OGC	Open Geospatial Consortium
REST	Representational State Transfer

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869227